International Journal of Distributed Sensor Networks

Wireless Machine-to-Machine Networks 2013

Guest Editors: JianHua He, Yan Zhang, Zhong Fan, Hsiao-Hwa Chen, and Lin Bai

Wireless Machine-to-Machine Networks 2013

International Journal of Distributed Sensor Networks

Wireless Machine-to-Machine Networks 2013

Guest Editors: Jianhua He, Yan Zhang, Zhong Fan,
Hsiao-Hwa Chen, and Lin Bai

Copyright © 2014 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "International Journal of Distributed Sensor Networks." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Miguel Acevedo, USA
Sanghyun Ahn, Korea
Mohammod Ali, USA
Jamal N. Al-Karaki, Jordan
Habib M. Ammari, USA
Masoud Ardakani, Canada
Javier Bajo, Spain
Prabir Barooah, USA
Alessandro Bogliolo, Italy
Richard R. Brooks, USA
James Brusey, UK
Erik Buchmann, Germany
Jian-Nong Cao, Hong Kong
João Paulo Carmo, Portugal
Jesús Carretero, Spain
Luca Catarinucci, Italy
Henry Chan, Hong Kong
Chih-Yung Chang, Taiwan
Periklis Chatzimisios, Greece
Ai Chen, China
Han-Hua Chen, China
Peng Cheng, China
Jinsung Cho, Korea
K.-K. R. Choo, Australia
Chi-Yin Chow, Hong Kong
W.-Y. Chung, Republic of Korea
Mauro Conti, Italy
Dinesh Datla, USA
Amitava Datta, Australia
Danilo De Donno, Italy
Ilker Demirkol, Spain
Der-Jiunn Deng, Taiwan
Longjun Dong, China
Chyi-Ren Dow, Taiwan
George P. Efthymoglou, Greece
Frank Ehlers, Italy
Melike Erol-Kantarci, Canada
Giancarlo Fortino, Italy
Luca Foschini, Italy
David Galindo, France
Deyun Gao, China
Weihua Gao, USA
Athanasios Gkelias, UK
Iqbal Gondal, Australia
Jayavardhana Gubbi, Australia
Cagri Gungor, Turkey
Song Guo, Japan
Andrei Gurtov, Finland
Qi Han, USA
Zdenek Hanzalek, Czech Republic
Tian He, USA
Junyoung Heo, Korea
Zujun Hou, Singapore
Yung-Fa Huang, Taiwan
Jiun-Long Huang, Taiwan
Xinming Huang, USA
Chin-Tser Huang, USA
Wei Huangfu, China
Mohamed Ibnkahla, Canada
Tan Jindong, USA
Ibrahim Kamel, UAE
Li-Wei Kang, Taiwan
Rajgopal Kannan, USA
Sherif Khattab, Egypt
Lisimachos Kondi, Greece
Marwan Krunz, USA
Kun-Chan Lan, Taiwan
Yee W. Law, Australia
Young-Koo Lee, Republic of Korea
Yong Lee, USA
JongHyup Lee, Korea
Sungyoung Lee, Republic of Korea
Seokcheon Lee, USA
Joo-Ho Lee, Japan
Kyung-Chang Lee, Korea
Shuai Li, USA
Shijian Li, China
Ye Li, China
Zhen Li, China
Shancang Li, UK
Jing Liang, China
Weifa Liang, Australia
Yao Liang, USA
Qilian Liang, USA
Wen-Hwa Liao, Taiwan
Alvin S. Lim, USA
Kai Lin, China
Ming Liu, China
Donggang Liu, USA
Yonghe Liu, USA
Zhong Liu, China
Chuan-Ming Liu, Taiwan
Zhigang Liu, China
Leonardo Lizzi, France
Giuseppe Lo Re, Italy
Jonathan Loo, UK
Juan Antonio Lpez Riquelme, Spain
Pascal Lorenz, France
Chun-Shien Lu, Taiwan
KingShan Lui, Hong Kong
Jun Luo, Singapore
fhlvaro Marco, Spain
Jose Ramiro Martinez-de Dios, Spain
Nirvana Meratnia, The Netherlands
Shabbir N. Merchant, India
Mihael Mohorcic, Slovenia
José Molina, Spain
Vallipuram Muthukkumarasamy, Australia
Kameswara Rao Namuduri, USA
George Nikolakopoulos, Sweden
Gregory O'Hare, Ireland
Suat Ozdemir, Turkey
Marimuthu Palaniswami, Australia
Ai-Chun Pang, Taiwan
Seung-Jong J. Park, USA
Soo-Hyun Park, Korea
Miguel A. Patricio, Spain
Wen-Chih Peng, Taiwan
Janez Perš, Slovenia
Dirk Pesch, Ireland
Shashi Phoha, USA
Antonio Puliafito, Italy
Hairong Qi, USA
Nageswara S.V. Rao, USA
Md. Abdur Razzaque, Bangladesh
Pedro Pereira Rodrigues, Portugal
Joel J. P. C. Rodrigues, Portugal
Mohamed Saad, United Arab Emirates
Sanat Sarangi, India
Stefano Savazzi, Italy
Marco Scarpa, Italy
Arunabha Sen, USA
Xiaojing Shen, China
Louis Shue, Singapore
Antonino Staiano, Italy

Tan-Hsu Tan, Taiwan
Guozhen Tan, China
Shaojie Tang, USA
Bulent Tavli, Turkey
Anthony Tzes, Greece
Agustinus B. Waluyo, Australia
Yu Wang, USA
Honggang Wang, USA
Ran Wolff, Israel
Wen-Jong Wu, Taiwan
Chase Qishi Wu, USA
Jianshe Wu, China
Feng Xia, China
Bin Xiao, Hong Kong
Qin Xin, Faroe Islands
Jianliang Xu, Hong Kong
Yuan Xue, USA
Ting Yang, China
Geng Yang, China
Hong-Hsu Yen, Taiwan
Li-Hsing Yen, Taiwan
Seong-eun Yoo, Korea
Ning Yu, China
Changyuan Yu, Singapore
Tianle Zhang, China
Yanmin Zhu, China
T. L. Zhu, USA
Yi-hua Zhu, China
Qingxin Zhu, China
Li Zhuo, China
Shihong Zou, China

Contents

Wireless Machine-to-Machine Networks 2013, Jianhua He, Yan Zhang, Zhong Fan, Hsiao-Hwa Chen, and Lin Bai
Volume 2014, Article ID 857139, 3 pages

A Performance Analysis Framework for WiFi/WiMAX Heterogeneous Metropolitan Networks Based on Cross-Layer Design, Martha Hernández Ochoa, Mario Siller, John Woods, and Hector Alejandro Duran-Limon
Volume 2014, Article ID 750971, 16 pages

DAG Based Multipath Routing Algorithm for Load Balancing in Machine-to-Machine Networks, Liang Liu, Tao Zhang, and Jun Zhang
Volume 2014, Article ID 457962, 12 pages

Secure Beamforming via Amplify-and-Forward Relays in Machine-to-Machine Communications, Zhongjian Liu, Xiaoning Zhang, Lin Bai, Chen Chen, and Haige Xiang
Volume 2013, Article ID 728532, 11 pages

SDMA-Based Aeronautical Machine-to-Machine Communications under SINR Constraints, Jindong Xie, Lin Bai, Jun Zhang, and Tao Zhang
Volume 2013, Article ID 475761, 9 pages

Elastic Information Management for Air Pollution Monitoring in Large-Scale M2M Sensor Networks, Yajie Ma, Yike Guo, Dilshan Silva, Orestis Tsinalis, and Chao Wu
Volume 2013, Article ID 251374, 14 pages

Design Considerations of Self-Adaptive Wireless M2M Network Communication Architecture, Xiao-jun Chen, Jia Ke, Yong-zhao Zhan, Man-rong Wang, and Ge Deng
Volume 2013, Article ID 979546, 11 pages

Collaborative Relay Beamforming Based on Minimum Power for M2M Devices in Multicell Systems, Xiaoning Zhang, Da Wang, Lin Bai, and Chen Chen
Volume 2013, Article ID 293565, 9 pages

Channel Estimation for Relay-Based M2M Two-Way Communications Using Expectation-Maximization, Xiaoyan Xu, Jianjun Wu, Chen Chen, Wenyang Guan, and Haige Xiang
Volume 2013, Article ID 676024, 11 pages

Random Access Control for M2M in LTE System, Zhide Chen and Yali Zeng
Volume 2013, Article ID 313797, 8 pages

Joint Access Control and Subchannel Allocation Scheme for Femtocell-Based M2M Network Using a Truthful Mechanism, Chengmei Li, Jianjun Wu, Ziqiang Feng, Xi Luan, and Haige Xiang
Volume 2013, Article ID 902574, 11 pages

Multiple Beam Selection for Combing M2M Communication Networks and Cellular Networks with Limited Feedback, Xiaoning Zhang, Lin Bai, and Wenyang Guan
Volume 2013, Article ID 540849, 11 pages

Multicast Capacity Analysis for High Mobility Social Proximity Machine-to-Machine Networks, Xin Guan
Volume 2013, Article ID 234728, 9 pages

Survey of the DASH7 Alliance Protocol for 433 MHz Wireless Sensor Communication, Maarten Weyn, Glenn Ergeerts, Luc Wante, Charles Vercauteren, and Peter Hellinckx
Volume 2013, Article ID 870430, 9 pages

The Performance of Multimessage Algebraic Gossip in a Random Geometric Graph, Gang Wang, Zun Lin, Wenyang Guan, and Feng Wang
Volume 2013, Article ID 545362, 7 pages

Trajectory Optimization of Packet Ferries in Opportunistic Social-Based Machine-to-Machine Networks, Xin Guan
Volume 2013, Article ID 895078, 11 pages

A Vector Algebraic Algorithm for Coverage Compensation in Hybrid Wireless Sensor Networks, Ning-ning Qin, Li-xia Guo, Zhi-guo Ding, and Bao-guo Xu
Volume 2013, Article ID 928528, 8 pages

Damping Based Traffic Allocation in Wireless Machine-to-Machine Communications Networks, Ruiguo Yu, Zhenxia Yuan, Mankun Zhao, Mei Yu, and Xu Lu
Volume 2013, Article ID 814267, 11 pages

Efficient and Secure Resource Management in Home M2M Networks, Yue Lai, Jiawen Kang, and Rong Yu
Volume 2013, Article ID 849572, 12 pages

Passive Synchronization Based Energy-Efficient MAC Protocol over M2M Wireless Networks, Pranesh Sthapit and Jae-Young Pyun
Volume 2013, Article ID 871607, 12 pages

A Game-Theoretic Analysis of Multiple Protocol Data Flows in Hierarchical M2M Communication Networks, Gang Wang, Jie Leng, and Lin Bai
Volume 2013, Article ID 584707, 8 pages

A Dual-Cycle Architecture in Cognitive M2M Wireless Networks, Ying Li, Xu Dong, Lifeng Wang, Xiaobo Guo, and Lin Bai
Volume 2013, Article ID 239170, 10 pages

Design and Evaluation of a Framework for Cooperative and Adaptive QoS Control of DSRC Network for Road Safety Applications, Wenyang Guan, Jianhua He, Zuoyin Tang, and Thomas M. Chen
Volume 2013, Article ID 946931, 10 pages

An Efficient Emergency Message Broadcasting Scheme in Vehicular Ad Hoc Networks, Hayder Salman Dawood and Yuming Wang
Volume 2013, Article ID 232916, 11 pages

EAP-Based Group Authentication and Key Agreement Protocol for Machine-Type Communications, Rong Jiang, Chengzhe Lai, Jun Luo, Xiaoping Wang, and Hong Wang
Volume 2013, Article ID 304601, 14 pages

Connectivity Analysis for Free-Flow Traffic in VANETs: A Statistical Approach, Chen Chen, Xiaobo Du, Qingqi Pei, and Yanan Jin
Volume 2013, Article ID 598946, 15 pages

Mixed Cooperation MAC Protocol with Sleep Mechanism for Data Acquisition in Wireless Machine-to-Machine Networks, Yulei Zhao, Bing Du, and Ning Ge
Volume 2013, Article ID 360267, 10 pages

An Energy-Heterogeneous Clustering Scheme to Avoid Energy Holes in Wireless Sensor Networks, Gong Bencan, Jiang Tingyao, Xu Shouzhi, and Chen Peng
Volume 2013, Article ID 796549, 8 pages

Outage Probability of Dual-Hop Relay System with Interference and Feedback Delay in Satellite M2M Networks, Yuan Lin, Yuanzhi He, Qinfei Huang, Hao Yin, and Shaodong Feng
Volume 2013, Article ID 373849, 11 pages

A Smart Electronic Tagging System Based on Context Awareness and Machine-to-Machine Interworking, Gil-Yong Lee, Nam-Yeol Yun, Sung-Chul Lee, and Soo-Hyun Park
Volume 2013, Article ID 392083, 17 pages

A Utility-Based Rate Allocation of M2M Service in Heterogeneous Wireless Environments, Yao Huang, Hui Tian, Jie Zhang, Cheng Qin, and Zhibo Wang
Volume 2013, Article ID 318407, 7 pages

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 857139, 3 pages
http://dx.doi.org/10.1155/2014/857139

Editorial

Wireless Machine-to-Machine Networks 2013

Jianhua He,[1] Yan Zhang,[2] Zhong Fan,[3] Hsiao-Hwa Chen,[4] and Lin Bai[5]

[1] *School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, UK*
[2] *Simula Research Laboratory, 1325 Lysaker, Norway*
[3] *Telecommunications Research Laboratory, Toshiba Research Europe Ltd., Bristol BS1 4NDl, UK*
[4] *National Cheng Kung University, Tainan City 70101, Taiwan*
[5] *Beihang University, Beijing 100191, China*

Correspondence should be addressed to Jianhua He; j.he7@aston.ac.uk

Received 23 February 2014; Accepted 23 February 2014; Published 24 March 2014

Copyright © 2014 Jianhua He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine-to-machine (M2M) technology allows direct communications between machine devices through wireless and/or wired systems without human intervention, which opens unique opportunities to businesses as well as consumers on cost reduction and services improvement. It holds huge potential for applications in a wide range of industries. It has been expected that there will be a huge increase in the number of machine devices enabled by wireless M2M technology and the amount of traffic generated by these devices in the next five years.

Due to the large diversities of the M2M applications (e.g., smart grids, smart home, wireless sensor networks, vehicle ad hoc networks, etc.) and M2M devices, various wireless network technologies (e.g., cellular networks, WiMAX, Wi-Fi, ZigBee, etc.) are likely to be applied jointly to effectively support M2M applications. However, the heterogeneous network technologies combined with the low-power machine devices and diverse QoS requirements of M2M applications present big challenges to the design and implementation of wireless M2M network protocols and algorithms.

Built on the last year's special issue on wireless M2M technologies and applications, this special issue continues to serve as a platform to gather researchers and engineers from different relevant areas, such as wireless sensor networks, RFID, WPAN, cellular networks and smart grid, to present the latest research advances on wireless M2M networks. Thirty papers have been accepted in this special issue. These papers are mainly focused on the following four research areas to address the main research challenges faced by wireless M2M networks, which are usually characterized by a large amount of low-power M2M devices, heterogeneous networks with limited bandwidth, and many coexisting applications with diverse QoS requirements: (1) developing advanced scalable wireless communication technologies and network architectures to improve network capacity; (2) designing and analyzing scalable wireless channel access and resource and interference management mechanisms for M2M applications; (3) designing and analyzing secure energy and bandwidth efficient network protocols; (4) designing and analyzing smart, adaptive, and self-organizing protocols and algorithms for dynamic wireless M2M networks.

In the paper entitled "*A dual-cycle architecture in cognitive M2M wireless networks*" by Y. Li et al. dual-cycle cognitive model (DCCM) applied develops large scale cognitive wireless networks, in which the structure and implementation of service-oriented radio architecture (SORA) are explored. Experiment results show that the proposed DCCM-based SORA is able to develop loose-coupled, reusable, and scalable heterogeneous wireless M2M networks. The paper entitled "*A performance analysis framework for WiFi/WiMAX heterogeneous metropolitan networks based on cross-layer design*" by M. H. Ochoa et al. propose a performance analysis framework (PAF) for heterogeneous wireless networks, which utilizes process algebra (PA) and the corresponding telecommunication traffic performance models and is applied to analysis of IEEE 802.16 and IEEE 802.11 protocols. In the paper entitled "*Multiple beam selection for combing M2M communication networks and cellular networks with limited feedback*" by

X. Zhang et al. a novel multiple beam selection (MBS) approach with limited feedback is proposed to provide network access to machine type devices in cellular networks by taking into account the orthogonal constraints. The paper entitled "*Random access control for M2M in LTE system*" by Z. Chen and Y. Zeng studied a random access control scheme for M2M devices in LTE system, which differentiates and controls access rates of users with different delay requirements. In the paper entitled "*Joint access control and subchannel allocation scheme for femtocell-based M2M network using a truthful mechanism*" by C. Li et al. a joint access control and subchannel allocation (JACSA) scheme is proposed for femtocell-based machine-to-machine (M2M) networks with both open access and hybrid access scenarios. In the paper entitled "*Collaborative relay beamforming based on minimum power for M2M devices in multicell systems*" by X. Zhang et al. relay beamforming strategies are studied to guarantee quality of service (QoS) requirements of multiple devices in multicell systems. In the paper entitled "*A utility-based rate allocation of M2M service in heterogeneous wireless environments*" by Y. Huang et al. a dynamic adaptive service rate allocation scheme is proposed to optimally distribute traffic from machine devices in heterogeneous wireless environments. The paper entitled "*A game-theoretic analysis of multiple protocol data flows in hierarchical M2M communication networks*" by G. Wang et al. proposes a new discriminatory user-based utility mechanism for data traffic from heterogeneous machine devices in hierarchical M2M networks.

In the paper entitled "*Mixed cooperation MAC protocol with sleep mechanism for data acquisition in wireless machine-to-machine networks*" by Y. Zhao et al. a mixed sleep-cooperative time division multiple access (TDMA) media access control (MAC) protocol is proposed for wireless M2M networks by exploring the idle state of each sensor node. In the idle state each sensor node dynamically goes into sleep state or cooperative state to maximize the network lifetime. The paper entitled "*A smart electronic tagging system based on context awareness and machine-to-machine interworking*" by G.-Y. Lee et al. presents a smart electronic tagging system, which connects a ubiquitous system network (USN) and Android mobile platforms to effectively track the locations of persons being monitored. In the paper entitled "*Design considerations of self-adaptive wireless M2M network communication architecture*" by X.-J. Chen et al. a concept of attribute assembly layer is proposed and an adaptive architecture for wireless M2M network is built, which unifies the data link layers for heterogeneous networks. In the paper entitled "*Passive synchronization based energy-efficient MAC protocol over M2M wireless networks*" by P. Sthapit and J.-Y. Pyun a new MAC protocol called passive synchronization based energy-efficient MAC (PSMAC) synchronizes nodes in their sleep states by using interrupts generated from a proposed radio-triggered hardware to provide diverse QoS with ultra-low power consumption and mobility support. In the paper entitled "*Multicast capacity analysis for high mobility social proximity machine-to-machine networks*" by X. Guan the multicast capacity of bus-assisted VANETs in forwarding scenario and routing scenario is studied. The paper entitled "*Trajectory optimization of packet ferries in opportunistic social-based machine-to-machine networks*" by X. Guan investigates the optimization problem of the moving trajectory by considering the minimum transmission delay for wireless M2M networks, which is studied by a semi-Markov decision process model. In the paper entitled "*Secure beamforming via amplify-and-forward relays in machine-to-machine communications*" by Z. Liu et al. collaborative amplify-and-forward relaying is applied to form a beamforming system and provide physical layer security for a wireless M2M communication network. The paper entitled "*EAP-based group authentication and key agreement protocol for machine-type communications*" by R. Jiang et al. presents a group authentication and key agreement protocol, called EG-AKA, for machine-type communications, which combines elliptic curve Diffie-Hellman (ECDH) based on EAP framework. The proposed protocol provides stronger security and better performance compared to conventional EAP-AKA scheme. The paper entitled "*The performance of multimessage algebraic gossip in a random geometric graph*" by G. Wang et al. proposed an information dissemination method based on algebraic gossip in a random geometric graph (RGG) to disseminate messages rapidly in large scale networks.

The paper entitled "*DAG based multipath routing algorithm for load balancing in machine-to-machine networks*" by L. Liu et al. presents a directed acyclic graph based multipath routing algorithm for congestion mitigation, where different routing paths are confined in a directed acyclic graph (DAG) under time delay constraint. In the paper entitled "*Channel estimation for relay-based M2M two-way communications using expectation-maximization*" by X. Xu et al. channel estimation method developed for relay-based M2M two-way communications is proposed, in which the expectation-maximization algorithm is utilized to provide the maximum likelihood solution in the presence of unobserved data through stable iterations.

In the paper entitled "*Damping based traffic allocation in wireless machine-to-machine communications networks*" by R. Yu et al. a new load balancing algorithm is proposed to distribute traffic over multiple available paths for wireless M2M networks.

The paper entitled "*Efficient and secure resource management in home M2M networks*" by Y. Lai et al. presents an efficient bandwidth measurement and resource management method for home M2M networks. In the paper entitled "*Elastic information management for air pollution monitoring in large-scale M2M sensor networks*" by Y. Ma et al. an elastic resource allocation strategy was proposed, which is a novel management technique based on elastic computing, for large-scale M2M sensor networks developed for air pollution monitoring. In the paper entitled "*A vector algebraic algorithm for coverage compensation in hybrid wireless sensor networks*" by N.-N. Qin et al. a vector algebra based algorithm is proposed by exploring redundant nodes as an extra dimension for coverage compensation to tackle the coverage hole problem in hybrid wireless sensor networks. In the paper entitled "*An energy-heterogeneous clustering scheme to avoid energy holes in wireless sensor networks*" by G. Bencan et al. energy consumption of nodes at different network areas is analysed and an energy-efficient clustering routing protocol

is proposed to improve network lifetime of wireless sensor networks. The paper entitled "*Survey of the DASH7 Alliance Protocol for 433 MHz wireless sensor communication*" by M. Weyn et al. presents a survey of an active RFID standard for wireless sensor communications, which covers comparisons to the general concepts, key techniques, and a software implementation of the standard.

The paper entitled "*Design and evaluation of a framework for cooperative and adaptive QoS control of DSRC network for road safety applications*" by W. Guan et al. presents a framework for cooperative QoS control for vehicle ad hoc networks and implementation, in which network feedback and cross-layer design are exploited to collaboratively achieve targeted QoS. In the paper entitled "*Connectivity analysis for free-flow traffic in VANETs: a statistical approach*" by C. Chen et al. statistical properties of direct and indirect connectivity in vehicular environment are analysed and exploited to provide insights to protocols design and improvement. In the paper entitled "*SDMA-based aeronautical machine-to-machine communications under SINR constraints*" by J. Xie et al. an SDMA-based aeronautical M2M communication system is proposed together with a low complexity user selection approach to achieve efficient use of limited spectrum for aeronautical M2M communications. In the paper entitled "*Outage probability of dual-hop relay system with interference and feedback delay in satellite M2M networks*" by Y. Lin et al. the outage probability of a dual-hop wireless M2M communication system using amplify-and-forward (AF) and decode-and-forward (DF) relaying schemes is analysed. The paper entitled "*An efficient emergency message broadcasting scheme in vehicular ad hoc networks*" by H. S. Dawood and Y. Wang presents an efficient emergency message broadcasting scheme with low overhead for road safety application to avoid multiple accidents and reduce traffic jams.

The thirty papers included in this special issue presented the latest research works on both protocol design and theoretic analysis for wireless M2M networks. We hope that this special issue can help our readers get a better understanding about the breadth and depth of the current research advances on wireless M2M networks and boost further research and practical developments in the field of wireless M2M networks.

Acknowledgments

The Guest Editors would like to thank all the authors for their contributions to this special issue. Our special thanks go to all the reviewers for their hard work in providing timely responses and constructive comments.

Jianhua He
Yan Zhang
Zhong Fan
Hsiao-Hwa Chen
Lin Bai

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 750971, 16 pages
http://dx.doi.org/10.1155/2014/750971

Research Article

A Performance Analysis Framework for WiFi/WiMAX Heterogeneous Metropolitan Networks Based on Cross-Layer Design

Martha Hernández Ochoa,[1] Mario Siller,[1] John Woods,[2] and Hector Alejandro Duran-Limon[3]

[1] *CINVESTAV, Unidad Guadalajara, Avenida del Bosque 1145, Col el bajío, 45019 Zapopan, JAL, Mexico*
[2] *University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK*
[3] *Information Systems Department, CUCEA, University of Guadalajara, Periférico Norte No. 799, Núcleo Belenes, 45100 Zapopan, Jalisco, Mexico*

Correspondence should be addressed to Martha Hernández Ochoa; mhernande@gdl.cinvestav.mx

Received 22 August 2013; Revised 30 November 2013; Accepted 14 December 2013; Published 3 March 2014

Academic Editor: Jianhua He

Copyright © 2014 Martha Hernández Ochoa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The communication between network nodes within different protocol domains is often regarded simply as a black box with unknown configuration conditions in the path. We address network heterogeneity using a white box approach and focus on its interconnection processes. To achieve this purpose, a *Performance Analysis Framework* (PAF) is proposed which is composed of the formalization of the latter using process algebra (PA) and the corresponding teletraffic performance models. In this contribution, we target the IEEE 802.16 and IEEE 802.11 protocols. For the teletraffic models, we extend previous models for such scenario with the inclusion of the following protocol operational parameters (metrics): bit error rate (BER), packet error ratio (PER), and packet length (pl). From the framework teletraffic models, the optimal packet length (OPL), end to end throughput, delay, and packet loss are obtained. The PAF outperforms previous modeling solutions in terms of delay and throughput relative to NS3 simulation results.

1. Introduction

Recently, the numbers of users and the quantity of information in heterogeneous networks have grown rapidly in metropolitan areas. The interconnection between nodes within different protocol domains is often regarded simply as a black box with unknown configuration conditions in the path. With increased demand, it has become necessary to analyze and develop new models to understand the performance of heterogeneous metropolitan area networks (HMAN). There are a number of factors that decrease the transmission efficiency between networks, such as (1) delay, (2) jitter, (3) medium access protocols, (4) packet loss ratio, (5) packet length (pl), (6) bandwidth, (7) bit error rate (BER), and (8) packet error ratio (PER). These factors are key measures, reflecting the network performance, but consideration must also be given to scalability, interoperability, and security to achieve the best network performance. HMAN can be considered as a Machine-to-Machine (M2M) heterogeneous network. The process of autonomously communicating two electronic systems (i.e., without human intervention) such as machines, devices, and actuators is called M2M communication. M2M communication's characteristics are large number of heterogeneous nodes, a gateway, low power, and low cost [1–4]. An M2M network is composed of five key elements as defined by the European Telecommunications Standards Institute (ETSI) [3]: (1) the M2M components (e.g., electrical devices, smart nodes, etc.), (2) the M2M gateway, which is responsible for the connection among the devices and the connection of the M2M communication network, (3) the M2M applications, which work as a middleware layer to pass data through various application services and are

(1) Identify the network heterogeneity. This includes all the involved communication protocols across the available communication paths and domains. Use an end to end communication perspective.
(2) Define the metrics to models.
(3) Survey available teletraffic models for the identified communication protocols from each domain.
(4) Extend and adapt the previous models according to the network design and corresponding performance metrics for each communication protocol or network domain. Use the CLD approach.
(5) Identify the gateway nodes and the corresponding interconnection tasks defined in the protocol specification and network design. Base this identification process on the roles established on Section 4.1
(6) Derive the interconnection teletraffic models for the gateway nodes.
(7) Based on the processes defined in Section 4.1, integrate all the involved teletraffic models across the communication paths under an end to end perspective.
(8) Validate the end to end performance models using test bed implementations or network simulation. Improve teletraffic models if necessary.

Box 1: White box approach.

used by specific business processing engines, (4) the M2M area network, which enables connectivity between M2M components and M2M gateways, and (5) the M2M communication network (network domain) that provides connection between M2M gateway(s) and M2M application(s). Also, the M2M communication requires operational stability and sustainability [5].

The 802.16 protocol supports M2M applications in the 802.16p version [5]. This protocol enables a range of M2M applications in which the communication device requires wide-area wireless coverage in licensed bands and is automated, rather than human initiated or human-controlled. This is for purposes such as observation and control. The requirements that 802.16 is intended to address include low power consumption, a large number of devices, short-burst transmissions, and device tampering, detection, and reporting. The 802.16 protocol acts as an aggregation point for 802.16 M2M devices and supports peer to peer (P2P) connectivity between these devices [5, 8].

The 802.11 protocol in the version of IEEE802.11ah standard has been found as an optimal candidate for M2M communication in wireless communication systems. This protocol considers sensing applications and will address required functions such as low power consumption, large number of devices, long-range and short-burst data transmissions [9].

The 802.16 and 802.11 protocols can be part of an M2M network where M2M devices (communication enabled) form an M2M area network. This M2M area network is based on the IEEE802.11 protocol, whilst the access network that connects the M2M gateway (nodes with two interfaces: 802.11 and 802.16) to the M2M core (e.g., the 3rd Generation Partnership Project (3GPP) Initiative for mobile communication technologies) is based on the IEEE802.16 protocol. In the M2M service platform, the M2M gateway is connected to the M2M management server (through the communication network) whereby the M2M applications (smart metering, smart transport, healthcare database, etc.) are reached [3].

Most network performance models consider only the homogeneous specification (common part) of the protocols that build the network rather than also including the differences among them. We believe that end to end performance analysis of heterogeneous networks should consider (i) the detailed specification (physical (PHY) and media access control (MAC) layers) and operational parameters of each protocol (BER, PER and pl, etc.), and (ii) the interconnection process that makes the heterogeneous network possible. Most research focuses on homogeneous behavior [10–17]; therefore, new and novel approaches are required to explore heterogeneous network performance specifically based on the IEEE802.16 and IEEE802.11 protocols. We define a heterogeneous network as interconnected networks with different software, protocols, hardware, operational speed, technology, and so forth but yet still capable of interoperability. We address network heterogeneity using white box approach and focus on its interconnection processes Box 1. For this purpose, a *Performance Analysis Framework* (PAF) is proposed which is composed of the formalization of the latter using process algebra (PA) and the corresponding teletraffic performance models. In this contribution, we target the IEEE802.16 and IEEE802.11 protocols.

The main contribution of this paper is the evaluation of end to end throughput and delay in a HMAN by considering the effect of different layers based on CLD (layer 2 and layer

1 of OSI model). For this purpose, a PAF is proposed. A WiFi/WiMAX HMAN is used as a case study. The proposed PAF consists of the following key elements.

(i) The first element is formalization of the end to end communication process of a HMAN using PA.

(ii) The second element is teletraffic performance models. We derived end to end performance models for heterogeneous environments which take into consideration the interconnection process of different protocol domain networks, bit error rate, packet error ratio, and packet length. In our proposed models, we consider realistic error-prone channel conditions which have not been considered in previous related works.

The paper is structured as follows. Section 2 presents the related work. In Section 3, we present the theoretical background of the IEEE802.11 and IEEE802.16 protocols. The PAF (the formalization of the end to end communication process of a HMAN using PA and teletraffic performance models) is presented in Section 4. In Section 5, the experimental work is presented. The simulation results and discussions are presented in Section 6. Finally, some conclusions and future work are drawn up in Section 7.

2. Related Work

In this section, we present a review of some relevant related work. The review is divided into two categories: homogeneous and heterogeneous approaches.

2.1. Homogeneous Approaches. The homogeneous approaches reviewed are described below.

Bianchi's analytical model [10] is used to estimate the throughput of an IEEE802.11 network using the Distributed Coordination Function (DCF) under saturated conditions. This assumes (i) any transmission queue that always has packets to be sent, (ii) an ideal channel, and (iii) a finite number of stations. The model considers two DCF techniques: basic and RTS/CTS (request to send/clear to send). The approach adopted is to analyze a single station modeled using a Markov Chain. The results demonstrate that better performance is achieved when the RTS/CTS mechanism is used.

Duffy et al. in [11] present an extension of Bianchi's model. They consider on-saturated network conditions. They assume a perfect PHY layer, so transmission errors are caused only by collisions and do not occur due to noise on the medium. The analysis is focused on the throughput, collision probability, delay, total offered load, and (the optimal) minimum contention window. They employ three load types: Poisson, conditional, and uniform.

In [12], Lin and Wong's model (IEEE802.11n) addresses a unidirectional and bidirectional RTS/CTS access mechanism in the presence of collisions and channel errors in the system. This model, which is an extension of Bianchi's model, considers BER probability, minimum contention window length, and a maximum back-off stage. Their model also includes the MAC Protocol Data Unit Aggregation (A-MPDU) and MAC Service Data Unit Aggregation (A-MSDU) techniques to improve the MAC protocol performance. Simulation and analytical results are presented for throughput and delay. This is done for a different number of aggregation MPDUs and BER conditions.

Hwang et al. present in [13] a teletraffic mathematical analysis for the delay of bandwidth requests based on unicast, multicast, and broadcast IEEE802.16d/e polling in IEEE802.16d/e under error-free/error-prone wireless channel conditions. They derived the distribution for delays and the truncated binary exponential back-off (adopted as a contention resolution) by means of analytical methods. The authors study bandwidth efficiency and the utilization of transmission opportunity which is defined as a ratio of successful transmission opportunities and total transmission opportunities within a frame. Based on numerical analysis, they obtained optimum values for parameters such as number of transmission opportunities (or slots), the initial back-off window size, and Quality of Service (QoS) requirement on delay and loss.

In [14], Tian et al. propose a novel MAC scheme used for the DCF named Scheduled Random Access protocol (SRAP). This scheme is dived in two parts: schedule and contention. For the former the protocol allows a throughput close to the transmission capacity in a saturate case while for the latter a low delay is observed in low traffic load conditions. The analysis of simulation results showed that SRAP can improve the throughput with low delay. The throughput analysis is based on a teletraffic model and they state that SRAP achieves throughput close to the theoretical upper bound.

In [15], Calì et al. develop as an analytical model for the throughput a p-persistent IEEE802.11 protocol, which differs from the standard protocol in terms of the selection of the back-off interval. The standard protocol uses the binary exponential back-off while in the p-persistent case, the back-off is sampled from a geometric distribution with a parameter p. Also, they demonstrated that (i) the standard protocol can operate very far from the theoretical throughput limit depending on the network conditions and (ii) the IEEE802.11 protocol is close to the theoretical throughput limit when the p-persistent (geometric distribution) back-off algorithm is employed.

In [16], Liu et al. proposed a scheduling algorithm based on Cross-Layer Design (CLD) between the MAC layer and PHY layer. Each connection employs adaptive modulation and coding (AMC) and considers the QoS requirements. The algorithm operational parameters are derived from a series of teletraffic models which consider performance metrics such as bandwidth efficiency, throughput, delay, PER, and Signal to Noise Ratio (SNR). Simulations were implemented for the IEEE802.16 standard.

Chang et al. in [17] analyze throughput using the Markov model of Vinel et al. using different window sizes and Subscriber Station (SS) numbers. The SS employs a polling MAC instead of random access control, based on the uplink subframe of a (Worldwide Interoperability for Microwave Access) WiMAX Network.

2.2. Heterogeneous Approaches. The heterogeneous approaches reviewed are described below.

In [18] El-Azouzi et al. study an HMAN formed using the IEEE802.11 and IEEE802.16 protocols. The aim is to study how to integrate different technologies cooperating to provide universal connectivity and opportunity for the best suited services to users at anytime from anywhere. This is envisioned as a common scenario for fourth generation (4G) networks. They believe that the integration of IEEE802.11 and IEEE802.16 is one likely solution for distribution of high data rate services for next generation wireless networks (NGWN). They study the stability of the nodes which extend the WiMAX cell, in particular, gateway nodes. The gateway nodes have two interfaces: IEEE802.11 and IEEE802.16. They develop a CLD mathematical model for the throughput and delay and assume stable conditions in the queues. It considers layers 2 and 3. In the latter, two queues are employed: forwarding and high layer traffic queues. The Weighted Fair Queueing (WFQ) [19] is used as the scheduling mechanism under an assumption of saturation. From the model, they conclude that WiMAX parameters do not impact the performance in terms of throughput of pure ad hoc nodes and vice versa. We believe that the approach followed by the authors fits well for modeling the integration of different technologies allowing combining strengths and making up individual limitations. The solution from [18] has been applied in several works as follows: (i) for studying stability-throughput tradeoff in wireless ad hoc networks in [20]; (ii) for performance analysis of delay, throughput, and energy consumption using a comprehensive analytical model of the IEEE802.11 [21]; and (iii) for end to end delay performance analysis in wireless ad hoc networks under CLD as presented in [22]. In this paper, the work from [18] is referred to as the reference model.

Yang et al. in [23] consider an 802.11 wireless local area network (WLAN) which shares a common set of multiradio devices with another network named CO-NETWORK which uses WiMAX. They assume saturated network conditions for all WLAN radios. They study how the throughput of a WLAN can be affected by scheduling the CO-NETWORK. Based on teletraffic modeling, they show that this issue can be minimized using proposed scheduling optimization criteria for the CO-NETWORK.

A convergence-bridge is proposed in [24]. It unifies the WiFi/WiMAX Frequency Bands. In other words, by modifying the WiFi Orthogonal frequency-division multiplexing (OFDM) PHY layer, WiFi devices are enabled to join the WiMAX-OFDM wireless network. The convergence-bridge is a thin layer in the WiFi OFDM PHY layer with 64 carriers. The WiMAX OFDM is fixed with 256 carriers. The main proposal for the convergence-bridge is to use multiple carriers which fit both technologies.

Different aspects of the interconnection process for both homogeneous and heterogeneous networks have been studied in the literature. There are many challenges to overcome before there is widespread adoption of heterogeneous technologies of this kind into MAN scenarios [25] (e.g., embedded system and devices). OPL and BER for WiFi/WiMAX still require further analysis. This is the focus of this work in terms of teletraffic modeling.

Table 1: Three PHY layers specified by the IEEE802.11 standard [6].

PHY	Slot time	CW_{min}	CW_{maxim}
Frequency hopping spread spectrum (FHSS)	50 μs	16	1024
Direct sequence spread spectrum (DSSS)	20 μs	32	1024
Infrared (IR)	8 μs	64	1024

3. IEEE802.16 and IEEE802.16 Protocols Theoretical Background

3.1. Overview of IEEE802.16

MAC Layer. We focused only on *Time Division Duplexing* (TDD) which is divided into two transmission periods: downlink (DL) and uplink (UP). The DL is generally broadcast. TDD handles a duplex scheme, where DL and UP transmissions occur in different times but share the same frequency. The maximum transition time (round trip time) between transmitter and receiver is 2 μs. The TDD is built from the base station (BS) and SS transmissions [7].

PHY Layer. The physical layer is based on wireless MAN-OFDM interface according to the standard IEEE802.16-2004 [7]. This interface uses 256 subcarriers of which 192 are data subcarriers, 8 are pilot subcarriers, and 56 are null. The pilot subcarriers are used to minimize frequency and phase shift. The 56 null carriers are used for guard bands and DC frequencies.

3.2. Overview of IEEE802.11

MAC Layer. The DCF is employed in this research. The DCF is the fundamental mechanism to access the medium based on carrier sense multiple accesses with collision avoidance (CSMA/CA). The DCF employs a binary exponential back-off scheme. When a station wants to transmit a new packet, it monitors the channel activity. If the channel is idle for a period equal to the distributed interframe space (DIFS), the station transmits the packet. On the other hand, if the channel is busy (either during or immediately after the DIFS), the station continues to monitor the channel until it is sensed idle for the DIFS.

The station generates a random back-off interval before it transmits the packet. After an idle DIFS, a time slot is available and a station is allowed to transmit only at the start of each time. The time slot depends on the PHY layer (see Table 1). The back-off time is chosen in the interval 0 to W-1 in each packet transmission. The value W represents the contention window (CW), that is, the amount of time available for the slots [26]. In the first attempt, the W is equal to CW_{min} (minimum CW); after each unsuccessful transmission, the W is doubled subject to a maximum of CW_{maxim} (maximum CW). $CW_{maxim} = 2^{max}CW_{min}$, max is the maximum backoff stage. The values of CW_{min} and CW_{maxim} are shown in Table 1. The back-off time counter

decreases when the channel is sensed as being idle but stops when there is a transmission in the channel.

The attempt rate is defined in [10] as the probability that a station transmits in a randomly chosen slot time.

PHY Layer. The PHY layer employed in this research is the IEEE802.11g protocol. This protocol was finalized until June 2003; 802.11g is a relative late-comer to the wireless marketplace. Despite the late start, 802.11g is now the de facto standard wireless networking protocol. This standard is used on most laptops and handheld devices. The 802.11g protocol uses the same industrial, scientific and medical (ISM) frequency range as the 802.11b protocol.

This physical layer is based on DSSS according to the IEEE Standard 802.11 [6]. This PHY operates in the 2.4 GHz ISM band and at a maximum raw data rate of 54 Mbit/s (with usable throughput of about 22 Mbps). Also, this physical layer can consider OFDM modulation. This makes it incompatible with 802.11b, and the higher frequency means shorter range compared to 802.11b/g at the same power.

The frequency range is 2.400–2.495 GHz, which is used by the 802.11b and 802.11g radio standards (corresponding to wavelengths of about 12.5 cm). A single 802.11g link may use 54 Mbps radios, but it will only provide up to 22 Mbps of actual throughput. The remaining bandwidth is the overhead that the radios need in order to coordinate their signals using the 802.11g protocol.

Since the 802.11g wireless equipment is half duplex (i.e., it only transmits or receives, never both at once), the required throughput must be doubled accordingly, for a total of 10 Mbps. The wireless links must provide that capacity every second, or conversations will lag.

4. The HMAN Performance Analysis Framework (PAF)

In this section, we propose the PAF which addresses the transmission performance in an HMAN and is composed by (i) the formalization of the end to end communication process of an HMAN using PA and (ii) teletraffic performance models. In (i), PA is used to formally define the communication between the homogeneous (single protocol domain) and the heterogeneous (multiprotocol domain) network sections whilst, in (ii), several teletraffic performance models are defined and represent the network's behavior across the transmission path. The PAF is a general framework for HMAN which will be deeply described on the following subsections based a case study for IEEE802.11 and IEEE802.16 heterogeneous networks. The PAF is depicted in Figure 1.

4.1. The Formalization of the HMAN Model Description

General Strategy. The end to end communication process of an HMAN can be modeled by PA. This formalism represents a mathematically rigorous framework for modelling system processes.

We define a HMAN as a *septuple* $\Phi = \{S, T, \rightarrow, s, d, i, N(i)\}$ where $S = \{S_{\text{dom}\,1}, S_{\text{dom}\,2}, S_g\}$ is a finite set whose elements are the total number of nodes: $S_{\text{dom}\,1}$ and $S_{\text{dom}\,2}$ are a finite set whose elements can be any HMAN protocol and are defined as $S_{\text{dom}\,1} = \{N_1, N_2, \ldots, N_x\}$, $x \in \mathbb{N}$ and $S_{\text{dom}\,2} = \{N_1, N_2, \ldots, N_y\}$, $y \in \mathbb{N}$, respectively. S_g is a finite set whose elements are gateway nodes which have two interfaces and is defined as $S_g = S_{\text{dom}\,1} \cap S_{\text{dom}\,2} = \{N_1, N_2, \ldots, N_z\}$, $z \in \mathbb{N}$, s is the traffic source which generates the packets. In other words, a node can have any of the following three roles: (1) source (transmitter), (2) destination (receiver), and (3) intermediate node which could also be a gateway.

$T = \{\alpha_k, \overline{\alpha}\}$ is the transition label set: the packet, which is sent from source to destination is labeled as $\overline{\alpha}$ while in the opposite direction it is labeled as α_{pk}. The transition relation is represented by $\rightarrow$. The destination is symbolized by d whilst i is an intermediate node on path $R_{s,d}$ and $N(i)$ is a finite set whose elements are the neighbors of node i. $|S| = n$ is the total number of nodes. Each node has two queues: the $F\langle e_1, \ldots, e_{\text{bumax}}\rangle$ forwarding queue which carries the packets from other nodes to their respective destinations and the $Q\langle e_1, \ldots, e_{\text{bumax}}\rangle$ queue that manages the local node packets. The sequence of packets in the buffer is represented by $\langle e_1, \ldots, e_{\text{bumax}}\rangle$ and bumax symbolizes its maximum size. The local buffer can have any of the following three states: (1) empty: $Q_\psi\langle\varepsilon\rangle$, (2) full: $Q_\psi\langle e_1, \ldots, e_{\text{bumax}}\rangle$, and (3) $Q_\psi\langle e_1, \ldots, e_k\rangle$, where $(0 < k < \text{bumax})$.

Figure 2 shows the process graph that models the end to end communication flows in an HMAN. The reception and transmission flows set are represented by $(\alpha_1, \ldots, \alpha_\psi)$ and $(\overline{\alpha}_1, \ldots, \overline{\alpha}_\psi)$ respectively. A nonsourcenode is symbolized as N_ψ, where $1 \le \psi < n$. A source node is represented as N. The HMAN end to end communication process (CP) is defined as follows:

$$\text{CP}_{e2e} \stackrel{\text{def}}{=} \sum \left\{ N \cdot N_\psi : 1 \le \psi < n \right\}. \tag{1}$$

All plausible processes in the network can be represented using the derivation tree graph from Figure 2.

Any communication process $N \xrightarrow{\overline{\alpha}_1\alpha_1} \cdots \xrightarrow{\overline{\alpha}_\psi\alpha_\psi} N_\psi$ where $(\overline{\alpha}_1 \cdot \alpha_1, \ldots, \overline{\alpha}_\psi \cdot \alpha_\psi, N_\psi)$ is a derivation of N and $\overline{\alpha}_1 \cdot \alpha_1, \ldots, \overline{\alpha}_\psi \cdot \alpha_\psi$ is a communication-sequence of, so that N_ψ is a $\overline{\alpha}_1 \cdot \alpha_1, \ldots, \overline{\alpha}_\psi \cdot \alpha_\psi$-derivative of N.

When a packet is transmitted, the source awaits an acknowledgment from the receiver. This acknowledgement packet is symbolized as $\overline{\xi}$ when it leaves the receiver and as ξ when it reaches the source.

The transmission process from the source is defined as

$$N \stackrel{\text{def}}{=} \sum_{pk \in \text{PK}} \overline{\alpha}_1 \cdot \xi_1 \cdot Q\left\langle e_1, \ldots, e_{k-1}\right\rangle, \tag{2}$$

where PK is the packet set to be sent in the end to end communication process and $0 < k \le \text{bumax}$.

The receiving process at the destination node is defined as

$$N_\psi \stackrel{\text{def}}{=} \alpha_\psi\left(pk\right) \cdot \overline{\xi}_\psi, \tag{3}$$

where $\psi = n - 1$.

Figure 1: The PAF.

Figure 2: Process Graph.

The communication process for any intermediate node is defined as

$$\begin{aligned} N_\psi \stackrel{\text{def}}{=} {} & \alpha_\psi(pk) \cdot \overline{\xi}_\psi \cdot F_\psi \langle e_1, \ldots, e_k \rangle \cdot N_\psi \\ & + \overline{\alpha}_{\psi+1} \cdot \xi_{\psi+1} \cdot F_\psi \langle e_1, \ldots, e_{k-1} \rangle \cdot N_{\psi+1} \\ & + \overline{\alpha}_{\psi+1} \cdot \xi_{\psi+1} \cdot Q_\psi \langle e_1, \ldots, e_{k-1} \rangle \cdot N_{\psi+1}, \end{aligned} \tag{4}$$

where $1 \leq \psi < n$ and $0 < k \leq \text{bumax}$.

The bridging process for a gateway node is defined as

$$\begin{aligned} N_\psi \stackrel{\text{def}}{=} {} & \alpha_\psi(pk) \cdot \overline{\xi}_\psi \cdot F_\psi \langle e_1, \ldots, e_k \rangle \cdot N_\psi \\ & + \overline{\alpha}_{\psi+1} \cdot \xi_{\psi+1} \cdot F_\psi \langle e_1, \ldots, e_{k-1} \rangle \cdot N_{\psi+1}, \end{aligned} \tag{5}$$

where $1 \leq \psi < n$ and $0 < k \leq \text{bumax}$.

Aggregation occurs at bridge node when two or more source packets are embedded into a single forwarding packet. This depends on the packet size source protocol and the forwarding payload size protocol. This aggregation process is defined as follows:

$$\begin{aligned} & F_\psi \langle e_1, \ldots, e_k \rangle \\ & \quad \stackrel{\text{def}}{=} \sum_{i \in K} F_\psi \langle e_1, \ldots, e_k \rangle \cdot \overline{\alpha}_{\psi+1}(e_1, \ldots, e_k) \cdot \xi_{\psi+1} \cdot N_{\psi+1}, \end{aligned} \tag{6}$$

where $(0 < k < \text{bumax})$ and it represents the number of embedded source packets that can fit into the forwarding payload.

The defined PA defines all the processes and entities involved in any HMAN. A second aspect, which is addressed by the PAF. is the network behavior. This is discussed next.

4.2. Network Behavior Modeling

Methodology. Teletraffic theory is considered as a tool to model and analyze the HMAN behavior. We propose white box approach modeling methodology.

4.3. Case Study: The 802.11 and 802.16 HMAN. We model the end to end communication process for the HMAN study case by PA as follows:

$$\begin{aligned} \mathrm{CP}_{e2e} \stackrel{\text{def}}{=} \sum \{ & \overline{\alpha}_1 \cdot \xi_1 \cdot Q\langle e_1, \ldots, e_{k-1}\rangle \cdot \alpha_\psi(pk) \\ & \cdot \overline{\xi}_\psi \cdot F_\psi\langle e_1, \ldots, e_k\rangle \cdot \overline{\alpha}_{\psi+1} \cdot \xi_{\psi+1} \\ & \cdot F_\psi\langle e_1, \ldots, e_{k-1}\rangle \cdot \alpha_\psi(pk) \cdot \overline{\xi}_\psi \\ & \cdot \sum_{i\in K} F_\psi\langle e_1, \ldots, e_k\rangle \cdot \overline{\alpha}_{\psi+1}(e_1, \ldots, e_k) \\ & \cdot \xi_{\psi+1} \cdot \alpha_\psi(pk) \cdot \overline{\xi}_\psi \}. \end{aligned} \tag{7}$$

The previous equation (7) is derived from (1), (2), (3), (4), (5), and (6).

The teletraffic performance models are derived from an HMAN based on the IEEE802.11 and IEEE802.16 standards considered as $S_{\mathrm{dom}\,1}$ and $S_{\mathrm{dom}\,2}$, respectively. The PAF teletraffic performance models for the case study are related to (i) Bianchi's performance model for IEEE802.11 DCF [10]; (ii) Lin and Wong's analytical model, which represents the performance under unidirectional and bidirectional data transfer 802.11 [12]; (iii) Fakhri et al.'s mathematical model, which studies throughput optimization for OFDM modulation in a 802.16 network [27]; and (iv) Ci and Sharif's model for an adaptive optimal frame length predictor for IEEE802.11 [28].

We extend the reference model with the inclusion of the following protocol operational parameters (metrics): bit error rate (BER), packet error ratio (PER), and packet length (pl). Thus, our HMAN model considers BER, PER, and pl; therefore we get end to end throughput and delay under error-prone channel conditions.

The CLD from [18] considers both network and MAC parameters. The network layer handles two queues scheduled using a WFQ scheme [19]; we modify the CLD model from [18]. We now conceive that the WFQ is between the network layer and the MAC Layer. This is done in order to reduce network bandwidth usage. Each node has the same network layer and WFQ. This design permits the exchange of communication and information between layers and allows greater flexibility. The queue controls two queues: *high layer packets* (hlp) and *forward packets* (fp) which have an infinite capacity. The fp is the forwarding queue which carries the packets from other nodes to their destinations and the hlp which contains packets generated by node i itself. Each queue has its own transmitted probability: fp_i is the probability to transmit from fp whereas $1-\mathrm{fp}_i$ is the probability to transmit from hlp. The HMAN is considered a *saturated system,* which means that each node always transmits packets from hlp while fp could be empty. The CLD for HMAN is shown in Figure 3.

4.3.1. Mathematical Model for IEEE802.16. The mathematical model for IEEE802.16 is based on Fakhri et al.'s model [27]. This model is focused on the *optimization of throughput*, BER, and OPL in a wireless system for OFDM modulation.

Figure 3: CLD for HMAN Model.

There are some assumptions made when developing this mathematical model. The transmitter sends packets of $L_{\mathrm{dom}\,2}$ bits in a continuous stream and the transmitter attaches a C bit as the CRC. The throughput is defined as the number of payload bits per second received correctly [27] (8):

$$T_{\mathrm{dom}\,2} = \sum_{l=1}^{N} \frac{P_{\mathrm{dom}\,2\,\mathrm{load}}}{L_{\mathrm{dom}\,2}} R_l f(\gamma_l). \tag{8}$$

$P_{\mathrm{dom}\,2\,\mathrm{load}} = L_{\mathrm{dom}\,2} - O_{\mathrm{bytes}}$, $L_{\mathrm{dom}\,2}$ is the total pl (bits), $O_{\mathrm{bytes}} = H_{\mathrm{MAC}} + S_{\mathrm{FSH}} + C$, H_{MAC} is the average MAC header size, S_{FSH} is the fragmentation subheader size, C is the CRC bit, R_l is the symbol rate assigned to the subcarriers l in bits per second, $f(\gamma_l)$ is the packet success rate (PSR) per user with m-Quadrature Amplitude Modulation (QAM) scheme and γ_l is the SNR in dB given by (9)

$$\gamma_l = \frac{P_l}{N_0 * R_l}, \tag{9}$$

where P_l is the received power in watts, N_0 is the one-sided noise power spectral density in watts/Hz.

A symbol error in the packet automatically results in a packet loss, and the PSR is given in terms of symbol error rate (SER) P_e by

$$f(\gamma_l) = (1 - P_e(\gamma))^{L_{\mathrm{dom}\,2}/b}, \tag{10}$$

where b is the number of bits per m-QAM symbol. In (11), the P_e of m-QAM in and additive white Gaussian noise (AWGN) channel is (approximately) given by [29]

$$P_e(\gamma) = 4\left(1 - \frac{1}{2^{b/2}}\right) Q(x), \tag{11}$$

where $x = \sqrt{(3/(2^b - 1))\gamma}$ and the $Q(x)$ function is defined as

$$Q(x) = \frac{e^{-x^2/2}}{4\mathrm{mod}} + \frac{1}{2\mathrm{mod}} \sum_{j=1}^{\mathrm{mod}-1} \exp\left(\frac{-x^2}{2\sin^2\theta_j}\right), \tag{12}$$

where $\theta_j = j\pi/2\mathrm{mod}$ and mod is the modulation type.

4.3.2. Mathematical Model for IEEE802.11. In our research, we consider the PER which is determined from the BER. The BER is defined as the number of bit errors divided by the total number of bits transferred in a time interval and the pl [12]. The PER is denoted by p_e whilst the BER is P_{BER}. The PER is defined as

$$p_e = 1 - (1 - P_{\mathrm{BER}})^{L_a}, \tag{13}$$

where $L_{\text{dom}\,1}$ is WiFi pl in bits, which includes the PHY layer header (PHY_{H}), the MAC layer header (MAC_{H}), and the packet payload. Let $\tau^{\text{dom}\,1}$ be the duration of WiFi slot (sec). The payload information (bits per second) is defined in

$$P_{\text{dom}\,1\,\text{load}} = \frac{L_{\text{dom}\,1} - H_{\text{total}}}{\tau^{\text{dom}\,1}}, \tag{14}$$

where

$$H_{\text{total}} = \left(\text{PHY}_{\text{H}} + \text{MAC}_{\text{H}}\right). \tag{15}$$

The PHY layer header and MAC layer header are defined in [6].

4.3.3. HMAN End to End Throughput Model. The expression for throughput in [18] is

$$\begin{aligned}\text{thp}_{s,B} = {} & \left(y_s(1-\varphi)^{n(|i,B|+1)}\right) \\ & \times \left(\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} \pi_{i,s,d} f(\varphi_i)^{-1} \right. \\ & + \pi_{g,s,d} f\left[\tau_{i,B}^m \frac{L_{\text{dom}\,1}}{L_{\text{dom}\,2}}\right] \phi(\vec{\gamma}_i) \\ & \left. + y_s P_{s,d}\left[\frac{1}{\varphi_s}\right]\right)^{-1},\end{aligned} \tag{16}$$

where $y_s = 1 - \pi_s f_s$ is a value of a N-dimensional row vector which contains stability values per node. φ is the transmission probability as defined in [10], $N(i)$ is the number of WiFi neighbors, and $|(i,B)|$ represents the number of intermediate nodes s and B which represents the base station. The throughput is defined as the payload (bits) per second received successfully and is measured in packets per second. In (16), the dividend is the average service time per packet at node i. Within the dividend, π_i represents the probability that F_i has at least one packet to be forwarded in the beginning of the start of each cycle (in [18] a cycle is referred to as total number of slots to transmit one packet until it's successful or dropped), $\pi_{i,s,d}$ is the probability that F_i has a packet ready in the first position of the queue to be forwarded to path $R_{s,d}$ in the beginning of each *cycle*, $\phi(\vec{\gamma}_i) = (1-e^{-\gamma_i})^{L_{\text{dom}\,2}/b}$ is the function of PSR, $\tau_{i,B}^m = L_{\text{dom}\,2}/\rho_{i,B}^m$ is the WiMAX packet transmission time (sec), $\rho_{i,B}^m = \sum_{l\in\mathscr{L}_i} \gamma_{i,B,l}^m \Delta_f$ is the aggregation transmission rate (bps), when nodes use an m-QAM modulation level, $\gamma_{i,B,l}^m \Delta_f$ is the transmit rate (bits per subcarrier), Δ_f is the bandwidth of one single subcarrier, K is maximum number of transmissions allowed by a gateway i per packet for all paths, and $P_{s,d}$ is the probability that a node s generates and transmits a packet to node d. Some assumptions are considered from [18] as follws: (i) in the uplink transmission, all nodes have the same destiny; thus, $P_{s,B} = 1$; (ii) the heterogeneous network is a symmetric mesh system; hence; each node has the same number of WiFi neighbors; and (iii) the forwarding probability is $f_i \equiv f$ and $\varphi_i \equiv \varphi$.

The throughput of the HMAN model is based on [12, 27, 28]. However, they address only homogeneous networks. In our proposed model (see (17)), we introduce the PER as dividend and consider the heterogeneity of the network (802.11 and 802.16). The end to end throughput under error-prone channel conditions is then estimated as follows:

$$\begin{aligned}\text{thpHMAN}_{s,B} & \\ = {} & \left(Y_s(1-\varphi)^{n(|i,B|+1)}\right) \\ & \times \left(\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} \pi_{i,s,d} f(\varphi_i)^{-1} + \pi_{g,s,d} f p_i \right. \\ & \times \left[\tau_{i,B}^m \frac{L_{\text{dom}\,1}}{(1-p_e) L_{\text{dom}\,2}}\right] f(\gamma_l) \\ & \left. + Y_s\left[\frac{1}{\varphi_s}\right]\right)^{-1},\end{aligned} \tag{17}$$

where $f(\gamma_l) = (1 - P_e(\gamma))^{L_{\text{dom}\,2}/b}$ (13) and p_e is taken from (13).

Now, using (14), the throughput of the HMAN model is rewritten as follows:

$$\begin{aligned}\text{thpHMAN}_{s,B} & \\ = {} & \left(Y_s(1-\varphi)^{n(|i,B|+1)}\right) \\ & \times \left(\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} \pi_{i,s,d} f(\varphi_i)^{-1} + \pi_{g,s,d} f p_i \right. \\ & \times \left[\sum_{j=1}^{N} \frac{P_{\text{dom}\,2\,\text{load}}}{L_{\text{dom}\,2}} \frac{P_{\text{dom}\,1\,\text{load}}}{\rho_{i,B}^m (1-p_e)}\right] \\ & \left. \times f(\gamma_l) + Y_s\left[\frac{1}{\varphi_s}\right]\right)^{-1}.\end{aligned} \tag{18}$$

4.3.4. HMAN End to End Throughput Optimization. We employed a variable change in the throughput equation (18) in order to differentiate this equation with respect to packet length $v, h, z(L_{\text{dom}\,1}, L_{\text{dom}\,2})$, and u. The $\text{thpHMAN}_{s,B}$ is then defined as follows:

$$\begin{aligned}\text{thpHMAN}_{s,B} & \\ = {} & \frac{v}{\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} h + z(L_{\text{dom}\,1}, L_{\text{dom}\,2}) + u},\end{aligned} \tag{19}$$

where $v = Y_s(1-\varphi)^n$, $h = \pi_{i,s,d} f(\varphi_i)^{-1}$, $z(L_{\text{dom}\,1}, L_{\text{dom}\,2}) = \pi_{g,s,d} f p_i T\lceil P_{\text{dom}\,1\,\text{load}}/\rho_{g,B}^m(1-p_e)\rceil$, $u = Y_s[1/\varphi_s]$, and $T = \sum_{j=1}^{N}(P_{\text{dom}\,2\,\text{load}}/L_{\text{dom}\,2}) f(\gamma_j)$.

4.3.5. Optimal WiMAX Packet Length. We get the optimal WiMAX pl $L_{\text{dom}\,2}$ by differentiating (19) with respect to $L_{\text{dom}\,2}$ and using (8), (9), and (10) produces

$$\frac{d\text{thpHMAN}_{s,B}}{dL_{\text{dom}\,2}} = -\frac{v\left[z'\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right)\right]}{\left[\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} h + z\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right) + u\right]^2}, \tag{20}$$

where

$$z\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right) = \pi_{g,s,d} fp_i \sum_{j=1}^{N} \frac{L_{\text{dom}\,2} - O_{\text{bytes}}}{L_{\text{dom}\,2}} \times \left(1 - P_e\left(\gamma_j\right)\right)^{L_{\text{dom}\,2}/b} \left[\frac{P_{\text{dom}\,1\,\text{load}}}{\rho_{g,B}^m\left(1 - p_e\right)}\right]. \tag{21}$$

The derivative of $z(L_{\text{dom}\,1}, L_{\text{dom}\,2})$ is calculated with respect to $L_{\text{dom}\,2}$ as

$$\frac{dz\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right)}{dL_{\text{dom}\,2}} = \pi_{g,s,d} fp_i \left[\frac{P_{\text{dom}\,1\,\text{load}}}{\rho_{g,B}^m\left(1 - p_e\right)}\right] \times \left[\frac{O_{\text{bytes}}}{L_{\text{dom}\,2}^2} f\left(\gamma_j\right) + \left(1 - \frac{O_{\text{bytes}}}{L_{\text{dom}\,2}}\right) \times \frac{f\left(\gamma_j\right)\ln\left(1 - P_e\left(\gamma_j\right)\right)}{b}\right]. \tag{22}$$

Setting this to zero produces an equation in $L_{\text{dom}\,2}$:

$$-\frac{v\left[z'\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right)\right]}{\left[\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} h + z\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right)\right]^2} = 0,$$

$$-\left(v\left[\pi_{g,s,d} fp_i \left[\frac{P_{\text{dom}\,1\,\text{load}}}{\rho_{g,B}^m\left(1 - p_e\right)}\right] \times \left[\frac{O_{\text{bytes}}}{L_{\text{dom}\,2}^2} f\left(\gamma_j\right) + \left(1 - \frac{O_{\text{bytes}}}{L_{\text{dom}\,2}}\right) \times \frac{f\left(\gamma_j\right)\ln\left(1 - P_e\left(\gamma_j\right)\right)}{b}\right]\right]\right) \times \left(\left[\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} h + z\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right)\right]^2\right)^{-1} = 0,$$

$$v\left[\pi_{g,s,d} fp_i \left[\frac{P_{\text{dom}\,1\,\text{load}}}{\rho_{g,B}^m\left(1 - p_e\right)}\right] \times \left[\frac{O_{\text{bytes}}}{L_{\text{dom}\,2}^2} f\left(\gamma_j\right) + \left(1 - \frac{O_{\text{bytes}}}{L_{\text{dom}\,2}}\right) \times \frac{f\left(\gamma_j\right)\ln\left(1 - P_e\left(\gamma_j\right)\right)}{b}\right]\right] = 0,$$

$$\frac{O_{\text{bytes}}}{L_{\text{dom}\,2}^2} f\left(\gamma_j\right) + \left(1 - \frac{O_{\text{bytes}}}{L_{\text{dom}\,2}}\right) \frac{f\left(\gamma_j\right)\ln\left(1 - P_e\left(\gamma_j\right)\right)}{b} = 0. \tag{23}$$

We adopt the notation $L_{\text{dom}\,2} = L_{\text{dom}\,2}^*$ for the optimal WiMAX pl that satisfies (23) then solving for $L_{\text{dom}\,2}$:

$$L_{\text{dom}\,2}^* = \frac{O_{\text{bytes}}}{2} + \frac{\sqrt{{O_{\text{bytes}}}^2 - \left(4bO_{\text{bytes}}/\ln\left(1 - P_e\left(\gamma\right)\right)\right)}}{2}. \tag{24}$$

Thus, in a WiMAX system, the OPL $L_{\text{dom}\,2}$ depends on the SNR per symbol γ, symbol error probability P_e, and the constellation size 2^b.

4.3.6. Optimal Ad Hoc Packet Length. We differentiate (18) with $L_{\text{dom}\,1}$ (using (13) and (14)) and set it to zero to obtain the following condition:

$$\frac{d\text{thpHMAN}_{s,B}}{dL_{\text{dom}\,1}} = -\left(v\left(\pi_{g,s,d} fp_i \frac{T}{\tau^{\text{dom}\,1}\rho_{g,B}^m}\right)\left(1 - P_{\text{BER}}\right)^{-L_{\text{dom}\,1}} \times \left[1 - \ln\left(1 - P_{\text{BER}}\right)\left(L_{\text{dom}\,1} - H_{\text{total}}\right)\right]\right) \times \left(\left[\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} h + z\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right) + u\right]^2\right)^{-1}. \tag{25}$$

Next we set the derivative to zero:

$$\left(v\left(\pi_{g,s,d} fp_i \frac{T}{\tau^{\text{dom}\,1}\rho_{g,B}^m}\right)\left(1 - P_{\text{BER}}\right)^{-L_{\text{dom}\,1}} \times \left[1 - \ln\left(1 - P_{\text{BER}}\right)\left(L_{\text{dom}\,1} - H_{\text{total}}\right)\right]\right) \times \left(\left[\sum_{s,d:i\in R_{s,d}\backslash i\in S_g} h + z\left(L_{\text{dom}\,1}, L_{\text{dom}\,2}\right) + u\right]^2\right)^{-1} = 0,$$

Figure 4: PyViz illustration on NS3.

$$v\left(\pi_{g,s,d} fp_i \frac{T}{\tau^{\text{dom}\,1}\rho^m_{g,B}}\right)(1-P_{\text{BER}})^{-L_{\text{dom}\,1}} \times \left[1-\ln(1-P_{\text{BER}})(L_{\text{dom}\,1}-H_{\text{total}})\right]=0,$$
$$1-\ln(1-P_{\text{BER}})(L_{\text{dom}\,1}-H_{\text{total}})=0. \tag{26}$$

We adopt the notation $L_{\text{dom}\,1} = L^*_{\text{dom}\,1}$ for the optimal WiMAX pl that satisfies (26) then solving for $L_{\text{dom}\,1}$:

$$L^*_{\text{dom}\,1} = H_{\text{total}} + \frac{1}{\left|\ln(1-P_{\text{BER}})\right|}. \tag{27}$$

Therefore, in an WiFi system, the OPL $L_{\text{dom}\,1}$ depends on the BER P_{BER}.

4.3.7. End to End Delay. The mean end to end delay $D_{s,d}$ of a packet on the path $R_{s,d}$ is the mean time taken from the instant that a packet reaches the MAC layer of the source to the time that is received in seconds. That delay time is for both successful and dropped packets. The expression for delay is the same as in [18]:

$$D_{s,d} = \frac{\Psi^{\text{succ}}_{s,s,d}}{\varphi_s} + \sum_{i=1}^{|R_{s,d}|}\left(\overline{W}^{pt}_i + \tau^{\text{succ}}_{i,s,d}\right), \tag{28}$$

where $\overline{W}^{pt}_i = (\overline{R}^{pt}_i + \tau^Q_i((1-f_i)/f_i))/(1-\text{thp}_{s,B}(\tau^F_i - \tau^Q_i((1-f_i)/f_i)))$ is the average waiting time in the forwarding queue F_i of a pt (WiMAX or WiFi protocols) arrival packet at node i, $\tau^F_i = \sum_{s,d}(\pi_{i,s,d}/\pi_i)T_{i,s,d}$ represents the mean service time of F_i, $\tau^Q_i = \sum_d \varphi_i T_{i,i,d}$ is the average service time of Q_i, and the mean residual time of a packet for a (s,d) connection is $\overline{R}^{pt}_i = \sum_{s,d}\pi_{i,s,d} f_i \overline{R}^{pt}_{i,s,d} + \sum_d P_{i,d}(1-\pi_i f_i)\overline{R}^{pt}_{i,i,d}$, where

$$\overline{R}^{pt}_{i,s,d} = \begin{cases} \dfrac{T^{(2)}_{i,s,B}}{2T_{i,s,B}} - \dfrac{1}{2}, & \text{if } i \in S_g \text{ and } d = B \\ \dfrac{T^{(2)}_{i,s,d}}{2T_{i,s,d}} + \dfrac{1}{2}, & \text{otherwise.} \end{cases} \tag{29}$$

The second moment of $T^{(2)}_{i,s,B}$ service time is given by

$$T^{(2)}_{i,s,B} = \begin{cases} \Psi^{(2)}_{i,s,d}\left[\dfrac{\tau^m_{i,B}}{\tau^a}\right]^2 & \text{if } i \in S_g \text{ and } d = B \\ \dfrac{\Psi^{(2)}_{i,s,d} + \Psi_{i,s,d}(1-\varphi_i)}{\varphi_i^2} & \text{otherwise,} \end{cases} \tag{30}$$

as $\tau^{\text{succ}}_{i,s,d}$ is the mean service time of a successfully transmitted packet on the same path $R_{s,d}$. pt is used for WiFi or WiMAX, $\tau^{\text{succ}}_{i,s,d}$ which has the same form as $\tau_{i,s,d}$ can be expressed as follows:

$$\tau^{\text{succ}}_{i,s,d} = \frac{\Psi^{\text{succ}}_{i,s,d}}{\varphi_i}, \tag{31}$$

where $\Psi^{\text{succ}}_{i,s,d} = \sum_{k=1}^{k_{i,s,d}} k(1-P_{i,s,d})^{k-1}P_{i,s,d}$ is the average number of attempts until it reaches a successful point.

The delay of the HMAN Model is derived using (18) as follows:

$$D\text{HMAN}_{s,d} = \frac{\Psi^{\text{succ}}_{s,s,d}}{\varphi_s} + \sum_{i=1}^{|R_{s,d}|}\left(\overline{W}\text{HMAN}^{pt}_i + \tau^{\text{succ}}_{i,s,d}\right). \tag{32}$$

Based on $\overline{W}^{pt}_i$, the HMAN average waiting time in the forwarding queue F_i is $\overline{W}\text{HMAN}^{pt}_i = (\overline{R}^{pt}_i + \tau^Q_i((1-fp_i)/f_{pi}))/(1-\text{thpHMAN}_{s,B}(\tau^F_i - \tau^Q_i((1-fp_i)/f_{pi})))$. The rest of the variables from (32) are defined above.

5. The Experimental Work

Both WiMAX and WiFi networks are used in the simulations. The objective is to evaluate the proposed HMAN model against the reference model [18]. The simulation experiments are described as follows.

The experimental work was carried out on the ns3 network simulator [30]. The simulation scenario, shown in Figure 4, is set for an M2M heterogeneous network of 9 SS of which 5 are WiFi nodes, 2 are gateways (multiple interfaces WiFi and WiMAX), and 2 are WiMAX nodes. There is a base station (BS) WiMAX and each node has an ID, from 1 to 9; node IDs are sorted as follows: 2 to 6 are the WiFi nodes, 8 and 9 are WiMAX nodes, and 1 and 7 are the gateways node (IEEE802.11 and IEEE802.16). The nodes are distributed based on Table 2. IEEE802.11 PHY uses Direct-Sequence Spread Spectrum (DSSS) [12]. IEEE802.11 MAC

Figure 5: HMAN topology.

Table 2: Nodes coordinates.

Node ID	1	2	3	4	5	6	7	8	9	B
X (m)	190	160	60	0	135	0	230	2300	1400	1000
Y (m)	10	80	60	80	0	30	60	10	60	30

was used as the MAC protocol. Some characteristics of the model were based on IEEE802.11 and IEEE802.16 standards. The simulation time was 500 s and the number of transmitted packets was 500 (based on the central limit theorem).

We consider a Constant Speed Propagation Delay Model and a Friss Propagation Loss Model which correspond well to our Model. The Friss propagation Loss Model considers a frequency of 5.5 GHz at 300 000 km/s. Optimized Link State Routing (OLSR) [31] was used for instantaneous updates for each routing table.

There are three data flows: a, b, and c, shown in Figure 5. Node 4 is considered as the source for all data flows. The destination nodes are: node 9 for flow a, node 8 for flow b, and node 9 for flow c. Nodes 9 and 8 are configured with Quadrature Phase Shift Keying (QPSK) modulation. We develop two scenarios in which both have the same simulation parameters (from Section 5.1). In scenario (1), we configured gateway 1 with one subcarrier and QPSK modulation (see Table 4) and gateway 7 with one subcarrier and 16-QAM (see Table 4) and the cross-traffic average for flow b at gateway 7 was 47.5% reception (Rx) and 52.5% transmission (Tx). In scenario (2), we configured gateway 1 with one subcarrier and 16-QAM modulation (see Table 4) and gateway 7 with one subcarrier and QPSK modulation (see Table 4); the cross-traffic average for flow b at gateway 7 was 16% Rx and 26.5% Tx. A total of 12 subscenarios were conducted each with different pl. The pl ranged from 100 to 1200 bytes with (increments) Δpl = 100 bytes. Figure 4 shows the NS3 Python Visualizer (PyViz) representation of the HMAN topology from Figure 5 (Cartesian plane).

The HMAN network topology is depicted in Figure 5.

5.1. Simulation Parameters. Some simulation parameters are summarized in Tables 3, 4, 5, and 6. The following parameters are used in both scenarios.

Table 4 shows the spectral efficiencies (rate) using IEEE802.16 adaptive coding and modulation (ACM) settings.

6. Simulation Results and Discussions

To validate the HMAN Model, we compare the obtained results with those obtained by the solution from [18]. We analyzed the following metrics: PSR, end to end throughput, end to end delay, BER, and OPL. The main goal for the analysis is to compare the HMAN performance against

Figure 6: (a) QPSK: PSR versus SNR in connection *a (scenario 1)*, *b (scenario 2)*, and *c (scenario 1)*, respectively. (b) 16-QAM: PSR versus SNR in connection *a (scenario 2)*, *b (scenario 1)*, and *c (scenario 2)*, respectively.

Table 3: Simulation parameters.

Parameter	Value
Simulator	NS-3-dev
Simulation length	500 s
Transmission start	0.6 s
PHY WiMAX layer	OFDM
PHY WiFi layer	DSSS
MAC WiFi layer	CSMA/CA
Code division multiplexing (CDMA) codes	256
$\tau^{\text{dom}\,2}$ and $\tau^{\text{dom}\,1}$	2 ms
Bandwidth	10 MHz
Automatic repeat reQuest (ARQ)	Selective Repeat

Table 4: ACM settings for IEEE802.16 [7].

Modulation order	Target SINR (db)	Coding order	Spectral efficiency (bits/symbol)
BPSK	6.4	1/2	0.5
QPSK	9.4	1/2	1
QPSK	11.2	3/4	1.5
16-QAM	16.4	1/2	2
16-QAM	18.2	3/4	3
64-QAM	22.3	2/3	4
64-QAM	24.4	3/4	4.5

Figure 7: PSR versus SER.

the solution from [18] and to verify that the HMAN model agrees with the NS3 simulation.

6.1. Packet Success Ratio (PSR). PSR was analyzed for ranged pl mentioned above in 12 subscenarios corresponding to scenarios 1 and 2. Figure 6(a) shows PSR versus SNR using QPSK for flows a, b, and c. Flows a and c employ the scenario 1 configuration whilst flow b uses the scenario 2. Figure 6(b) shows PSR versus SNR using 16-QAM modulation results for flows a, b, and c. Flows a and c employ the scenario 2 configuration whilst flow b uses the scenario 1 configuration.

Figure 8: End to end throughput versus pl (bytes) in connections (a), (b), and (c), respectively.

The SNR values are derived from the obtained PSR using $\phi(\vec{\gamma}_i) = (1 - e^{-\gamma_i})^{L/b}$ and solving it for γ_i (employing a single subcarrier). It is observed from Figure 6(a) that when the same modulation scheme (equal baud rate) is employed for both the source and destination nodes, the PSR is higher than the PSR using a different scheme as shown in Figure 6(b) (different baud rate). It is also observed that as the pl increases, the SNR is changed.

6.2. BER. The BER and SER values are obtained from (10), (11), and (12) using the PSR simulation results. Table 7 presents the average values for the 12 subscenarios corresponding to scenarios 1 and 2. We observed that when the same modulation scheme is employed for both WiFi and WiMAX domains the BER value is lower than the BER value using a different scheme.

Table 5: Attempt rate probability (for each node i).

$P1$	$P2$	$P3$	$P4$	$P5$	$P6$	$P7$	$P8$	$P9$
0.5	0.7	0.4	0.3	0.7	0.4	0	0	0

Figure 9: End to end delay (ms) versus pl (bytes) in connections (a), (b), and (c), respectively.

Table 6: IEEE802.16 MAC headers [7].

Header	Size
General mac header (GMH)	6 bytes
Grant manager subheader (GMSH)	2 bytes
Packing subheader (PSH)	3 bytes
Fragmentation subheader (FSH)	2 bytes
Cyclic redundancy check (CRC)	4 bytes

Table 7: BER values.

Scenario	Flow	SER	BER
1	a, c	$1.3e-5$	$6.5e-6$
1	b	$1.86e-4$	$4.7e-5$
2	a, c	$4e-4$	$9.5e-5$
2	b	$9.15e-5$	$4.6e-5$

6.3. Symbol Error Rate (SER). Two important metrics for an HMAN end to end performance analysis are the PSR and SER. In the case of the WiMAX domain, these metrics are derived from (10), (11), and (12). Figure 7 shows the PSR and SER relation under different modulation schemes. It is observed that as the number of bits increases per symbol for a given SER value, the PSR decreases significantly. Note that the corresponding BER and SER simulation result values (Table 7) are shown within the BER range of Figure 7.

6.4. End to End Throughput. End to end throughput was analyzed for both scenarios in an error-prone channel with different BER values from Table 7 for CSMA/CA with RTS/CTS. Figure 8 shows the throughput versus variable pl results for flows a and c using scenario 2, whilst flow b used scenario 1. As shown in Table 7 the chosen BER values are representative of the protocols under test and have been selected to evaluate the models under diverse network conditions. Reference model [18] and HMAN model results are obtained from (10) and (9), respectively.

From the above experimental results, we calculated the mean square error (MSE) for both models in each communication flow. We can observe in Figure 8 that the throughput obtained by the HMAN model is 11.46% more accurate in all the flows than the throughput obtained by the reference

Figure 10: OPL versus PER in connection: (a) and (b).

model. Also, the results of HMAN are closer to the results of NS3 as can be seen in Figure 8.

6.5. End to End Delay. End to end delay was analyzed in both scenarios. Figure 9 shows end to end delay versus variable pl in flows a, b, and c. From Figure 9, it is predictably observed that the delay increases as the pl is increased. The average absolute error (AAE) is obtained between the HMAN model and NS3 simulation results. The AAE is 4.14% for flow a, 4.39% for flow c, and 8.75% for flow c (100–1200 bytes). Flow b employs scenario 1 whilst flows a and c employ scenario 2. We consider an error-prone channel with the BER values of Table 7 in both scenarios. It is observed that the HMAN model and reference model are smooth; by contrast the NS3 simulation results are undulating. This is because NS3 is directly modeling the processing entities, for example, network interfaces. The HMAN model outperforms previous solutions in terms of delay and throughput relative to the NS3 results and is largely indifferent to pl. Reference model and the HMAN model results are obtained from (28) and (32), respectively.

From the above experimental results, we calculated the MSE for both models in each flow communication. The delay obtained by the HMAN model is 34.21% more accurate in all the flows than the delay obtained by the reference model. Also, the results of HMAN are closer to the results of NS3 as can be seen in Figure 9.

6.6. OPL. Efforts were made to find the OPL for both scenarios; there will be pl that maximizes the throughput in an error-prone channel. We used the PER obtained from both scenarios as shown in Table 7. In Figure 10(a), it is observed that for flow a from scenario 2 the OPL is 135 bytes for a PER of 0.1 and the OPL decreases as the PER is increased. Figure 10(b) shows flow b from scenario 1 where the OPL is 293 bytes for a PER of 0.01. Again the OPL decreases as the PER is increased. These results are obtained from (16) (a) for $L^*_{\text{dom}\,1}$ and 17 for $L^*_{\text{dom}\,2}$ (b).

7. Conclusions

We analyzed a heterogeneous network composed of a WiMAX cell and a WiFi network. The WiMAX protocol shares many characteristics with cellular networks such as architectural support for billing, mobility, and QoS. The main contribution in this paper is the evaluation of end to end throughput and delay in a HMAN by considering the effect of different layers within the CLD (layer 2 and layer 1 of OSI model). We extended previous models for such a scenario with the inclusion of the following protocol operational parameters (metrics): BER, PER, pl, and OPL. Further, numerical and simulation results were performed to validate our HMAN model. The HMAN model outperforms previous modeling solutions in terms of delay and throughput relative to the NS3 results and is largely indifferent to pl. In a WiMAX system, the OPL depends on the SNR per symbol, error probability, and the constellation size. In a WiFi system, the OPL depends on the BER. By using our HMAN expression, we can compute the OPL for a given set of network conditions to improve network adaptability; this could be applied dynamically. The HMAN can be further extended to consider other network metrics such as jitter and frame segmentation and other protocols such as CAN, Zigbee, and Bluetooth. The scenarios can also be applied to a number of different modulation schemes and node densities.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, 2011.

[2] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: architectures, standards, and QoS improvement," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, 2011.

[3] ETSI. Etsi ts 102 690 v1.1.1: Machine-to-machine communications (m2m); functional architecture, 2011.

[4] I. Bojic, G. Jezic, D. Katusic, S. Desic, M. Kusek, and D. Huljenic, "Communication in machine-to-machine environments," in *Proceedings of the 5th Balkan Conference in Informatics*, pp. 283–286, ACM, 2012.

[5] J. Kim, J. Lee, J. Kim, and J. Yun, "M2M service platforms: survey, issues, and enabling technologies," *IEEE Communications Surveys & Tutorials*, 2013.

[6] IEEE 802.11 Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Computer Society Sponsored by the LAN/MAN Standards Committee, IEEE Standard 802.11, 2007.

[7] IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Std 802.16, 2002.

[8] K. Chang, A. Soong, M. Tseng, and Z. Xiang, "Global wireless machine to machine standardization," *IEEE Journals & Magazines: Internet Computing*, vol. 15, no. 2, pp. 64–69, 2011.

[9] S. Aust, R. V. Prasad, G. M. M. Ignas, and N. Niemegeers, "IEEE 802.11ah: advantages in standards and further challenges for Sub 1 GHz Wi-Fi," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 6885–16889, 2012.

[10] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.

[11] K. Duffy, D. Malone, and D. J. Leith, "Modeling the 802.11 distributed coordination function in non-saturated conditions," *IEEE Communications Letters*, vol. 9, no. 8, pp. 715–717, 2005.

[12] Y. Lin and V. W. S. Wong, "Frame aggregation and optimal frame size adaptation for IEEE 802.11n WLANs," in *Proceedings of the Global Telecommunications Conference (IEEE GLOBECOM '06)*, pp. 1–6, San Francisco, Calif, USA, December 2006.

[13] E. Hwang, K. J. Kim, A. Lyakhov, and B. D. Choi, "Performance analysis of bandwidth requests under unicast, multicast and broadcast pollings in IEEE 802.16d/e," *Telecommunication Systems*, vol. 50, pp. 15–30, 2012.

[14] X. Tian, X. Chen, T. Ideguchi, and T. Okuda, "Improving protocol capacity by scheduling random access on WLANs," *Telecommunication Systems*, vol. 37, no. 1-3, pp. 19–28, 2008.

[15] F. Calì, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Transactions on Networking*, vol. 8, no. 6, pp. 785–799, 2000.

[16] Q. Liu, X. Wang, and G. B. Giannakis, "A cross-layer scheduling algorithm with QoS support in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, pp. 839–847, 2006.

[17] B.-J. Chang, C.-M. Chou, and Y.-H. Liang, "Markov chain analysis of uplink subframe in polling-based WiMAX networks," *Computer Communications*, vol. 31, no. 10, pp. 2381–2390, 2008.

[18] R. El-Azouzi, E. Sabir, S. K. Samanta, R. El-Khoury, and E.-H. Bouyakhf, "An end-to-end QoS framework for IEEE 802.16 and ad-hoc integrated networks," in *Proceedings of the 6th International Conference on Mobile Technology, Application and Systems (Mobility '09)*, ACM, September 2009.

[19] B. Partridge, *Gigabit Networking*, Addison-Wesley Publishing, 1994.

[20] A. Kherani, R. El-Khoury, R. El-Azouzi, and E. Altman, "Stability-throughput tradeoff and routing in multi-hop wireless ad hoc networks," *Computer Networks*, vol. 52, no. 7, pp. 1365–1389, 2008.

[21] E. Sabir, R. El-Azouzi, and El-Houssinebouyakhf, *Cross-Layered QoS Framework for Next Generation Wireless Networks*, Universitaires Europeennes (EUE), 2011.

[22] R. El-Azouzi, E. Sabir, S. K. Samanta, and R. El-Khoury, "Asymptotic delay analysis and timeout-based admission control for ad hoc wireless networks with asymmetric users," *Computer Communications*, vol. 33, no. 17, pp. 2057–2069, 2010.

[23] X. Yang, J. Zhu, X. Guo, and T. Wang, "Intermittent WLAN and interactions across heterogeneous wireless networks," *Telecommunication Systems*, vol. 43, no. 1-2, pp. 13–24, 2010.

[24] A. Al-Sherbaz, C. Adams, and S. Jassim, "WiMAX-WiFi convergence with OFDM bridge," in *Mobile Multimedia/Image Processing, Security, and Applications*, S. S. Agaian and S. A. Jassim, Eds., vol. 7351, May 2009.

[25] B. Li, Y. Qin, C. P. Low, and C. L. Gwee, "A Survey on mobile WiMAX (Wireless broadband access)," *IEEE Communications Magazine*, vol. 45, no. 12, pp. 70–75, 2007.

[26] B. A. Forouzan, *Data Communication and Networking*, McGraw Hill, 3rd edition, 2004.

[27] Y. Fakhri, B. Nsiri, D. Aboutajdine, and L. J. Vidal, "Throughput optimization Via the Packet length and transmission rate for wireless OFDM system in downlink transmission," *International Journal of Computer Science and Network Security B*, vol. 6, no. 3, pp. 41–46, 2006.

[28] S. Ci and H. Sharif, "Adaptive pptimal frame length predictor for IEEE 802.11 wireless LAN," in *Proceedings of the 6th IEE International Symposium Digital Signal Processing for Communication Systems (IEE DSPCS '02)*, Sydney, Australia.

[29] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2000.

[30] "The ns-3 network simulator," ns-3.10, 2011, http://www.nsnam.org/.

[31] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," Request for Comment 3626, Network Working Group, Project Hipercom INRIA, 2003.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 457962, 12 pages
http://dx.doi.org/10.1155/2014/457962

Research Article

DAG Based Multipath Routing Algorithm for Load Balancing in Machine-to-Machine Networks

Liang Liu, Tao Zhang, and Jun Zhang

School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Tao Zhang; zhtao@buaa.edu.cn

Received 17 August 2013; Revised 27 November 2013; Accepted 27 November 2013; Published 6 January 2014

Academic Editor: Jianhua He

Copyright © 2014 Liang Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In M2M networks, most nodes are powered by battery; hence the overused nodes may easily be out of power, which causes the reduction of network lifetime. To solve this problem, it is helpful to balance network load into more nodes and links, so as to reduce network congestion. Multipath routing is a useful tool to reduce congestion, since data flow can be dispersed into multiple paths. However, most of the previous multipath routing algorithm is based on the path disjoint constraint, which leads to the lack of routing paths and the failure to disperse data flow into more links. In this paper, a directed acyclic graph based multipath routing algorithm for congestion minimization (DAGMR) is proposed, where different routing paths are confined in a directed acyclic graph (DAG) under time delay constraint. Furthermore, the data flow distribution is accomplished through partial capacity network to obtain the minimal multipath routing congestion factor. Simulation indicates that our algorithm can obtain lower congestion factor and formulates multipath routing graph with more nodes and links than algorithm with path disjoint constraint. In addition, our algorithm is a polynomial time complexity algorithm with nodes and links number of the entire network.

1. Introduction

At present, Machine-to-Machine (M2M) network has emerged as an innovative topic and is receiving huge attention [1]. One of the biggest advantages of M2M networks is that nodes directly communicate with each other without human intervention, so that costs can be reduced and services can be improved in a wide range of industries. However, performance of M2M network suffers a lot from low-power character of network nodes. Since most nodes in wireless M2M networks are powered by batteries, the overused nodes can easily be out of power, which leads to the reduction of lifetime of entire networks [2]. To solve above problems, network load should be balanced into different nodes and links so as to reduce the congestion of certain links and ease the burden of certain overused nodes.

The objective to balance network load is to reduce network congestion. To achieve this objective, multipath routing is a useful method, since network data flow can be dispersed into multiple paths, so as to reduce network congestion. Former researches of this aspect are as follows. Reference [3] represents multipath routing as an optimization problem of congestion minimization and proposes a polynomial time algorithm to solve the problem of congestion minimization under path number constraint and time delay constraint. Reference [4] proposes a bandwidth aware multipath routing algorithm for Ad Hoc network. References [5, 6] propose low complexity link state multipath routing algorithm. Reference [7] proposes a tree search based multipath routing algorithm. Reference [8] proposes a network decomposition and traffic balancing based multipath routing algorithm.

Those algorithms above are mostly based on the principle that different paths should be disjoint (including node disjoint and link disjoint). Although the probability of data session failure can be reduced through the path disjoint principle based multipath routing, network congestion cannot be minimized, due to the following two reasons. First, the paths disjoint constraint leads to the reduction of paths number, which is a negative impact on the balance of data flow. Second, while trying to find disjoint paths, some extra links with relatively low link capacity use rate would be ignored. For the example illustrated in Figure 1, although four parallel paths

Figure 1: An example of multipath routing, where additional links besides disjoint paths can be used to reduce network congestion.

Figure 2: An example of layer structure for DAG, where entire graph is divided into 5 layers.

from S to D, which are depicted in solid line, exist, it is also helpful if some extra links crossing those paths can be utilized. For instance, when $v_2 \rightarrow v_3$ and $v_3 \rightarrow D$ are both in high capacity use rate and the capacity use rate of $v_6 \rightarrow D$ is relatively lower, congestion of $v_2 \rightarrow v_3$ and $v_3 \rightarrow D$ can be reduced, if v_2 disperses some data flow to v_6.

In order to further reduce network congestion, this paper proposes a *directed acyclic graph (DAG) based multipath routing algorithm*. Our algorithm abandons the path disjoint principle, and the final multipath routing graph is obtained through *maximum directed acyclic graph under time delay constraint (ML-DAG)*. Comparing with the former multipath routing algorithms, our algorithm constructs multipath routing graph with more nodes and links, and the network congestion can be efficiently reduced. The rest of this paper is organized as follows. Basic concept and the *problem of multipath routing for congestion minimization under time delay constraint (MCMTD)* are proposed in Section 2. In Section 3, the explicit description of three parts of entire algorithm with the corresponding time complexity analysis is proposed. Section 4 illustrates the simulation results, compared with the split multipath routing with maximally disjoint paths (SMR) [9] algorithm. Section 5 concludes the entire paper.

2. Basic Model and Problem Formulation

2.1. Basic Model. The topology of M2M network is modeled as a directed graph, $G(V, E)$, where V and E are set of nodes and set of links, respectively, with the following properties.

(i) For each link $e = (v_i, v_j)$, v_i, v_j refer to head and tail of e, respectively, which can be denoted as $v_i = T(e)$ and $v_j = H(e)$, besides v_i is called the *parent node* of v_j. For any node $v \in V$, $O_G(v)$ and $I_G(v)$ denote the set of out links and set of input link of v in G, respectively.

(ii) G is a graph with bidirectional edges, namely, for any $e = (v_i, v_j) \in E$, $\overleftarrow{e} = (v_j, v_i) \in E$.

(iii) For any $e = (v_i, v_j) \in E$, $d(v_i, v_j)$ (or d_{ij}, $d(e)$) and c_e denote *time delay* and *capacity* of the link, respectively. Given one data flow session $F_{S,D}(\gamma)$, where γ refers to total data quantity per time unit and S, D refer to the source node and destination node, respectively, let c_e^u $(0 < c_e^u < c_e)$ be the capacity of e occupied by other flow sessions besides $F_{S,D}$.

(iv) Given $G'(V_{G'}, E_{G'})$, if $V_{G'} \subset V$ and $E_{G'} \subset E$, then G' is one *subgraph* of G, namely, $G' \subset G$. Given $G_1, G_2 \subset G$, and G^u satisfies that $V_{G^u} = V_{G_1} \bigcup V_{G_2}$ and $E_{G^u} = E_{G_1} \bigcup E_{G_2}$, then G^u is also a subgraph of G and can be denoted as $G^u = G_1 \bigcup G_2$.

(v) Given $S, D \in G$, a *simple path*, $P_{S,D}^G(V_p, E_p)$, is a subgraph of G, namely, $P_{S,D}^G(V_p, E_p) \subset G$, such that:

(1) $V_p = \{S, v_{p_1}, v_{p_2}, \ldots, v_{p_k}, D\}$;
(2) $E_p = \{(S, v_{p_1}), (v_{p_1}, v_{p_2}), \ldots, (v_{p_{k-1}}, v_{p_k}), (v_{p_k}, D)\}$.

Furthermore, if for all $1 \le i < j \le k$, $v_{p_i} \neq v_{p_j}$, then $p_{S,D}$ is a *simple path*, which is denoted as $p_{S,D}^*$.

(vi) A *circle*, $P^c(V_{p^c}, E_{p^c})$, is a directed graph, such that

(1) $V_{p^c} = \{S, v_{p_1}, v_{p_2}, \ldots, v_{p_k}\}$;
(2) $E_{p^c} = \{(S, v_{p_1}), (v_{p_1}, v_{p_2}), \ldots, (v_{p_{k-1}}, v_{p_k}), (v_{p_k}, S)\}$.

In addition, *directed acyclic graph (DAG)* [10] is one special kind of graph, which is defined as follows.

Definition 1. A directed graph (as shown in Figure 2), $G^N(V_{G^N}, E_{G^N})$, is called a *directed acyclic graph (DAG)*, such that for all $v_i, v_j \in V_{G^N}$, if simple path from v_i to v_j, $P_{v_i,v_j}^* \subset G^N$, exists, then no path from v_j to v_i exists in G^N.

According to above definition, DAG has two important properties.

First, DAG can be divided into several layers, which is defined as follows.

Definition 2. Given one DAG, G^N, the *layer structure* of G^N is denoted as layer $(G^N) = \{G^N_{l_1}, \ldots, G^N_{l_1}\}$ such that

(1) $G^N_{l_1}, \ldots, G^N_{l_1} \subset G^N$, such that $\bigcup_{i=1}^{k} V_{G^N_{l_i}} = V_{G^N}$, and for all $1 \le i < j \le k$, $V_{G^N_{l_i}} \bigcap V_{G^N_{l_j}} = \emptyset$.

(2) Nodes of the first layer have no input links, namely, $V_{G^N_{l_1}} = \{v \mid v \in V_{G^N}, I_{G^N} = \emptyset\}$; and $\forall 1 < i \le k$, nodes of layer i have parent node in layer $i-1$, namely, $V_{G^N_{l_i}} = \{v \mid v \in V_{G^N} \setminus \bigcup_{j=1}^{i-1} V_{G^N_{l_j}}, \exists e \in I_{G^N}(v), T(e) \in V_{G^N_{l_{i-1}}}\}$.

(3) Links in each layer connect nodes in the same layer, namely, $E_{G^N_{l_i}} = \{e \in G^N \mid T(e), H(e) \in V^N_{G_{l_i}}\}$.

Second, longest path [11] between two nodes can be defined in DAG as follows.

Definition 3. Given one DAG, G^N, and $v_i, v_j \in V_{G^N}$, if simple path exists from v_i to v_j in G^N, then *longest path* from v_i to v_j in G^N, $\overline{P^{G^N}_{v_i,v_j}}$, satisfies for any $P^{G^N}_{v_i,v_j} \subset G^N$, $d(\overline{P^{G^N}_{v_i,v_j}}) \ge d(P^{G^N}_{v_i,v_j})$.

Furthermore, following two types of DAGs are presented in this paper, which are useful throughput the paper.

Definition 4. A DAG, $G^L_{S,D}(V_{G^L}, E_{G^L})$, is called an *L delay constraint directed acyclic graph (L-DAG)*, if the following properties satisfy:

(1) $G^L_{S,D}$ satisfies closure property, namely, $I_{G^L_{S,D}}(S) = \emptyset$, $O_{G^L_{S,D}}(S) = \emptyset$, $\forall v \in V \setminus \{S\} I_{G^L_{S,D}}(v) = \emptyset$, for all $v \in V \setminus \{D\} O_{G^L_{S,D}}(v) = \emptyset$;

(2) time delay of the longest path from S to D is smaller than L, namely, $d(\overline{P^{G^u}_{S,D}}) \le L$.

Definition 5. Given G, an L-DAG, $\widehat{G}^L_{S,D}$ is called a *maximum L-DAG (ML-DAG)*, if the following property holds: for all $v_i, v_j \in V_{\widehat{G}^L_{S,D}}$, if there exists $p_{v_i v_j} \subset G \setminus \widehat{G}^L_{S,D}$, then one of the following two conditions holds.

(1) $G_n = \widehat{G}^L_{S,D} \bigcup p_{v_i v_j}$ is a subgraph with circle.

(2) If G_n is a DAG, $d(\overline{P^{G_n}_{S,D}}) > L$.

2.2. Multipath Routing for Congestion Minimization under Time Delay Constraint. As mentioned in [3], the multipath routing strategy for data transmission can be denoted as follows.

Definition 6. Multipath routing strategy (MRS) from S to D is denoted as $T(S, D, \gamma, L) = \{G^L_{S,D}(V_{G^L}, E_{G^L}), X(E_{G^L})\}$, where γ denotes total data transferred in one time unit, L denotes the upper bound of time delay for data send, $G^L_{S,D}$ is an L-DAG and $X(E_{G^L}) = \{x(e) : e \in E_{G^L})\}$ denotes the amount of data to send on each link of $G^L_{S,D}$ per time unit, which satisfies the following.

(1) Link capacity constraint, namely, for all $e \in E_{G^L}$, $0 < x(e) \le c_e - c^u_e$.

(2) Network flow constraint, namely, $\sum_{e \in O(S)} x(e) = \sum_{e \in I(D)} x(e) = \gamma$, for all $v \in V_{G^L} \setminus \{S, D\}$, $\sum_{e \in O(v)} x(e) = \sum_{e \in I(v)} x(e)$.

As mentioned before, the efficiency of multipath routing strategy is evaluated by the congestion factor [3], which is defined as below.

Definition 7. Given $T(S, D, \gamma, L) = \{G^L_{S,D}(V_{G^L}, E_{G^L}), X(E_{G^L})\}$, *congestion factor of each link* is

$$\alpha(e) = \begin{cases} \dfrac{c^u_e + x(e)}{c_e}, & \text{if } e \in E_{G^L}, \\ \dfrac{c^u_e}{c_e}, & \text{if } e \in E_G \setminus E_{G^L}. \end{cases} \tag{1}$$

Congestion factor of each link is $\alpha_G = \max\{\alpha e, e \in E_G\}$. Congestion factor of MRS is $\alpha_T = \max\{\alpha e, e \in E_{G^L}\}$.

In above definition, $\alpha(e)$ is the judgement of congestion on each link and α_G is the judgement of congestion of entire network, according to [3]. However, α_G is not a good evaluation for MRS. For one thing, if some link of G has relatively large usage rate, α_G, the congestion factor of whole network is not able to judge the good MRS from the bad ones. For another, to further reduce the network minimization, it is needed to disperse the data flow into those links with relatively low usage rate. Hence, α_T is a better judgement for MRS than α_G. According to above analysis, *problem of multipath routing for congestion minimization under time delay constraint (MCMTD)* can be proposed as follows.

Problem MCMTD:
Minimize α_T, s.t:

(1) $T(S, D, \gamma, L) = \{G^L_{S,D}(V_{G^L}, E_{G^L}), X(E_{G^L})\}$ satisfies Definition 6.

(2) For all $e \in E_{G^L}$, $x(e) + c^u_e \le \alpha_T c_e$.

3. Multipath Routing Algorithm for Congestion Minimization

Problem MCMTD can be summarized as constructing MRS to minimize α_T. To achieve this aim, we need the following two phases.

(1) Obtain one L-DAG $\widehat{G}^L_{S,D}$, which includes as many links as possible.

(2) Allocate data flow into links of $\widehat{G}^L_{S,D}$, which has relatively lower link capacity use rate.

3.1. Obtain ML-DAG. Let $G^L_{S,D}$ be an initial L-DAG then, according to Definitions 4 and 5, the ML-DAG, $\widehat{G}^L_{S,D}$, can be

obtained through gradually adding nodes and links to $G^L_{S,D}$, until the no circle property, closure property, or time delay constraint cannot be maintained. The procedure of obtaining ML-DAG includes two phases.

3.1.1. Network Simplification. Network simplification is helpful to exclude nodes and links with relatively long distance from S and D, so that the scale of original network can be reduced and further algorithm of obtaining ML-DAG can be simplified. To achieve this aim, the concept of distance between two nodes should be defined.

Definition 8. Given G, for all $v_i, v_j \in V_G$, *the distance from* v_i *to* v_j *in* G, $d_G(v_i, v_j)$ is defined as follows.

(1) If no path exists from v_i to v_j in G, $d_G(v_i, v_j) = \infty$.

(2) Else, if $\widehat{p}^G_{v_i v_j}$ is the shortest path from v_i to v_j in G, then $d_G(v_1, v_2) = d(\widehat{p}^G_{v_i v_j})$.

According to above definition, the distance between two nodes corresponds to the length of shortest path between two nodes, which leads to the following two theorems.

Theorem 9. *If* $G_1 \subset G$, $v_i, v_j \in V_{G_1}$, *then* $d_{G_1}(v_i, v_j) \geq d_G(v_i, v_j)$.

Proof. Consider the following three situations.

(1) If $d_G(v_i, v_j) = \infty$, then no path exists from v_i to v_j in G. Since $G_1 \subset G$, no path exists from v_i to v_j in G_1. Hence, $d_{G_1}(v_i, v_j) = \infty$, and $d_{G_1}(v_i, v_j) \geq d_G(v_i, v_j)$ holds.

(2) If $d_G(v_i, v_j) \neq \infty$ and $d_{G_1}(v_i, v_j) = \infty$, then $d_{G_1}(v_i, v_j) \geq d_G(v_i, v_j)$ holds.

(3) If $d_G(v_i, v_j) \neq \infty$ and $d_{G_1}(v_i, v_j) \neq \infty$, let $\widehat{p}^{G_1}_{v_i v_j}$ be the shortest path from v_i to v_j in G_1, then $\widehat{p}^{G_1}_{v_i v_j} \subset G$, which denotes that $d_G(v_i, v_j) = d(\widehat{p}^G_{v_i v_j}) \leq d(\widehat{p}^{G_1}_{v_i v_j}) = d_{G_1}(v_i, v_j)$.

In summation, Theorem 9 is proved. □

Theorem 10. *Given ML-DAG,* $\widehat{G}^L_{S,D}(V_{\widehat{G}}, E_{\widehat{G}})$, *and* $v \in (V_G)$, *if* $d_G(S, v) + d_G(v, D) > L$, *then* $v \notin V_{\widehat{G}}$.

Proof. Assume that $v \in V_{\widehat{G}}$; then $d_{\widehat{G}^L_{S,D}}(S, v) + d_{\widehat{G}^L_{S,D}}(v, D) \leq L$, which conflicts with $d_{\widehat{G}^L_{S,D}}(S, v) + d_{\widehat{G}^L_{S,D}}(v, D) \geq d_G(S, v) + d_G(v, D) > L$. So the theorem is proved. □

According to above two theorems, if $d_G(S, v) + d_G(v, D) > L$, then v does not belong to $\widehat{G}^L_{S,D}$ and should be excluded. Hence, the algorithm of network simplification is as follows.

Algorithm 11. Network simplification is as follows.

Input: $G(V, E)$, L, S, D.

Step 1: If $d_G(S, D) \leq L$, let $G_r = G$, go to Step 2; else, algorithm fails.

Step 2: For any $v \in V_{G_r}$, if $d_{G_r}(S, v) + d_{G_r}(v, D) > L$, $V_{G_r} = V_{G_r} \setminus \{v\}$, $E_{G_r} = E_{G_r} \setminus (I_{G_r}(v) \bigcup O_{G_r}(v))$. When all notes have been traversed, output G_r.

3.1.2. Expanding Algorithm to Obtain ML-DAG. As mentioned before, ML-DAG can be obtained through gradually adding nodes and links. Each adding step is called expanding operation, as defined below.

Definition 12. Expanding operation (OR) from G^L_i to G^L_{i+1}, namely, $G^L_{i+1} = EO(G^L_i)$, satisfies the following.

(1) G^L_i and G^L_{i+1} are both L-DAG.

(2) Simple path $p^{\text{add}}_{v_i,v_j} \subset G_r$, such that $v_i, v_j \in V_{G^L_i}$, $p^{\text{add}}_{v_i,v_j} \setminus \{v_i, v_j\} \subset V_{G_r} \setminus V_{G^L_i}$, $E^{\text{add}}_{v_i,v_j} \subset E_{G_r} \setminus E_{G^L_i}$.

(3) $G^L_{i+1} = G^L_i \bigcup p^{\text{add}}_{v_i,v_j}$.

Expanding operation satisfies the following theorem.

Theorem 13. *Given* G^L_i, G^L_{i+1}, *which satisfies* $G^L_{i+1} = EO(G^L_i)$, *let* $\overline{P^{G^L_i}_{v_i,v_j}}$ *and* $\overline{P^{G^L_{i+1}}_{v_i,v_j}}$ *be the longest path from* S *to* D *in* G^L_i *and* G^L_{i+1}, *respectively, then* $d(\overline{P^{G^L_i}_{v_i,v_j}}) \geq \overline{P^{G^L_{i+1}}_{v_i,v_j}}$.

Proof. Since $\overline{P^{G^L_i}_{v_i,v_j}} \subset G^L_i \subset G^L_{i+1}$ and $\overline{P^{G^L_{i+1}}_{v_i,v_j}}$ is the longest path of G^L_{i+1}, according to the concept of longest path, $d(\overline{P^{G^L_i}_{v_i,v_j}}) \geq \overline{P^{G^L_{i+1}}_{v_i,v_j}}$. □

Let $G^L_0 = p^{G_r}_{\min}(S, D)$, namely, the shortest path from S to D is chosen to be the initial L-DAG. According to Theorem 13, $\widehat{G}^L_{S,D}$ can be obtained through finite times of EOs, since G_r is a graph with finite links and each EO adds new links to G^L_i. In order to make sure of the completeness and efficiency of the entire process to obtain ML-DAG, we have to figure out a principle of expanding operation. Considering that each expanding may probably increase the length of the longest path from S to D, which is a negative influence on the time delay constraint, our expanding operation follows the principle that the increase length of longest path after each expanding can be minimized. To achieve this objective, several examples in Figure 3 are proposed to show a phenomenon.

Assuming time delay of all links equals 1, then each graph in above figure is described as follows.

(a) Original L-DAG, G^L_i is a graph with two paths from S to D. Nodes of G^L_i can be divided into layers as, L1: S; L2: v_1, v_2; L3: v_3, v_4; L4: v_5, v_6; L5: D. One of the longest path from S to D is $S \to v_1 \to v_3 \to v_5 \to D$, time delay of which is 4.

(b) Let (v_2, v_3) be the adding path, where layer of v_2 is lower than layer of v_3, then longest path does not change.

Figure 3: Examples of adding path to L-DAG.

(c) Let (v_2, v_1) be the adding path, where layer of v_2 and v_1 are in the same layer, then the longest path becomes $S \to v_2 \to v_1 \to v_3 \to v_5 \to D$, time delay of which is 5.

(d) Let (v_4, v_1) be the adding path, where layer of v_4 is lower than layer of v_1, then longest path becomes $S \to v_2 \to v_4 \to v_1 \to v_3 \to v_5 \to D$, time delay of which is 6.

These examples show that adding one path from lower layer to the higher layer is a better choice than adding one path from the higher layer to the lower one, according to the principle mentioned above. Our algorithm to obtain the ML-DAG is based on this phenomenon. Algorithm of acquiring ML-DAG (Algorithm 18) is illustrated through the flow chart in Figure 4.

Some modular of the flow chat is explained as shown in Figure 4.

M1 Two layers, l_{i_1} and l_{i_2}, are chosen according to the following procedure. For each G_i^L, let q be the layer gap, such that $q = i_2 - i_1$. Initially, set $q = k - 1$. Then, gradually reduce q from $k-1, k-2, \ldots, 0, -1$, to $-(k-1)$, when two new layers need to be chosen. If $q = -(k-1)$ and no appropriate layers can be found so that adding path exists from i_1 to i_2, then M1 returns false.

M2 M2 is used to find $v_1 \in V_{G_{l_1}}$ and $v_2 \in V_{G_{l_2}}$. If all nodes that belong to l_{i_1}, l_{i_2} have been traversed and no appropriate expanding path can be obtained, then M2 returns false.

M3 G_T satisfies that $V_{G_T} = (V_{G_r} \setminus G_i^L) \bigcup \{v_1, v_2\}$ and $E_{G_T} = E_{G_i^L} \bigcup \{e \mid e \in G_r, H(e), I(e) \in V_{G_T}\}$.

M4 M4 judges whether path from v_1 to v_2 in G_T exists.

M5 M5 judges whether adding one path from v_1 to v_2 generates one circle.

M6 $G^c = G_i^L \bigcup \widehat{P}_{v_1,v_2}^{G_T}$, and $\overline{P_{S,D}^{G^c}}$ is the longest path from v_1 to v_2 in G_T.

M7 M7 judges whether time delay constraint still holds after expanding operation.

The completeness of above algorithm is guaranteed by next theorem.

Theorem 14. *ML-DAG, $\widehat{G}_{S,D}^L$, can be obtained through Figure 4.*

Proof. To prove the theorem, it should be shown that all possible adding paths have been considered when algorithm ends. In the beginning of each expanding operation, it can be guaranteed that G_i^L is a DAG and can be divided into layers. Hence, adding paths can be obtained through the following steps. First, find two layers l_{i_1} and l_{i_2} through M2. Second, choose two nodes v_1, v_2 belonging to those two layers, respectively, through M3. Third, obtain one adding path from v_1 to v_2 through M3, M4, M5, and M6. Since M1 traverses all layer pairs, M2 traverses all nodes pairs belonging to those two layers, and the correctness of adding path can be ensured through M5, M6 and M7; when M1 returns false, all possible adding paths have been considered. Therefore, Theorem 14 is proved. □

3.2. Obtain Minimul Congestion Factor. After obtaining $\widehat{G}_{S,D}^L$, the minimal congestion factor of MRS, α_T, can be obtained through allocating data flow into links with low capacity use rate. To achieve this objective, partial capacity graph needs to be considered.

Definition 15. Given one L-DAG, $G_{S,D}^L(V_{G^L}, E_{G^L})$, *$\beta$-partial capacity graph*, $G_{S,D}^{\beta}(V_{G^\beta}, E_{G^\beta})$, satisfies the following:

Figure 4: Entire flow chart of Algorithm 18.

(1) $G_{S,D}^{\beta}$ and $G_{S,D}^{L}(V_{G^L}, E_{G^L})$ have the same topology, namely, $V_{G^\beta} = V_{G^L}$, $E_{G^\beta} = E_{G^L}$;

(2) for all $e \in E_{G^\beta}$

$$c_e^{\beta} = \begin{cases} \beta c_e - c_e^u, & \text{if } c_e^u / c_e < \beta, \\ 0, & \text{if } c_e^u / c_e \geq \beta. \end{cases} \tag{2}$$

Furthermore, the residual graph after data flow distribution is defined as follows.

Definition 16. Given one L-DAG, $G^{\beta}(V_{G^\beta}, E_{G^\beta})$, and the data flow distribution, $X(G^{\beta}) = \{x^{\beta}(e) \mid 0 \leq x^{\beta}(e) \leq c\}$, the *residual graph*, $G_X^{\beta}(V_{G_X^\beta}, E_{G_X^\beta})$, satisfies

(1) $E_{G_X^\beta} = \{e \in E_{G^\beta} x^{\beta}(e) > 0\}$,

(2) $V_{G_X^\beta} = \{v \in V_{G^\beta} \mid \exists e \in I_{G^\beta}(v), x^{\beta}(e) > 0, \text{or } \exists e \in O_{G^\beta}(v), x^{\beta}(e) > 0\}$.

Based on above definitions, α_T can be obtained through gradually adjusting β, until $\mu(\beta)$ sufficiently approaches γ, namely, $\mu(\beta) \approx \gamma$, where $\mu(\beta)$ denotes the max-flow [12] from S to D in $G_{S,D}^{\beta}$. The feasibility of above adjusting process can be guaranteed through the following theorem.

Theorem 17. *$\mu(\beta)$ is a monotone increasing and continuous function with $\beta \in [0, 1]$.*

Proof. Let $X(\beta) = \{x^{\beta}(e) \mid x^{\beta}(e) \leq c_e^{\beta}, e \in E_{G_{S,D}^L}\}$ be the distribution of $\mu(\beta)$ on each link $G_{S,D}^L$.

(1) Proof of monotone increasing is as follows:

for all $0 \leq \beta_1 < \beta_2 \leq 1$, $c_e^{\beta_1} \leq c_e^{\beta_2}$, then $\forall e \in E_{G_{S,D}^{\beta}}$, $x^{\beta_1} \leq c_e^{\beta_1} \leq c_e^{\beta_1}$. So $X(\beta_1)$ is also the distribution of one available network flow on $G_{S,D}^{\beta_2}$. Hence, $\mu(\beta_1) \leq \mu(\beta_2)$.

(2) Proof of continuous is as follows:

according to Definition 15, for all $\beta^* \in [0, 1]$, if $\beta \to \beta^*$, $c_e^{\beta} \to c_e^{\beta^*}$, which eventually leads to $\mu(\beta) \to \mu(\beta^*)$.

□

Above theorem indicates that α_T can be obtained through adjusting β. In order to accelerate the adjustment, an algorithm analogous to binary search [13] is proposed as follows.

Algorithm 18. Obtain minimal congestion factor of MRS.

Input $\widehat{G}_{S,D}^L$, S, D, γ, $\{c_e | e \in E_{\widehat{G}}\}$, $\{c_e^u | e \in E_{\widehat{G}}\}$, η.

Step 1 Let $\beta_L = 0$ and $\beta_R = 1$. If $\mu(\beta_R) < \gamma$, return false; else, let $\beta = \beta_R$ go to Step 3.

Step 2 Let $\beta = (1/2)\beta_L + \beta_R$, go to Step 3.

Step 3 Obtain $G_{S,D}^{\beta}$, $\mu(\beta)$, and the corresponding data flow distribution, $X(G_{S,D}^{\beta})$, go to Step 4.

Step 4 If $\mu(\beta) > (1 + \eta) \cdot \gamma$, let $\beta_R = \beta$, go to Step 2;
else if $\mu(\beta) < (1 - \eta) \cdot \gamma$, let $\beta_L = \beta$, go to Step 2;
else if $(1 - \eta) \cdot \gamma \leq \mu(\beta) \leq (1 + \eta) \cdot \gamma$, go to Step 5.

Step 5 Let $\alpha_T = \beta$.
Let $G_{S,D}^f(V_{G^f}, E_{G^f}) = G_X^{\beta}$, where G_X^{β} is the residual graph corresponding with $G_{S,D}^{\beta}$ and $X(G_{S,D}^{\beta})$, according to Definition 16.
Let $X(G_{S,D}^f) = \{x(e) \mid e \in E_{G^f}\}$.
Return α_T, $G_{S,D}^f(V_{G^f}, E_{G^f})$, and $X(G_{S,D}^f)$.

3.3. Time Complexity Analysis. Let N_V and N_E be the entire nodes number and links number, respectively. Time complexity analysis of each algorithm is as follows. Algorithm 11 can be completed via the construction of two Dijkstra trees [12], with complexity $O(N_V^2)$, and the traversing of all nodes on the graph, with complexity $O(N_V)$, hence time complexity

Figure 5: Simulation result of congestion factor with the change of network nodes number. (a) denotes final congestion factor of two algorithms. (b) denotes final congestion factor ratio of our algorithm to SMR.

of Algorithm 11 is $O(N_V^2)$. In Figure 4, each expanding operation consists of layer algorithm ($O(N_V + N_E)$), shortest path algorithm ($O(N_V^2)$) [12], and longest path algorithm ($O(N_V + N_E)$) [11]. Hence, each expanding procedure can be finished in $O(N_V^2 \cdot (N_V^2 + N_E))$, which is the worst case when no node and link are excluded after Algorithm 11 and each pair of nodes has been traversed. Considering that at least one link can be added in each expanding operation, entire algorithm can be finished in $O(N_V^2 \cdot N_E \cdot (N_V^2 + N_E))$ in the worst case. In Algorithm 18, let $\mu(\beta^*) = \gamma$, then, after $\lceil \log_2(1/\eta) \rceil$ iterations, $\beta \in [(1-\beta)\cdot\beta^*, (1+\beta)\cdot\beta^*]$ satisfies, and Algorithm 18 ends. Each iteration needs a max-flow algorithm, with time complexity $O(N_V^2 \cdot N_E)$ in the worst case. Hence, Algorithm 18 can be finished in $O(N_V^2 \cdot N_E \cdot \lceil \log_2(1/\eta) \rceil)$. In summary, time complexity of entire algorithm is

$$\begin{aligned} &O\left(N_V^2\right) + O\left(N_V^2 \cdot N_E \cdot \left(N_V^2 + N_E\right)\right) \\ &\quad + O\left(N_V^2 \cdot N_E \cdot \left\lceil \log_2\left(\frac{1}{\eta}\right)\right\rceil\right) \\ &= O\left(N_V^2 \cdot N_E \cdot \left(N_V^2 + N_E + \left\lceil \log_2\left(\frac{1}{\eta}\right)\right\rceil\right)\right), \end{aligned} \tag{3}$$

which is polynomial time complexity with N_V and N_E.

4. Simulation and Analysis

4.1. Introduction to Basic Network Environment. The simulation network, $G(V, E)$, is formulated as a Transit-Stub Domain [14], where the connection probability between any two nodes is initially set to be 0.01. Basic parameters of capacity and time delay are as follows.

(i) For $\forall e \in E$, let link capacity, c_e, and average used capacity, $\overline{c_e^u}$, be 1 M bps and 0.5 M bps, respectively. Let current used capacity, c_e^u, be randomly chosen between $(1-\theta)\overline{c_e^u}$ and $(1+\theta)\overline{c_e^u}$, where $\theta = 0.3$. Let the data sent quantity from S to D be $\gamma = 0.3$ M bps.

(ii) For $\forall e \in E$, let $d(e) = 1$. For any given source node, S, and destination node, D, $d_G(S, D) = 4$ holds. Time delay constraint $L = 6$.

4.2. Simulation and Comparison Analysis. As mentioned before, our DAG based multipath routing algorithm (DAGMR) does not satisfy the path disjoint constraint; hence, performance of DAGMR should be compared with "Split Multipath Routing Algorithm with Maximally Disjoint Paths (SMR)" [9], which is a typical algorithm that follows the path disjoint constraint.

First, the most important performance for the load balancing in Machine-to-Machine network is the congestion factor of MRS, which is illustrated in Figure 5.

Figure 5 indicates that DAGMR obtains lower congestion factor than SMR, and the congestion factor gap of two algorithms increases as the nodes number in whole network increases. Moreover, when more than 200 nodes exist in network, congestion factor obtained by DAGMR is 5% smaller than that obtained by SMR.

Figure 6: Simulation result of nodes and links number in final routing graph with the change of entire network nodes number. (a) denotes the final multipath graph nodes number of DAGMR and SMR. (b) denotes the ratio of final multipath graph nodes number of DAGMR to SMR. (c) denotes the final multipath graph links number of DAGMR and SMR. (d) denotes the ratio of final multipath graph links number of DAGMR to SMR.

Second, DAGMR is helpful to distribute the data flow into more nodes and links, so as to ease the burden of certain overused nodes, as mentioned before. Hence, the number of nodes and links in the final multipath graph of DAGMR should be checked and compared with that of SMR.

Figure 6 indicates that nodes and links number in final routing graph obtained by DAGMR is larger than that obtained by SMR, and the gap of nodes and links number obtained in final routing graph by two algorithms increases as the network nodes number increases. Moreover, when entire network has more than 200 nodes, DAGMR subsumes 20% more nodes and links than SMR.

In above simulations, the connection probability of any two nodes is a fixed value, and performance about different nodes number is illustrated. Next, let the nodes number of whole network be fixed as 300 and let connection probability

Figure 7: Simulation result of congestion factor with the change of connection probability between any two nodes when entire network nodes number is 300. (a) denotes the congestion factor of final multipath graph of DAGMR and SMR. (b) denotes the congestion factor ratio of DAGMR to SMR.

of any two nodes be changed from 0.008 to 0.015. Then, congestion factor of MRS is illustrated in Figure 7.

Figure 7 indicates that DAGMR obtains lower congestion factor than SMR, and the congestion factor gap of two algorithms increases as the connection probability increases. Moreover, when connect rate is larger than 0.09, congestion factor obtained by DAGMR is 5% lower than that obtained by SMR.

Furthermore, the throughput performance of DAGMR should be analyzed. Given source node, S, destination node, D, and time delay constraint L, let α^R be the restricted congestion factor of MRS, namely, for any $T(S, D, \gamma, L)$, according to Definition 6, $\alpha_T \leq \alpha^R$ holds. Then, maximum data throughput, $\gamma_{\max}$, can be obtained through DAGMR, where second phase of the algorithm should be changed to obtain the maximum throughput under restricted congestion factor of α^R. Let the basic parameter of network be the same as before, where $d_G(S, D) = 4$ and time delay constraint $L = 6$. Let the restrict congestion factor be changed from 0.5 to 1. Then throughput performance of network with 200, 300, 400, 500 nodes is illustrated in Figures 8, 9, 10, and 11.

Figures 8, 9, 10, and 11 indicate the following aspects.

(i) Throughput increases as the network nodes number increases, which is reasonable, since the more the nodes exist in network, the more the opportunity is to get extra routing path.

(ii) Throughput obtained by DAGMR is larger than that obtained by SMR, and the gap of throughput increases as the restricted congestion factor increases.

(iii) The throughput ratio between DAGMR and SMR is larger than 1 and reaches the peak value when restricted congestion factor is about 0.65.

In summation, our DAG based multipath routing algorithm (DAGMR) achieves lower congestion factor and subsumes more nodes and links in the final multipath subgraph than SMR, when data sent quantity is fixed in a flow session. Meanwhile, when restricted congestion factor is fixed, DAGMR obtains larger data throughput than SMR. Therefore, DAGMR is a better choice for network load balancing in Machine-to-Machine networks.

5. Conclusion

Multipath routing is a useful method to balance network load and reduce congestion and has received extensive research. However, most of previous multipath routing algorithms are not able to fully utilize the advantage of multipath, due to the premise that routing paths should be disjoint. In this paper, a new multipath algorithm is proposed, in which the path disjoint constraint is replaced by a principle that multipath should be confined into a directed acyclic graph under time delay constraint. Meanwhile, the partial capacity network is proposed, in order to obtain the final multipath congestion factor. Simulation shows that our algorithm achieves lower

Figure 8: Simulation result of data throughput with the change of restricted congestion factor, when entire nodes number is 200.

Figure 9: Simulation result of data throughput with the change of restricted congestion factor, when entire nodes number is 300.

Figure 10: Simulation result of data throughput with the change of restricted congestion factor, when entire nodes number is 400.

Figure 11: Simulation result of data throughput with the change of restricted congestion factor, when entire nodes number is 500.

congestion factor and subsumes more nodes and links than the classical split multipath routing algorithm with disjoint paths. In addition, our algorithm is a polynomial time complexity algorithm with nodes number and links number of the entire network.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (nos. 61171069, 61231013, and 60933012), the 973 Program of China under Grant 2011CB707000, and the National 863 Program of China no. 2011AA110102.

References

[1] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[2] S. K. Singh, M. P. Singh, and D. K. Singh, "A survey of energy-efficient hierarchical cluster-based routing in wireless sensor networks," *International Journal of Advanced Networking and Applications*, vol. 2, no. 2, pp. 570–580, 2010.

[3] R. Banner and A. Orda, "Multipath routing algorithms for congestion minimization," *IEEE/ACM Transactions on Networking*, vol. 15, no. 2, pp. 413–424, 2007.

[4] W. Yang, X. Yang, G. Liu, and C. Dong, "A bandwidth aware multi-path routing protocol in mobile Ad Hoc networks," *Journal of Computational Information Systems*, vol. 7, no. 3, pp. 685–696, 2011.

[5] P. Mérindol, J.-J. Pansiot, and S. Cateloin, "Low complexity link state multipath routing," in *Proceedings of the IEEE INFOCOM Workshops (GIS '09)*, Rio de Janeiro, Brazil, April 2009.

[6] P. Mérindol, P. Francois, O. Bonaventure, S. Cateloin, and J.-J. Pansiot, "An efficient algorithm to enable path diversity in link state routing networks," *Computer Networks*, vol. 55, no. 5, pp. 1132–1149, 2011.

[7] W. Jiang, Q. Dai, Y. Zhou, and R. Liu, "A novel MPR approach based on inverse-traversing of Multi-Labeled tree," in *Proceedings of the IEEE International Conference on Intelligent Computing and Integrated Systems (ICISS '10)*, pp. 273–276, Guilin, China, October 2010.

[8] C. Bruni, F. Delli Priscoli, G. Koch, A. Pietrabissa, and L. Pimpinella, "Multipath routing by network decomposition and traffic balancing," in *Proceedings of the Future Network and Mobile Summit*, Florence, Italy, June 2010.

[9] S.-J. Lee and M. Gerla, "Split multipath routing with maximally disjoint paths in Ad hoc networks," in *Proceedings of the International Conference on Communications (ICC '01)*, pp. 3201–3205, June 2000.

[10] K.-S. Hong and L. Choi, "DAG-based multipath routing for mobile sensor networks," in *Proceedings of the International Conference on ICT Convergence (ICTC '11)*, pp. 261–266, Seoul, Korea, September 2011.

[11] K. Ioannidou and S. D. Nikolopoulos, "The longest path problem is polynomial on cocomparability graphs," *Algorithmica*, vol. 65, no. 1, pp. 177–205, 2013.

[12] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice- Hall, Upper Saddle River, NJ, USA, 1993.

[13] R. Nowak, "Generalized binary search," in *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 568–574, Urbana-Champaign, Ill, USA, September 2008.

[14] K. L. Calvert, M. B. Doar, and E. W. Zegura, "Modeling internet topology," *IEEE Communications Magazine*, vol. 35, no. 6, pp. 160–163, 1997.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 728532, 11 pages
http://dx.doi.org/10.1155/2013/728532

Research Article

Secure Beamforming via Amplify-and-Forward Relays in Machine-to-Machine Communications

Zhongjian Liu,[1] Xiaoning Zhang,[1] Lin Bai,[2] Chen Chen,[1] and Haige Xiang[1]

[1] *State Key Laboratory of Advanced Optical Communication Systems and Networks, Peking University, Beijing 100871, China*
[2] *School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Xiaoning Zhang; zhangxn@pku.edu.cn and Lin Bai; l.bai@buaa.edu.cn

Received 18 August 2013; Accepted 18 November 2013

Academic Editor: Jianhua He

Copyright © 2013 Zhongjian Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the collaborative use of amplify-and-forward relays to form a beamforming system and provide physical layer security for a wireless machine-to-machine (M2M) communication network. We investigate two objectives: (i) the achievable secrecy rate maximization subject to the relay power constraint and (ii) the relay transmit power minimization under a secrecy rate constraint. For the first objective, we propose a secrecy rate maximization (SRM) beamforming scheme. The secrecy rate maximization problem can be formed into a two-level optimization problem and we solve it using semidefinite relaxation (SDR) techniques. To reduce the complexity of the SRM beamforming scheme, a virtual eavesdropper-based SRM (VE-SRM) beamforming scheme is proposed, in which we hypothesize a virtual eavesdropper instead of all eavesdroppers and maximize the secrecy rate according to the virtual eavesdropper. In addition, for the second objective, we design a relay power minimization (RPM) beamforming scheme, in which an iterative algorithm combining the SDR technology and the gradient-based method is devised by studying the convexity of the RPM problem. By relaxing the constraints of the RPM beamforming scheme, we propose a virtual eavesdropper-based RPM (VERPM) beamforming scheme, which reduces the multivariate RPM problem to a problem of a single variable, and thus an analytical solution is obtained. Our proposed beamforming designs can work well even if the number of eavesdroppers is larger than that of relays, while the existing schemes, for example, the null-space beamforming schemes, cannot work under this condition. Simulation results are presented to demonstrate the significance of performance improvements with the SRM and RPM beamforming schemes. It is also shown that the virtual eavesdropper approaches significantly reduce the complexity with acceptable performance degradation.

1. Introduction

The demands of wireless personal communications drive the development of abundant advanced wireless technologies and systems such as the cognitive radio network (CRN) and Third Generation Partnership Project (3GPP) Long Term Evolution-Advanced (LTE-Advanced) [1]. In addition to human-to-human (H2H) communications, an emerging technology empowering full mechanical automation (e.g., the Internet of Things and the smart grid) that may change our living styles is vigorously being developed. Such communications among machine-type devices are known as M2M communications [2–4]. Due to the use of machine-type devices and the broadcast nature of wireless media, privacy and security issues play an important role in wireless M2M networks. Although security is typically addressed via cryptographic approaches, there have been several attempts at addressing security at the physical layer, following the pioneering work of Wyner [5]. Physical layersecurity can be considered to solve the security problem from the point of view of information theory. The basic principle of information-theoretic security calls for the combination of cryptographic schemes with channel coding techniques. Physical layer security exploits the randomness of communication channels to guarantee that the transmitted messages cannot be decoded by a third party maliciously eavesdropping on the wireless media. The following work by Leung-Yan-Cheong and Hellman characterized the secrecy capacity of scalar Gaussian wiretap channel [6]. Along the same lines, the Gaussian multiple-input-multiple-output (MIMO) wiretap channel was investigated and its secrecy capacity was established in terms of

an optimization problem over all possible input covariance matrices [7–9]. There are some recent works focusing on secrecy rates based on partial channel state information (CSI) [10–12] and multiuser scenarios [13–15].

The secrecy rate based on single-antenna systems is hampered by channel conditions. When the channel condition between the source and the destination is worse than the source-to-eavesdropper channel, positive secrecy rate cannot be achieved unless multiple antennas are employed at the transmitter. Since the use of multiple antennas on a node involves a potentially significant hardware cost, node cooperation is an alternative low-cost scheme that enables single-antenna nodes to enjoy the benefits of multiple-antenna systems. In particular, node cooperation via relays can increase the achievable secrecy rate by exploiting/mitigating the channel effects [16–20]. There are mainly two relaying protocols for the cooperative secure transmission: decode-and-forward (DF) [16] and amplify-and-forward (AF) [17, 18]. Dong et al. [18] proposed a two-stage amplify-and-forward cooperative approach. In [18], the source and relays are located in the same cluster, while the destination and eavesdroppers are at far away locations outside this cluster. A system design was proposed to maximize the upper and lower bounds of secrecy capacity in the presence of a single eavesdropper. For the multiple eavesdroppers case, Dong et al. [19] proposed suboptimal and conservative null-space beamforming schemes for both the secrecy rate maximization problem subject to a power constraint and the relay power minimization problem subject to a secrecy rate constraint, which completely nulled out the signal leakage to all eavesdroppers in Stage 2. However, this kind of null-space beamforming scheme requires that the number of relays is larger than that of eavesdroppers. In addition, Yang et al. [20] obtained the secrecy rate maximization (SRM) weights design in the presence of multiple eavesdroppers, but they assumed that there are no direct links between the source and the destination, as well as the source and the eavesdroppers.

In this paper, we consider a more general case of multiple relays and multiple eavesdroppers, where there are possible direct links between the source and the destination as well as the source and the eavesdroppers. Global CSI is assumed to be available and even the eavesdroppers' channels are known. CSI on the eavesdroppers' channels can be obtained when the eavesdroppers are active in the network as their transmissions can be monitored. This is applicable particularly in wireless networks, in which terminals play dual roles as legitimate receivers for some signals and eavesdroppers for others [19]. We focus on the cooperative secure beamforming designs for both the secrecy rate maximization problem subject to a power constraint, and the relay power minimization problem subject to a secrecy rate constraint. The contributions of this paper can be summarized as follows.

(i) For the first optimization problem, we propose an SRM beamforming scheme. Using the semidefinite relaxation (SDR) technique, we perform the two-level optimization SRM problem to a one-dimensional search problem. Its performance approaches that of the optimal beamforming scheme.

(ii) From a practical point of view, it is necessary to design a more simplified solution. Null-space beamforming has low complexity to be implemented, while it is out of operation when the eavesdroppers outnumber the relays. In order to overcome the drawback of null-space beamforming scheme, we propose a virtual eavesdropper-based SRM (VE-SRM) beamforming scheme by hypothesizing a virtual eavesdropper instead of all the eavesdroppers and maximize the secrecy rate according to the virtual eavesdropper.

(iii) For the second optimization problem, we propose a relay power minimization (RPM) beamforming scheme to minimize the relay power subject to a secrecy rate constraint. To solve the multivariate problem, an iterative algorithm combining SDR and gradient-based method is devised.

(iv) To reduce the complexity of the RPM problem, we relax the eavesdropper constraint and propose a virtual eavesdropper-based RPM (VE-RPM) beamforming scheme, which reduces the multivariate RPM problem to a problem of a single variable, and thus an analytical solution is obtained.

Moreover, the beamforming schemes we proposed can work well even if the number of eavesdroppers is larger than that of relays.

The rest of this paper is organized as follows. In Section 2, an AF-based cooperative network with multiple eavesdroppers scenario is described and the corresponding secure beamforming problems are formulated. In Section 3, we propose an SRM beamforming scheme and a practical VE-SRM beamforming scheme for the SRM problem subject to a power constraint. In Section 4, we propose an RPM beamforming scheme and an analytical VE-RPM beamforming scheme for the RPM problem subject to a secrecy rate constraint. Simulation results are provided in Section 5 and the conclusions are drawn in Section 6.

2. System Model and Problem Formulation

2.1. Channel Model. In this paper, we consider a communication unit of wireless M2M network which consists of one source (Alice), N ($N \geq 1$) relay nodes, one destination node (Bob), and J ($J > 1$) eavesdroppers (Eves). Define the set of the relays $\mathscr{R} \triangleq \{R_1, R_2, \ldots, R_N\}$ and the set of the Eves $\mathscr{E} \triangleq \{E_1, E_2, \ldots, E_J\}$. All channels are assumed to be flat fading and quasistatic. In particular, denote by $\mathbf{h}_{SR}$ the channel vector between Alice and the relays in $\mathscr{R}$, by h_{SD} the channel coefficient between Alice and Bob, by $h_{SE,j}$ the channel vector between Alice and the jth Eve, by $\mathbf{h}_{RD}$ the channel vector between the relays in $\mathscr{R}$ and Bob, and by $\mathbf{h}_{RE,j}$ the channel vector between the relays in $\mathscr{R}$ and jth Eve. The noise at each node is assumed to be zero-mean white complex Gaussian with variance σ^2. Each node is equipped with a single omnidirectional antenna. The configuration is illustrated in Figure 1.

There are two stages in the AF protocol. In Stage 1, Alice broadcasts its n encoded symbols to its relays. For the purpose

Figure 1: AF relay assisted wireless M2M network in the presence of eavesdroppers.

of evaluating the achievable secrecy rate, it is assumed that the codewords used at Alice are Gaussian inputs. In a time unit, let x be the transmitted symbol, and the average power of an symbol is normalized; that is, $\mathbb{E}\{|x|^2\} = 1$. P_0 is the total power budget for transmitting one symbol in the two stages. When the source transmits the symbol x, the received signals at the N relays, stacked in vector $\mathbf{y}_R(N \times 1)$, become

$$\mathbf{y}_R = \sqrt{P_S}\mathbf{h}_{SR}x + \mathbf{n}_R, \tag{1}$$

where P_S is the transmit power of Alice. It holds that $0 < P_S \leq P_0$. $\mathbf{n}_R$ is the noise vector at the relays.

In Stage 2, the N relays forward the signals that they received during Stage 1 to the destination. More specifically, each relay transmits a weighted version of the noisy signal that they received during Stage 1. Thus, the transmitted signals of all the relays are $\mathbf{D}(\mathbf{w})\mathbf{y}_R$, where $\mathbf{D}(\mathbf{a})$ denote the diagonal matrix with the vector $\mathbf{a}$ on its main diagonal and $\mathbf{w}$ is the weight vector and $\mathbf{y}_R$ is given in (1). The transmit power budget for Stage 2 is $P_R \triangleq P_0 - P_S$.

The received signal at the destination is

$$\begin{aligned} y_D &= \sqrt{P_S}\mathbf{h}_{RD}^{\dagger}\mathbf{D}(\mathbf{w})\mathbf{h}_{SR}x + \mathbf{h}_{RD}^{\dagger}\mathbf{D}(\mathbf{w})\mathbf{n}_R + n_D \\ &= \sqrt{P_S}\mathbf{h}_{RD}^{\dagger}\mathbf{D}(\mathbf{h}_{SR})\mathbf{w}x + \mathbf{n}_R^T\mathbf{D}\left(\mathbf{h}_{RD}^{\dagger}\right)\mathbf{w} + n_D \\ &= \mathbf{a}^{\dagger}\mathbf{w}x + \mathbf{n}_R^T\mathbf{D}\left(\mathbf{h}_{RD}^{\dagger}\right)\mathbf{w} + n_D, \end{aligned} \tag{2}$$

where $(\cdot)^{\dagger}$ denotes the conjugate transpose, $(\cdot)^T$ denotes the transpose, $\mathbf{a} \triangleq \sqrt{P_S}\mathbf{D}(\mathbf{h}_{SR}^{\dagger})\mathbf{h}_{RD}$, and n_D is the noise at Bob.

The received signal at the jth Eve is

$$\begin{aligned} y_E^j &= \sqrt{P_S}\mathbf{h}_{RE,j}^{\dagger}\mathbf{D}(\mathbf{w})\mathbf{h}_{SR}x + \mathbf{h}_{RE,j}^{\dagger}\mathbf{D}(\mathbf{w})\mathbf{n}_R + n_{E,j} \\ &= \sqrt{P_S}\mathbf{h}_{RE,j}^{\dagger}\mathbf{D}(\mathbf{h}_{SR})\mathbf{w}x + \mathbf{n}_R^T\mathbf{D}\left(\mathbf{h}_{RE,j}^{\dagger}\right)\mathbf{w} + n_{E,j} \\ &= \mathbf{b}_j^{\dagger}\mathbf{w}x + \mathbf{n}_R^T\mathbf{D}\left(\mathbf{h}_{RE,j}^{\dagger}\right)\mathbf{w} + n_{E,j}, \end{aligned} \tag{3}$$

where $\mathbf{b}_j \triangleq \sqrt{P_S}\mathbf{D}(\mathbf{h}_{SR}^{\dagger})\mathbf{h}_{RE,j}$ and $n_{E,j}$ is the noise at the jth Eve.

2.2. Achievable Secrecy Rate. Recalling that P_0 is the transmit power in Stage 1, and the achievable rate at Bob is

$$R_D = \frac{1}{2}\log_2\left(\alpha + \frac{\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w} + 1\right)\sigma^2}\right), \tag{4}$$

where $\alpha \triangleq 1 + P_S|h_{SD}|^2/\sigma^2$, $\mathbf{R}_a \triangleq \mathbf{a}\mathbf{a}^{\dagger}$, and $\mathbf{U} \triangleq \mathbf{D}(\mathbf{h}_{RD})\mathbf{D}(\mathbf{h}_{RD}^{\dagger})$. Note that $P_S|h_{SD}|^2/\sigma^2$ is the received signal-noise ratio (SNR) in Stage 1 at Bob. Similarly, the achievable rate at the jth Eve is

$$R_{E,j} = \frac{1}{2}\log_2\left(\beta_j + \frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w} + 1\right)\sigma^2}\right), \tag{5}$$

where $\beta_j \triangleq 1 + P_S|h_{SE,j}|^2/\sigma^2$, $\mathbf{R}_{b,j} \triangleq \mathbf{b}_j\mathbf{b}_j^{\dagger}$, and $\mathbf{V}_j \triangleq \mathbf{D}(\mathbf{h}_{RE,j})\mathbf{D}(\mathbf{h}_{RE,j}^{\dagger})$. Note that $P_S|h_{SE,j}|^2/\sigma^2$ is the SNR in Stage 1 at the jth Eve.

With the presence of J Eves, the achievable secrecy rate is given by [19]

$$R_s = \left(R_D - \max\left(R_{E,1}, R_{E,2}, \ldots, R_{E,J}\right)\right)_+, \tag{6}$$

where $(x)_+$ represents the maximal one between x and 0. The achievable secrecy rate in (6) can be interpreted as a worst-case result; that is, the Eve with the best channel dominates the secrecy rate. For convenience, we drop the operator $(\cdot)_+$ and the constant coefficient 1/2.

Our beamforming design is based on amplify-and-forward cooperative scheme, and the optimization criterion is the maximization of the achievable secrecy rate subject to a relay power constraint or the minimization of the relay transmit power subject to a secrecy rate constraint. The specific problems addressed in this paper are as follows.

2.2.1. Maximize the Secrecy Rate under a Relay Power Constraint. Given a source power P_S and a relay power P_R, design the cooperative beamforming vector $\mathbf{w}$ to maximize the secrecy rate; that is,

$$\begin{aligned} \max_{\mathbf{w}} \min_{E_j \in \mathcal{E}} \quad & \left\{\log_2\left(\alpha + \frac{\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w} + 1\right)\sigma^2}\right)\right. \\ & \left. - \log_2\left(\beta_j + \frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w} + 1\right)\sigma^2}\right)\right\} \\ \text{s.t.} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w} \leq P_R, \end{aligned} \tag{7}$$

where $\mathbf{T} \triangleq P_S\mathbf{D}\{\mathbf{h}_{SR}\}\mathbf{D}\{\mathbf{h}_{SR}^{\dagger}\} + \mathbf{I}$.

2.2.2. Minimize the Relay Transmit Power under a Secrecy Rate Constraint. Given a secrecy rate constraint Γ_0, design the cooperative beamforming vector $\mathbf{w}$ to minimize the relay power; that is,

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w} \\ \text{s.t.} \quad & R_s \geq \Gamma_0, \end{aligned} \tag{8}$$

where R_s is given by (6). Here we point out that the constraints $R_s \geq \Gamma_0$ and $R_s = \Gamma_0$ are equivalent. The reason is that R_s is a nondecreasing function of the relay power P_R.

3. Secrecy Rate Maximization Beamforming Design under Relay Power Constraint

3.1. Existing Beamforming Scheme. In this subsection, we demonstrate the existing beamforming scheme under the relay power constraint. In [19], a simple null-space beamforming design was proposed when the number of relays is larger than that of eavesdroppers. Under the assumption of no signal leakage to all eavesdroppers in Stage 2, the associated SRM problem (7) can be simplified to

$$\begin{aligned} \max_{\mathbf{w}} \quad & \left\{\log_2\left(\alpha + \frac{\mathbf{w}^\dagger \mathbf{R}_a \mathbf{w}}{(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1)\sigma^2}\right) - \log_2(\beta_{\max})\right\} \\ \text{s.t.} \quad & \mathbf{h}_{RE,j}^\dagger \mathbf{w} = 0, \quad \forall E_j \in \mathscr{E} \\ & \mathbf{w}^\dagger \mathbf{T}\mathbf{w} \le P_R, \end{aligned} \tag{9}$$

where $\beta_{\max} \triangleq 1 + \max_{E_j \in \mathscr{E}}(P_0 |h_{SE,j}|^2/\sigma^2)$. Under the relay power constraint, the null-space SRM problem was studied in [19], where a closed-form solution was derived.

3.2. Secrecy Rate Maximization Beamforming. In order to maximize the secrecy rate under the total power constraint at relays, a SRM beamforming scheme is developed. By introducing a slack variable τ, problem (7) can be equivalently rewritten as

$$\begin{aligned} \max_{\mathbf{w},\tau>0} \quad & \left\{\log_2\left(\alpha + \frac{\mathbf{w}^\dagger \mathbf{R}_a \mathbf{w}}{(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1)\sigma^2}\right) - \log_2\left(\frac{1}{\tau}\right)\right\} \\ \text{s.t.} \quad & \log_2\left(\beta_j + \frac{\mathbf{w}^\dagger \mathbf{R}_{b,j} \mathbf{w}}{(\mathbf{w}^\dagger \mathbf{V}_j \mathbf{w} + 1)\sigma^2}\right) \le \log_2\left(\frac{1}{\tau}\right), \\ & \qquad \forall E_j \in \mathscr{E} \\ & \mathbf{w}^\dagger \mathbf{T}\mathbf{w} \le P_R. \end{aligned} \tag{10}$$

Since the logarithm function is monotonically increasing, we remove it and further obtain

$$\begin{aligned} \min_{\mathbf{w},\tau>0} \quad & \frac{(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1)\sigma^2}{(\mathbf{w}^\dagger (\alpha\sigma^2 \mathbf{U} + \mathbf{R}_a)\mathbf{w} + \alpha\sigma^2)\tau} \\ \text{s.t.} \quad & \frac{(\mathbf{w}^\dagger \mathbf{V}_j \mathbf{w} + 1)\sigma^2}{\mathbf{w}^\dagger (\beta_j \sigma^2 \mathbf{V}_j + \mathbf{R}_{b,j})\mathbf{w} + \beta_j \sigma^2} \ge \tau, \quad \forall E_j \in \mathscr{E} \\ & \mathbf{w}^\dagger \mathbf{T}\mathbf{w} \le P_R. \end{aligned} \tag{11}$$

To solve problem (11), following the idea of SDR [21], define $\mathbf{W} \triangleq \mathbf{w}\mathbf{w}^\dagger$ and drop the nonconvex rank-one constraint to get a relaxed problem of (11) as follows:

$$\begin{aligned} \min_{\mathbf{W}\succeq\mathbf{0},\tau>0} \quad & \frac{(\operatorname{Tr}(\mathbf{U}\mathbf{W}) + 1)\sigma^2}{(\operatorname{Tr}((\alpha\sigma^2 \mathbf{U} + \mathbf{R}_a)\mathbf{W}) + \alpha\sigma^2)\tau} \\ \text{s.t.} \quad & \frac{(\operatorname{Tr}(\mathbf{V}_j \mathbf{W}) + 1)\sigma^2}{\operatorname{Tr}((\beta_j \sigma^2 \mathbf{V}_j + \mathbf{R}_{b,j})\mathbf{W}) + \beta_j \sigma^2} \ge \tau, \quad \forall E_j \in \mathscr{E}, \\ & \operatorname{Tr}(\mathbf{T}\mathbf{W}) \le P_R. \end{aligned} \tag{12}$$

While problem (12) is still nonconvex due to the variable τ, one may check that, for a given τ, problem (12) changes into a quasiconvex problem [22]. By this observation, treat (12) as a two-level optimization problem—the inner level (problem (12) with a fixed τ) is a quasiconvex problem, and the outer level (optimization over τ) is a single-variable optimization problem.

Let us start with the inner-level optimization problem. Though this class of quasiconvex problems can be solved by using the bisection method [22], one usually needs to deal with many semidefinite programs (SDPs) during the bisection search [22]. Here, we adopt a more efficient approach by reformulating quasiconvex problem into a convex SDP through the Charnes-Cooper transformation [23]. For some $\mathbf{Z} \succeq \mathbf{0}$ and $\eta > 0$, we let $\mathbf{W} = \mathbf{Z}/\eta$ and recast (12) as

$$\begin{aligned} \min_{\mathbf{Z}\succeq\mathbf{0},\eta>0} \quad & \{\operatorname{Tr}(\mathbf{U}\mathbf{Z}) + \eta\} \\ \text{s.t.} \quad & \tau\left(\operatorname{Tr}\left((\alpha\sigma^2 \mathbf{U} + \mathbf{R}_a)\mathbf{Z}\right) + \alpha\sigma^2\eta\right) = 1 \\ & \tau\left(\operatorname{Tr}\left((\beta_j \sigma^2 \mathbf{V}_j + \mathbf{R}_{b,j})\mathbf{Z}\right) + \beta_j \sigma^2 \eta\right) \\ & \quad \le \left(\operatorname{Tr}(\mathbf{V}_j \mathbf{Z}) + \eta\right)\sigma^2, \quad \forall E_j \in \mathscr{E}, \\ & \operatorname{Tr}(\mathbf{T}\mathbf{Z}) \le \eta P_R. \end{aligned} \tag{13}$$

Problem (13) is an SDP whose optimal solution can be efficiently obtained by available solvers, for example, CVX. In the following, we turn our attention to the outer-level part, which is a single-variable optimization problem as follows:

$$\begin{aligned} \min_{\tau} \quad & \phi(\tau), \\ \text{s.t.} \quad & \tau_{\text{lb}} \le \tau \le \tau_{\text{ub}}, \end{aligned} \tag{14}$$

where $\phi(\tau)$ denotes the optimal value of (13) and τ_{lb} and τ_{ub} are the lower and upper bound of τ hidden in (10)–(12), respectively. Upon carefully examining the feasibility of the constraints in (12), we have

$$\begin{aligned} \tau &\le \min_{E_j \in \mathscr{E}} \frac{(\operatorname{Tr}(\mathbf{V}_j \mathbf{W}) + 1)\sigma^2}{\operatorname{Tr}((\beta_j \sigma^2 \mathbf{V}_j + \mathbf{R}_{b,j})\mathbf{W}) + \beta_j \sigma^2} \\ &< \min_{E_j \in \mathscr{E}} \left(\frac{1}{\beta_j}\right) \triangleq \tau_{\text{ub}}. \end{aligned} \tag{15}$$

The effective τ_{lb} arises from the requirement of nonnegative secrecy rate in (10); that is,

$$\begin{aligned}\tau &\geq \frac{\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2}{\mathbf{w}^{\dagger}\left(\alpha\sigma^2\mathbf{U}+\mathbf{R}_a\right)\mathbf{w}+\alpha\sigma^2}\\ &= \frac{1}{\alpha+\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}/\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2}\\ &\geq \frac{1}{\alpha+\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}/\mathbf{w}^{\dagger}\left(\mathbf{U}+\mathbf{T}/P_R\right)\mathbf{w}\sigma^2}\\ &\geq \frac{1}{\alpha+\lambda_{\max}\left(\left(\mathbf{U}+\mathbf{T}/P_R\right)^{-1}\mathbf{R}_a\right)/\sigma^2} \triangleq \tau_{\text{lb}},\end{aligned} \tag{16}$$

where (16) is a generalized Rayleigh Quotient (GRQ) problem [24] and $\lambda_{\max}(\mathbf{A})$ represents the maximal eigenvalue of the matrix $\mathbf{A}$.

The lower bound implies that, if $\tau < \tau_{\text{lb}}$, the secrecy rate will be negative and the design does not make sense. Since (13) is a single-variable problem ($\tau_{\text{lb}} \leq \tau \leq \tau_{\text{ub}}$), one-dimensional optimization techniques can be used to search for the solution.

So far, we can cope with (12) through performing a one-dimensional search during which we solve a sequence of SDPs. We denote the solution of (12) by $(\mathbf{W}^*, \tau^*)$. As we have dropped the rank one constraint of $\mathbf{W}$, the $\mathbf{W}^*$ can achieve a upper bound of the optimal secrecy rate. If the rank of $\mathbf{W}^*$ is equal to 1, the feasible $\mathbf{w}$ of (10) can be obtained from $\mathbf{W}^*$ via eigenvalue decomposition. If the rank of $\mathbf{W}^*$ is larger than 1, we can extract an approximate solution $\mathbf{w}$ from $\mathbf{W}^*$ using the Gaussian Randomization procedure [21]. We will compare the achievable secrecy rate of the SRM beamforming scheme and the upper bound of the optimal secrecy rate in the simulation section.

3.3. Virtual Eavesdropper-Based Secrecy Rate Maximization Beamforming. As mentioned above, the SRM beamforming is still difficult to implement due to the existence of multiple SDPs. To further reduce the complexity of the implementation, we propose a VE-SRM beamforming scheme. The SRM problem (7) can be rewritten as

$$\begin{aligned}\max_{\mathbf{w}} \quad & \left\{\log_2\left(\alpha+\frac{\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2}\right)\right.\\ & \left.-\max_{E_j\in\mathscr{E}}\log_2\left(\beta_j+\frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2}\right)\right\}\\ \text{s.t.} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\leq P_R.\end{aligned} \tag{17}$$

Obviously, it is the two-level optimization that makes the SRM problem complicated to solve. The key to address problem (17) is to solve $\max_{E_j\in\mathscr{E}}\log_2(\beta_j+\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}/(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1)\sigma^2)$. The eavesdropping rate at E_j is related to its received SNR, and the eavesdropper that has the largest SNR is the one that has the maximal eavesdropping rate. Thus, the largest SNR in all Eves for a given $\mathbf{w}$ in problem (17) can be expressed as

$$\text{SNR}_{\max}^{\mathbf{w}} = \max_{E_j\in\mathscr{E}}\left\{\beta_j-1+\frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2}\right\}. \tag{18}$$

In order to drop the max expression in (18), we need to find an upper bound of the largest SNR (corresponding to the largest eavesdropping rate). Note that $\text{SNR}_{\mathbf{w}}^{\max} \leq \beta_{\max}-1+\sum_{E_j\in\mathscr{E}}(\text{SNR}_{2,j}^{\mathbf{w}})$ is always valid, where $\text{SNR}_{2,j}^{\mathbf{w}}$ is the received SNR of E_j in Stage 2. Problem (18) can be relaxed as

$$\text{SNR}_{\max}^{\mathbf{w}} \leq \beta_{\max}-1+\sum_{E_j\in\mathscr{E}}\frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2}. \tag{19}$$

Notice that

$$\frac{\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}} \geq \lambda_{\min}\left(\mathbf{T}^{-1}\mathbf{V}_j\right), \quad \forall E_j\in\mathscr{E},\ \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\leq P_R \tag{20}$$

and the upper bound of $\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1$ can be expressed as

$$\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1 \geq \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\left(\lambda_{\min}\left(\mathbf{T}^{-1}\mathbf{V}_j\right)+\frac{1}{P_R}\right). \tag{21}$$

According to the inequality in (21), the formula in (19) can be rewritten as

$$\text{SNR}_{\max}^{\mathbf{w}} \leq \beta_{\max}-1+\frac{\mathbf{w}^{\dagger}\mathbf{R}_g\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\sigma^2}, \tag{22}$$

where $\mathbf{R}_g \triangleq \sum_{E_j\in\mathscr{E}}(\mathbf{R}_{b,j}/\Omega_j)$ and $\Omega_j = \lambda_{\min}(\mathbf{T}^{-1}\mathbf{V}_j)+1/P_R$, for all $E_j \in \mathscr{E}$. It seems that there is a virtual eavesdropper whose channel gain is $\beta_{\max}-1$ in Stage 1 and instantaneous channel gain matrix is $\mathbf{R}_g$ in Stage 2. By using the virtual eavesdropper instead of all the Eves, the two-level optimization SRM problem becomes a one-dimension optimization problem. Substituting (22) into (17), we can get

$$\begin{aligned}\max_{\mathbf{w}} \quad & \left\{\log_2\left(\alpha+\frac{\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2}\right)\right.\\ & \left.-\log_2\left(\beta_{\max}+\frac{\mathbf{w}^{\dagger}\mathbf{R}_g\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\sigma^2}\right)\right\}\\ \text{s.t.} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\leq P_R.\end{aligned} \tag{23}$$

The VE-SRM beamforming scheme relaxes the constraint of selecting the Eve with the highest eavesdropping rate, while the null-space beamforming scheme introduces a strong constraint that nulls out all signals at Eves. Thus, the VE-SRM beamforming design is a better solution than the null-space beamforming scheme. Since the logarithm function is monotonically increasing, remove it and obtain

$$\begin{aligned}\max_{\mathbf{w}} \quad & \left\{\frac{\alpha\sigma^2\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+\alpha\sigma^2+\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1}\,\frac{\mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}}{\beta_{\max}\mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\sigma^2+\mathbf{w}^{\dagger}\mathbf{R}_g\mathbf{w}}\right\}\\ \text{s.t.} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w}\leq P_R.\end{aligned} \tag{24}$$

Then problem (24) can be further simplified as

$$\max_{\mathbf{w}} \quad \left\{ \frac{\mathbf{w}^{\dagger}\widetilde{\mathbf{R}}_h\mathbf{w}}{\mathbf{w}^{\dagger}\widetilde{\mathbf{R}}_g\mathbf{w}} \frac{\mathbf{w}^{\dagger}\mathbf{R}_v\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{R}_u\mathbf{w}} \right\} \tag{25}$$

$$\text{s.t.} \quad \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w} \le P_R, \tag{26}$$

where $\widetilde{\mathbf{R}}_h \triangleq \alpha\sigma^2\mathbf{U} + (\alpha\sigma^2/P_R)\mathbf{T} + \mathbf{R}_a$, $\widetilde{\mathbf{R}}_g \triangleq \beta_{\max}\sigma^2\mathbf{T} + \mathbf{R}_g$, $\mathbf{R}_v \triangleq \mathbf{T}/P_R$, and $\mathbf{R}_u \triangleq \mathbf{U} + \mathbf{T}/P_R$. The objective function in (25) is a product of two GRQ problems, and it is a difficult problem in general. To simplify the analysis, a gradient-based method is used in order to obtain the updating equation for the beamforming vector. We denote the product of two GRQ problems in (25) by $\mathbf{J}(\mathbf{w})$, and its gradient with respect to $\mathbf{w}$ is

$$\begin{aligned}\nabla\mathbf{J}(\mathbf{w}) = {} & \frac{2}{\mathbf{w}^{\dagger}\widetilde{\mathbf{R}}_g\mathbf{w}}\left(\widetilde{\mathbf{R}}_h\mathbf{w} - \widetilde{\mathbf{R}}_g\mathbf{w}\frac{\mathbf{w}^{\dagger}\widetilde{\mathbf{R}}_h\mathbf{w}}{\mathbf{w}^{\dagger}\widetilde{\mathbf{R}}_g\mathbf{w}}\right)\frac{\mathbf{w}^{\dagger}\mathbf{R}_v\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{R}_u\mathbf{w}} \\ & + \frac{\mathbf{w}^{\dagger}\widetilde{\mathbf{R}}_h\mathbf{w}}{\mathbf{w}^{\dagger}\widetilde{\mathbf{R}}_g\mathbf{w}}\frac{2}{\mathbf{w}^{\dagger}\mathbf{R}_u\mathbf{w}}\left(\mathbf{R}_v\mathbf{w} - \mathbf{R}_u\mathbf{w}\frac{\mathbf{w}^{\dagger}\mathbf{R}_v\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{R}_u\mathbf{w}}\right).\end{aligned} \tag{27}$$

Considering the simple steepest gradient method [25], the equation for updating the weight vector regarding the function $\mathbf{J}(\mathbf{w})$ is

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu_k\nabla\mathbf{J}(\mathbf{w}(k)), \tag{28}$$

where $\mathbf{w}(k)$ denotes the weight vector at the kth iteration and μ_k is a real scalar step size at the kth iteration which is determined for the convergence of the iterative procedure. (28) is the updating equation for finding the new weight vector $\mathbf{w}(k+1)$. The convergence analysis of the gradient method has been proposed in [22].

As shown in Algorithm 1, we construct an iterative VE-SRM beamforming algorithm. The proposed algorithm requires an initial beamforming vector $\mathbf{w}(0)$. The first GRQ in (26) is related to the received signals both in Bob and Eves, while the second GRQ is mainly corresponding to the noise amplification at the relays. Therefore, we let the initial weight vector be $\rho\mathbf{v}$, where $\mathbf{v}$ is the unit-norm eigenvector of the matrix $\widetilde{\mathbf{R}}_g^{-1}\widetilde{\mathbf{R}}_h$ corresponding to the largest eigenvalue of the matrix and ρ is a scalar chosen to satisfy the power constraint and it equals

$$\rho = \sqrt{\frac{P_R}{\mathbf{v}^{\dagger}\mathbf{T}\mathbf{v}}}. \tag{29}$$

4. Relay Power Minimization Beamforming Design under Secrecy Rate Constraint

4.1. Existing Beamforming Scheme. In this subsection, we demonstrate the existing beamforming scheme under the secrecy rate constraint. In [19], a simple null-space RPM beamforming design was proposed when there were more relays than eavesdroppers. Under the assumption of no information leakage to all eavesdroppers in Stage 2, the associated RPM problem (8) can be simplified to

$$\begin{aligned}\min_{\mathbf{w}} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w} \\ \text{s.t.} \quad & \left(\mathbf{h}_{RE}^{j}\right)^{\dagger}\mathbf{w} = 0, \quad \forall E_j \in \mathscr{E} \\ & \frac{\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1} \ge \left(\beta_{\max}4^{\Gamma_0} - \alpha\right)\sigma^2.\end{aligned} \tag{30}$$

Under the secrecy rate constraint, the null-space RPM problem was studied in [19], where a closed-form solution was derived.

4.2. Relay Power Minimization Beamforming. In order to minimize the transmit power at relays under the secrecy rate constraint, an RPM beamforming scheme is developed. Problem (8) can be reformulated into

$$\begin{aligned}\min_{\mathbf{w}} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w} \\ \text{s.t.} \quad & \frac{\alpha + \mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}/\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2}{\max_{E_j\in\mathscr{E}}\left\{\beta_j + \mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}/\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2\right\}} \ge 4^{\Gamma_0}.\end{aligned} \tag{31}$$

Introducing a list of auxiliary slacking variables $\mathbf{r} \triangleq \{r_1, \ldots, r_J\}$, which satisfies

$$\beta_j + \frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2} \le r_j, \quad j = 1, 2, \ldots, J. \tag{32}$$

Then we can further rewrite problem (31) as

$$\begin{aligned}\min_{\mathbf{w},\mathbf{r}} \quad & \mathbf{w}^{\dagger}\mathbf{T}\mathbf{w} \\ \text{s.t.} \quad & \alpha + \frac{\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2} \ge r_j4^{\Gamma_0} \\ & \beta_j + \frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2} \le r_j, \quad \forall E_j \in \mathscr{E}.\end{aligned} \tag{33}$$

The equivalence of the problems (33) and (31) can be shown as follows. Let $\mathbf{w}$ be a feasible solution to (31), and we can get

$$\frac{\alpha + \mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}/\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2}{\max_{E_j\in\mathscr{E}}\left\{\beta_j + \mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}/\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2\right\}} \ge 4^{\Gamma_0}. \tag{34}$$

If we select

$$r_j \in \left[\beta_j + \frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}+1\right)\sigma^2}, \left(\alpha + \frac{\mathbf{w}^{\dagger}\mathbf{R}_a\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}+1\right)\sigma^2}\right)4^{-\Gamma_0}\right]. \tag{35}$$

Then $\{\mathbf{w}, r_1, \ldots, r_J\}$ is a feasible solution to (33). On the other hand, let $\{\mathbf{w}, r_1, \ldots, r_J\}$ denote a feasible solution to (33), it is easy to certify that $\mathbf{w}$ is also feasible for (31). Therefore problems (33) and (31) are equivalent.

4.2.1. Optimization of $\mathbf{w}$ *for Fixed* $\mathbf{r}$. Just as the SRM beamforming design, we introduce $\mathbf{W} \triangleq \mathbf{w}\mathbf{w}^{\dagger}$ based on the SDR idea and drop the nonconvex constraint $\operatorname{rank}(\mathbf{W}) = 1$. We rewrite (33) as

$$\begin{aligned} \min_{\mathbf{W}\succeq\mathbf{0}} \quad & \operatorname{Tr}(\mathbf{T}\mathbf{W}) \\ \text{s.t.} \quad & \operatorname{Tr}\left(\left(\alpha\sigma^2\mathbf{U} + \mathbf{R}_a - r_j 4^{\Gamma_0}\sigma^2\mathbf{U}\right)\mathbf{W}\right) \\ & \quad \geq \left(r_j 4^{\Gamma_0} - \alpha\right)\sigma^2 \\ & \operatorname{Tr}\left(\left(\beta_j\sigma^2\mathbf{V}_j + \mathbf{R}_{b,j} - r_j\sigma^2\mathbf{V}_j\right)\mathbf{W}\right) \\ & \quad \leq \left(r_j - \beta_j\right)\sigma^2, \quad \forall E_j \in \mathscr{E}. \end{aligned} \tag{36}$$

The advantage of (36) is that it is an SDP problem and can be optimally solved [22]. Note that the obtained matrix $\mathbf{W}$ might not be rank 1; therefore, (36) is a relaxation of (34) for fixed $\mathbf{r}$. If the rank of the solution $\mathbf{W}$ is larger than 1, the Gaussian Randomization method could be employed to find an approximated solution.

4.2.2. Optimization Over $\mathbf{r}$. Denote the objective value of (36) as a function of $\mathbf{r}$ by $\mathbb{P}_{\mathbf{r}}$. It can be seen that solving the original problem (33) is equivalent to minimizing $\mathbb{P}_{\mathbf{r}}$ with respect to $\mathbf{r}$. The optimization of $\mathbf{W}$ has been embedded in evaluating $\mathbb{P}_{\mathbf{r}}$. Since r_j is a nonnegative variable with a range shown in (35), the optimal solution can be found by global optimization algorithms [26]. However, we will study the analytical property of $\mathbb{P}_{\mathbf{r}}$ to devise more efficient algorithms. In the following we focus on finding the optimal $\mathbf{r}$ to minimize $\mathbb{P}_{\mathbf{r}}$.

Theorem 1. $\mathbb{P}_{\mathbf{r}}$ *is a convex function of* $\mathbf{r}$.

Proof. The Lagrangian of problem (36) is

$$\begin{aligned} & L(\mathbf{W}, \mathbf{\Phi}, \mathbf{r}, \mathbf{y}, \mathbf{z}) \\ & \quad = \operatorname{Tr}(\mathbf{T}\mathbf{W}) - \operatorname{Tr}(\mathbf{\Phi}\mathbf{W}) \\ & \qquad - \sum_{E_j\in\mathscr{E}} y_j \left\{\operatorname{Tr}\left(\left(\alpha\sigma^2\mathbf{U} + \mathbf{R}_a - r_j 4^{\Gamma_0}\sigma^2\mathbf{U}\right)\mathbf{W}\right)\right. \\ & \qquad\qquad \left. - \left(r_j 4^{\Gamma_0} - \alpha\right)\sigma^2\right\} \\ & \qquad + \sum_{E_j\in\mathscr{E}} z_j \left\{\operatorname{Tr}\left(\left(\beta_j\sigma^2\mathbf{V}_j + \mathbf{R}_{b,j} - r_j\sigma^2\mathbf{V}_j\right)\mathbf{W}\right)\right. \\ & \qquad\qquad \left. - \left(r_j - \beta_j\right)\sigma^2\right\}, \end{aligned} \tag{37}$$

where $\mathbf{\Phi} \succeq \mathbf{0}, \mathbf{y}, \mathbf{z} \geq 0$ are dual variables associated with the constraints in (36).

The Lagrange dual problem can be written as

$$\begin{aligned} \max_{\mathbf{y},\mathbf{z}} \quad & \sum_{E_j\in\mathscr{E}} \left\{y_j\left(r_j 4^{\Gamma_0} - \alpha\right)\sigma^2 + z_j\left(\beta_j - r_j\right)\sigma^2\right\} \\ \text{s.t.} \quad & \mathbf{T} + \sum_{E_j\in\mathscr{E}} \left\{z_j\left(\beta_j\sigma^2\mathbf{V}_j + \mathbf{R}_{b,j} - r_j\sigma^2\mathbf{V}_j\right)\right. \\ & \qquad \left. - y_j\left(\alpha\sigma^2\mathbf{U} + \mathbf{R}_a - r_j 4^{\Gamma_0}\sigma^2\mathbf{U}\right)\right\} \succeq \mathbf{0}. \end{aligned} \tag{38}$$

Due to the strong duality between the primary problem (36) and the dual problem (38), the objective value of (38) is also $\mathbb{P}_{\mathbf{r}}$. For given $\mathbf{y}, \mathbf{z}$, the objective function of (38) is linear and convex about $\mathbf{r}$; therefore, its maximum over $\{\mathbf{y}, \mathbf{z}\}$, $\mathbb{P}_{\mathbf{r}}$, is convex about $\mathbf{r}$ [22]. This completes the proof. □

Since $\mathbb{P}_{\mathbf{r}}$ is a convex function, a gradient-based algorithm can find the global optimum of $\mathbb{P}_{\mathbf{r}}$, and thus greatly reduces the complexity. We have the following straightforward results to facilitate the design of gradient-based algorithm [27].

The derivative of $\mathbb{P}_{\mathbf{r}}$ with respect to $\mathbf{r}$ is given by

$$\begin{aligned} \frac{\partial\mathbb{P}_{\mathbf{r}}}{\partial r_j} &= \frac{\partial L}{\partial r_j} \\ &= y_j 4^{\Gamma_0}\sigma^2\left(\operatorname{Tr}(\mathbf{U}\mathbf{W} + 1)\right) - z_j\sigma^2\left(\operatorname{Tr}\left(\mathbf{V}_j\mathbf{W}\right) + 1\right). \end{aligned} \tag{39}$$

The main iterative algorithm can be summarized as follows. For given $\mathbf{r}$, solve both the primary problem (36) and the Lagrange dual problem (38) to obtain the primary variable $\mathbf{W}$ and dual variables $\mathbf{y}, \mathbf{z}$. Then update $\mathbf{r}$ using the gradient-based algorithm in (39). The overall iterative algorithm to solve (31) is given in Algorithm 2.

4.3. Virtual Eavesdropper-Based Relay Power Minimization Beamforming. Similarly to the VE-SRM beamforming design, we propose a practical VE-RPM beamforming scheme to simplify the complexity of the RPM beamforming scheme. The eavesdroppers constraint in (31) can be relaxed as

$$\begin{aligned} & \max_{E_j\in\mathscr{E}} \left\{\beta_j + \frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w} + 1\right)\sigma^2}\right\} \\ & \qquad \leq \beta_{\max} + \sum_{E_j\in\mathscr{E}} \frac{\mathbf{w}^{\dagger}\mathbf{R}_{b,j}\mathbf{w}}{\left(\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w} + 1\right)\sigma^2}. \end{aligned} \tag{40}$$

Similar to (20),

$$\frac{\mathbf{w}^{\dagger}\mathbf{V}_j\mathbf{w}}{\mathbf{w}^{\dagger}\mathbf{U}\mathbf{w}} \geq \lambda_{\min}\left(\mathbf{U}^{-1}\mathbf{V}_j\right), \quad \forall E_j \in \mathscr{E}. \tag{41}$$

(1) **Initial State:**
Initialize $\widetilde{\mathbf{R}}_h$, $\widetilde{\mathbf{R}}_g$, $\mathbf{R}_v$, $\mathbf{R}_u$ and the initial beamforming vector $\mathbf{w}(0)$.
(2) **Iterative VE-SRM Beamforming Algorithm**:
(i) Clock index $k = 0$.
(ii) **Repeat**
(a) *Linear search*. Choose step size μ_k via backtracking line search [22].
(b) *Update*. $\mathbf{w}(k+1) = \mathbf{w}(k) + \mu_k \nabla \mathbf{J}(\mathbf{w}(k))$.
(c) *Calculate*. $\mathbf{J}(\mathbf{w}(k+1))$.
(d) *Increase*. clock index $k = k + 1$.
(iii) **Until** $\mathbf{J}(\mathbf{w}(k)) \leq \mathbf{J}(\mathbf{w}(k-1))$ and $\mu_k < \mu_{\text{thresh}}$.
(3) **Secure Transmission:**
Distribute the beamforming weights to the relays, and begin the secure data transmission.

Algorithm 1: The proposed iterative VE-SRM beamforming algorithm.

(1) **Initial State:**
Initialize α, $\mathbf{R}_a$, β_j, $\mathbf{R}_{b,j}$, $\forall E_j \in \mathscr{E}$.
(2) **Iterative RPM Beamforming Algorithm**:
(i) Clock index $k = 0$.
(ii) **Repeat**
(a) *Calculate*. calculate $\mathbf{W}$ in (34) for a fixed $\mathbf{r}_k$
(b) *Update*. update $\mathbf{r}_{k+1}$ using the gradient-based algorithm in (39).
(c) *Increase*. clock index $k = k + 1$.
(iii) **Until** convergence.
(3) **Secure Transmission:**
Distribute the beamforming weights to the relays, and begin the secure data transmission.

Algorithm 2: The proposed iterative RPM beamforming algorithm.

The formula in (40) can be rewritten as

$$\begin{aligned} &\leq \beta_{\max} + \sum_{E_j \in \mathscr{E}} \frac{\mathbf{w}^\dagger \mathbf{R}_{b,j} \mathbf{w}}{\lambda_{\min}\left(\mathbf{U}^{-1}\mathbf{V}_j\right)\left(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1/\lambda_{\min}\left(\mathbf{U}^{-1}\mathbf{V}_j\right)\right)\sigma^2} \\ &\leq \beta_{\max} + \frac{\mathbf{w}^\dagger \mathbf{R}_g \mathbf{w}}{\left(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1\right)\sigma^2}, \end{aligned} \tag{40}$$

$$\lambda_j \triangleq \begin{cases} \lambda_{\min}\left(\mathbf{U}^{-1}\mathbf{V}_j\right), & \lambda_{\min}\left(\mathbf{U}^{-1}\mathbf{V}_j\right) < 1, \\ 1, & \lambda_{\min}\left(\mathbf{U}^{-1}\mathbf{V}_j\right) \geq 1, \end{cases} \tag{42}$$

where

and $\mathbf{R}_g \triangleq \sum_{E_j \in \mathscr{E}} (\mathbf{R}_{b,j}/\lambda_j)$. Through the relaxation, it seems that there was a virtual eavesdropper, whose received SNR is

$$\text{SNR}_{\text{VE}} = \beta_{\max} - 1 + \frac{\mathbf{w}^\dagger \mathbf{R}_g \mathbf{w}}{\left(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1\right)\sigma^2}, \tag{44}$$

and the relay power minimization problem (31) under the secrecy rate constraint can be expressed as

$$\begin{aligned} &\min_{\mathbf{w}} \quad \mathbf{w}^\dagger \mathbf{T}\mathbf{w} \\ &\text{s.t.} \quad \frac{\alpha + \mathbf{w}^\dagger \mathbf{R}_a \mathbf{w}/\left(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1\right)\sigma^2}{\beta_{\max} + \mathbf{w}^\dagger \mathbf{R}_g \mathbf{w}/\left(\mathbf{w}^\dagger \mathbf{U}\mathbf{w} + 1\right)\sigma^2} \geq 4^{\Gamma_0}. \end{aligned} \tag{45}$$

Note that the minimum relay power is achieved when the constraint in (45) is an equality; that is,

$$\begin{aligned} &\min_{\mathbf{w}} \quad \mathbf{w}^\dagger \mathbf{T}\mathbf{w} \\ &\text{s.t.} \quad \frac{\mathbf{w}^\dagger\left(\mathbf{R}_a + \alpha\sigma^2\mathbf{U}\right)\mathbf{w} + \alpha\sigma^2}{\mathbf{w}^\dagger\left(\mathbf{R}_g + \beta_{\max}\sigma^2\mathbf{U}\right)\mathbf{w} + \beta_{\max}\sigma^2} = 4^{\Gamma_0}. \end{aligned} \tag{46}$$

Defining $\mathbf{w} = \mathbf{T}^{-1/2}\widetilde{\mathbf{w}}$, the optimization problem in (46) is equivalent to

$$\begin{aligned} &\min_{\widetilde{\mathbf{w}}} \quad \widetilde{\mathbf{w}}^\dagger \widetilde{\mathbf{w}} \\ &\text{s.t.} \quad \widetilde{\mathbf{w}}^\dagger \widetilde{\mathbf{R}} \widetilde{\mathbf{w}} = \varsigma, \end{aligned} \tag{47}$$

where

$$\widetilde{\mathbf{R}} \triangleq \mathbf{T}^{-1/2}\left(\mathbf{R}_a + \alpha\sigma^2\mathbf{U} - 4^{\Gamma_0}\mathbf{R}_g - 4^{\Gamma_0}\beta_{\max}\sigma^2\mathbf{U}\right)\mathbf{T}^{-1/2},$$
$$\varsigma \triangleq \left(\beta_{\max}4^{\Gamma_0} - \alpha\right)\sigma^2. \quad (48)$$

Note that if Γ_0 is chosen such that $\varsigma > 0$ but $\widetilde{\mathbf{R}}$ is negative definite, then the optimization problem in (46) will be infeasible and the transmit power in Stage 2 is zero. In the simulation section we will discuss the infeasibility probabilities of different beamforming schemes. If $\varsigma \leq 0$, the secrecy rate constraint is satisfied in Stage 1, and there is no need to waste power in Stage 2. Above all, let us assume that $\varsigma > 0$ and $\widetilde{\mathbf{R}}$ is positive definite. By using the method of Lagrange multipliers, we obtain

$$\widetilde{\mathbf{R}}\widetilde{\mathbf{w}} = \frac{1}{\lambda}\widetilde{\mathbf{w}}, \quad (49)$$

where λ is the Lagrange multiplier. Multiplying both sides of (49) with $\lambda\widetilde{\mathbf{w}}^\dagger$ yields

$$\widetilde{\mathbf{w}}^\dagger\widetilde{\mathbf{w}} = \lambda\widetilde{\mathbf{w}}^\dagger\widetilde{\mathbf{R}}\widetilde{\mathbf{w}} = \lambda\varsigma. \quad (50)$$

Then minimizing $\widetilde{\mathbf{w}}^\dagger\widetilde{\mathbf{w}}$ is equivalent to minimizing λ, so $1/\lambda$ corresponds to the largest eigenvalue of $\widetilde{\mathbf{R}}$, and thus $\widetilde{\mathbf{w}}$ is related to the largest eigenvector of $\widetilde{\mathbf{R}}$. Finally, the solution of (46) is

$$\mathbf{w} = \mu\mathbf{T}^{-1/2}\mathbf{v}_{\max}, \quad (51)$$

where $\mathbf{v}_{\max}$ is the largest unit-norm eigenvector of $\widetilde{\mathbf{R}}$ and the scalar μ is

$$\mu = \sqrt{\frac{\varsigma}{\mathbf{v}_{\max}^\dagger\widetilde{\mathbf{R}}\mathbf{v}_{\max}}}. \quad (52)$$

5. Simulation Results

In this section, we carry out some simulations to evaluate the secrecy rate performance of the proposed designs. For simplicity and without loss of generality, we consider a cooperative M2M network with N relays located in a (20 m × 20 m) square area whose center is (0 m, 0 m). Alice and Bob are located at (−20 m, 0 m) and (20 m, 0 m), respectively. The eavesdroppers are located in a 10 m × 10 m square area surrounding Bob. The network scenario is shown in Figure 2. The other simulation parameters are set up as follows. The noise variance is normalized; that is, $\sigma^2 = 1$. We also assume that the power transmitted in Stage 1 is 10 dB; that is, $P_S = 10$ dB. Rayleigh fading channel is assumed, where the channel gain consists of both the path loss (i.e., free-space loss) and the Rayleigh fading coefficient. The path loss factor is 3.5 [28]. The Monte Carlo simulations consisting of 1000 independent trails are performed to obtain the average results.

In Figure 3, we show the secrecy rates of different schemes versus the total transmit power budget when the number of relays is 8. The number of Eves J is 2 or 7. The figure clearly demonstrates the advantages of the proposed schemes. It can be observed that the performance of SRM beamforming scheme is always best in the three schemes, and its performance is almost the same as the upper bound of the optimal beamforming scheme when the number of relays is larger than that of Eves. Due to the relaxation in the VE-SRM beamforming scheme, there is a performance degradation compared to the SRM beamforming design. The performance of VE-SRM beamforming scheme is still better than that of the null-space beamforming scheme, especially when the number of Eves approaches the number of relays. It is because that the degree of freedom (DoF) of the null-space beamforming design becomes small as the number of eavesdroppers increases. When the number of Eves is larger than that of relays, there will be no extra DoF for the null-space beamforming scheme.

Figure 4 illustrates the relationship between the secrecy rate and the number of Eves. The relay number N is 8. The source power P_0 is set to 15 dB or 20 dB. The proposed SRM beamforming scheme approaches the upper bound of

Figure 2: Network scenario for simulations.

Figure 3: Secrecy rate versus total power constraint.

Figure 4: Secrecy rate versus number of Eves.

Figure 5: Relay transmit power versus secrecy rate constraint.

optimal beamforming even if the number of Eves is larger than that of relays. Also, the proposed designs exhibit great robust performance compared to the null-space beamforming scheme. As shown in Figure 4, the null-space method suffers a serious secrecy rate degradation compared to the proposed designs. When the number of Eves approaches that of the relays, the secrecy rate of the null-space beamforming drops quickly, while the proposed beamforming schemes can still yield consistent secrecy rate even when the number of Eves exceeds that of the relays.

Figure 5 shows the results for the relay transmit power against the secrecy rate constraint. The relay number N is 8. The number of Eves is set to 2, 4, or 6. It is observed

Figure 6: Infeasibility probability versus number of Eves.

that the relays require more and more power for a fixed secrecy rate as the number of Eves increases. However, the performance of the proposed beamforming schemes, RPM and EV-RPM beamforming schemes, is more efficient than the performance of the null-space beamforming scheme. The gaps between the proposed designs and the null-space beamforming scheme grows as the number of Eves increases, because the DoF has a serious impact on the null-space approach but less impact on our beamforming methods.

In Figure 6, we illustrate the infeasibility probability against the number of Eves. We assume that there are $N = 8$ relays and the secrecy rate constraints are $\Gamma_0 = 0.5$ bps/Hz or $\Gamma_0 = 1$ bps/Hz. Here, the infeasibility probability is obtained by counting the failed number of trials among the 100000 channel realizations that each algorithm is able to find the solution to meet the secrecy rate constraint. It can be seen that all the schemes will not work sometimes when secrecy rate target Γ_0 is large or the number of Eves is large. It is also shown that the proposed beamforming schemes are feasible under the considered secrecy rate constraint. Here, the reason is that the proposed schemes are not forced to null out all the signal at all Eves.

6. Conclusion

In this paper, we have considered an AF-based cooperative protocol to improve the performance of secure wireless M2M communications in the presence of multiple eavesdroppers. The optimization criterion is the maximization of the achievable secrecy rate subject to a transmit power constraint or the minimization of the transmit power subject to a secrecy rate constraint. In previous works, the null-space beamforming schemes were imposed, at the expense of optimality, in order to obtain a simple solution. In our works, we propose beamforming schemes that do not rely on the null-space

conditions, and thus our proposed beamforming designs can work well even if the number of eavesdroppers is larger than that of relays. Moreover, in the cooperative M2M network we have assumed that the global CSI is available, and even the eavesdroppers channels are known. System design of cooperative schemes that uses partial CSI or channel statistics will be considered in the future work.

Acknowledgments

This work has been supported by the Specialised Research Fund for the Doctoral Program of the Ministry of Education of China (Grant no. 20120001120125) and the National Natural Science Foundation of China (Grant nos. 61250001, 91338106, 61231011, and 61231013).

References

[1] "Evolved Universal Terrestral Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Netwoek (EUTRAN)," 3GPP TS 36. 300 V10.0.0, 2010.

[2] Q. Bi, G. I. Zysman, and H. Menkes, "Wireless mobile communications at the start of the 21st century," *IEEE Communications Magazine*, vol. 39, no. 1, pp. 110–116, 2001.

[3] J. Xie, J. Zhang, and L. Bai, "Multilayer orthogonal beamforming for priority-guaranteed wireless communications," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 307467, 8 pages, 2012.

[4] C. Chen, L. Liu, X. Du, Q. Pei, and X. Zhao, "Improving driving safety based on safe distance design in vehicular sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 469067, 13 pages, 2012.

[5] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[6] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, 1978.

[7] P. Parada and R. Blahut, "Secrecy capacity of SIMO and slow fading channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '05)*, pp. 2152–2155, Adelaide, Australia, September 2005.

[8] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: the MISOME wiretap channel," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3088–3104, 2010.

[9] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '08)*, pp. 524–528, Ontario, Canada, July 2008.

[10] Y. Liang, H. V. Poor, and S. Shamai, "Secure communication over fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2470–2492, 2008.

[11] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4687–4698, 2008.

[12] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2453–2469, 2008.

[13] Y. Liang and H. V. Poor, "Generalized multiple access channels with confidential messages," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 952–956, Seattle, Wash, USA, July 2006.

[14] R. Liu, T. Liu, H. V. Poor, and S. Shamai, "Multiple-input multiple-output gaussian broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4215–4227, 2010.

[15] E. Tekin and A. Yener, "The Gaussian multiple access wire-tap channel," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5747–5755, 2008.

[16] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Secure wireless communications via cooperation," in *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1132–1138, Urbana Champaign, Ill, USA, September 2008.

[17] D. Wang, L. Bai, X. Zhang, W. Guan, and C. Chen, "Collaborative relay beamforming strategies for multiple destinations with guaranteed QoS in wireless machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 525640, 10 pages, 2012.

[18] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Amplify-and-forward based cooperation for secure wireless communications," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 2613–2616, Taipei, Taiwan, April 2009.

[19] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Improving wireless physical layer security via cooperating relays," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1875–1888, 2010.

[20] Y. Yang, Q. Li, W. K. Ma, J. Ge, and P. C. Ching, "Cooperative secure beamforming for AF relay networks with multiple eavesdroppers," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 35–38, 2013.

[21] Z. Q. Luo, W. K. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[23] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z. Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4306–4316, 2008.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1987.

[25] S. Haykin, *Adaptive Filter Theory*, Pearson Education, Upper Saddle River, NJ, USA, 4th edition, 2002.

[26] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *Journal of Optimization Theory and Applications*, vol. 79, no. 1, pp. 157–181, 1993.

[27] J. Nocedal and S. J. Wright, *Numerical Optimization*, Operations Research and Financial Engineering, Springer, Berlin, Germany, 2nd edition, 2006.

[28] R. Zhang, L. Song, Z. Han, and B. Jiao, "Distributed coalition formation of relay and friendly jammers for secure cooperative networks," in *Proceedings of the IEEE International Conference on Communications (ICC '11)*, Kyoto, Japan, June 2011.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 475761, 9 pages
http://dx.doi.org/10.1155/2013/475761

Research Article

SDMA-Based Aeronautical Machine-to-Machine Communications under SINR Constraints

Jindong Xie, Lin Bai, Jun Zhang, and Tao Zhang

School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Lin Bai; l.bai@buaa.edu.cn

Received 13 August 2013; Accepted 18 November 2013

Academic Editor: Jianhua He

Copyright © 2013 Jindong Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to achieve an efficient usage on a limited spectrum for aeronautical machine-to-machine (M2M) communications, in this paper, a space division multiple access- (SDMA-) based architecture is proposed, where the random access of short messages noisy (SMN) M2M communications is considered between autopilot systems onboard and the ground control center for the air traffic precise control, that is, full 4-dimensional trajectory management. It shows that using a zero-forcing beamforming combined with a low complexity aircraft grouping strategy, the SDMA-based mechanism is able to improve the spectrum efficiency under signal to interference plus noise ratio (SINR) constraints.

1. Introduction

To ensure the continuous safety and efficiency improvement of air traffic flows, the accurate trajectory surveillance and control abilities are highly required for the advanced air traffic control (ATC), that is, 4-dimensional trajectory- (4DT-) based management (The "4D-trajectory management" concept improves the ATC by using "time" with early traffic sequencing based on the aircrafts way-point passage predictions. It is planned to be deployed from 2017 and allowed to enter into services beyond 2020), which has already been scheduled and will be employed in the following decade [1, 2]. Therefore, the future air to ground (A/G) datalink is foreseen to support continuous interactions between the flight management system (FMS) onboard and the ground ATC system for real-time automatic monitoring and control [3–6]. Since an ATC message payload generally contains only hundreds of bits (e.g., position coordinates and timestamps) [7] and needs to be updated frequently (e.g., 16 times per second) [8, 9], future aeronautical communications can be considered as typical short messages noisy (SMN) machine-to-machine (M2M) communications.

In order to support 4DT services, the next generation aeronautical datalinks will be developed by taking into account the SMN-M2M communications. For the sake of safety-of-life related issues [10], two crucial characteristics have to be considered as follows:

(i) all aircraft should maintain seamless connections with the control center;

(ii) each connection has to support a primary rate for real-time interactions.

Here, a primary rate is defined as the basic achievable data rate to provide essential communication services for 4DT management.

Unfortunately, the existing aeronautical datalinks may not be able to satisfy both two objectives from above efficiently [11, 12]. On the other hand, considering the limited aeronautical spectrum [13], the spectrum efficient strategy of aeronautical M2M communications can be reconfigured under the primary rate constraint. In summary, the future aeronautical M2M communication system can be demonstrated as follows:

(i) a spectrum should be reused by multiple airborne station (AS) nodes;

(ii) aircraft could transmit or receive short messages at real time, without extra resource allocations;

(iii) a primary rate should be guaranteed for reliable M2M communications.

Since an interval in distance has to be guaranteed between different aircrafts (e.g., 4~8 nautical miles separation along with a flight course [14]), space division multiple access (SDMA) could be an efficient means [15, 16] to improve the aeronautical spectrum efficiency. Based on SDMA, a spectrum will be multiplexed by multiple AS nodes naturally. Also, real-time A/G interactions could be supported by SDMA assisted random access. With the help of SDMA, potential conflicts of random access will be mitigated easily.

In particular, beamforming over multiple-antenna systems can be an efficient approach to perform the SDMA [17]. Among different beamforming techniques, dirty paper coding (DPC) [18] has been developed to provide the optimal performance by separating combined spatially selective signals. However, DPC is impractical due to its prohibitively high complexity.

As a low complexity linear beamforming approach, orthogonal beamforming [19] is able to support hierarchical priority guaranteed SDMA in M2M communications over cellular networks [20], while it may not be appropriate for aeronautical SMN-M2M communications as an equal, performance of different 4DT management services is required on safety-of-life level. Combined with a semiorthogonal users selection (SUS) scheme [21], zero-forcing beamforming (ZFBF) has been proposed in [22] to provide equitable performance for all selected users. Thus, we may consider ZFBF as a good candidate technique for SDMA-based aeronautical M2M communications. However, unfortunately, SUS cannot achieve the SNR or rates to a certain desired level at each iteration of user selection. Therefore, ZFBF-SUS cannot guarantee the reliability of connections with the safety-of-life related 4DT management.

In this paper, we propose an SDMA-based approach for the future L-band continental A/G communication systems [23, 24]. By using ZFBF over antenna arrays equipped in a ground station (GS), one frequency spectrum is able to support M2M random accesses without conflicts for multiple AS nodes, simultaneously. Moreover, in order to guarantee the primary rate for A/G communications, we propose a low complexity user selection strategy for the ZFBF under signal to interference plus noise ratio (SINR) constraints to exploit the SDMA gain.

The rest of the paper is organized as follows. Section 2 provides the system model. Our proposed SDMA-based mechanism for aeronautical M2M communications is introduced in Section 3. After the impacts of feedback reliability discussion in Section 4, the simulation results are presented in Section 5 to analyze the performance. Finally, we conclude this paper in Section 6.

Notation. The superscripts T and H stand for the transpose and Hermitian transpose, respectively. Denote by $\|\cdot\|$ the 2-norm, and $|\cdot|$ is the total elements number of a certain set. The statistical expectation is represented by $\mathbb{E}[\cdot]$. The distribution of a circularly symmetric complex Gaussian (CSCG) random variable with mean a and covariance b is denoted by $\mathcal{CN}(a,b)$. Besides, $\mathbf{I}_{(M\times N)}$ stands for an $M\times N$ identity matrix, and $\mathbf{I}_M = \mathbf{I}_{(M\times M)}$.

Figure 1: The geometry model for A/G communications over the GS equipped with multiple antennas array in a macrocell.

2. System Model

Consider a section of the continental A/G communication system with K aircraft randomly located in a macrocell. Let $\mathcal{K} = \{1, 2, \ldots, K\}$ be the index set of all AS nodes. We suppose that each AS is equipped with a single antenna, while the GS is equipped with a uniform linear array (ULA) of M elements. Due to the reciprocity, we only focus on the reverse link of A/G communications (i.e., from AS to GS), while the main results are easily extended to forward link scenarios.

Let the index of a certain spectrum be $f \in \mathcal{F} = \{1, 2, \ldots, F\}$ with $F \le K$. Assume that AS-m and AS-n are assigned as a group to access the fth spectrum, simultaneously. The resulting geometry model can be described in Figure 1.

Let $\Theta_{\mathcal{S}_f} = \{\theta_m, \theta_n\}$ represent the angle of arrival (AoA) set of AS-m and AS-n with respect to the array normal, where $\mathcal{S}_f = \{m, n\}$ denotes the set of all selected AS nodes in group f. For the sake of simplicity, we do not consider the effect of azimuth angles in the paper, since the ULA is a 2-dimensional distinguish array and insensitive to the azimuth angle (A 3-dimensional distinguish array such as a circular or planar array can be considered as an extension of this work, where the azimuth angles are taken into account).

In aeronautical A/G communication scenario, a wireless channel is generally characterized by a Rician fading with a strong line of sight (LoS) component and a weak diffusion path [25, 26]. If the Rician factor is given by $K_{\text{Rice}} = 15\,\text{dB}$ in a typical en-route scenario [24], the contribution of LoS component becomes far greater than that of the diffusion part. Consequently, an additive white Gaussian noise (AWGN) LoS channel can be used to sketch the A/G channel, which is further modeled as an array response vector (ARV) since an antenna array is employed at the GS.

Considering all M elements of the ULA are placed at the direction of $\mathcal{Z}$-axis, the ARV of the AS-m is given by

$$\mathbf{a}_m = \mathbf{a}(\theta_m) = \begin{bmatrix} 1 \\ e^{-j2\pi(d/\lambda)\cos\theta_m} \\ \vdots \\ e^{-j2\pi(M-1)(d/\lambda)\cos\theta_m} \end{bmatrix}, \quad (1)$$

where d is the antenna spacing and λ denotes the wavelength of the carrier. Since AoAs of AS nodes could be seen unchanged within a short period and can be easily obtained

from positioning systems or efficient AoA estimations, it can be assumed that the ARV of each AS is constant over the duration of a time slot and is well known by the GS.

Denote by $N_f = |\mathcal{S}_f|$ is the total number of selected AS nodes on the fth spectrum resource, for example, $N_f = 2$ in Figure 1. Moreover, we have $\mathcal{S}_1 \cup \cdots \mathcal{S}_f \cdots \cup \mathcal{S}_F = \mathcal{K}$ and $N_1 + \cdots + N_f + \cdots + N_F = K$, where F is the total number of required frequency resources.

Let $\mathbf{H}_{\mathcal{S}_f} = [\mathbf{a}_m, \mathbf{a}_n] \in \mathcal{C}^{M\times N_f}$ and let $\mathbf{S}_{\mathcal{S}_f} = [s_m, s_n]^T \in \mathcal{C}^{N_f\times 1}$ be the channel matrix and expected signal vector from AS nodes in $\mathcal{S}_f$ to GS, respectively. Then the received signal vector $\mathbf{y}_f$ at the GS over spectrum f is given by

$$\mathbf{y}_f = \mathbf{H}_{\mathcal{S}_f}\mathbf{S}_{\mathcal{S}_f} + \mathbf{z}_f, \tag{2}$$

where $\mathbf{z}_f$ is the additive white Gaussian noise and $\mathbf{z}_f \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_{N_f})$.

The transmission power of s_i from AS-i is assumed to be identical for different i; that is, $\mathbb{E}[|s_i|^2] = P_t$. However, due to the propagation path loss, the received powers could be different. Since the channels between AS and GS include strong LoS components, for the sake of simplicity, the received power can be described using the following simple path-loss model [27]:

$$\frac{1}{\gamma} = \frac{P_t}{P_r} = \left(\frac{4\pi l_0}{\lambda}\right)^2\left(\frac{l}{l_0}\right)^n, \tag{3}$$

where P_t and P_r denote transmit and receive power, respectively. In (3), λ is propagation wavelength, n is the path-loss exponent, l is the distance between the transmitter and receiver, and l_0 is a close-in reference distance. Since λ, n, and l_0 are considered as constants, the path-loss is a function of l with the known parameters λ, n, and l_0.

3. Primary Rate Guaranteed SDMA

In this section, the main solution of SDMA-based mechanism for aeronautical M2M communications is presented. As aforementioned, antenna array beamforming and multiple users selection are coupled issues. Lets first consider the beamforming problem.

3.1. Zero-Forcing Beamforming. Let

$$\begin{aligned}\mathbf{W}_f &= \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{N_f}\right] \\ &= \left[\frac{\widetilde{\mathbf{w}}_1}{\|\widetilde{\mathbf{w}}_1\|}, \frac{\widetilde{\mathbf{w}}_2}{\|\widetilde{\mathbf{w}}_2\|}, \ldots, \frac{\widetilde{\mathbf{w}}_{N_f}}{\|\widetilde{\mathbf{w}}_{N_f}\|}\right],\end{aligned} \tag{4}$$

where $\widetilde{\mathbf{w}}_i$ is the ith column vector of $\mathbf{H}_{\mathcal{S}_f}(\mathbf{H}^{\mathrm{H}}_{\mathcal{S}_f}\mathbf{H}_{\mathcal{S}_f})^{-1}$, that is, the Hermitian transpose of the Moore-Penrose pseudoinverse of $\mathbf{H}_{\mathcal{S}_f}$.

Since $\mathbf{W}_f$ is generated by the pseudoinverse of $\mathbf{H}^{\mathrm{H}}_{\mathcal{S}_f}$, the column vector $\mathbf{w}_i$ in $\mathbf{W}_f$ becomes orthogonal with the column vectors $\mathbf{a}_j$ in $\mathbf{H}_{\mathcal{S}_f}$, when $i \neq j$ and $1 \leq \{i, j\} \leq N_f$. According to Figure 2, the estimated signal of the ith selected user in $\mathcal{S}_f$ is given by

Figure 2: The ZFBF beamforming for A/G communications.

$$\widetilde{s}_i = \mathbf{w}^{\mathrm{H}}_{k_i}\mathbf{y}_f = \frac{\sqrt{\gamma_{k_i}P_t}}{\|\widetilde{\mathbf{w}}_i\|}s_i, \tag{5}$$

where γ_k is the path-loss parameter in (3) with respect to AS-k. Besides, $1/\|\widetilde{\mathbf{w}}_i\|$ is termed as equivalent channel gain (ECG). The received SINR expression of the ith selected user in $\mathcal{S}_f$ can be described as

$$\mathrm{SINR}_{k_i} = \frac{\gamma_{k_i}P_t}{\|\widetilde{\mathbf{w}}_i\|^2\sigma^2}, \tag{6}$$

where $k_i \in \mathcal{K}$ is the index of the ith selected AS in $\mathcal{S}_f$. Then the achievable rate of AS-k_i is given by

$$r_{k_i} = \log_2\left(1 + \frac{\gamma_{k_i}P_t}{\|\widetilde{\mathbf{w}}_i\|^2\sigma^2}\right). \tag{7}$$

Since ATC services are safety-of-life related, the proposed system should be able to provide a primary rate for each selected AS node. Here, the primary rate is defined as

$$\overline{r} = \log_2\left(1 + \overline{\mathrm{SNR}}\right), \tag{8}$$

where $\overline{\mathrm{SNR}}$ is the threshold SNR of an edge-located AS node without performing the SDMA. Then SINR constraints can be applied to all selected AS nodes as follows:

$$\begin{aligned} r_{k_i} &\geq \overline{r}, \\ \mathrm{SINR}_{k_i} &\geq \overline{\mathrm{SNR}}. \end{aligned} \tag{9}$$

By taking into account the path loss in a macrocell, assuming all the AS have the same noise variance (i.e., $\sigma^2 = 1$), the threshold transmitting power is obtained by

$$\overline{P}_t = \left(\frac{4\pi l_0}{\lambda}\right)^2\left(\frac{R}{l_0}\right)^n\overline{\mathrm{SNR}}, \tag{10}$$

where $\overline{\mathrm{SNR}}$, R, l_0, λ, and n are given.

Figure 3: The received SNR with $\overline{P}_t$ transmitting power versus various potential AoAs, where $\overline{\text{SNR}} = 0\,\text{dB}$, $l_0 = 1\,\text{m}$, $\lambda = 0.3\,\text{m}$, and $n = 2.0$.

Apparently, $\overline{P}_t$ is essential and desired by all AS transmitters, since each AS always fights over the edge of a macrocell. However, if one AS transmits signals using power of $P_t = \overline{P}_t$, the received SNR, which is affected by the path loss, can be described in Figure 3.

As Figure 3 illustrates, the received SNR of AS nodes with smaller absolute AoAs is higher than those nodes located at the edge area of a macrocell. Since $\overline{\text{SNR}}$ is sufficiently large to hold the primary rate, the higher received SNR of inner AS nodes becomes unnecessary with respect to supporting the 4DT management services. Therefore, the conventional power control schemes have shown the disadvantages for the sake of the spectral efficiency.

Using ZFBF, a spectrum can be multiplexed among inner and fringe AS nodes. In order to measure the advantage of ZFBF-SDMA, the improvement of performance can be described as an SDMA gain [28–30]. Let the SDMA gain be the ratio of the spectrum efficiency with and without the SDMA applied. If the primary rate constraint of M2M communications for different aircraft is identical and represented by $\overline{r}$, the SDMA gain is given by

$$G_{\text{SDMA}} = \frac{K\overline{r}/F}{K\overline{r}/K} = \frac{K}{F}. \tag{11}$$

It shows in (11) that smaller F leads to higher SDMA gain. Then it is desired to develop a user selection method to perform ZFBF-SDMA with a small F. Since positions of AS nodes (including AoAs and propagation distances) may cause significant impacts on SINR, AS nodes cannot be grouped randomly due to the SINR constraint. Therefore, the multiuser selection plays a crucial role in the primary rate guaranteed SDMA, which results in improved and reliable overall performance.

3.2. Low Complexity User Selection. The optimal user selection is the exhaustive search, which is impractical due to the prohibit high complexity. The greedy algorithms, such as semiorthogonal user selection (SUS) [22] could be efficient but suboptimal. However, since the SUS may not secure the SINR of each candidate, the primary rate or SINR constraint cannot be guaranteed. If the pseudoinverse operation is carried out on channel matrix of selected users iteratively, SNR may be traceable but the computational complexity, that is, $\mathcal{O}(KM^3)$, is considerably high for the time varying scenarios of aeronautical communications. In this subsection, we propose a low complexity user selection with complexity of $\mathcal{O}(KM^2)$, which is able to guarantee the rate or SINR constraint for different users.

In order to avoid pseudoinverse operations (which lead to high complexity) in the process of user selection, we focus on the expression of the received SINR in (6). Apparently, the best pairing candidate is to choose the one that maximize the SINR as follows:

$$k_i^* = \arg\max_{i\in\mathcal{K}} \frac{\gamma P_t}{\left\|\widetilde{\mathbf{w}}_i\right\|^2 \sigma^2}. \tag{12}$$

Note that if $\|\widetilde{\mathbf{w}}_i\|$ is obtained without taking pseudoinverse, user selection under the SINR constraint can be simplified. Next steps will illustrate the dedicated way of inverse-free user selection.

The orthonormal basis of the subspace spanned by $\mathbf{H}_{\mathcal{S}_f}$ can be generated as

$$\mathbf{g}_n = \frac{\widetilde{\mathbf{g}}_n}{\left\|\widetilde{\mathbf{g}}_n\right\|}, \quad 1 \le n \le N_f, \tag{13}$$

where N_f, $(N_f \le M)$, is the number of accepted AS nodes in Group f, and

$$\widetilde{\mathbf{g}}_n = \left[\mathbf{I}_M - \sum_{i=1}^{n-1} \frac{\widetilde{\mathbf{g}}_i \widetilde{\mathbf{g}}_i^{\text{H}}}{\left\|\widetilde{\mathbf{g}}_i\right\|^2}\right] \mathbf{a}_{k_n}. \tag{14}$$

Therefore, the channel vector of the nth selected user for the fth spectrum can be rewritten as

$$\mathbf{a}_{k_n} = \widetilde{\mathbf{g}}_n + \sum_{i=1}^{n-1} \frac{\widetilde{\mathbf{g}}_i \widetilde{\mathbf{g}}_i^{\text{H}}}{\left\|\widetilde{\mathbf{g}}_i\right\|^2} \mathbf{a}_{k_n}. \tag{15}$$

Then the channel matrix $\mathbf{H}_{\mathcal{S}}$ is decomposed as

$$\mathbf{H}_{\mathcal{S}_f} = \mathbf{QRD}. \tag{16}$$

In (16), the unitary matrix $\mathbf{Q}$ is given by

$$\mathbf{Q} = \left[\frac{\widetilde{\mathbf{g}}_1}{\left\|\widetilde{\mathbf{g}}_1\right\|}, \ldots \frac{\widetilde{\mathbf{g}}_n}{\left\|\widetilde{\mathbf{g}}_n\right\|}, \ldots, \frac{\widetilde{\mathbf{g}}_{N_f}}{\left\|\widetilde{\mathbf{g}}_{N_f}\right\|}\right]. \tag{17}$$

The upper triangular matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \begin{bmatrix} 1 & \xi_{1,2} & \cdots & \xi_{1,N_f} \\ 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & 1 & \xi_{N_f-1,N_f} \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{18}$$

where $\xi_{i,j} = \tilde{\mathbf{g}}_i^{\mathrm{H}}\mathbf{a}_j/(\|\tilde{\mathbf{g}}_i\|\|\tilde{\mathbf{g}}_j\|)$, and the diagonal matrix $\mathbf{D}$ is written as

$$\mathbf{D} = \mathrm{diag}\left[\|\tilde{\mathbf{g}}_1\|, \ldots, \|\tilde{\mathbf{g}}_n\|, \ldots, \|\tilde{\mathbf{g}}_{N_f}\|\right]. \tag{19}$$

According to (4) and (6), the denominator of the SINR expression, which is directly affected by ECG, can be written as

$$\begin{aligned}\|\tilde{\mathbf{w}}_n\|^2 &= \left(\mathbf{W}_f^{\mathrm{H}}\mathbf{W}_f\right)_{n,n} \\ &\overset{(a)}{=} \left(\mathbf{D}^{\mathrm{H}}\mathbf{R}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{Q}\mathbf{R}\mathbf{D}\right)_{n,n}^{-1} \\ &\overset{(b)}{=} \left(\mathbf{D}^{\mathrm{H}}\mathbf{R}^{\mathrm{H}}\mathbf{R}\mathbf{D}\right)_{n,n}^{-1} \\ &\overset{(c)}{=} \frac{\left(\mathbf{R}^{\mathrm{H}}\mathbf{R}\right)_{n,n}^{-1}}{\|\tilde{\mathbf{g}}_n\|^2},\end{aligned} \tag{20}$$

where $(\mathbf{X})_{i,j}$ denotes the (i, j)th element of $\mathbf{X}$. Note that equality (a) is based on the fact that $\mathbf{W}_f = \mathbf{H}_{\mathcal{S}_f}(\mathbf{H}_{\mathcal{S}_f}^{\mathrm{H}}\mathbf{H}_{\mathcal{S}_f})^{-1}$ and (16). Since $\mathbf{Q}$ is a unitary matrix; that is, $\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{I}_M$, equality (b) can be readily obtained. Finally, equality (c) is verified by noticing that $\mathbf{D}$ is a diagonal matrix.

Besides, the numerator of (20), based on [22], can be decomposed as

$$\begin{aligned}\left(\mathbf{R}^{\mathrm{H}}\mathbf{R}\right)_{n,n}^{-1} &= \sum_{j=1}^{N}\left|\left(\mathbf{R}^{-1}\right)_{n,j}\right|^2 \\ &= \left|\left(\mathbf{R}^{-1}\right)_{n,n}\right|^2 + \sum_{j=1, j\neq n}^{N}\left|\left(\mathbf{R}^{-1}\right)_{n,j}\right|^2.\end{aligned} \tag{21}$$

In order to obtain the $\mathbf{R}^{-1}$, the $\mathbf{R}$ can be decomposed by

$$\mathbf{R}_n = \begin{bmatrix}\mathbf{R}_{n-1} & \mathbf{r}_{n-1} \\ \mathbf{0} & 1\end{bmatrix}, \tag{22}$$

where $\mathbf{R}_n$ denotes the $n \times n$ sized submatrix of $\mathbf{R}$ and $\mathbf{r}_{n-1} = [\xi_{1,n}, \ldots, \xi_{n-1,n}]^{\mathrm{T}}$. Thus, the $\mathbf{R}^{-1}$ can be readily obtained by

$$\mathbf{R}_n^{-1} = \begin{bmatrix}\mathbf{R}_{n-1}^{-1} & -\mathbf{R}_{n-1}^{-1}\mathbf{r}_{n-1} \\ \mathbf{0} & 1\end{bmatrix}. \tag{23}$$

Since $\mathbf{R}_1^{-1} = \mathbf{R}_1 = 1$ and $\mathbf{R}_n^{-1}$ is the function of $\mathbf{R}_{n-1}^{-1}$, the inverse of $\mathbf{R}$ can be calculated by taking $n-1$ iterations of vector-matrix multiplication. Note that $\mathbf{R}^{-1}$ is an upper triangular matrix and $|(\mathbf{R}^{-1})_{n,n}| = 1$, $1 \leq n \leq N_f$. Hence, the received SINR of the nth selected user $k_n \in \mathcal{S}_f$ can be rewritten as

$$\mathrm{SINR}_{k_n} = \frac{\gamma_{k_n}P_t}{\sigma^2}\frac{\|\tilde{\mathbf{g}}_n\|^2}{\left(1 + \sum_{j=n+1}^{N_f}\left|(\mathbf{R}^{-1})_{n,j}\right|^2\right)}. \tag{24}$$

It is shown that the denominator of (24) can be obtained only by taking vectors multiplication. Based on (24), the matrix-inverse free selection (MFS) for all K AS nodes is summarized in Algorithm 1.

Based on steps {5, 7, and 9} in Algorithm 1, it is easy to observe that the complexity of proposed MFS is $\mathcal{O}(KM^2)$, which is one order lower than those with matrix-inverse on each iteration, that is, $\mathcal{O}(KM^3)$. From steps 15~19, it shows that the received SINR of all accepted AS nodes have to be larger than $\overline{SNR}$, where the SINR constraints are confirmed for all selected AS nodes, that is, a reliable SDMA for spectrum multiplexing.

4. Impacts of AoA Reliability

So far, we have assumed that perfect AoA information can be obtained at the GS and the AoAs is unchanged within a short time slot. However, the AoAs actually change during the path of AS nodes flying over the GS. Since imperfect AoAs can seriously influence the SDMA performance, impacts of the AoA reliability (i.e., AoA accuracy and updating frequency) will be discussed in this section.

4.1. The Imperfect AoA Estimation. Note that any error or inaccurate AoA estimation will result in a degraded beamforming performance and lead to the interference in 4DT A/G data transmission, which is not desirable. It is important to investigate the impact of imperfect AoA estimations on the performance of the proposed aeronautical SDMA architecture.

While AoA estimation errors have been studied in the literature [31–33], analytical formulations to see the impact of AoA estimation errors are difficult for multiple user broadcasting with SDMA. Thus, in this section, we present a preliminary simulation result on the sensitivity of sum rates to imperfect AoAs. The estimated AoA by the GS, denoted by $\hat{\theta}$, can be characterized as a random variable with a uniform distribution as follows [31]:

$$f\left(\hat{\theta}\right) = \frac{1}{2\Delta}, \quad -\Delta \leq \left(\hat{\theta} - \theta\right) \leq \Delta, \tag{25}$$

where θ is the true AoA and Δ^2 is the variance of the AoA estimation error in $\hat{\theta}$. Let Δ be normalized as $\Delta = \tau\Delta_{\max}$, where τ is the proportionality factor and $\Delta_{\max}$ is the maximum AoA error obtained by the maximum null-to-null beam width (i.e., from null-to-null at the both sides of the $\theta_m = 90^\circ$). Based on (1) and the array factor expression of ULA in [16], $\Delta_{\max}$ can be found as

$$\Delta_{\max} = 0.5\pi - \arccos\left(\frac{2}{M}\right). \tag{26}$$

We can evaluate the impact of imperfect AoA information by defining SINR degradation as

$$\delta_{\mathrm{SINR}} = \mathrm{SINR}\left(\hat{\theta}\right) - \mathrm{SINR}\left(\theta\right). \tag{27}$$

In (27), SINR($\hat{\theta}$) denotes the degraded SINR, which is obtained by beamforming with $\hat{\theta}$, and SINR(θ) means the ideal result with θ. Then, the average SINR degradation can be found in Figure 4.

Input: $\{\mathbf{A}, \mathbf{H}, K, M, \overline{\text{SNR}}, \overline{P}_t\}$
$f = 1$;
while $f < K$ **do**
  $k_1 = \arg\min(\text{abs}(\mathbf{A}))$;
  $\mathbf{H}_{S_f} = \mathbf{H}(:, k_1)$;
  for $i = 2 : M$
    Generate $\mathbf{g}_{i-1}$ by (13);
    for $k = 1 : K$
      Generate $\mathbf{R}_i$ by (18);
      for $j = 1 : i - 1$
        Generate $\mathbf{R}_j^{-1}$ by (23);
      end for j
      Calculate $\text{SINR}(k_i)$ by (24);
    end for k
    $k_i^* = \arg\max \text{SINR}(k_i)$;
    if $\text{SINR}(k_i^*) > \overline{\text{SNR}}$
      $\mathbf{H}_{S_f} = [\mathbf{H}_{S_f}, \mathbf{H}(:, k_i^*)]$;
    else
      break;
    end if
  end for i
  $f = f + 1$;
end while
$F = f$;
Output: $\{F, \{\mathbf{H}_{S_1}, \ldots, \mathbf{H}_{S_F}\}\}$

Algorithm 1: The matrix-inverse free selection: **MFS**.

Figure 4: Average SINR degradation as a function of τ, where $M = 10$ and $K = 50$.

As shown in Figure 4, the δ_{SINR} decreases with Δ up to $\Delta_{\max}$, which results from the fact that the proposed SDMA highly depends on the phase accuracy of ARVs. Although a noticeable mitigation of the performance loss can been seen at the case of a large AoA error, the overall degradation of SINR is unacceptable. Therefore, in order to provide a reliable A/G communication service, the GS has to obtain AoA information of all AS nodes precisely.

If the AoA information is obtained over AS feedback, the updating frequency will be crucial considering the position of an aircraft changes rapidly. This issue will discuss in the next subsection.

4.2. AoA Update Frequency. In the implementation of the primary rate guaranteed SDMA, it is possible to obtain the positions or flight tracks of all AS nodes by AS feedback or positioning function of the ATC system. Consequently, the frequency of information update will affect the AoA accuracy directly and be discussed in this subsection.

Based on the geographic model in Figure 1, the AoA of an AS flying at inner area changes much more bigger than others far away from the GS; even they have the same altitude and speed. However, according to Figure 4, the AS node locates at inner area; for example, $R = 50$ km. It is less sensitive to AoA errors than those located in the edge of the macrocell; for example, $R = 200$ km. Therefore, the tolerance of AoA deviations can be roughly estimated, when $\Delta \leq 0.1\Delta_{\max}$, the performance becomes negligible ($\delta_{\text{SINR}} \leq 3$ dB, as shown in Figure 4). By using macrocell model illustrated in Figure 1, the AoA deviations can be expressed in terms of the position drift, denoted by δ_x, which is given by

$$\tan(\theta_m \pm \Delta) = \frac{x \pm \delta_x^{\pm}}{h_a}, \tag{28}$$

where x denotes the horizontal distance between AS-m and the GS, and

$$\delta_x^+ = h_a\left(\tan\left(\theta_m + \Delta\right) - \tan\left(\theta_m\right)\right),$$
$$\delta_x^- = h_a\left(\tan\left(\theta_m\right) - \tan\left(\theta_m - \Delta\right)\right). \quad (29)$$

It is noteworthy that the position tolerance for beamforming is quite different between inner and boundary areas of a macrocell. With $M = 10$, for example, we have $\Delta_{\max} = 0.2014$ rad (11.5394°). For the case of $R = 200$ km, $\theta_{\max} = 1.5208$ rad, and $\theta_{\min} = 0$ rad, it can be shown that

$$\begin{aligned}\delta_{\max}^+ &= h_a\left(\tan\left(\theta_{\max} + 0.1\Delta_{\max}\right) - \tan\left(\theta_{\max}\right)\right)\\ &= 134.9898\,\text{km},\\ \delta_{\max}^- &= h_a\left(\tan\left(\theta_{\max}\right) - \tan\left(\theta_{\max} - 0.1\Delta_{\max}\right)\right)\\ &= 57.5024\,\text{km},\\ \delta_{\min}^+ = \delta_{\min}^- &= h_a\left(\tan\left(0.1\Delta_{\max}\right)\right)\\ &= 0.2014\,\text{km}.\end{aligned} \quad (30)$$

Apparently, with the position drift $\delta_{\min}^+$, that is, approximately 200 meters, for a turbine powered jetliner or a corporate jet cruising at 450–500 Knots in general, which is 231.3–257.2 meters per second, one AS node should update its position no more than 1 second. Note that the updating frequency can be lower at the edge of a macrocell, which will result in a less burden in implementation.

Based on the promising A/G datalink protocols (i.e., L-DACS1 and L-DACS2), it is required that every AS should feedback its position or distance away from the GS in every superframe [24, 34]. Note that the superframe duration is 0.24 second in L-DACS1 or 1 second in L-DACS2, which means the feedback channel is already designed in future systems and the updating frequency will also fulfill all requirements of the proposed aeronautical SDMA system.

5. Numerical Results

In this section, numerical results are presented under assumptions that the unitary noise variance $\sigma^2 = 1$, the standard ULA with $d = 0.5\,\lambda$, the path-loss exponent $n = 2.0$, and the close-in reference distance $l_0 = 1$ m.

Figure 5 shows the SDMA gain versus total number of AS nodes K for different number of antenna elements M, where 10^6 Monte Carlo simulations are carried out to plot each point. It demonstrates that with the overall AS nodes number K increased, the performances of SDMA is improved, which is termed as the multiuser diversity. On the other hand, higher SDMA gain can also be obtained by employing large M, where the dimension of ARVs is extended to enhance the performance of SDMA. Hence, we can conclude that the system capacity can be improved by increasing the number of antenna elements and the number of aircraft, theoretically. Besides, comparing with the OBF under the same SINR constraint, the proposed ZFBF-MFS algorithm could obtain apparent advantages.

Figure 5: The SDMA gain versus total number of AS nodes K.

Figure 6: The SDMA gain comparison of increasing number M of antenna elements in the different macrocell, where $R = \{50, 200\}$.

Simulation results of SDMA gain versus number of antenna elements (M) are presented in Figure 6, where $R = \{50, 200\}$ (in km) are considered. It shows that the SDMA gain can be strongly improved by increasing the number of antenna array elements, since the orthogonality between different AS nodes can be improved by employing more array elements. It also shows that the performance gap is limited between the systems with different values of overall AS nodes K. In this case, the number of antennas becomes the main factor of SDMA gain when a practical density of aircraft environment is considered in a macrocell, for example,

Figure 7: SDMA gain as a function of R, where $M = \{10, 30\}$.

less than a hundred. Note that the OBF with the same SINR constraint also has the same increasing trend but lower performance.

In Figure 7, simulation results of our proposed primary rate guaranteed SDMA and OBF with SINR constraints are presented with different array elements numbers; that is, $M = \{10, 30\}$. It shows that the SDMA gain degrades as R increases. If the location of AS follows the uniform distribution in a macrocell, more AS nodes will result in similar AoAs when the radius of macrocell increases. The detailed AoA distribution could be found in [35]. Consequently, we can observe that the SDMA strategy becomes efficient as the macrocell is relatively small. Furthermore, the performance gaps between $K = 80$ and $K = 20$ are insignificant, which indicates the proposed SDMA approach is able to provide stable performance with different K.

In general, the simulations above illustrate that the proposed ZFBF with MFS is a proper SDMA approach for 4DT services. And the ZFBF with MFS will obtain a better SDMA gain compared with OBF in [19, 20, 35].

6. Conclusion

In this paper, an SDMA-based aeronautical M2M communication system has been proposed together with a low complexity user selection approach. It has been shown that a primary rate is guaranteed for all selected AS nodes, while the same frequency spectrum can be reused by ZFBF. Using the proposed matrix-inverse free user selection, the SINR of each AS candidate is calculated with acceptable complexity, that is, $\mathcal{O}(KM^2)$.

Through simulations, we have confirmed that the SDMA gain can be strongly improved by increasing the number of antenna elements, while the SDMA can be more efficient when the radius of macrocell is relatively small. Furthermore, our proposed SDMA approach is able to provide stable performance in a high density AS nodes scenario, which is desired by the future aeronautical SMN-M2M communications.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 61250001, 91338106, 61231011, 61171069, and 61231013) and by the National High Technology Research and Development Program (Grant no. 2011AA110102).

References

[1] International Civil Aviation Organization, "Global air navigation capacity & efficiency plan," in *Proceedings of the 12th Air Navigation Conference (ANC '12)*, vol. 9750, Montreal, Canada, 2012.

[2] SESAR, *System Interoperability With Air/Ground Data Sharing*, European ATM Master Plan, 2nd edition, 2012.

[3] International Civil Aviation Organization, "Data communication issues," in *Proceedings of the 12th Air Navigation Conference (ANC '12)*, AN-conf/12-WP37.1.1, Montreal, Canada, 2012.

[4] EUROCONTROL, *Initial 4D- 4D Trajectory Data Link (4DTRAD) Concept of Operations*, 1st edition, 2008.

[5] S. Plass, "Seamless networking for aeronautical communications: one major aspect of the SANDRA concept," *IEEE Aerospace and Electronic Systems Magazine*, vol. 27, no. 9, pp. 21–227, 2012.

[6] M. Schnell and S. Scalise, "NEWSKY: concept for networking the sky for civil aeronautical communications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 22, no. 5, pp. 25–29, 2007.

[7] International Civil Aviation Organization, *Global Operational Data Link Document (GOLD)*, 1st edition, 2010.

[8] SESAR, "I-4D- flying a new dimension," SESAR Factsheet, 2012.

[9] M. R. C. Jackson, J. Gonda, R. Mead, and G. Saccone, "The 4D trajectory data link (4DTRAD) service: closing the loop for air traffic control," in *Proceedings of the Integrated Communications, Navigation and Surveillance Conference (ICNS '09)*, pp. 1–10, Arlington, Va, USA, May 2009.

[10] EUROCONTROL, "Air traffic requirements," EEC Note No. 28/98, 1998.

[11] N. Neji, R. D. Lacerda, A. Azoulay, and T. Letertre, "Survey on the future aeronautical communication system and its development for continental communications," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 182–1191, 2013.

[12] R. Jain and F. Templin, "Requirements, challenges, and analysis of alternative for wireless datalinks for unmanned aircraft systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 5, pp. 852–860, 2012.

[13] International Civil Aviation Organization, "Report on the results of the International Telecommunications Union (ITU) World Radiocommunication Conference (WRC '12)," in *Proceedings of the 12th Air Navigation Conference (ANC '12)*, AN-conf/12-IP/7, Montreal, Canada, 2012.

[14] International Civil Aviation Organization, "Air traffic management," ATM 501, 2007, No. 4444, 15th edition.

[15] J. Choi, *Optimal Combining and Detection: Statistical Signal Processing for Communications*, Cambridge University Press, 2010.

[16] H. van Trees, *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*, Wiley-Interscience, New York, NY, USA, 2002.

[17] D. C. Chang and C. N. Hu, "Smart antennas for advanced communication systems," *Proceedings of IEEE*, vol. 100, no. 7, pp. 2233–2249, 2012.

[18] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[19] J. Choi and J. Ha, "Orthogonal beamforming for overlay mode of OFDMA-based rural broadband wireless access," in *Proceedings of the IEEE Wireless Communications and Networking Conference: PHY and Fundamentals (WCNC '12)*, pp. 508–512, 2012.

[20] J. Xie, J. Zhang, and L. Bai, "Multilayer orthogonal beamforming for priority-guaranteed wireless communications," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 307467, 8 pages, 2012.

[21] T. Yoo, G. J. Foschini, R. A. Valenzuela, and A. J. Goldsmith, "Common rate support in multi-antenna downlink channels using semi-orthogonal user selection," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3449–3461, 2011.

[22] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.

[23] F. Hoffmann, U. Epple, M. Schnell, and U. Fiebig, "Feasibility of LDACS1 cell planning in European airspace," in *Proceedings of the IEEE/AIAA 31st Digital Avionics Systems Conference (DASC '12)*, Williamsburg, Va, USA, October 2012.

[24] EUROCONTROL, "L-DACS1 system definition proposal deliverable D2," Tech. Rep, 2009.

[25] E. Haas, "Aeronautical channel modeling," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 2, pp. 254–264, 2002.

[26] D. Stacey, *Aeronautical Radio Communication Systems and Networks*, Wiley, 2008.

[27] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, John Wiley & Sons, New York, NY, USA, 1981.

[28] G. M. Galvan-Tejada and J. G. Gardiner, "Theoretical model to determine the blocking probability for SDMA systems," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 5, pp. 1279–1288, 2001.

[29] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[30] M. Kountouris and J. G. Andrews, "Downlink SDMA with limited feedback in interference-limited wireless networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2730–2741, 2012.

[31] H. Li, Y.-D. Yao, and J. Yu, "Outage probabilities of wireless systems with imperfect beamforming," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 5, pp. 1503–1515, 2006.

[32] A. Elnashar, S. M. Elnoubi, and H. A. El-Mikati, "Further study on robust adaptive beamforming with optimum diagonal loading," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 12, pp. 3647–3658, 2006.

[33] V. Vakilian, J.-F. Frigon, and S. Roy, "Effects of angle-of-arrival estimation errors, angular spread and antenna beamwidth on the performance of reconfigurable SISO systems," in *Proceedings of the 13th IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM '11)*, pp. 515–519, August 2011.

[34] EUROCONTROL, "L-DACS2 system definition proposal: deliverable D2," Tech. Rep, 2009, Version 1.0.

[35] "Dual priority SDMA in aeronautical communications," in *Integrated Communications, Navigation and Surveillance Conference (ICNS '13)*, pp. 1–49, Herndon, Va, USA, April 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 251374, 14 pages
http://dx.doi.org/10.1155/2013/251374

Research Article

Elastic Information Management for Air Pollution Monitoring in Large-Scale M2M Sensor Networks

Yajie Ma,[1] Yike Guo,[2] Dilshan Silva,[2] Orestis Tsinalis,[2] and Chao Wu[2]

[1] *College of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China*
[2] *Department of Computing, Imperial College London, London SWT 2BW, UK*

Correspondence should be addressed to Yajie Ma; mayajie@wust.edu.cn

Received 20 August 2013; Accepted 24 October 2013

Academic Editor: Jianhua He

Copyright © 2013 Yajie Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In large-scale machine-to-machine sensor networks, the applications such as urban air pollution monitoring require information management over widely distributed sensors under restricted power, processing, storage, and communication resources. The continual increases in size, data generating rates, and connectivity of sensor networks present significant scale and complexity challenges. Traditional schemes of information management are no longer applicable in such a scenario. Hence, an elastic resource allocation strategy is introduced which is a novel management technique based on elastic computing. With the discussion of the challenges of implementing real-time and high-performance information management in an elastic manner, an air pollution monitoring system, called EIMAP, was designed with a four-layer hierarchical structure. The core technique of EIMAP is the elastic resource provision scheduler, which models the constraint satisfaction problem by minimizing the use of resources for collecting information for a defined quality threshold. Simulation results show that the EIMAP system has high performance in resource provision and scalability. The experiment of pollution cloud dispersion tracking presents a case study of the system implementation.

1. Introduction

Recently, an increasing amount of research interest has been drawn towards data management in large-scale machine-to-machine (M2M) sensor networks [1–3], where a large number of high-throughput autonomous sensor nodes communicate directly with each other without human intervention and can be distributed over wide areas. M2M sensor networks have found their applications ranging from home monitoring to industrial sensing, including environment and habitat monitoring, traffic control, and health care. Such networks are usually characterised by a large number of sensors, wide coverage areas, a huge amount of data, complicated connectivity, and increasingly stringent response-time requirements. Their applications normally require data management over widely distributed sensors under restricted power, processing, storage, and communication resources. The continual increases in size, data rates, and connectivity of sensor networks present significant scale and complexity challenges. This is especially true when the computational resources available are limited. Thus, efficient support from sensor data management for data acquisition, transmission, storage, and retrieval becomes critical [4].

1.1. Motivation. Current research on information management for sensor networks has increasingly focused on real-time sensor data collection and sharing of computational and storage resources for sensor data processing and management. Technologies that support the building of large-scale infrastructures, integrating heterogeneous sensors, data, and computational resources deployed over a wide area accelerate the integration of sensor networks with distributed computing, grid computing, and even the state-of-the-art cloud computing. Recent researches on large-scale M2M sensor networks bring about some instructive designs of information management frameworks. For example, our former work in [5] designed a grid-based sensor information management platform. It obtains a high resolution of pollution characteristics in urban environment by high-density distributed sensors. Research in [6] investigates a feedback-based model-driven push approach to support user queries.

It presents a two-tier sensor architecture which develops sensor proxies at higher tier that each proxy controls tens of sensors at lower layer. In order to support energy-efficient query processing, sensors only transmit deviations to proxies compared with model-predicted values, which makes the system depend highly on the model calculated from past observations. Methods of parallel processing of data in sensor networks are discussed in [3], where the authors investigated how much degree existing distributed database solutions and programming models (e.g., MapReduce) are suitable for large-scale sensor network data processing. Based on the analysis, a general architecture for different data processing applications is developed. A similar data processing approach is proposed in [7]. The idea is to use distributed databases to store sensory data and MapReduce programming model for large-scale sensory data parallel processing. An interesting and useful implementation in this approach is that it employs cloud-based storage and computing infrastructure, which is remarkable for the research and design in this paper.

However, as the amount of information monitored by an M2M sensor network increases, two key issues arise in this context that cannot be addressed by existing approaches.

1.1.1. Effectively Avoiding Information Overload. This is done by organizing information collection and processing to focus on analysing only information relevant to the user needs. This includes deciding what information each of the sensor units should be collecting, at what rates, how and where it is processed, summarized, and stored, what information should be exchanged between the sensors, and how all such information should flow within the network. Tradeoffs between whether only local information collected from one sensor or global information collected from all sensors arise when addressing these decisions. This kind of on-demand processing has to eliminate overprovisioning when used with utility pricing. It is also expected to be able to remove the need to overprovision in order to meet the demands of millions of users.

1.1.2. Efficiently Maximizing the Value of Collected Information under Resource and Real-Time Constraints. A finite number of sensor units exist, and each has finite processing capacity, memory, and storage size, communication bandwidth, and battery power available to it. Tradeoffs occur since allocating a group of sensors to explore a particular geographic region would mean fewer resources available for exploring other areas. Similarly, within one region, assigning more processing or memory capacity to explore the features of a particular event in detail would mean less capacity available to explore other events. However, for real-time applications, the underlying algorithms require that the services quality improvement be monotonic to the consumption of the resource needed. In this case, how to identify the critical information needed and intelligently allocate resources to get it are a key point that deserves further research.

Based on the consideration of these issues, a vehicle/person-mounted air pollution monitoring system EIMAP (the acronym for "Elastic Information Management for Air Pollution") is proposed in this paper. This system has a four-layer architecture which can contain thousands of sensors distributed over an entire urban area to monitor airborne pollutants including SO_2, NO, NO_2, benzene, and ozone. The data volume that needs to be processed varies from several bytes (individual readings per sensor per minute that are used to identify irregularities and anomalies in real time) to 8 GB (whole readings per sensor per day that are used to capture high-resolution urban air pollution distribution resulting from transportation down to the single building level). In order to provide flexible resource for such large volume-variant data flows, an elastic resource allocation mechanism was introduced to EIMAP system. The key difference between our system and existing approaches is that EIMAP is endowed with data-aware QoS-driven capability for sensor management. Whether a sensor is active or not does not only depend on the energy-efficient consideration, but also, and more importantly, rely on the environment that the sensor resides, and in how much degree it is required to provide resource to the task.

1.2. Research Contributions. Our design of EIMAP has led to the following main contributions.

(1) Introducing Elasticity to Large-Scale M2M Sensor Information Management. Elasticity captures a fundamental aspect of cloud computing: when limited resources are offered for potentially unlimited use, providers must manage them elastically by scaling up and down as needed [8]. Elastic information management (EIM) is a technique that is based on elastic computing (EC), which is a feature of cloud computing. In [9], EC is defined as the use of computer resources which vary dynamically to meet a variable workload. The mathematical definition of elasticity in economics is

$$E_{y,x} = \left|\frac{\partial \ln y}{\partial \ln x}\right| = \left|\frac{\partial y}{\partial x} \cdot \frac{x}{y}\right|, \tag{1}$$

where $E_{y,x}$ denotes the elasticity of y with respect to x; $\partial y/\partial x$ is the derivative of y with respect to x. This formula reveals the ratio of the percent change in one variable to the percent change in another variable [10]. Based on this concept, we designed a four-layer architecture for M2M sensor network information management. In comparison with other architectures, the novelty is that a special layer, elastic management layer, is embedded which provides a scalable, resource-aware infrastructure for processing environmental streaming data produced by a range of heterogeneous mobile sensors.

(2) Developing a Scheduling Algorithm for Real-Time Resource Allocation. This algorithm overcomes the disadvantage of fixed resource provision strategy, which is not adaptive in changing environments. It also takes into account both of the resource provision and environmental feature detection by modeling it as a constraint satisfaction problem of minimizing the use of resources in a sensor network for collecting monitoring information at a defined quality threshold. The experimental results show that this algorithm performs well in elastic resource allocation.

1.3. Paper Layout. The remainder of this paper is organized as follows. In Section 2, we discuss the related work in the areas of information management in M2M sensor networks and resource allocation strategies for information management. Section 3 addresses the challenges by introducing EIMAP which comprises a four-layer hierarchical information management architecture. In Section 4, the design of the scheduler for elastic management is presented with the pseudocode for each part of the scheduling algorithm. Section 5 analyses the performance of the scheduling algorithm and simulates the EIMAP system in the WikiSensing sensor data management platform to evaluate the capability of the concurrent streaming management. Section 6 presents a case study of air pollution monitoring in East London. Section 7 concludes the paper with a summary of the research and a discussion of future work.

2. Related Work

2.1. Information Management in M2M Sensor Networks. Information management for sensor networks has been drawn much research attention for decade [11–13]. In a large-scale sensor network for air pollution monitoring, although one would often be more interested in highly polluted areas in environmental monitoring, a quick response to a pollution event related to other areas or time snapshots would usually be highly desirable. Information management for such an application will focus more on defining the events, or the feature of interests, which involves the techniques of data representation, summarization, and organization. To do so, techniques of improving the performance of query under resource constraints have been developed in recent years. For example, several approaches have focused on adaptive sampling techniques which aim to restrict, in an intelligent way, the amount of information gathered within the network. A popular approach is based on using Kalman filters [14] which enable a group of sensors to respond to and track fast moving signals in a slowly changing background. For the QoS management, the Aurora system [11] developed by Brown University and the TinyDB [15, 16] system developed by MIT all provide QoS managing strategies to support reliable services.

Other systems have focused on efficient data summarisation to facilitate query propagation/processing and to improve distributed data storage within the networks. For instance, the BBQ system [17] maintains a correlation-aware probabilistic model in a base station to provide a robust interpretation of sensor readings. Data acquisition from sensors happens only when the model is not able to offer approximate answers to certain queries with acceptable confidences. StonesDB [18] applies a multiresolution scheme, generating two summary streams (wavelet-based summary and subsampled summary streams) from input data streams. In [19], the authors addressed the problem of the information-driven management criteria for sensor networks. They proposed a novel measurement for usefulness of information: uncertainty reduction rather than information gain is used to justify the performance of the information-driven performance. Other techniques, such as synopses-based approximate answers [20], histogram analysis [21], and wavelet-based data summaries [22, 23], can all be used to investigate how to guarantee high accuracy and speed of data summarisation.

2.2. Resource Allocation for Information Management. Resource allocation is a key technique for information management. Many attentions have been paid to this area in recent years. The research generally has two categories: one is to develop novel system infrastructures to meet different resource allocation demands; another is to design high performance algorithms to support fast resource allocation computation.

For the first category, [24] investigated the challenges of resource allocation for reconfigurable multisensor networks. The authors discussed the problems of resource allocation under environmental and technical constraints. A hierarchical model was proposed in this paper but not concrete resource allocation strategy was presented. For efficient resource allocation, hierarchical or layered system architectures can be studied in many papers. For example, in [25], the researchers designed a layered distributed system and extended the existing disjoint coalition formation protocol to solve the multi-sensor task allocation problem which aims to automatically decide the best sensors to specified task; in [26], a market-based 4-layer architecture was presented for adaptive task allocation which formulises the pricing mechanism to achieve a fair energy balance among sensor nodes while minimizing delay; [27] designed a two-tiered on-demand resource allocation strategy which is specially designed for VM-based computing environments.

For the second category, most of the algorithms treat the resource allocation problem as an optimisation. They usually handle tradeoffs between system performance and resource constraints. [28] discussed two tradeoffs in health monitoring sensor system. The authors addressed two optimisation problems in this paper, one of which is to obtain sustainable power supply, while another one is to achieve high quality of service. The solutions to two optimisation formulas were given as well. Groot et al. analysed an adaptive optimisation for the tradeoff between resource allocation and the reconfigurable resources within the multi-sensor network in [29]. The resource allocation algorithm aims to maximize the system utility by finding the optimal set of services. [30] considers sensor assignment problems in both static and dynamic sensing environments. Heuristic algorithms were studied to address the NP-hard optimisation problems. Other schemes for the category can also be seen, such as agent-based algorithm [31] and posterior-based decision making scheme [32].

3. Elastic Sensor Network Architecture

The key feature of the EIMAP architecture is using autonomous sensors, whether fixed or mobile, to provide coverage of a specific geographical area to collect real-time

pollution data on key aspects such as traffic conditions, vehicle emissions, ambient pollutant concentration, and human exposure. The focus of constructing an EIMAP system is related to the data management, computation management, information management, and knowledge discovery management associated with the sensors and the data they generate, and how they can be addressed in real-time within an open computing environment. To do so, in this section, we first analyse the challenges in implementing such a system, and then we propose a four-layer architecture to address these challenges.

3.1. Challenges in Implementing EIMAP System. Considering the resource characteristics of large-scale M2M sensor networks, the main issues and challenges related to constructing an elastic system are as follows.

3.1.1. Dynamic Interactivity via M2M Architecture. Within a mobile sensor network, the sensors themselves naturally form an M2M network and communicate with each other through it. In order to satisfy the real-time analysis requirements, the sensors themselves will have to store part of the information and communicate with each other within the M2M network. The measurements from the sensors, both mobile and static, will be filtered and processed using a set of specialized algorithmic processes, before being warehoused in a repository. The design and implementation of a suitable M2M sensor architecture will need to satisfy the real-time analysis requirements as well as decide the data storage/communication tradeoffs. The sensors in such a system will need to be equipped with sufficient computational capabilities to participate in the elastic environment and to feed data to the warehouse as well as perform analysis tasks and communicate with their peers.

3.1.2. Elastic Resource Allocation under Resource Constraints. In such a scenario, strategies of allocating or scheduling finite sensing resources for exploring surveillance regions in more detail have to be proposed. One also has to take into consideration the dynamic changes that occur in the sensed environments. We model this scheduling problem as a constraint satisfaction problem for selecting a particular resource allocation strategy for maximizing the value of information collected at any time step. Such resource allocation needs to take into account constraints on the resources, the decision-making time (e.g., the value of information may diminish if its transmission is delayed), and other problem-dependent constraints (e.g., a need to keep full coverage of a particular area or particular events using a minimum number of sensors). Hence, the strategies of allocating or scheduling have to be able to (a) define the resource and application constraints together with the associated solvers and (b) estimate the increased information (information gain) every time step for the different strategies through the selection of the appropriate measures in terms for its completeness, quality, and reliability.

3.2. EIMAP Hierarchical Architecture. Considering the challenges analysed above, we introduce the elastic computing capability of EIMAP, which aims to provide a reliable, scalable infrastructure for elastic management of streams of environmental data produced by a range of heterogeneous mobile sensors. Therefore, a four-layer architecture was designed as shown in Figure 1. This architecture is also well suited to the dynamic, on-demand, pay-per-use nature of the emerging utility computing platforms.

3.2.1. Sensor Layer. This layer manages all the raw hardware level resources in the system, such as the environmental characteristics, different types of sensors, network connection, storage, sensing activities, and distributed raw data. Sensors within the environment are heterogeneous and may be mobile or static. Hence, the wireless connectivity can provide different access protocols to the IP backbone including WiFi (802.11.g), ZigBee (802.15.4), and WiMAX (802.16). The sensors have the capability to sample one or more pollutants or other environmental properties such as noise or temperature. This data will then be transported to the data store in the upper layer (which will be introduced in the following paragraphs). Since potentially the volume of sensory data is significant whereas, the processing resource is limited, the key of the sensing activity is efficiency—salient regions should be paid more attention to and, consequently, consume more processing resource. In this case, an attention-based sensing mechanism [33–36] is preferred, which can extract irregularities and anomalies from massive background noise. To do so, an intelligent control strategy supported by the elastic management from higher layer is necessary.

3.2.2. Elastic Management Layer. This is the core layer of the EIMAP architecture. The purpose of this layer is to provide an elastic resource provision infrastructure for the whole system. It contains resources that have been abstracted/encapsulated so that they can be exposed to the upper layer and the end users as integrated resources, for instance, repositories, resource catalogue services, resource scheduler, and specialized services such as sensor registry/activity management. The resource supply and its supply infrastructure can scale up and down dynamically based on application resource need, which is able to deliver software application environments with a resource usage-based computing model. The resource scheduling service, which is critical for the system performance, is the core service of this layer as well as the whole EIMAP architecture. It enables virtual organization management, resource management, and load balancing in order to guarantee an easy access to sensor data in heterogeneous physical sensors. We will discuss it in next section in detail.

3.2.3. Data Analysis Layer. This layer (whether centralized or distributed) is concerned with information comprehension, including how to summarize the data and how to develop and use models representing the data to control the operation of the sensing activities, such as adjusting sampling rates of specific sensors or making decision of allocating more sensing resources to a particular geographic area to gain further information about it. Centralized and decentralized

Figure 1: EIMAP hierarchical architecture.

data mining algorithms are developed in this layer to meet the needs of different data analysis tasks. The analysis results are delivered to the application layer according to different user requirements.

3.2.4. Application Layer. This layer retrieves information from the data analysis layer and uses this information as the input to different applications, not only for the air pollution monitoring. Because the lower layers are designed to be application-independent, the framework is universal for different applications, such as traffic optimisation, security surveillance, mental training, and city planning. Besides, a user-defined service module makes the system extensible so that the users can take advantage of new services that become available.

4. Scheduler for Elastic Sensing

Resource allocation is a key issue in EIMAP which affects not only the sensing activities regarding specific events but also the performance of the whole system including speed and accuracy of response, fairness of queries, and experience for users. Suppose such an application scenario: using sensors to track moving objects such as the pollution cloud (due to dispersion, the pollution cloud always moves and changes its shapes/size). A fixed resource provision strategy is not preferred especially in a resource-restricted environment. Hence, elastic resource provision is a better choice to improve the system performance. In a sensor network, the available underlying resources are sensors, including the sensing behaviours, distributed computational capabilities, and communication links (connectivity, bandwidth, radio power, etc.) that sensors or sensor peers can provide. In consideration of the resource constraints in sensor networks, a resource awareness mechanism is essential to provide a strategy of allocating or scheduling finite sensing resources in exploring potential regions of interest and to take into consideration the dynamic changes that occur in the sensed environment.

A scheduler in the elastic management layer is designed for such an elastic sensing requirement, which aims to model a constraint satisfaction problem of selecting a particular resource allocation strategy for maximizing the value of information collected at any time step, or, minimizing the use of resources in a sensor network for collecting monitoring information at a defined quality threshold.

In order to model this constrained optimisation problem, we feature the surveillance area as follows.

(1) The entire geographical area is divided into grid units and each grid has a predefined size to cover a reasonable region of the area according to the specific requirements of air monitoring.

Table 1: Scheduling algorithm description.

Step	Description
1	Generate a candidate set of nodes
2	Define objective function
3	Identify constraints
4	Find the solution of scheduling

(2) There is a sensor in the centre of each grid which collects and maintains a series of sensor readings for historical or real-time query.

And according to the physical property of the resource provider, the resource constraints can be classified into two categories:

(1) hardware resource constraints, including

(a) size of monitored area/number of grids,
(b) storage capability,
(c) surplus energy,
(d) communication distance,
(e) available bandwidth;

(2) software resource constrains, including

(a) measuring accuracy requirement,
(b) pollutant diffusion model,
(c) sensory data attributes.

Suppose now we have identified the feature of interest in an area as an m-dimensional vector $A = (a_1, a_2, \ldots, a_m)$ (A can be achieved by attention-based mechanisms and the computational detail is out of the scope of this paper). Suppose the available set of nodes in this area is V($|V|$ is the number of nodes in V). Each node $i \in V$ has a reading $Y_i = (y_1, y_2, \ldots, y_m)$ that describes the feature of the node or the grid where the node resides. $a_j \in A$ is an attribute corresponding to an element $y_j \in Y$. The resource constrained scheduling strategy can be described by the steps that are shown in Table 1.

4.1. Generate a Candidate Set of Nodes. The following GC algorithm shown in Algorithm 1 is used to find the candidate nodes for resource provision where the feature of interest A is likely to be detected. The algorithm returns a list of candidates by matching every reading in every node with a given feature.

The algorithm starts with a given number of iterations and a null set of candidate nodes C (line 1). For a given number of sampling times, the Euclidean distance between the mean value of the readings in each node and the given feature is calculated (lines 3 to 5) (the Euclidean distance between two readings P and Q can be calculated as Euclidean $(P, Q) = \sqrt{|p_1 - q_1|^2 + |p_2 - q_2|^2 + \cdots + |p_m - q_m|^2}$). If the distance is no larger than a predefined threshold ε, then the node providing this reading satisfies the constraint of data similarity and has to be added into C as a candidate (line 6).

4.2. Define the Objective Function. In order to describe whether a candidate node is chosen to be a resource provider or not, we define a decision variable x_i:

$$x_i = \begin{cases} 1, & \text{if candidate node } i \text{ is chosen,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The scheduler tries to find an optimal set of nodes from the candidate nodes given the resource constraints. In a sensor system, a vital resource constraint is the node energy. An energy-aware system will have better performance in system life time [37–39]. Hence, in our system, we select the surplus energy as the optimisation objective and the aim of the optimisation is to minimize the rate of the energy consumption, which can be formulated as

$$\min \sum_{i=1}^{|C|} w_i x_i, \qquad w_i = \frac{\mathrm{RE}(i)}{\mathrm{SE}(i)}, \tag{3}$$

where RE(i) is the required energy for node i to execute the current task. SE(i) is the surplus energy in node i. Then w_i ($w_i > 0$) is a weight to measure what percentage of the surplus energy of the sensor the current task will consume. For a sensor network, small value of w_i will bring better energy performance, which means the nodes with less energy consumption rate are chosen and the lifetime of the network is prolonged.

4.3. Identify Constraints. In air pollution monitoring, considering the pollutant diffusion model, we cannot let a sensor monitor an arbitrary size of area in order to guarantee the measurement accuracy. Furthermore, a single sensor is less possible to provide enough storage and computation capacity for the whole task. Therefore, the scheduler has to find a set of nodes with reasonable number of nodes for resource provision. To simplify the analysis, suppose all sensors have the same storage space to cache data; all the links have the same bandwidth; the communication distance is adequate for data transmitting from one grid to a neighbour grid. And we suppose that all the pollution data analysed in this paper are generated and diffused under the same model. Hence, the hardware/software resource constraints that need to be taken into account are reduced to the number of grids, surplus energy, measuring accuracy, and data attributes. According to the guidance of environmental data collection [40], the minimum number of nodes N in a sampling unit has to satisfy

$$N = \frac{t^2 s^2}{D^2}, \quad N \le |C|, \tag{4}$$

where t is the critical value for 2-tailed t-test with a specified degree of freedom; s is the standard deviation of the samples; D is the absolute deviation. If the square of a monitoring unit is S, the maximum distance L between two sampling nodes is

$$L = \sqrt{\frac{S}{N}}. \tag{5}$$

```
(1) Given A; NS = NUM_SAMPLES; C = Ø;
(2) for (i = 1 to |V|) {
(3)     for (j = 1 to NS)
(4)         Y_ij^- = (1/NS) Σ_{j=1}^{NS} Y_ij
(5)     if (Eulidean (Y_ij^-, A) <= ε)
(6)         C = C ∪ {i};
(7) }
(8) return C;
```

Algorithm 1: GC algorithm.

```
Input  RE, w^LB and distance measurements between any node pair
Output A resource provider set P
(1)     P = Φ; δ = RE/SE_max / w^LB;      /* SN is null at beginning */
(2)     for each i: i ∈ C parallel_do {     /* Parallel process for each i*/
(2.1)     x_i = 0;                          /* Node i is a candidate */;
(2.2)     calculate w_i;
(2.3)     if (w_i ≥ 1) x_i = −1;            /*i is no longer a candidate */
        } /* end for */
(3)     for each i: i ∈ C & (i is a candidate) parallel_do {
(3.1)     p_i = min{1, (w^LB/w_i)^(w_i/w^LB)};
(3.2)     if p_i > δ   x_i = 1;             /*i becomes a provider */
        } /* end for */
(4)     for each i: i ∈ C & (i is a candidate) parallel_do
(4.1)     if (x_j = = 0 for all j with distance (i, j) ≤ L)
(4.2)       x_i = 1;                        /*i becomes a provider */
(5)     P = {i | x_i = 1, i ∈ C};
```

Algorithm 2: PAS procedure.

Therefore, the constraint optimisation can be formulated as a 0-1 integer linear programming (ILP) problem as shown in the following.

0-1 ILP for Resource Allocation

OPT1:

$$\min \quad \sum_{i=1}^{|C|} w_i x_i \tag{6}$$

$$\text{s.t.} \quad \sum_{j=1}^{|C|} a_{ij} x_j > 1, \quad \forall 1 \le i \le |C| \tag{7}$$

$$x_i \in \{0, 1\}, \quad \forall 1 \le i \le |C|, \tag{8}$$

where a_{ij} is a decision variable related to the geography distance between node i and node j:

$$a_{ij} = \begin{cases} 1, & \text{if distance}(i, j) \le L, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

In OPT1, the number of grids constraint is explicitly represented by inequality (7), the surplus energy constraint is formulized by w_i, the measuring accuracy is considered by ε and t, and hence represented by L, and the sensory data attributes constraints are examined by D and s, and also represented by L.

Finding the optimal solution for ILP is NP-hard and may be solved in linear time as an LP-type problem with a constant number of variables [41, 42]. Approaches, such as enumeration, cutting plane, and branch and bound, are unacceptable for real-time scheduling in the scenario of air pollution monitoring because the time complexity of them exponentially increases with the number of variables. Hence, approximate solutions are compromised for such problems.

4.4. Find Out the Solution of Scheduling. Here we give a proximate algorithm for scheduling (PAS) to find the resource provider set P. In this algorithm, a parameter w^{LB} is used which is defined as follows

w^{LB} is a lower bound for all w_i, which is a predefined constant satisfying $0 < w^{\mathrm{LB}} < 1$.

Algorithm 2 shows the pseudocode of the parallel procedure of PAS in each candidate. In the procedure, δ is a threshold where $\mathrm{SE}_{\max}$ is the maximum surplus energy in the whole network which can be simply set as the initial energy value. Step 2.3 deletes all the nodes that have less surplus

energy than the required energy from the candidate set. Step 3.1 is the key processing where each node calculates the probability of becoming a provider. The probability function makes the nodes with weights comparatively closer to w^{LB} have higher probability to become providers (in the case that w_i is smaller than w^{LB}, node i will become a provider with $p_i = 1$). While Step 4 is a complementary process which guarantees that the set P satisfies the requirement of measurement accuracy, for any node, if there is no other provider within distance L, this node becomes a provider.

5. Performance Analysis

5.1. Scheduling Algorithm Performance Analysis

5.1.1. Complexity Analysis. The time complexity of FC algorithm is $O(|V|)$. In the PAS procedure, each of the steps 2, 3, and 4 has the time complexity $O(|C|)$. Therefore, the time complexity of the whole algorithm is $O(|V|)$.

For the message complexity, suppose the maximum degree of the sensor network topology is Δ. The algorithm only requires the message exchange in PAS step 4. Hence, the message complexity is $O(\Delta|C|) = O(\Delta|V|)$.

5.1.2. Size of Provider Set. In this experiment, we calculate the average size of the provider set P and the calculation time of PAS. We compare both of the values with the results calculated by ILP.

We use a topology generator to generate random topologies in an area with radius = 100. In consideration of the purpose of this experiment, we simply assume that all the nodes are candidates and the maximum distance L is given different values instead of calculated by formula (4) (the selection of candidates and the calculation of L will not affect the results in this experiment). For different topology parameter values, the random graph is generated and simulated until a predefined confidence interval for the population mean is reached and, then, simulation results are measured by simply taking the average of all cases. Here, we achieve a precision of 1% with the 90% confidence interval of the provider set. In the experiment, $w^{\mathrm{LB}} = 0.01$ and $\mathrm{RE} = 0.9$. Each node is randomly assigned a surplus energy value between 0 and 100. Then $\delta = (\mathrm{RE}/\mathrm{SE}_{\max})/w^{\mathrm{LB}} = 0.9$, which means if the value p_i in PAS step 3.1 is larger than 0.9, then node i becomes a provider. The total number of nodes ranges from 40 to 130. The experiment investigates the impact of different distance limitation L on the size of P. The results are shown in Figure 2.

In the figure, we can see that the size of provider set P generated by PAS approximately increases linearly with the total number of nodes. Larger L corresponds to smaller P because a single node can cover a larger geographical area. The size of P generated by PAS is about 1 to 2 times of that generated by ILP. As OPT1 matches the classic minimum independent set problem, according to [30], the size of any independent set in a unit-disk graph is at most 4opt + 1 our algorithm gives a reasonable result.

Figure 2: Size of provider set with different distance limitation.

Figure 3: Comparison of running time ($L = 25$).

5.1.3. Running Time. This experiment compares the calculation times of PAS and ILP with $L = 25$. The result is shown in Figure 3. From the figure we can see that the convergence time of our algorithm is much less than that of ILP, and our algorithm is network scale independent while the running time of ILP increases with the increasing total number of nodes. Hence, our algorithm has better performance in scalability.

5.1.4. Average Surplus Energy. This experiment calculates the average surplus energy (SE) of each node in provider set. PAS algorithm is an optimisation solution aiming to minimize the ratio of the energy consumption, in other word, maximize the surplus energy of the provider set with a given required

Table 2: Comparison of number of providers generated by PAS Steps 3.2 and 4.2.

Number of nodes	PAS step		Number of nodes	PAS step	
	3.2	4.2		3.2	4.2
40	7.0700	12.6667	90	16.3800	12.9133
50	9.1200	13.0700	100	17.9633	12.5867
60	11.0633	13.1833	110	19.9667	12.1800
70	13.0100	13.2200	120	22.4867	11.5933
80	14.3367	13.2267	130	23.8700	11.5333

Figure 4: Comparison of surplus energy.

Table 3: Air pollution monitoring scheduling results.

	08:30	15:30	17:30
L (meters)	199.92	128.30	187.44
Number of candidates	17	3	9
Number of providers	6	2	4
AR	0.35	0.67	0.44

energy for a task. Therefore, we expect that the provider set generated by PAS has higher average SE in comparison with that of the whole network.

The result is shown in Figure 4. For the whole network, as the SE of each node is randomly assigned from 0 to 100, the average SE is about 50. For the providers, the curve in the figure presents two features. First, the average SE is much larger than 50 as we expected. Second, SE approximately linearly increases with the number of nodes. To explain this, let us check the providers generated by PAS. In PAS, a node has two chances to become a provider: in step 3.2 and step 4.2. Step 3.2 is a mandatory criterion for a node to become a provider if $p_i > \delta$ (i.e., this node has very low energy consumption rate or very high SE). And step 4.2 is a complementary processing to satisfy the distance constraint. So the more proportion of providers the selected by step 3.2, the higher average SE is achieved. From Table 2 we can see that, in PAS the number of providers selected by step 4.2 is about a constant around 12, whereas the number of providers selected by step 3.2 increases with increasing total number of nodes. This statistics explains the result in Figure 4 well and this experiment proves that our system has high performance in energy consumption.

5.2. EIMAP System Performance Measurement. In this experiment, we use WikiSensing [43] and Siege benchmarking utility [44] to simulate our EIMAP system. WikiSensing is an online collaborative platform for sensor data management. It can simulate as many sensors as the system being tested requires, including sensor registration, data sampling, user query response and database management. We use WikiSensing to simulate the lower 2 layers of EIMAP: the sensor layer is simulated by generating 140 nodes records with specified location IDs. Each sensor has a sequence of readings stored in the database. The database is maintained on the IC cloud computing infrastructure [45]. Each node has the capability of receiving quires and sending response. The elastic management layer is realized by integrating our scheduling algorithm into the optimization module of WikiSensing. As the data analysis functions are not essential for this experiment, we can treat the data analysis layer as a layer that executes nothing but transmits the user queries from the interface between 3rd/4th layer to the interface between 2nd/3rd layer directly. And the application layer is simulated by the Siege benchmarking. It can simulate the users' behavior of accessing a web server with a configurable number of concurrent simulated users. The duration of the "siege" is measured in transactions, the sum of simulated users, and the number of times each simulated user repeats the process of accessing the server. With Siege benchmarking, it is possible for us to measure the performance of EIMAP to see how it will stand up to load on the internet. The simulation environment is illustrated as shown in Figure 5.

The experiment uses Siege to simulate concurrent users from 100 to 1000. The elapsed time of each test is 60 seconds. In WikiSensing we simulated 30 sensors and different aggregation ratio AR. Here we define AR as follows:

$$\mathrm{AR} = \frac{\text{Number of providers}}{\text{Number of candidates}}. \tag{10}$$

The data stored in IC Cloud is air pollution data which will be described in detail in the next section. The performance evaluation calculates the average response time of the

Figure 5: EIMAP system performance testing environment.

Figure 6: Average response time of EIMAP.

queries, which is the round trip time of sending a request and receiving a response. The results are shown in Figure 6.

In Figure 6, the response time presents linear increase as the number of concurrent users increases. AR = 1 means the system collects data from all the candidates (in most of existing approaches [2, 3, 6, 7] including our former research [5]; although the system architectures and resource providing schemes are different, they all can be categorised into the design with a scheduler that AR = 1), while AR = 0.1 means 1/10 of sensors are chosen to be providers and the system will only collect data from them. As AR increases, the response time increases (the response time of AR = 0.1 is much shorter than that of AR = 1), and hence, providing a better system performance for clients.

6. Air Pollution Scenario

In this section, we introduce a case study for our algorithm by applying it to the air pollution scenario. The experiment, based on our former research [5], uses the air pollution data collected from 140 sensors (in a 100-metre rectangular grid) distributed in a 1 km × 1.4 km area represented as red dots in the map of Figure 7(a). The map shows an urban area around the Tower Hamlets and Bromley areas in East London. There are some of the typical urban landmarks such as the main road extending from A6 to L10, the hospitals around C5 and K4, the schools in B7, C8, D6, F10, G2, H8, K8, and L3, the train stations at D7 and L5, and Gas Works between D2 and E1. 140 sensors collect data from 8:00 to 17:59 at a 1-minute interval to monitor the pollution volumes of NO, NO_2, SO_2, and Ozone. Then there are 600 data items for each node and totally 84000 data items for the whole network. Each data item is identified by a time stamp, a location, and a four-pollutant volume reading. The time-plot profiles of four pollutants over 10 hours are shown in Figure 7(b). Each profile is the overlap time plots of all the 140 sensors for one pollutant over 10 hours. For example, the upright figure shows the volume of NO from 08:00 to 17:59. At 8:30, 140 sensors generate three typical readings: over 200 ppm, between 60 ppm and 80 ppm, and less than 20 ppm. However, this figure cannot tell us which sensor generates what readings.

The case study will investigate the resource provision for tracking a given feature of interest. For this purpose, we specify the feature with high volume of NO + NO_2 + SO_2, which is defined as a vector A= (170, 180, 150). And we pick up 3 time stamps 08:30, 15:30 and 17:30 for data analysis (according to Figure 7(b), around these 3 time stamps there exist fairly high level pollution volumes of NO, NO_2, and SO_2 in some of the locations that are distinct compared to other locations). As feature A is the saliency of the pollutants concentration which stands out against their neighbours/surroundings, according to air pollution dispersion characteristics (the concentration of traffic emissions on highway decayed 50% at 150 m location and further 30% at 400 m location [46]), we define $\varepsilon = \|A\|\cdot 30\%$, which means a sampled value matches A if it falls into the intervals of $[\|A\| - \varepsilon, \|A\| + \varepsilon]$. And we delimit a sampling unit as an area that is covered by 25 grid units/nodes (about 500 m × 500 m). The maximum distance L is calculated as follows:

$$L = \sqrt{\frac{S}{N_{\max}}},$$
$$N_{\max} = \arg\max\left\{N_{\mathrm{NO}}, N_{\mathrm{NO}_2}, N_{\mathrm{SO}_2}\right\}, \tag{11}$$

which means we calculate N for each pollutant in each sampling unit and the maximum N is used to calculate L. Other parameters are given the same values as described in Section 5.

Table 3 summarises the results of executing the scheduling algorithm in this area. The values of L are different because the values of N are different according to formula (4). Figure 6 visualises the results of the feature tracking. Figure 8(a)(A)–(C) highlight the areas of interest monitored

(a) 140 sensors distributed in an area of East London

(b) Time plots profiles of four pollutants over 10 hours

Figure 7: Sensor distribution and data profiles in an area of East London.

Figure 8: Visualisation of feature tracking.

by all the candidates at different time stamps. For example, in the morning the feature is located at the main road and schools (A8, B7, H8, and K8). At 15:30, the feature is only found at two schools and at 17:30 the feature only covers the main road. This characteristic matches the traffic property during a weekday (people going to school and work in the morning rush hours by vehicles makes both the main road and school areas have high pollution, while people off school at about 15:30 and off work at about 17:30, respectively, make the pollution distribution different). These three figures show the resource provision without scheduling scheme, where all the nodes that match the feature are active. Figure 8(b)(A)–(C) illustrate the resource providers chosen by our scheduling algorithm. Each provider is represented by a black node and the yellow circle is the corresponding sensor coverage area with radius L. From the figures we can see that the areas of interest are all picked up and monitored by fewer sensors. Hence with our scheduling algorithm, the resource is elastically provided, whereas the feature tracking still performs well.

7. Conclusion

In this paper, we discussed the main challenges in information management and real-time resource allocation when applying large-scale M2M sensor networks to air pollution monitoring. An elastic information management architecture was proposed to address those challenges by using pervasive roadside and vehicle/person-mounted sensors by

combining and extending state-of-the-art elastic computing and data management techniques. The experiments results on the elastic resource allocation scheme, the entire system performance, and the air pollution monitoring case study show that our design provides higher performance in energy efficiency and system response speed, as well as an effective saliency detection and coverage in the scenario of pollutant distribution with less sensors.

Direct further work on the algorithm improvement includes continued research on other resource constraints not taken into account in this paper, such as the storage capability and available bandwidth. Our long-term work will focus on the development of the management platform to allow demonstration and further analysis of other applications. The integration with sensor hardware is also a key step which can provide the collection of the application data from the real world to support the real-time data analysis.

Acknowledgments

This work was jointly supported by National Science Foundation of China, Grant no. 61104215, and Engineering and Physical Science Research Council (EPSRC), Grant no. EP/H042512/1. This work is also partly supported by project "Digital City Exchange," Grant no. EP/I038837/1, funded by Research Council UK.

References

[1] S. Wylie, J. Heide, B. Avci, D. D. Vaccaro, O. Ghica, and G. Trajcevski, "Distributed data management for large-scale wireless sensor networks simulations," in *Proceedings of the 15th International Conference on Extending Database Technology*, pp. 626–629, Berlin, Germany, March 2012.

[2] M. Balazinska, A. Deshpande, M. J. Franklin et al., "Data management in the worldwide sensor web," *IEEE Pervasive Computing*, vol. 6, no. 2, pp. 30–40, 2007.

[3] C. Jardak, J. Riihijärvi, F. Oldewurtel, and P. Mähönen, "Parallel processing of data from very large-scale wireless sensor networks," in *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing (HPDC '10)*, pp. 787–794, Chicago, Ill, USA, June 2010.

[4] M. Welsh, "Where do we go from here? The big problems in sensor networks," in *Proceedings of the Wireless Sensing Solutions Conference*, Chicago, Ill, USAA, Keynote Talk, September 2005.

[5] Y. Ma, M. Richards, M. Ghanem, Y. Guo, and J. Hassard, "Air pollution monitoring and mining based on sensor Grid in London," *Sensors*, vol. 8, no. 6, pp. 3601–3623, 2008.

[6] M. Li, D. Ganesan, and P. Shenoy, "PRESTO: feedback-driven data management in sensor networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 4, pp. 1256–1269, 2009.

[7] B. Tang and Y. Wang, "Design of large-scale sensory data processing system based on cloud computing," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 4, no. 8, pp. 1004–1009, 2012.

[8] S. Dustdar, Y. Guo, B. Satzger, and H.-L. Truong, "Principles of elastic processes," *IEEE Internet Computing*, vol. 15, no. 5, pp. 66–71, 2011.

[9] R. Buyya, J. Broberg, and A. Goscinski, *Cloud Computing: Principles and Paradigms*, John Wiley and Sons, 2011.

[10] E. Dowling, *Introduction to Mathematical Economics*, McGraw-Hill, 3rd edition, 2000.

[11] D. Carney, U. Cetintemel, M. Cherniack et al., "Monitoring streams: a new class of data management applications," in *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB '02)*, pp. 215–226, Hong Kong, 2002.

[12] S.-H. Baek, E.-C. Choi, J.-D. Huh, and K.-R. Park, "Sensor information management mechanism for context-aware service in ubiquitous home," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1393–1400, 2007.

[13] T. Kawakami, B. L. N. Ly, S. Takeuchi, Y. Teranishi, K. Harumoto, and S. Nishio, "Distributed sensor information management architecture based on semantic analysis of sensing data," in *Proceedings of the International Symposium on Applications and the Internet (SAINT '08)*, pp. 353–356, Turku, Finland, August 2008.

[14] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee, "Using adaptive tracking to classify and monitor activities in a site," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 22–29, Los Alamitos, Calif, USA, June 1998.

[15] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on Database Systems*, vol. 30, no. 1, pp. 122–173, 2005.

[16] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "The design of an acquisitional query processor for sensor networks," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 491–502, San Diego, Calif, USA, June 2003.

[17] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB '04)*, pp. 588–599, Toronto, Canada, 2004.

[18] Y. Diao, D. Ganesan, G. Mathur, and P. Shenoy, "Rethinking data management for storage-centric sensor networks," in *Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research (CIDR '07)*, pp. 22–32, Asilomar, Calif, USA, January 2007.

[19] E. H. Aoki, A. Bagchi, P. Mandal, and Y. Boers, "A theoretical look at information-driven sensor management criteria," in *Proceedings of the 14th International Conference on Information Fusion (Fusion '11)*, Chicago, Ill, USA, July 2011.

[20] S. Acharya, P. B. Gibbons, V. Poosala, and S. Ramaswamy, "The Aqua approximate query answering system," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 574–576, Philadephia, Pa, USA, June 1999.

[21] Y. E. Ioannidis and V. Poosala, "Histogram-based approximation of set-valued query answers," in *Proceedings of the 25th International Conference on Very Large Data Bases (VLDB' 99)*, pp. 174–185, Edinburgh, Scotland, September 1999.

[22] K. Chakrabarti, M. Garofalakis, R. Rastogi, and K. Shim, "Approximate query processing using wavelets," in *Procedings of the 26th International Conference on Very Large Databases (VLDB' 00)*, pp. 111–122, Cairo, Egypt, September 2000.

[23] D. Ganesan, B. Greenstein, D. Perelyubskiy, D. Estrin, and J. Heidemann, "An evaluation of multi-resolution storage for sensor network," *SIGCOMM Computer Communication Review*, vol. 34, no. 1, pp. 125–130, 2004.

[24] T. H. De Groot, R. F. Tigrek, O. A. Krasnov, A. Huizing, and A. Yarovoy, "Resource allocation challenges for reconfigurable

multi-sensor networks," in *Proceedings of the 8th European Radar Conference (EuRAD '11)*, pp. 142–145, Manchester, UK, October 2011.

[25] D. Pizzocaro, A. Preece, F. Chen, T. L. Porta, and A. Bar-Noy, "Demo: a distributed architecture for heterogeneous multi sensor-task allocation," in *Proceedings of the 7th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS '11)*, Barcelona, Spain, June 2011.

[26] N. Edalat, W. Xiao, C.-K. Tham, E. Keikha, and L.-L. Ong, "A price-based adaptive task allocation for wireless sensor network," in *Proceedings of the 6th International Conference on Mobile Adhoc and Sensor Systems (MASS '09)*, pp. 888–893, October 2009.

[27] Y. Song, Y. Sun, and W. Shi, "A two-tiered on-demand resource allocation mechanism for VM-based data centers," *IEEE Transactions on Services Computing*, vol. 6, no. 1, pp. 116–129, 2013.

[28] Y. He, W. Zhu, and L. Guan, "Optimal resource allocation for pervasive health monitoring systems with body sensor networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1558–1575, 2011.

[29] T. H. de Groot, O. A. Krasnov, and A. Yarovoy, "Adaptive optimization algorithms for utility-driven resource allocation in reconfigurable multi-sensor networks," in *Proceedings of the 9th European Radar Conference*, pp. 330–333, Amsterdam, The Netherlands, November 2012.

[30] M. P. Johnson, H. Rowaihy, D. Pizzocaro et al., "Sensor-mission assignment in constrained environments," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 11, pp. 1692–1705, 2010.

[31] M. Şensoy, T. Le, W. W. Vasconcelos, T. J. Norman, and A. D. Preece, "Resource determination and allocation in sensor networks: a hybrid approach," *Computer Journal*, vol. 54, no. 3, pp. 356–372, 2011.

[32] R. Mattikalli, R. Fresnedo, P. Frank, S. Locke, and Z. Thunemann, "Optimal sensor selection and placement for perimeter defense," in *Proceedings of the 3rd IEEE International Conference on Automation Science and Engineering (IEEE CASE '07)*, pp. 911–918, Washington, DC, USA, September 2007.

[33] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention and visual search: an information theoretic approach," *Journal of Vision*, vol. 9, no. 3, article 5, pp. 1–24, 2009.

[34] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, 2011.

[35] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Overt visual attention for free-viewing and quality assessment tasks: impact of the regions of interest on a video quality metric," *Signal Processing*, vol. 25, no. 7, pp. 547–558, 2010.

[36] X.-P. Hu, L. Dempere-Marco, and E. R. Davies, "Bayesian feature evaluation for visual saliency estimation," *Pattern Recognition*, vol. 41, no. 11, pp. 3302–3312, 2008.

[37] M. Younis, M. Youssef, and K. Arisha, "Energy-aware management for cluster-based sensor networks," *Computer Networks*, vol. 43, no. 5, pp. 649–668, 2003.

[38] T. P. Huynh, Y. K. Tan, and K. J. Tseng, "Energy-aware wireless sensor network with ambient intelligence for smart LED lighting system control," in *Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society (IECON '11)*, pp. 2923–2928, Victoria, Australia, November 2011.

[39] I. Papadimitriou and L. Georgiadis, "Energy-aware routing to maximize lifetime in wireless sensor networks with mobile sink," *Journal of Communications Software and Systems*, vol. 2, no. 2, pp. 141–151, 2006.

[40] The United States Environmental Protection Agency document, "Guidance on Choosing A Sampling Design for Environmental Data Collection," EPA QA/G-5S, EPA/240/R-02/005, 2002.

[41] J. Matoušek, M. Sharir, and E. Welzl, "A subexponential bound for linear programming," *Algorithmica*, vol. 16, no. 4-5, pp. 498–516, 1996.

[42] K. Aardal and F. Eisenbrand, "Integer programming, lattices, and results in fixed dimension," *Handbooks in Operations Research and Management Science*, vol. 12, pp. 171–243, 2005.

[43] D. Silva, M. Ghanem, and Y. Guo, "WikiSensing: an online collaborative approach for sensor data management," *Sensors*, vol. 12, no. 10, pp. 13295–13332, 2012.

[44] "Siege Benchmark," http://www.joedog.org/siege-home/.

[45] Y.-K. Guo and L. Guo, "IC cloud: enabling compositional cloud," *International Journal of Automation and Computing*, vol. 8, no. 3, pp. 269–279, 2011.

[46] T. Reponen, S. A. Grinshpun, S. Trakumas et al., "Concentration gradient patterns of aerosol particles near interstate highways in the Greater Cincinnati airshed," *Journal of Environmental Monitoring*, vol. 5, no. 4, pp. 557–562, 2003.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 979546, 11 pages
http://dx.doi.org/10.1155/2013/979546

Research Article

Design Considerations of Self-Adaptive Wireless M2M Network Communication Architecture

Xiao-jun Chen,[1] Jia Ke,[2,3] Yong-zhao Zhan,[3] Man-rong Wang,[3] and Ge Deng[4]

[1] *Affiliated Hospital of Jiangsu University, Zhenjiang, Jiangsu 212013, China*
[2] *School of Management, Jiangsu University, Zhenjiang, Jiangsu 212013, China*
[3] *School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China*
[4] *University of Alabama, 1108 14th Avenue Apartment 308, Tuscaloosa, AL 35487, USA*

Correspondence should be addressed to Jia Ke; kejia@ujs.edu.cn

Received 23 June 2013; Revised 8 November 2013; Accepted 21 November 2013

Academic Editor: Jianhua He

Copyright © 2013 Xiao-jun Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For the existing heterogeneous wireless M2M network architectures, different data link layers lack a common structure. In this paper, the concept of attribute assembly layer has been proposed, and an adaptive architecture for wireless M2M network has been built, based on the hierarchical heterogeneous modeling of Ptolemy II. Finally, saturated and unsaturated conditions network experiments have been designed in the part of system test. Experimental results show that this architecture has a low memory occupation and less time cost. It unifies the data link layers for heterogeneous networks and is compatible well with the existing platforms, communication protocols, and network mechanisms. The self-adaptive network has more connected sets. The proposed architecture has good adaptive capacity and it can apply to potential communication protocols.

1. Introduction

With the rapid development of wireless communication technology and wireless M2M networks, there emerge lots of typical wireless communication technologies, mainly including 3G [1] based wireless WAN, IEEE802.16 [2] based wireless MAN [3], IEEE802.11 [4] based wireless LAN [5], wireless PAN based on IEEE802.15 [6] and blue tooth technique [7], ad hoc networks [8], sensor networks [9], wireless mesh networks [10], NGEO satellite mobile communications [11], and vehicle area networks [12]. For the diversity of existing wireless communication technology, applications and protocols of the wireless network would run over different link layers and data transferring also cannot depend on the same link-layer mechanism. Therefore, we need to consider, over the existing communication protocol architecture for wireless M2M network, if there is a mechanism independent of link layers to develop applications and protocols.

Communication protocol architecture with common operability has a good code reuse and it can meet the needs of compatibility, interoperability, and scalability for the link layer. Thus, such architecture has a great advantage in the development for wireless M2M network. Many organizations associated with M2M network have put forward various communication protocol architectures for heterogeneous networks. For instance, for the heterogeneous network composed of UMTS and WLAN, the European Telecommunication Standards Institute has defined two kinds of fusion structure, namely, the loosely coupled one [13] and the tightly coupled one [14]. The loosely coupled structure is a research focus at present [15], while the tightly coupled one is rarely applied in actual projects, which is due to the high complexity and large investment for implementing tight coupling. In the research on flight information system, US Johns Hopkins University Applied Physics Laboratory has studied how to make the mobile ad hoc networks, wireless local area networks, and satellite communication networks within the American national aviation system work synergistically on the basis of an internet-like core network [16]. In the study of utilizing the general interfaces for the wireless honeycomb communication network to extract and send information for wireless M2M networks, Athens University

Figure 1: Comparison of network architectures.

put forward a framework and provided the business implementation based on application [17]. Taking GSM/DECT and GSM/WLAN as the examples, Attalah and Ismail have analyzed the problems caused by multimode operation in the terminal design and proposed corresponding solutions [18]. In the research of Japan's e-Japan plan, a structure for future mobile communication systems MIRAI has been proposed [19]. This method attempts to realize a universal access to the local network for end users, but it has not considered the universal access to multiple networks. In addition to the previous method. Sachs has proposed the concept of general link layer (GLL) [20]. This technique designs the wireless link layer to be a general mode with compatibility, but the link layer protocols for all the wireless networks need to be changed. Finally, IETF has designed the architecture of 6LowPAN [21], shown as Figure 1(b). This architecture can be well extended on the application and protocol. However, it has not fundamentally solved the problem of automatic matching for protocols, but just adds a 6LowPAN adaptation layer to perform the protocol format conversion from IPv6 to IEEE802.15.4.

In order to ensure the adaptive capacity of communication architectures, this paper has proposed an adaptive general structure for M2M networks, shown as Figure 1(c), which is based on traditional network architecture, shown as Figure 1(a), and inspired by 6LowPAN. The proposed architecture has reassembled the network structure and brought in attribute assembly layer. The attribute assembly layer consists of two parts, which are the attribute factory and the assembly factory. The proposed architecture has been designed to adapt to different types of potential communication protocols, and such ability is revealed by the compatibility with the existing different link layers.

Saturated and unsaturated conditions network experiments have been designed in a part of system test [22]. The simulation experiments show that the adaptive network architecture can be well mapped to traditional wireless M2M network protocols, including packet distribution, packet collection, and mesh routing. With the attribute assembly layer, this proposed architecture can effectively analyze the packet attributes conveyed from the upper communication mechanisms and adapt them to the specific link layer protocol.

Figure 2: Attribute factory of the prototype.

2. Adaptive Communication Architecture

Adaptive adaptation layer is the core of the adaptive architecture and it represents the general communication architecture of wireless M2M, shown as Figure 2. The adaptive architecture has unified and simplified the specific communication protocols for various M2M networks. Therefore, it can keep good interoperability between the protocols above the adaptation layer and the layers below (MAC layer and link layer) [23], which is due to attribute factory. The adaptive adaptation layer separates communication protocols and applications from the header formats of output packets, and the standard header format of packets is generated by the independent packet format conversion module of the assembly factory. The adaptive communication architecture consists of three parts, namely: the attribute factory, the application and communication protocol, and the assembly factory. The attribute factory provides a serial of communication interfaces for above applications, so the application and communication protocol are running over the attribute factory. The assembly factory is designed to encapsulate the packets delivered from the attribute factory.

2.1. Attribute Factory. The protocol stack of attribute factory has defined a serial of communication prototypes, and the specific communication prototypes are shown in Table 1. Applications and protocols above the attribute factory can call any one or more communication prototypes. The protocol stack has a hierarchical structure, and a complicated protocol is composed of relatively simple protocols. The hierarchical structure for the protocol stack of attribute factory is shown as Figure 3.

The attribute factory supports the transmission prototypes of single hop and multihop. The multihop prototype does not specify how the packets through the network perform routing, and the routing work is assumed by the communication protocol above. The communication protocol chooses an appropriate next-hop address according to the header field of a packet. The separation of multihop model

Table 1: Communication protocol prototype.

Communication prototype	Abbreviation	Way of working
Anonymous single-hop broadcast	asb	Single-hop broadcast to all neighboring nodes, no source address and destination address
Single-hop broadcast with source address	ssb	On the basis of anonymous single-hop broadcast, adding source address to the output packet
Single-hop transmission	sh	On the basis of single-hop broadcast with source address, adding destination address to the output packet
Single-hop transmission with retransmission	shr	On the basis of single-hop transmission, introducing retransmission mechanism
Single-hop transmission with affirmance	sha	Sending a packet to the neighboring node by reliable transmission (confirmation and retransmission)
Nonrepetitive single-hop broadcast	nr-sb	Stopping forwarding the packets with the same packet attributes sent from other neighboring nodes so as to reduce the total number of packet transmission
Nonrepetitive single-hop broadcast with source address	nr-ssb	On the basis of nonrepetitive single-hop broadcast, adding the packet attribute of source address
Multihop transmission	mh	With routing protocols, sending the packet to a specific destination node in the way of hop-by-hop
Multihop transmission with affirmance	amh	On the basis of multihop transmission, introducing confirmation mechanism for adjacent forwarding
Network flooding	fld	On the basis of nonrepetitive single-hop broadcast with source address, introducing the mechanism of gradual transmission

Figure 3: Self-adaptive communication architecture.

from routing protocol contributes to implementing any one or more types of routing protocol. When multiple types of routing protocol are needed, we only need to add them, and the specific transmission of multihop is done by communication prototype. In a certain sense, the communication model has provided a high abstraction of data transmission and it has separated the transmission model from the protocol logic.

The attribute factory substitutes the header of a packet with some packet attributes, which contain the same information required in communication as the traditional packet header. Therefore, packet attributes are abstract representation for the data of header field. The attribute factory has predefined some packet attributes, including id of a packet, type of a packet, addresses of the sender and receiver, hop count, lifetime, number of retransmission, reliability, and link conditions. The application and protocol can also define additional attributes.

2.2. Assembly Factory. The assembly factory can not only generate the packet header encapsulated in the way of bit, but also generate the packet header for a specific link layer. The application and protocol pass data of the application layer to the attribute factory, which is a response for adding packet attributes, then the attribute factory passes the application data and packet attributes to the assembly factory. According to the packet attributes and the specific link layer, the packet header format conversion module of the assembly factory will generate a corresponding header and then deliver it to the sending module. Finally, the MAC layer of the sending module will decide how to transmit a packet according to the meaning of the packet header field.

Packet headers are generated by independent third party header conversion modules. These modules encapsulate corresponding data and packet attributes and consequently generate packets with typical formats. All the work mentioned above is done by the various assembly factories of the adaptive adaptation layer. Different assembly factories make different standards for the packet header and generate corresponding packets. For example, through the adaptive adaptation layer, data of the application layer can be easily encapsulated into packets with 802.15.4 MAC layer headers and those with UDP/IP headers.

Compared with the traditional protocol architecture, the advantage of adaptive adaptation layer is that applications running over it have no need to consider whether to be compatible with the underlying communication protocols

Figure 4: The transfer processes of protocol data unit between BRPfunC and BC.

or potential ones. This is because the assembly factory can generate packet headers required according to the communication protocol of link layer.

2.3. Data Flow Process. The main purpose of this paper is to implement the protocol in the platform integration and assembly. In order to achieve "reads Properties" services and protocol stack, they were separated in the platform. The former is provided by BRPfunC component, and application layer protocol stack data encapsulation, network layer, and the virtual link layer are implemented by the BC components. Figure 4 describes the transfer processes of protocol data unit between BRPfunC and BC. BRPfunC components are used to collect data and attribute assignment. BRPfunC component calls AdcReadClientC system components read interface for reading data by the sensing device, while the data is assigned to the corresponding BRPfunC object's attribute. BRPfunC component called rpfunc interface again, for the data transmission between application layer and the network layer.

In the system, BC components have the following main functions. Firstly, it provides the rpfunc interface to receive the "Read Properties" services generated packets; secondly, to support application-layer protocol stack, network layer, and the virtual link layer functions. In the application layer, control information APCI is added to the front of the original data, forming an application layer protocol data unit APDU. In the network layer, control information NPCI is added to the APDU packets, forming a network layer protocol data unit NPDU. BC components add virtual link layer protocol control information BVLC to NPDU, forming a BACnet protocol data unit BPDU. Data transmission has been achieved between BC component and IPC component, by calling udpclient interface provided with the IPC component.

3. Ptolemy II Based Adaptive Communication Architecture for Wireless M2M Network

Ptolemy is a platform for the study of heterogeneous modeling, simulation, and design for parallel or real-time embedded systems [24], which is developed by UC Berkeley University, USA. The design of wireless M2M network system in this paper has followed the requirements for role-oriented modeling and simulation. This paper has analyzed the principle and mechanism for the fusion of heterogeneous architectures in the wireless M2M network system and it has completed the system modeling based on the modeling and simulation platform of Ptolemy II.

3.1. Establishment of the Whole Model for Wireless M2M Network. Based on the wireless M2M network prototype proposed in the literature [25], this paper has extended the processing module of the heterogeneous architecture for wireless M2M network. The main UML class structure is shown as Figure 4. A complete model for wireless M2M network system is composed of four parts, namely, the wireless director (WirelessDirector), the node models (NodeModels), the wireless channel (WirelessChannel), and the wireless IO port (WirelessIOPort), and all other parts can be extended by these four main classes. The wireless director is a response for commanding all of the system behaviors in the wireless M2M network system and it determines the nodes listening to broadcast by wireless channel. The node models take charge of creating wireless I/O port and can check whether the input port has received wireless signals or not. If no signal is at present, this node will broadcast to all the nodes in its reachable area. Node model is an atomic role. Therefore, for the wireless M2M network systems with different characteristics, the feature of a node depends on the synthesis of such role. The wireless channel determines the transmission of wireless signals among nodes. It not only judges which nodes within the transmission range, but also considers the delay and stability of signals in the transmission. The delay and transmission range of channels are determined by the signal range parameters and signal stability of the sender. The wireless channel is the backbone of the wireless M2M network system and it is a response for the transmission of all information in the network. The wireless I/O port is one type of the Ptolemy II I/O ports and it is designed to send and receive wireless signals.

The logic layer structures for attribute factory and assembly factory based on wireless channel are shown as the red boxes in Figure 5. The attribute factory encapsulates behaviors of the upper applications, communication protocols, and routing protocols, while the assembly factory is a response for producing packet headers and distributing the packets to corresponding networks.

3.2. Establishment of Heterogeneous System Model. As shown in Figure 6, based on the system-level behavior description of Ptolemy II for the heterogeneous network communication system of wireless M2M, the attribute factory model of the heterogeneous network communication system of wireless M2M can be divided into four channel modules, wherein applications correspond to the module of ApplicationChannel, communication protocols correspond to that of CommunicationProtocolChannel, and routing protocols

Figure 5: UML modeling for the wireless M2M network architecture based on Ptolemy II.

correspond to that of RoutingProtocolChannel, respectively. The assembly factory has simulated the communication among the heterogeneous networks of 802.15.4, 802.11 and Ethernet.

3.3. Network Connectivity. To establish the coordinate system, set the length of a region as $x(m)$, $x \geq 0$, right-to-left position of the node denoted by V_i, where $i = 1, 2, \ldots$ node density is denoted by ρ (ρ is the number of nodes per meter). Node distribution satisfies the Poisson distribution of $\lambda = \rho x$; k nodes in the area are the probability mass function which is $p(k) = \lambda^k e^{-\lambda}/k!$; a Poisson process with the above model corresponds to a counting process; that is at certain time x, the total number of nodes in the region $K(x)$ is a Poisson process counting process. According to the nature of the Poisson distribution function, x the total number of

Figure 6: Heterogeneous system model.

regional nodes purpose $K(x)$ expectation and variance is λ; that is,

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{(\rho x)^k e^{-\rho\lambda}}{k!}, \tag{1}$$

$$p(k, x) = \frac{(\rho x)^k e^{-\rho\lambda}}{k!}, \tag{2}$$

$$E[K] = \lambda = \rho x, \tag{3}$$

$$D[K] = E\left[K^2\right] - (E[K])^2 = \lambda = \rho x. \tag{4}$$

Definition 1. R is the node supporting wireless LAN protocol maximum transmission distance. So that $x = R$, then among the distance R, at least one communication node probability of $\Pr\{k \geq 1\} = 1 - \Pr\{k = 0\} = 1 - e^{-\rho R}$. Let $d(i, i+1) = V_{i+1} - V_i$ as any two adjacent distances between nodes in x sections, representing nodes successively entering the area $d(i, i+1)$ close to the arrival time. For ease description, the d_i written as $d_{(i-1,i)}$. Setting the origin of coordinates as V_0, there is $d_1 = |V_1 - V_0|$, in the sequence $\{d_i, i = 1, 2, \ldots\}$.

Because of $V_0 = 0$, $V_i > 0$, it is possible to dispense with the absolute value of the operator; that is, $d_1 = V_1 - V_0 = V_1$; $d_2 = V_2 - V_1 = V_2 - d_1$. In x sections, the emergence of the events of a node $\{d_1 > x\}$, using formula (1), the Poisson process in the interval $[0, x]$ has no node; that is,

$$\Pr\{d_1 > x\} = \Pr\{k(x) = 0\} = e^{-\rho x}. \tag{5}$$

So, the exponential distribution is d_1 with mean $1/\rho$. Consider $\Pr\{d_2 > x\} = E[\Pr\{d_2 > x d_1\}]$. However,

$$\begin{aligned} \Pr\{d_2 > x \mid d_1 = v\} &= \Pr \text{ node not in } (v, v+x] \mid d1 \\ &= v = \Pr\{\text{node not in } (v, v+x]\} = e^{-\rho x}. \end{aligned} \tag{6}$$

In formula (6), with d_2 also as the exponential distribution $1/\rho$ with mean random variable, an inference can be drawn.

Inference 1. $\{d_i, i = 1, 2, \ldots\}$ are independent and identically distributed exponential random variable with mean $1/\rho$. Consider the following:

$$\begin{aligned} \Pr\{d_1 \leq x\} &= 1 - e^{-\rho x}, \\ \Pr\{d_2 \leq x\} &= 1 - e^{-\rho x}, \\ \Pr\{d_i \leq x\} &= 1 - e^{-\rho x}, \\ E[d_i] &= \frac{1}{\rho}, \\ D[d_i] = E\left[d_i^2\right] - (E[d_i])^2 &= \frac{1}{\rho^2}. \end{aligned} \tag{7}$$

For a particular distance x in region L, if there are k nodes, the connectivity probability between node i and node $i+1$ is presented as

$$\Pr\{d_{(i,i+1)} \leq R\} = 1 - e^{-\rho R}. \tag{8}$$

For the distance of the node in the region L, the connectivity probability between nodes is presented as

$$E[d_{(i,i+1)}] = \frac{1}{\rho}. \tag{9}$$

The connectivity probability of nodes in L region is presented as

$$\begin{aligned} P_\text{Connection}(L) &= \prod_{i=1}^{k} \Pr\{d_i \leq R\} \\ &= \prod_{i=1}^{k} \left(1 - e^{-\rho R}\right), \\ k = \rho L = \left(1 - e^{-\rho R}\right)^{k-1} &= \left(1 - e^{-\rho R}\right)^{\rho L - 1}. \end{aligned} \tag{10}$$

Analyze the distribution of the position coordinates of the V_k node.

As $d_{(i,i+1)} = V_{i+1} - V_i$, $V_k = \sum_{i=1}^{k} d_i$, $k \geq 1$; k is total number of nodes in the distance x; V_k is derivation of the cumulative distribution function as follows: k nodes appear in the distance x if only if interval $[0, x]$ is at least the number of nodes k; that is,

$$K(x) \geq k \Longleftrightarrow V_k \leq x, \tag{11}$$

$$\Pr\{V_k \leq x\} = \Pr\{K(x) \geq k\}. \tag{12}$$

Take formula (1) into formula (12), a cumulative distribution function about V_k is

$$\Pr\{V_k \leq x\} = \Pr\{K(x) \geq k\} = \sum_{j=k}^{\infty} e^{-\rho x} \frac{(\rho x)^j}{j!}. \tag{13}$$

From the above equation, the probability mass function of V_k is

$$p(V_k, x) = \rho e^{-\rho x} \frac{(\rho x)^{k-1}}{(k-1)!}. \tag{14}$$

(a) Initial nodes (b) Initial topological structure (c) Operation

(d) Attenuation topological structure (e) Operation (attenuation)

Figure 7: Experimental simulation for the self-adaptive network.

With formula (14), V_k is a parameter k and $\rho\Gamma$ is a gamma distribution. Obey the gamma distribution; expectations and variance were

$$E[V_k] = \frac{k}{\rho},$$
$$D[V_k] = E\left[V_k^2\right] - (E[V_k])^2 = \frac{k}{\rho^2}. \quad (15)$$

These conclusions will be applied to the network under conditions of saturated and nonsaturated conditions model simulation experiments.

3.4. Saturated and Unsaturated Conditions. In the traditional methods, the paper [22] proposed using a two-dimensional Markov chain model for modeling wireless network. By the steady-state calculation of Markov chain model, transmission probability of nodes and transmit data collision probability were obtained in any slot, which lead to major system performance indicators: saturation throughput. However, the radio channel supposed as an ideal channel in this theory, which does not consider the problem of hidden node, exposed node and channel error conditions in multihop wireless networks. Traffic nodes are at saturation, which means data has been in the queue. Capacity of each node's MAC layer transmission queue is infinite for overflow condition. Under construction modeling process, the situation are not be considered, which are the limited number of retransmissions, the back off freezing, difference of maximum number of retransmissions and the maximum backward progression.

In order to solve the above problems, researchers have proposed the corresponding improvement model to study the system throughput, latency, and energy consumption performance. Establish improved nonsaturated conditions wireless network analysis model, using the two-dimensional Markov model simplicity, considering the practical application of saturated and unsaturated traffic situation, and consider the protocol bakeoffs timer frozen state and limited number of retransmissions. Combining the steady-state solving Markov model and queuing model, we deduce the type throughput and delay performance model under different loads.

4. System Test

The experiments were done on a machine, which is configured with Pentium Dual-Core 2.5 G CPU, 1 G SDRAM, and 5400RPM IDE, with the operation system of Windows2000 Server. In order to facilitate comparison of the topology, we have fixed the network scale with 100 m ∗ 100 m ∗ 100 m and the number of nodes with 200 m. Meanwhile, the node radius has been fixed with the best coverage radius of the network.

4.1. Analysis of Network Communication. As shown in Figure 7(a), 200 M2M nodes are randomly deployed in the test area. The simulation diagrams, shown as Figures 7(b)–7(e), show that the context of the virtual topology is very clear

Table 2: Parameter values for simulation and model.

Parameters	Values
Channel frequeney	24 GHz
Channel bit rate	1 Mbit/s
PHY header	192 bits
MAC header	224 bits
LlcSnap header	64 bits
ACK	112 bits
Packet payload (Ldata)	8192 bits
CWmin	31
CWmax	1023
Retransmission limit (m)	5
Propagation delay (δ)	1 μs
Slot time (σ)	20 μs
SIFS	10 μs
DIFS (=2σ + SIFS)	50 μs
tdata	8672 μs
tack	304 μs

and there is good connectivity among the nodes. In Figures 7(b) and 7(d), the black line represents the communication link based on 802.15.4, the blue one represents that based on 802.11, and the red one represents that based on Ethernet. Whenever it is in the initial condition or after the attenuation of the communication capacity, effective network communication architecture can be well built.

To set the scene of a unified simulation and modeling, modify Ptolemy simulation platform chosen at random bake-offs value function and the propagation delay function and suppose that all the network nodes can communicate with each other that there are no hidden nodes in the network. Nodes using the basic access, the node's physical layer mechanism for debugging, network simulation parameters, and the model parameters are shown in Table 2.

Define that the normalized throughput S is the unit of data transmission slots. Payload length ratio and the time slot [26], which means that the formula is

$$S = \frac{E\,[\text{data}]}{E\,[\text{slot}]} = \frac{L_{\text{data}} p_s'}{p_s' T_s + p_I' T_I + p_c' T_c}. \tag{16}$$

p_s' is the probability of success sending the radio channel condition in any time slot, p_I' is the probability of idle radio channel, and p_c' is the radio channel collision probability; the formula is

$$\begin{aligned} p_s' &= n\tau(1-\tau)^{n-1}\left(1-p_e\right), \\ p_I' &= (1-\tau)^n \\ p_c' &= 1-(1-\tau)^n - n\tau(1-\tau)^{n-1}. \end{aligned} \tag{17}$$

4.2. Memory Cost. Since the memory of a M2M node is limited, space occupation of the adaptive adaptation layer becomes an important measure of the feasibility and validity. The total memory occupation can be divided into two parts, wherein one part is used by the precompiled code, which is stored in on-chip ROM, and the other is the data segment for code running, including various kinds of buffers. Adaptive buffer contains applications and various packet attributes, and the drivers have system priority while writing. The queue buffer for the physical layer is allocated dynamically by the queue pool and its initial content is replicated from the adaptive buffer. The specific number and management of the queue buffers are decided by the corresponding application. The loading conditions of all the modules are shown in Table 3.

Table 3: Module load status.

Module	ROM (bytes)	RAM (bytes)
Adaptive buffer	416	168
Queue buffers	320	730
asb	134	2
ssb	114	2
sh	124	2
shr	272	16
sha	254	4
nr-sb	390	15
nr-ssb	406	15
mh	246	2
amh	266	3
fld	356	7

The experimental result shows that the cost of the queue buffers is the largest, which depends on the number of queue buffers and the allocation of the application and protocol.

4.3. Data Round-Trip and Execution Time. To measure the impact of the cost of packet attributes in the adaptive adaptation layer on packet transmission, we have compared the communication performances of 802.11, 802.15.4, Ethernet, and self-adaptive network. For the link layer protocols, we have tested the transmission between two nodes with unicast and broadcast, respectively. As shown in Figure 8, the round-trip time of the above four protocol stacks is nearly the same, which is because that the round-trip time largely depends on the time waiting for the waking up of another node.

The execution time of a prototype is calculated from an application calling the performance function of the adaptive adaptation layer to the successful sending of packets. In this paper, a comparison of asb, ssb, sh, sha, and mh was made. Considering the impact of data load size, we have tested the prototypes above using 802.11, 802.15.4, Ethernet, and self-adaptive network, respectively. The execution time for different prototypes is shown in Figure 9.

As the experimental results show, there is no significant difference between the execution time of the underlying prototypes, such as asb, ssb, and sh, and that of communication prototypes for the traditional protocol stack, while the execution time of the upper prototypes, such as sha and mh, is more than that of the traditional protocol stack, usually 0.5 ms. The reason for the above phenomenon is that the upper communication prototypes need to perform packet

Figure 8: Comparison for the round-trip time.

Figure 9: Comparison for the execution time.

attributes conversion, which the protocol stacks of 802.11 and 802.15.4 do not need, and in such process they need additional time to transfer data from the adaptive buffer to the queue buffer.

4.4. Network Connectivity Test. To verify the model, this paper generates multiple sequences in consistent Poisson

Figure 10: The relationships between probability of connectivity and transmission distance.

process; the maximum transmission distance between nodes is 100 m. Figure 10 shows the connectivity probability between any two nodes. In the case of constant node density, number of nodes into the same set connected increases, the probability of internode communication increasing with maximum communication distance between nodes increasing. With node density increasing in the region, the connecting probability transition width, of any two nodes in this region, becomes narrower. Self-adaptive network connectivity has higher degree, in order to make significant changes to this phenomenon, self-adaptive network connectivity has higher degree.

Figure 11 shows the connected sets diameter (length) and the maximum communication distance between nodes relationship, as connected sets diameter maximum communication distance between nodes increases exponentially in a growing trend, and with the increase in node density and connected sets diameter growing faster, self-adaptive network has more connected sets.

4.5. Transmission Rate and Throughput. To compare the performance of the adaptive architecture protocol stack with that of IEEE802.15.4, 802.11, and Ethernet, their channel transmission capacity and throughput have been measured, respectively, and we have made a comparison of them. As shown in Figure 12, in the initial stage, channel transmission rates of the four architectures have similar linear changes. With the increase in the node load, the transmission rates of the architectures have slight changes but still maintain at a high level. Owing to the channel adaptation of protocols, the channel utilization of the adaptive architecture can reach more than 98%, which is higher than that of the other two architectures. As shown in Figure 13, the node throughput capacity will decrease with the increase in node

Figure 11: The relationships between connected sets and node density.

Figure 12: Compared with the transmission.

number, which is because of the channel congestion caused by the increase in node number. As the protocol stack of IEEE802.15.4 is simple and it needs much less buffers, the throughput of IEEE802.15.4 is the highest among the four and it remains relatively stable. Both the adaptive structure and 802.11 add the adaptation layer, and for the implementation

Figure 13: Compared with the throughput.

of congestion control, the protocol complexity is increased. In addition, the common operability and matching support for different channels need to be considered in certain situations. For the above reasons, the node throughput of 802.11 and the adaptive structure would reduce accordingly. However, their node throughputs tend to be unchanged with the increase in node number.

5. Conclusion

In view of current development situation of the wireless M2M networks and the existing various architectures, this paper has proposed novel adaptive communication architecture. The architecture with adaptive adaptation mechanism can well solve the problem of cross-layer communication. For the protocol stack with hierarchy, the division for different protocol logics can ensure the application to mask the underlying platform and it can also provide unified attribute interface services for the upper layer and there is no need to break the existing protocol rules. Moreover, using the lightweight adaptive communication architecture can reduce the developing complexity, which is caused by the need for changing the code frequently to implement different communication protocols. The architecture proposed in this paper has good scalability and operability. Not only can it be well compatible with the MAC layer of 802.11, Ethernet, and IEEE802.15.4, but also be independent of the link layer and can interconnect with any network seamlessly. In order to specify the traffic loads of the nodes, we make definition and inference about network connectivity and it has been applied to the network under conditions of saturated and

nonsaturated conditions model simulation experiments. As proved by the simulation tests, the memory occupation and time cost of the proposed architecture are relatively low. Because of self-adaptive network for the model with a more flexible regulation, its connectivity has higher degree and has more connected sets. What is more, it has high channel utilization and moderate throughput. In a word, the proposed architecture can adapt to all kinds of wireless M2M network applications.

Acknowledgments

This research has partially been supported by the National Natural Science Foundation of China under Grant nos. 60673190, and 61203244, National Natural Science Foundation of Jiangsu Province under Grant no. BK2009199, MOE (Ministry of Education in China) Project of Humanities and Social Sciences under Grant no. 12YJCZH020, and Science and Technology Support Program of ZhenJiang City under Grant no. GY2010023.

References

[1] "Security Architecture," 3GPP TS 23. 234 V6. 1. 0. 3G, 2004.

[2] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE Std 802. 16-2004, 2004.

[3] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, vol. 43, no. 9, pp. S23–S30, 2005.

[4] IEEE 802.11s Workgroup, "Joint SEE-Mesh/Wi-Mesh Proposal to 802.11 TGs," IEEE Document IEEE 802 11-06/328r0.

[5] "Draft Supplement to Standard for Telecommunications and Information Exchange between Systems-LAN/MAN Specific Requirements-Part II: MAC enhancements for quality of service (QoS)," IEEE-802. 11WG. IEEE 802. 11e Standard Draft/D8. 0, 2004.

[6] "Part 15. 4: Wireless MAC and PHY layer specifications for low-rate wireless personal area networks," IEEE Std 802. 15. 4-2006, 2006.

[7] J. Jaros and R. Vrba, "Bluetooth Technology for Smart Sensor Implementation," Czech Republic 15/11/2001, 1SSN, 1213-161X. Department of Microelectronics, 2002.

[8] X. Sun and E. J. Coyle, "The effects of motion on distributed detection in mobile ad hoc sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 460924, 14 pages, 2012.

[9] J. Ren, Y. Zhang, K. Liu et al., "An energy-efficient cyclic diversionary routing strategy against global eavesdroppers in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 834245, 15 pages, 2013.

[10] M. Ibnkahla, Q. M. Rahman, A. I. Sulyman, H. A. Al-Asady, J. Yuan, and A. Safwat, "High-speed satellite mobile communications: technologies and challenges," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 312–338, 2004.

[11] D. G. Oh, P. Kim, Y. J. Song, S. I. Jeon, and H.-J. Lee, "Design considerations of satellite-based vehicular broadband networks," *IEEE Wireless Communications*, vol. 12, no. 5, pp. 91–97, 2005.

[12] J. P. Conti, "The long road to WiMax," *IEE Review*, vol. 51, no. 10, pp. 38–42, 2005.

[13] A. Doufexi, E. Tameh, A. Nix, S. Armour, and A. Molina, "Hotspot wireless LANs to enhance the performance of 3G and beyond cellular networks," *IEEE Communications Magazine*, vol. 41, no. 7, pp. 58–65, 2003.

[14] ETSI TR and Broadband Radio Access Networks (BRAN), "HIPERLAN Type 2. Requirements and Architectures for Interworking between HIPERLAN/2 and Third Generation Cellular Systems. v. 1.1.1," 2001.

[15] M. Stemm and R. H. Katz, "Vertical handoffs in wireless overlay networks," *Mobile Networks and Applications*, vol. 3, no. 4, pp. 335–350, 1998.

[16] J. Burbank, R. Nichols, S. Munjal, R. Pattay, and W. Kasch, "Advanced communications networking concepts for the national airspace system," in *Proceedings of the IEEE Aerospace Conference*, pp. 1905–1923, March 2005.

[17] V. Tsetsos, G. Alyfantis, T. Hasiotis, O. Sekkas, and S. Hadjiefthymiades, "Commercial wireless sensor networks: technical and business issues," in *Proceedings of the 2nd Annual Conference on Wireless onDemand Network Systems and Services (WONS '05)*, pp. 166–173, St. Moritz, Switzerland, 2005.

[18] J. G. Attalah and M. Ismail, "Future 4G front-ends enabling smooth vertical handovers," *IEEE Circuits and Devices Magazine*, vol. 22, no. 1, pp. 6–15, 2006.

[19] G. Wu, M. Mizuno, and P. J. M. Havinga, "MIRAI architecture for heterogeneous network," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 126–134, 2002.

[20] J. Sachs, "A generic link layer for future generation wireless networking," in *Proceedings of the International Conference on Communications (ICC '03)*, pp. 834–838, May 2003.

[21] C. Bormann, "Draft-bormann-6lowpan-cbhc-00.txt. Context-based Header Compression for 6lowpan," The 6LoWPAN Working Group, 2008.

[22] S. Kim, S. Kim, and D. S. Eom, "A robust and space-efficient stack management method for wireless sensor network OS with scarce hardware resources," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 962151, 18 pages, 2012.

[23] J. Kabara and M. Calle, "MAC protocols used by wireless sensor networks and a general method of performance evaluation," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 834784, 11 pages, 2012.

[24] J. Eker, J. W. Janneck, E. A. Lee et al., "Taming heterogeneity: the ptolemy approach," *Proceedings of the IEEE*, vol. 91, no. 1, pp. 127–143, 2003.

[25] E. A. Lee, "Finite State Machines and Modal Models in Ptolemy II," Technical Report UCB/EECS-2009-151, EECS Department, University of California, Berkeley, Calif, USA, 2009.

[26] V. Gau and J.-N. Hwang, "Adaptive probabilistic broadcasting over dense wireless ad hoc networks," *International Journal of Digital Multimedia Broadcasting*, vol. 2010, Article ID 741792, 12 pages, 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 293565, 9 pages
http://dx.doi.org/10.1155/2013/293565

Research Article

Collaborative Relay Beamforming Based on Minimum Power for M2M Devices in Multicell Systems

Xiaoning Zhang,[1] Da Wang,[1] Lin Bai,[2] and Chen Chen[1]

[1] *State Key Laboratory of Advanced Optical Communication Systems and Networks, Peking University, Beijing 100871, China*
[2] *School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Lin Bai; l.bai@buaa.edu.cn and Chen Chen; c.chen@pku.edu.cn

Received 17 August 2013; Revised 7 November 2013; Accepted 21 November 2013

Academic Editor: Jianhua He

Copyright © 2013 Xiaoning Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, machine-to-machine (M2M) communication has been studied in the single cell system. However, in the multicell system multiple M2M type devices at the edge of a cell may suffer from the strong interference that consists of the intercell interference from other cells and the intracell interference from other M2M devices in the local cell. In this paper, we study the relay beamforming strategy to guarantee the quality of service (QoS) requirements of the multiple destination devices in multicell systems. We minimize the transmit power of the base stations (BSs) and relays to save the power of M2M devices, while guaranteeing the receive signal-to-interference-and-noise ratio (SINR) of the destination devices. The main contribution of this paper is that we propose an iterative algorithm to jointly optimize the BS and relay beamforming weights with minimizing the BS and relay power under the receive SINR constraints in the perfect channel state information (CSI). Using the semidefinite relaxation (SDR) technology, the optimization problems for the BS and relay beamforming weights can be effectively solved. In addition, we also discuss the issue of imperfect CSI in practice. Simulation results validate our theoretical analysis and demonstrate that our proposed iterative scheme can achieve near-optimal performance within a few iterations.

1. Introduction

Recently, machine-to-machine (M2M) communication has been studied by many standards as the European Telecommunications Standards Institute (ETSI) [1] and 3rd Generation Partnership Project (3GPP) [2, 3]. Machine-type communication devices, such as smart meters, automotive applications, and smart phones, can communicate in M2M systems, which do not necessarily need human intervention [4]. According to the 3GPP standards, the base station (BS) can be used to manage the M2M devices and many BSs can divide the M2M devices into multiple cells [5]. The M2M devices communicate with each other through the BS in the cell system [6].

To offer the high data-rate for the M2M systems, multiple-input multiple-output (MIMO) antenna systems can increase the channel capacity for given transmit power and bandwith in previous studies [7]. However, the performance of MIMO systems may degrade due to correlated signals among antennas and cannot offer high coverage in large areas for M2M devices. Recently, many literatures research on the relay devices to guarantee the quality of service (QoS) requirements of the destination devices [8]. The BS transmit signals to the destination devices with the help of one or more relay devices in wireless M2M networks in Figure 1.

In general, there are three types of relay schemes, including the amplify-and-forward (AF) [9], decode-and-forward (DF) [10], and compress-and-forward (CF) [11] schemes. In the AF scheme, relay devices receive the signals transmitted from the source and forward the amplified signals to the destinations. In the DF scheme, relay devices decode and reencoded the received signals to the destinations. In the CF scheme, relay devices send quantized and compressed signals to the destinations. Among them the AF scheme is the most attractive for the M2M relay communication due to its low implementation complexity. With the aid of channel state information (CSI), relay devices can be designed to work collaboratively to guarantee the QoS requirements by forming a virtual beamforming system [12, 13].

Figure 1: The wireless multicell M2M system.

Most existing contexts have studied relay beamforming strategies in a single cell. Relay strategies in sensor networks with minimum mean square error (MMSE) performance are designed in [8], which is subject to local and global relay power constraints, respectively. Elkheir et al. [14] presented an M2M relay communication scheme, which is subject to the total relay power under specific performance constraints and the individual power of each relay. Wang et al. [15] proposed the iterative strategies to jointly optimize the source antenna selection and the collaborative relay beamforming weights in wireless M2M networks. Li et al. [16] proposed two algorithms with random user scheduling and greedy user scheduling in maximizing the achievable user number by dynamical power assignment in the cellular system. Chen et al. [17] considered signal transmission with the aid of multiple half-duplex single-antenna relay nodes using the AF strategy for a multiuser wireless M2M communication system and proposed two suboptimal relay beamforming schemes that only require local CSI to minimize mean square error (MSE) for all the users with nonorthogonal channels. Zheng et al. proposed the relay beamforming schemes that achieve maximum the received signal-to-noise ratio (SNR) for a single destination under both total and individual power constraints with the aid of perfect CSI [18] and imperfect CSI [19]. In addition, Choi investigated the distributed beamforming for AF relay nodes when a consensus algorithm was employed for cooperative beamforming in [20] and the minimum mean square error (MMSE) criteria were used in [21]. Nguyen et al. studied the relay beamforming schemes to minimize the total relay power in a multiple AF relay network with multiple source-destination pairs under the signal-to-interference-and-noise ratio (SINR) [22] and SNR requirement [23] at each destination, respectively. However, in the multicell system destination devices at the edge of a cell may suffer from the high signal attenuation, as well as the strong interference that consists of the intercell interference from other cells and the intracell interference from other M2M devices in the local cell. In such a case, coordination multicell processing of the BSs and relay devices in adjacent cells is needed to cope with the severe intercell and intracell interference [24, 25].

In this paper, we study the collaborative relay beamforming for multiple M2M destination devices which suffered intercell and intracell interference in the multicell systems. The relay devices can coordinately form virtual beams to transmit beamforming signals from a BS towards multiple destination devices in the AF scheme. Due to the power limitation of the M2M devices, we optimize the BS and relay beamforming weight with minimizing the BS and relay power to guarantee the receive SINR for multiple destination devices with the aid of perfect CSI. We propose an iterative scheme to jointly optimize the BS and relay beamforming weights to deal with the severe intercell and intracell interference for multiple destination devices in the multicell system. Using the semidefinite relaxation (SDR) technique [26], we show that the optimization problems for the BS and relay beamforming weights can be converted into semidefinite programming (SDP) problems, which can be effectively solved by interior-point methods.

The rest of this paper is organized as follows. In Section 2, we describe a multicell system model with multiple destination devices in each cell which suffered from the intercell and intracell interference. Sections 3 and 4 optimize the BS and relay beamforming weights to minimize the BS and relay power under the receive SINR constraints, respectively. Then, we propose an iterative scheme to jointly optimize the BS and relay beamforming weights to deal with the severe intercell and intracell interference for multiple destination devices in Section 5. Simulation results are presented and discussed in Section 6. The issue of imperfect CSI is discussed in Section 7. Finally, conclusions are drawn in Section 8.

Notation. Vectors are written in boldface lowercase letters, for example, $\mathbf{x}$, while matrices are denoted by boldface uppercase letters for example, $\mathbf{X}$. $\mathrm{Tr}(\mathbf{S})$ represents the trace of a matrix $\mathbf{S}$, and $\mathrm{diag}(s_1, s_2, \ldots, s_k)$ denotes a diagonal square matrix with $s_1, s_2, \ldots, s_k$ as the diagonal elements. The superscripts $*$, T, and H stand for the conjugate, transposition, and Hermitian of a complex vector or matrix, respectively. $\mathbb{E}[z]$ calculates the expectation of a random entity z, while $|z|$ takes the modulus of a complex number z. $\mathbf{A} \succeq \mathbf{0}$ means that $\mathbf{A}$ is a hermitian semidefinite matrix and O means the upper bound of the computational complexity. In addition, the real and complex number fields are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively.

2. System Model

In this section, we describe a multicell system with L separate cells. In each cell, there are one BS with D_L antennas, M single-antenna AF relay devices, and K destination M2M devices, where $K \leq D_L$. If $K > D_L$, the BS can select destinations to transmit the signals to exploit selection

diversity among destinations [27]. The M2M devices may not only suffer the inter-cell interference from other cells, but also suffer the intracell interference from other M2M devices in the local cell. The system model of the lth cell is shown in Figure 2. We assume that the L cells operate on the same frequency band over independent flat fading channels. Since we focus on the performance enhancement at the edge of a cell, we do not consider the direct link between the BS and destination M2M devices. The k destination devices receive the interfering signal from the relay devices of the lth cell, which is denoted by $i_{l,k}$ in Figure 2.

In Figure 2, we consider the two-step AF relay protocol. In the first step, the lth BS broadcasts the K destinations' signal $s_{l,k}$, $k = 1, 2, \ldots, K$, to the relay devices using beamforming vectors in the lth cell. We assume that destinations' signals are independent with $\mathbb{E}[s_{l,k}] = 0$ and $\mathbb{E}[|s_{l,k}|^2] = P_{l,k}$. In the second step, the mth relay devices forward the received signals $r_{l,m}$ to the destinations with certain AF weights of the coordinated relay beamforming. These relay forwarded signals are the ICI for the other users in adjacent cells.

Thus, during the first step, the relay device m in the lth cell receives the signal $r_{l,m}$, which can be given by

$$r_{l,m} = \mathbf{g}_{l,m} \sum_{k=1}^{K} \mathbf{w}_{l,k} s_{l,k} + u_{l,m}, \tag{1}$$

where $\mathbf{g}_{l,m}$ is the channel coefficient row vector from the lth BS to the mth relay and $u_{l,m}$ is the background noise at the mth relay in the lth cell with zero mean and variance of $N_{l,m}$. $\mathbf{w}_{l,k}$ denotes the lth BS beamforming column vector intended for the kth destination in the lth cell.

Then, during the second step, the transmitted signal by the mth relay devices is

$$x_{l,m} = v_{l,m} r_{l,m}, \tag{2}$$

$$\begin{aligned} y_{l,k} &= \sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} r_{l,m} + \sum_{l' \neq l}^{L} \sum_{m=1}^{M} h_{l'\to l,m,k} v_{l',m} r_{l',m} + n_{l,k} \\ &= \sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} \mathbf{g}_{l,m} \mathbf{w}_{l,k} s_{l,k} \\ &\quad + \sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} \mathbf{g}_{l,m} \sum_{k' \neq k}^{K} \mathbf{w}_{l,k'} s_{l,k'} \\ &\quad + \sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} u_{l,m} \\ &\quad + \sum_{l' \neq l}^{L} \sum_{m=1}^{M} h_{l'\to l,m,k} v_{l',m} \mathbf{g}_{l',m} \sum_{k'=1}^{K} \mathbf{w}_{l',k'} s_{l',k'} \\ &\quad + \sum_{l' \neq l}^{L} \sum_{m=1}^{M} h_{l'\to l,m,k} v_{l',m} u_{l',m} + n_{l,k}, \end{aligned} \tag{3}$$

where $v_{l,m}$ is the complex-valued weight at the mth relay in the lth cell. Thus, the received signal at the kth destination in the lth cell is shown in (3) at the bottom of the page, where $h_{l'\to l,m,k}$ is the channel coefficient from the mth relay in the l'th cell to the kth destination in the lth cell and $n_{l,k}$ is the background noise at the kth destination in the lth cell with zero-mean and variance of $N_{l,k}$. Note that $\sum_{l' \neq l}^{L} \sum_{m=1}^{M} h_{l'\to l,m,k} v_{l',m} r_{l',m}$ is the intercell interfering signals from the relay devices of the other cells and $\sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} \mathbf{g}_{l,m} \sum_{k' \neq k}^{K} \mathbf{w}_{l,k'} s_{l,k'}$ is the intracell interfering signals from other destination devices in the local cell. Thus, the received SINR at the kth destination in the lth cell is given by

$$\mathrm{SINR}_{l,k} = \frac{P_s^{l,k}}{P_{\text{inter}}^{l,k} + P_{\text{intra}}^{l,k} + P_n^{l,k}}, \tag{4}$$

where $P_s^{l,k}$ is desired signal power at the kth destination in the l cell,

$$P_s^{l,k} = P_{l,k} \left| \sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} \mathbf{g}_{l,m} \mathbf{w}_{l,k} \right|^2, \tag{5}$$

$P_{\text{inter}}^{l,k}$ is the inter-cell interfering power from the relay devices of the other cells,

$$P_{\text{inter}}^{l,k} = \sum_{l' \neq l}^{L} \sum_{k'=1}^{K} P_{l',k'} \left| \sum_{m=1}^{M} h_{l'\to l,m,k} v_{l',m} \mathbf{g}_{l',m} \mathbf{w}_{l',k'} \right|^2, \tag{6}$$

$P_{\text{intra}}^{l,k}$ is the intracell interfering power from other destination devices in the local cell,

$$P_{\text{intra}}^{l,k} = \sum_{k' \neq k}^{K} P_{l,k'} \left| \sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} \mathbf{g}_{l,m} \mathbf{w}_{l,k'} \right|^2, \tag{7}$$

and $P_n^{l,k}$ is the noise power at the kth user in the lth cell,

$$\begin{aligned} P_n^{l,k} &= \sum_{m=1}^{M} \left| h_{l\to l,m,k} \right|^2 \left| v_{l,m} \right|^2 N_{l,m} \\ &\quad + \sum_{l'=1, l' \neq l}^{L} \sum_{m=1}^{M} \left| h_{l'\to l,m,k} \right|^2 \left| v_{l',m} \right|^2 N_{l',m} + N_{l,k} \\ &= \sum_{l'=1}^{L} \sum_{m=1}^{M} \left| h_{l'\to l,m,k} \right|^2 \left| v_{l',m} \right|^2 N_{l',m} + N_{l,k}. \end{aligned} \tag{8}$$

3. Minimum BS Power with SINR Constraints

In this section, we optimize the BS beamforming vectors to minimize the BS power in the multicell system under the destinations' SINR constraints. The lth BS total transmit power is

$$P_B^l = \sum_{k=1}^{K} P_{l,k} \mathbf{w}_{l,k}^H \mathbf{w}_{l,k}. \tag{9}$$

FIGURE 2: The system model of the lth cell with multiple destinations in multicell systems.

Thus, we can formulate the problem as

$$\begin{aligned}
&\min_{\mathbf{w}_{l,k}} \quad \sum_{l=1}^{L} P_B^l \\
&\text{subject to} \quad \mathrm{SINR}_{l,k} \geq \Gamma_{l,k}, \\
&\qquad l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K,
\end{aligned} \tag{10}$$

$$\begin{aligned}
\mathrm{SINR}_{l,k} = {} & P_{l,k} \mathbf{w}_{l,k}^H \mathbf{A}_{l\to l,k} \mathbf{w}_{l,k} \\
& \times \left(\sum_{k' \neq k}^{K} P_{l,k'} \mathbf{w}_{l,k'}^H \mathbf{A}_{l\to l,k} \mathbf{w}_{l,k'} \right. \\
& \left. + \sum_{l' \neq l}^{L} \sum_{k'=1}^{K} P_{l',k'} \mathbf{w}_{l',k'}^H \mathbf{A}_{l'\to l,k} \mathbf{w}_{l',k'} + P_n^{l,k} \right)^{-1}.
\end{aligned} \tag{11}$$

The SINR at the kth destination in (4) can be rewritten as (11), where $\mathbf{A}_{l\to l,k} = (\sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} \mathbf{g}_{l,m})^H (\sum_{m=1}^{M} h_{l\to l,m,k} v_{l,m} \mathbf{g}_{l,m})$.

Thus, Problem (10) can be formulated as

$$\begin{aligned}
&\min_{\mathbf{w}_{l,k}} \quad \sum_{l=1}^{L} \sum_{k=1}^{K} P_{l,k} \mathbf{w}_{l,k}^H \mathbf{w}_{l,k} \\
&\text{subject to} \quad \left(\frac{P_{l,k}}{\Gamma_{l,k}} + P_{l,k} \right) \mathbf{w}_{l,k}^H \mathbf{A}_{l\to l,k} \mathbf{w}_{l,k} \\
&\qquad - \sum_{l'=1}^{L} \sum_{k'=1}^{K} P_{l',k'} \mathbf{w}_{l',k'}^H \mathbf{A}_{l'\to l,k} \mathbf{w}_{l',k'} \geq P_n^{l,k}, \\
&\qquad l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K.
\end{aligned} \tag{12}$$

Using $\mathbf{W}_{l,k} \triangleq \mathbf{w}_{l,k} \mathbf{w}_{l,k}^H$, Problem (12) can be represented into

$$\begin{aligned}
&\min_{\mathbf{W}_{l,k}} \quad \sum_{l=1}^{L} \sum_{k=1}^{K} P_{l,k} \operatorname{Tr}(\mathbf{W}_{l,k}) \\
&\text{subject to} \quad \mathbf{W}_{l,k} \succeq \mathbf{0}, \quad \operatorname{rank}(\mathbf{W}_{l,k}) = 1, \\
&\qquad \left(\frac{P_{l,k}}{\Gamma_{l,k}} + P_{l,k} \right) \operatorname{Tr}(\mathbf{A}_{l\to l,k} \mathbf{W}_{l,k}) \\
&\qquad - \sum_{l'=1}^{L} \sum_{k'=1}^{K} P_{l',k'} \operatorname{Tr}(\mathbf{A}_{l'\to l,k} \mathbf{W}_{l',k'}) \geq P_n^{l,k}, \\
&\qquad l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K,
\end{aligned} \tag{13}$$

where $\mathbf{W}_{l,k} \succeq \mathbf{0}$ means that $\mathbf{W}_{l,k}$ is a Hermitian semidefinite matrix. Note that the rank constraint $\operatorname{rank}(\mathbf{W}_{l,k}) = 1$ comes from $\mathbf{W}_{l,k} \triangleq \mathbf{w}_{l,k} \mathbf{w}_{l,k}^H$,which is not convex. Thus, Problem (13) is not convex. We can use the SDR technique to relax Problem (13) [26]. Here, we drop the rank-one constraint and obtain the following SDP problem:

$$\begin{aligned}
&\min_{\mathbf{W}_{l,k}} \quad \sum_{l=1}^{L} \sum_{k=1}^{K} P_{l,k} \operatorname{Tr}(\mathbf{W}_{l,k}) \\
&\text{subject to} \quad \mathbf{W}_{l,k} \succeq \mathbf{0}, \\
&\qquad \left(\frac{P_{l,k}}{\Gamma_{l,k}} + P_{l,k} \right) \operatorname{Tr}(\mathbf{A}_{l\to l,k} \mathbf{W}_{l,k}) \\
&\qquad - \sum_{l'=1}^{L} \sum_{k'=1}^{K} P_{l',k'} \operatorname{Tr}(\mathbf{A}_{l'\to l,k} \mathbf{W}_{l',k'}) \geq P_n^{l,k}, \\
&\qquad l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K.
\end{aligned} \tag{14}$$

The SDP problem in (14) can be solved by using interior-point methods. According to [28], the computational complexity of Problem (14) is $O((LK)^3 D_l^{3.5} \log(1/\sigma))$, where σ is the solution accuracy of interior-point methods.

The solution of Problem (14) provides a lower bound on the objective function in the original problem (13) due to excluding the rank-one constraint. However, in our extensive simulations, we have never observed that the optimal solution of Problem (14) has a rank higher than one. Thus, the original problem in (13) can be optimally solved. The similar observation was also reported in [29] to design the optimal beamforming schemes. For the cases where the solution of Problem (14) has a rank higher than one, several randomization techniques [30] can be used to provide a good approximate solution $\widetilde{\mathbf{w}}_{l,k}$ for Problem (13) from the optimal solution $\mathbf{W}_{l,k}$ of Problem (14).

4. Minimum Relay Power with SINR Constraints

In this section, we optimize the relay beamforming weights to minimize the relay power with guaranteeing the receive SINR in the multicell system. Due to the power limitation of the M2M relay devices, the relay power is important to be minimized. The total relay power in the lth cell is given by

$$\begin{aligned} P_t^l &= \sum_{m=1}^{M} \mathbb{E}\left[\left|x_{l.m}\right|^2\right] \\ &= \sum_{m=1}^{M} \left|v_{l,m}\right|^2 \mathbb{E}\left[\left|r_{l,m}\right|^2\right] \\ &= \sum_{m=1}^{M} \left|v_{l,m}\right|^2 \left(\sum_{k=1}^{K} P_{l,k}\left|\mathbf{g}_{l,m}\mathbf{w}_{l,k}\right|^2 + N_{l,m}\right) \\ &= \mathbf{v}_l^H \mathbf{E}_l \mathbf{v}_l, \end{aligned} \tag{15}$$

where

$$\begin{aligned} \mathbf{v}_l &= \left[v_{l,1}, v_{l,2}, \ldots, v_{l,m}\right]^T, \\ \mathbf{E}_l &= \operatorname{diag}\left(\sum_{k=1}^{K} P_{l,k}\left|\mathbf{g}_{l,1}\mathbf{w}_{l,k}\right|^2 + N_{l,1}, \ldots, \right. \\ &\qquad \left. \sum_{k=1}^{K} P_{l,k}\left|\mathbf{g}_{l,M}\mathbf{w}_{l,k}\right|^2 + N_{l,M}\right). \end{aligned} \tag{16}$$

Thus, we can formulate the optimization problem as

$$\begin{aligned} \min_{\mathbf{v}_l} \quad & \sum_{l=1}^{L} P_t^l \\ \text{subject to} \quad & \mathrm{SINR}_{l,k} \geq \Gamma_{l,k}, \\ & l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K, \end{aligned} \tag{17}$$

$$\begin{aligned} \mathrm{SINR}_{l,k} &= P_{l,k}\left|\mathbf{b}_{l\to l,k}^T \mathbf{v}_l\right|^2 \\ &\quad \times \left(\sum_{k' \neq k}^{K} P_{l,k'}\left|\mathbf{b}_{l\to l,k'}^T \mathbf{v}_l\right|^2 \right. \\ &\quad + \sum_{l' \neq l}^{L} \sum_{k'=1}^{K} P_{l',k'}\left|\mathbf{b}_{l'\to l,k'}^T \mathbf{v}_{l'}\right|^2 \\ &\quad \left. + \sum_{l'=1}^{L} v_{l'}^H \mathbf{H}_{l'\to l,k} \mathbf{v}_{l'} + N_{l,k}\right)^{-1}. \end{aligned} \tag{18}$$

The receive SINR at the kth destination in (4) can be rewritten as (18), where

$$\begin{aligned} \mathbf{b}_{l\to l,k} &= \left[h_{l\to l,1,k}\mathbf{g}_{l,1}\mathbf{w}_{l,k}, \ldots, h_{l\to l,m,k}\mathbf{g}_{l,M}\mathbf{w}_{l,k}\right]^T, \\ \mathbf{b}_{l\to l,k'} &= \left[h_{l\to l,1,k}\mathbf{g}_{l,1}\mathbf{w}_{l,k'}, \ldots, h_{l\to l,m,k}\mathbf{g}_{l,M}\mathbf{w}_{l,k'}\right]^T, \\ \mathbf{H}_{l\to l,k} &= \operatorname{diag}\left(\left|h_{l\to l,1,k}\right|^2 N_{l,1}, \ldots, \left|h_{l\to l,m,k}\right|^2 N_{l,M}\right). \end{aligned} \tag{19}$$

Therefore, problem (17) can be formulated as

$$\begin{aligned} \min_{\mathbf{v}_l} \quad & \sum_{l=1}^{L} \mathbf{v}_l^H \mathbf{E}_l \mathbf{v}_l \\ \text{subject to} \quad & \left(\frac{P_{l,k}}{\Gamma_{l,k}} + P_{l,k}\right)\left|\mathbf{b}_{l\to l,k}^T \mathbf{v}_l\right|^2 \\ & - \sum_{l'=1}^{L} \mathbf{v}_{l'}^H \mathbf{H}_{l'\to l,k} \mathbf{v}_{l'} \\ & - \sum_{l'=1}^{L} \sum_{k'=1}^{K} P_{l',k'}\left|\mathbf{b}_{l'\to l,k'}^T \mathbf{v}_{l'}\right|^2 \geq N_{l,k}, \\ & l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K. \end{aligned} \tag{20}$$

Using $\mathbf{V}_l \triangleq \mathbf{v}_l \mathbf{v}_l^H$ and $\mathbf{B}_{l\to l,k} = P_{l,k}\mathbf{b}_{l\to l,k}\mathbf{b}_{l\to l,k}^H$, problem (20) can be changed into the matrix form

$$\begin{aligned} \min_{\mathbf{V}_l} \quad & \sum_{l=1}^{L} \operatorname{Tr}\left(\mathbf{E}_l \mathbf{V}_l\right) \\ \text{subject to} \quad & \mathbf{V}_l \succeq \mathbf{0}, \quad \operatorname{rank}\left(\mathbf{V}_l\right) = 1, \\ & \left(\frac{1}{\Gamma_{l,k}} + 1\right) \operatorname{Tr}\left(\mathbf{B}_{l\to l,k}\mathbf{V}_l\right) \\ & - \sum_{l'=1}^{L} \operatorname{Tr}\left(\mathbf{H}_{l'\to l,k}\mathbf{V}_{l'}\right) \\ & - \sum_{l'=1}^{L} \sum_{k'=1}^{K} \operatorname{Tr}\left(\mathbf{B}_{l'\to l,k'}\mathbf{V}_{l'}\right) \geq N_{l,k}, \\ & l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K, \end{aligned} \tag{21}$$

where $\mathbf{V}_l \succeq 0$ means that $\mathbf{V}_l$ is a Hermitian semidefinite matrix. Note that the rank constraint $\operatorname{rank}(\mathbf{V}_l) = 1$ comes from $\mathbf{V}_l \triangleq \mathbf{v}_l \mathbf{v}_l^H$, which is not convex. Thus, Problem (21) is not convex. We use the SDR technique to relax the nonconvex problem (21) into the following SDP problem [26]:

$$\begin{aligned} \min_{\mathbf{V}_l} \quad & \sum_{l=1}^{L} \operatorname{Tr}\left(\mathbf{E}_l \mathbf{V}_l\right) \\ \text{subject to} \quad & \mathbf{V}_l \succeq \mathbf{0}, \\ & \left(\frac{1}{\Gamma_{l,k}} + 1\right) \operatorname{Tr}\left(\mathbf{B}_{l\to l,k}\mathbf{V}_l\right) \\ & - \sum_{l'=1}^{L} \operatorname{Tr}\left(\mathbf{H}_{l'\to l,k}\mathbf{V}_{l'}\right) \\ & - \sum_{l'=1}^{L} \sum_{k'=1}^{K} \operatorname{Tr}\left(\mathbf{B}_{l'\to l,k'}\mathbf{V}_{l'}\right) \geq N_{l,k}, \\ & l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K. \end{aligned} \tag{22}$$

(1) Initialize the relay beamforming weights $\mathbf{v}_l = \mathbf{v}_0$, $l = 1, 2, \ldots, L$, where $\mathbf{v}_0$ is an M-dimension random vector.
(2) Using current relay beamforming weights, optimally solve Problem (14) to get the update BS beamforming weights.
(3) If Problem (14) is feasible, then obtain the optimal solution $\mathbf{W}_{l,k}^{\text{opt}}$.
(4) **If** $\text{rank}(\mathbf{W}_{l,k}^{\text{opt}}) = 1$ **then**
(5) Obtain the BS beamforming vector $\mathbf{w}_{l,k}$, which is the optimal solution of the original problem (10).
(6) **else**
(7) Obtain the BS beamforming vector $\mathbf{w}_{l,k}$ using the randomization techniques [30].
(8) **end if**
(9) Optimally solve Problem (22) with current BS beamforming weights to get the update relay beamforming weights.
(10) If Problem (22) is feasible, then obtain the optimal solution $\mathbf{V}_l^{\text{opt}}$.
(11) **if** $\text{rank}(\mathbf{V}_l^{\text{opt}}) = 1$ **then**
(12) Obtain the relay beamforming vector $\mathbf{v}_l$, which is the optimal solution of the original problem (17).
(13) **else**
(14) Obtain the relay beamforming vector $\mathbf{v}_l$ using the randomization techniques [30].
(15) **end if**
(16) If the total relay power P_t^l is sufficiently close to a fixed point, or the iteration number exceeds a predetermined number, then stop the iteration and go to *Step 17*. Otherwise go back to *Step 2*.
(17) **return** the BS and relay beamforming vectors, that is, $\mathbf{w}_{l,k}$ and $\mathbf{v}_l$, respectively.

Algorithm 1: Strategy 1. Iterative algorithm for the multicell systems.

Using the interior-point methods, we can solve the SDP problem (22). As shown in Section 3, the computational complexity of Problem (22) is $O((LK)^3 M^{3.5} \log(1/\sigma))$, where σ is the solution accuracy of interior-point methods.

In our extensive simulations, we also have never observed that the optimal solution of Problem (22) has a rank higher than one. This means that the optimal solution of Problem (22) is also optimal for the original problem in (17). If the solution of Problem (22) has a rank higher than one, we can also use the randomization techniques [30] to provide a good approximate solution $\tilde{\mathbf{v}}_l$ for Problem (21) from the optimal solution $\mathbf{V}_l$ of Problem (22).

5. Iterative Algorithm

Based on the previous results, we propose an iterative scheme (see Algorithm 1) to jointly optimize the BS and relay beamforming weights to minimize the BS and relay power with guaranteeing the receive SINR in the multicell system. This proposed algorithm firstly optimizes the BS beamforming vector according to the current relay beamforming vector. Then the relay beamforming vector is optimized on the basis of the current BS beamforming vector. Circulate the above two steps, until the total relay power is sufficiently close to a fixed point or the iteration number exceeds a predetermined number. The details of Algorithm 1 is in the following.

According to [28], the computational complexity of Algorithm 1 is $O((LK)^3(M^{3.5} + D_1{}^{3.5}) \log(1/\sigma))$. Since the BS and relay power are minimized at Step 2 and Step 9, after a few iterations from Step 2 to Step 16, our algorithm can obtain near-optimal performance, which can be confirmed by simulation results in Section 6.

6. Simulation Results

In this section, we present simulation results to validate the performance of our proposed iterative scheme for multiple destination M2M devices in a multicell system. It is assumed that the flat-fading channel coefficients $\mathbf{g}_{l,m}$, $h_{l' \to l,m,k}$, $l = 1, 2, \ldots, L$, $m = 1, 2, \ldots, M$, and $k = 1, 2, \ldots, K$, are independent Rayleigh random variables with mean zero and unit variance. For convenience, it is assumed that the noise at all the devices, including relays and destinations, has the same power spectral density; that is, $N_{l,m} = N_{l,k} = N_0$, $l = 1, 2, \ldots, L$, $m = 1, 2, \ldots, M$, and $k = 1, 2, \ldots, K$. The BS transmission SNR is set to $P_{l,k}/N_0 = 10$ (dB), which is the same for all the BSs to each destination in the multicell system. In the simulations, we assume that all destination devices have the same SINR constraint; that is, $\gamma_{l,k} = \gamma_0$, $l = 1, 2, \ldots, L$, and $k = 1, 2, \ldots, K$. We use CVX [31] to numerically solve the SDP problems.

Figure 3: Expected total relay power of our proposed scheme for multiple destinations in the multicell system.

Figure 4: Comparison with one destination in single cell system.

Figure 5: Expected total relay power of our proposed scheme for various scenarios in multicell systems.

In Figure 3, we show the total relay power $\sum_{l=1}^{L} P_t^l / N_0$ (dB) of our proposed scheme in the multicell system for different iteration numbers. Suppose that the system has $L = 3$ BSs with $D_l = 8$ antennas, $M = 12$ relays, and $K = 5$ destination devices in each cell which suffered the intercell and intracell interference (the system model is in Figure 2). In the simulations, we can see that the performance of our proposed scheme with 4 iterations is near the performance with the unlimited iterations. Thus, 4 iterations could be enough for the proposed iteration strategies in multicell system scenarios. In practice, we can stop the iterations when the total relay power is sufficiently close to a fixed point. The simulation results demonstrate that the performance of our proposed iterative scheme increases as the number of iterations increases and with only 4 iterations the proposed scheme can obtain near-optimal performance.

In Figure 4, we compare our proposed algorithm with the relay beamforming algorithm in [32]. Note that Ubaidulla and Chockalingam [32] only consider the single-user case in a single cell system (without considering the intercell interference). The scheme in [32] can be seen as a special case of our proposed scheme. For convenience, we assume that there are $L = 3$ BSs with $D_l = 8$ antennas, $M = 20$ relays, and $K = 1$ destination device in each cell of Figure 4. We choose the same channel conditions as in [32] to run the simulations for our proposed iterative scheme with 4 iterations. Note that Ubaidulla et al.[32] denoted δ_α and δ_β as the error norm bounds to determine the level of additive CSI error in the channel from the BS to the relay devices and from the relay nodes to the destination device, respectively. $\delta_\alpha = \delta_\beta = 0$ represents that each device, including relay and destination, knows the perfect CSI in the system. The results in Figure 4 indicate that the total relay power consumption of both the algorithms increases as the receive SINR constraint increases. However, the total relay power consumption of our proposed scheme is more than that of the proposed algorithm in [32]. The reason is that in our proposed algorithm the relay power consumption needs to overcome the inter-cell interference in order to improve the performance of the destination device in the multicell system.

Figure 5 shows the total relay power of our proposed iterative algorithm to jointly optimize the BS and relay beamforming weights for multiple destinations in various scenarios of multicell systems. We simulate three scenarios of multicell systems, where $(L, D_l, M, K) = (2, 4, 6, 2), (3, 8, 8, 3), (3, 10, 15, 3)$. We run simulations for the proposed iterative scheme with 4 iterations. Note that when the number of destinations and cells increases, the intercell and intracell interference will become severer. We need to increase the transmission power of BSs and relays to reach the receive SINR constraints. The simulation results in Figure 5 show that we can increase the number of antennas or

relays to cope with the severer interference under the same power constraint. As such, the performance in the case of $(L, D_l, M, K) = (3, 10, 15, 3)$ is close to that in the case of $(L, D_l, M, K) = (2, 4, 6, 2)$, as the number of antennas and relays is increased.

7. Discussion

In the above results, the perfect CSI at relay and destination devices is essential for our proposed iterative algorithm. Thus, the perfect CSI from the BSs to relays and from relays to the destinations needs to feedback to the central processor calculating the BS and relay beamforming weights. This becomes a drawback of the proposed relay-aided transmission schemes in practice wireless M2M networks, which will decrease frequency efficiency and could be costly if the number of relays and users is large. If the channel estimate error exists or the capacity of feedback channel is limited, the central processor only knows the imperfect CSI. In order to overcome this problem in practice, Huang et al. [33] considered the orthogonal beamforming systems with limited feedback under the per user unitary and rate control (PU2RC) technology, where the beamforming weights are selected from a codebook of multiple orthonormal bases in the multiuser system. The concept of the PU2RC technology has been included in the LTE standard. We can use the PU2RC technology to our collaborative relay beamforming algorithm for wireless M2M devices in multicell systems to deal with the problem of the imperfect CSI, where the BS and relay beamforming weights can be selected from a codebook in the central processor.

As the relay devices are distributed, we can use the distributed relay beamforming strategies as in [18, 19] to reduce the CSI feedback to deal with the issue of imperfect CSI in practice. It is assumed that each relay device can learn the local CSI from the source by training and from destinations by feedback, respectively, and measure its noise level. Note that Zheng et al. [18, 19] proposed a distributed implementation algorithm with local CSI under SNR constraints based on the Karush-Kuhn-Tucker (KKT) analysis. Another distributed algorithm was proposed in [21, 34], which used MMSE criterion to construct distributed relay beamforming strategies. However, they focused on the distributed beamforming schemes for the single cell system. Choi [21] investigated the distributed beamforming strategies for the single destination in the multiple relay system and Verdu [34] developed the distributed beamforming strategies into the scenario of multiple destinations. The distributed algorithm for multiple destinations in the multicell system will be investigated in our future work.

8. Conclusion

In this paper, we proposed an iterative scheme to jointly optimize the BS and relay beamforming weights under the receive SINR constraints to cope with the strong inter-cell and intracell interference in the multicell systems. We minimized the BS and relay beamforming power, while guaranteeing the receive SINR for the multiple destination devices by using the SDR technology. We showed that the optimization problems for BS and relay beamforming weights can be converted into SDP problems, which can be effectively solved by interior-point methods. Simulation results demonstrated that our proposed iterative scheme can achieve near-optimal performance with only 4 iterations.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments that have helped improving the overall quality of the paper. This work has been supported by the Specialised Research Fund for the Doctoral Program of the Ministry of Education of China (Grant no. 20120001120125), and the National Natural Science Foundation of China (Grant nos. 61250001, 61231011, and 61231013).

References

[1] "Machine-to-machine communications (M2M); smart metering use cases," ETSI TR 102 691 V1.1.1, May 2010.

[2] "System improvements for machine-type communications (MTC)," 3GPP TR 23.888, V1.0.0, September 2010.

[3] "Service requirements for machine-type communications (MTC)," 3GPP TS 23.368, V10.2.0, July 2010.

[4] J. Zhang, L. Shan, H. Hu, and Y. Yang, "Mobile cellular networks and wireless sensor networks: toward convergence," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 164–169, 2012.

[5] B. H. Lee and S. L. Kim, "Mobility control for machine-to-machine LTE systems," in *Proceedings of the 11th European Wireless Conference*, pp. 1–5, Vienna, Austria, April 2011.

[6] C. Chen, L. Liu, X. Du, Q. Pei, and X. Zhao, "Improving driving safety based on safe distance design in vehicular sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 469067, 13 pages, 2012.

[7] K. Yu and B. Ottersten, "Models for MIMO propagation channels: a review," *Wireless Communications and Mobile Computing*, vol. 2, no. 7, pp. 653–666, 2002.

[8] N. Khajehnouri and A. H. Sayed, "Distributed MMSE relay strategies for wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 7 I, pp. 3336–3348, 2007.

[9] Y. W. Liang and R. Schober, "Amplify-and-forward multi-antenna beamforming with joint source-relay power constraint," in *Proceedings of the IEEE 72nd Vehicular Technology Conference Fall (VTC '10-Fall)*, pp. 1–5, Ottawa, Canada, September 2010.

[10] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[11] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[12] A. K. Sadek, W. Su, and K. J. R. Liu, "Multinode cooperative communications in wireless networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 341–355, 2007.

[13] M. Alatossava, A. Taparugssanagorn, V. M. Holappa, and J. Ylitalo, "Measurement based capacity of distributed MIMO

antenna system in Urban microcellular environment at 5.25 GHz," in *Proceedings of the IEEE 67th Vehicular Technology Conference-Spring (VTC '08)*, pp. 430–434, Singapore, May 2008.

[14] G. A. Elkheir, A. S. Lioumpas, and A. Alexiou, "Energy efficient AF relaying under error performance constraints with application to M2M networks," in *Proceedings of the IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '11)*, pp. 56–60, Toronto, Canada, September 2011.

[15] D. Wang, X. N. Zhang, L. Bai et al., "Collaborative relay beamforming strategies for multiple destinations with guaranteed QoS in wireless machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 525640, 10 pages, 2012.

[16] Y. Li, C. Chen, L. Bai, and Y. Jin, "On the achievable user number of the downlinks in cellular-based machine-to-machine communications," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 714971, 11 pages, 2012.

[17] C. Chen, L. Bai, M. Feng, M. Huang, and T. Tian, "Distributed beamforming for relay assisted multiuser machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 213727, 9 pages, 2012.

[18] G. Zheng, K. K. Wong, A. Paulraj, and B. Ottersten, "Collaborative-relay beamforming with perfect CSI: optimum and distributed implementation," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 257–260, 2009.

[19] G. Zheng, K. K. Wong, A. Paulraj, and B. Ottersten, "Robust collaborative-relay beamforming," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3130–3143, 2009.

[20] J. Choi, "Distributed beamforming using a consensus algorithm for cooperative relay networks," *IEEE Communications Letters*, vol. 15, no. 4, pp. 368–370, 2011.

[21] J. Choi, "MMSE-based distributed beamforming in cooperative relay networks," *IEEE Transactions on Communications*, vol. 59, no. 5, pp. 1346–1356, 2011.

[22] D. H. N. Nguyen, H. H. Nguyen, and H. D. Tuan, "Distributed beamforming in relay-assisted multiuser communications," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, pp. 1–5, Dresden, Germany, June 2009.

[23] D. H. N. Nguyen, H. H. Nguyen, and T. T. Pham, "Distributed beamforming in multiuser multi-relay networks with guaranteed QoS," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pp. 1–6, Honolulu, Hawaii, USA, December 2009.

[24] J. Park and S. Lee, "MIMO beamforming for QoS enhancement via analog, digital and hybrid relaying," *IEEE Transactions on Broadcasting*, vol. 56, no. 4, pp. 494–503, 2010.

[25] D. Wang, L. Bai, C. Chen, Y. Jin, and J. Choi, "Coordinated relay beamforming based on the worst-case SINR in multicell wireless systems," in *Proceedings of the Global Telecommunications Conference (GLOBECOM '12)*, pp. 3941–3945, Anaheim, Calif, USA, December 2012.

[26] Z. Q. Luo, W. K. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[27] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478–1491, 2007.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[29] D. H. N. Nguyen, H. H. Nguyen, and H. D. Tuan, "Distributed beamforming in relay-assisted multiuser communications," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, pp. 1–5, Dresden, Germany, June 2009.

[30] N. D. Sidiropoulos, T. N. Davidson, and Z. Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Transactions on Signal Processing*, vol. 54, no. 6 I, pp. 2239–2251, 2006.

[31] M. Grant and S. Boyd, "CVX users' guide for CVX version 1.21 (build 808)," 2011, http://cvxr.com/.

[32] P. Ubaidulla and A. Chockalingam, "Robust distributed beamforming for wireless relay networks," in *Proceedings of the IEEE 20th Personal, Indoor and Mobile Radio Communications Symposium (PIMRC '09)*, pp. 2345–2349, Tokyo, Japan, September 2009.

[33] K. Huang, J. G. Andrews, and R. W. Heath Jr., "Performance of orthogonal beamforming for SDMA with limited feedback," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 152–164, 2009.

[34] S. Verdu, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 676024, 11 pages
http://dx.doi.org/10.1155/2013/676024

Research Article

Channel Estimation for Relay-Based M2M Two-Way Communications Using Expectation-Maximization

Xiaoyan Xu, Jianjun Wu, Chen Chen, Wenyang Guan, and Haige Xiang

State Key Laboratory of Advanced Optical Communication Systems and Networks, Peking University, Beijing 100871, China

Correspondence should be addressed to Jianjun Wu; just@pku.edu.cn and Chen Chen; chen.chen@pku.edu.cn

Received 18 August 2013; Accepted 6 November 2013

Academic Editor: Jianhua He

Copyright © 2013 Xiaoyan Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The growing popularity of machine-to-machine (M2M) communications in wireless networks is driving the need to update the corresponding receiver technology based on the characteristics of M2M. In this paper, an expectation-maximization-based maximum likelihood cascaded channel estimation method is developed for relay-based M2M two-way communications. As the closed-form solution of maximum likelihood channel estimation does not exist, and the superimposed signal structure at the receiver is conducive to the expectation-maximization application, the expectation-maximization algorithm is utilized to provide the maximum likelihood solution in the presence of unobserved data through stable iterations. Even in the absence of the training sequence, the cascaded channel estimates are obtained through the expectation-maximization iterations. The Bayesian Cramér-Rao lower bounds are derived under random parameters for the channel estimation, and the simulation demonstrates the validity of the proposed studies.

1. Introduction

Machine-to-machine (M2M) communication is expected to be one of the major derivers of cellular networks [1, 2] and has become one of the focuses in 3GPP [3]. M2M technology allows machine devices to communicate directly with each other without human intervention, which has been attracting more and more interests for their wide applications in many popular wireless communication systems, such as mobile ad hoc networks [1–3]. Wireless M2M networks supporting M2M-enabled machine devices are pivotal to the success of M2M. Sensor nodes in wireless M2M networks could be connected by a wide range of wireless network technologies, for example, cellular networks.

Cooperation is crucial in wireless networks as it can greatly attribute to ensuring connectivity, reliability, and performance. Relaying is promising in a wide variety of network types for the cooperation purpose. According to the recent IEEE 802.16p proposals, a wireless M2M device may act as an aggregation point and communicate data packets on behalf of the other M2M devices, which may have a poor communication link to the network. Relay-based schemes have been proposed to improve the link reliability for devices with weak links [4–7].

As a basic model of the relay-based M2M two-way communications, the three-node cooperation model, which is also referred to as the two-way relay channel, where two source nodes exchange information via a relay node using amplify-and-forward (AF) or decode-and-forward (DF) strategy, has attracted a great deal of research interest because of its improved spectral efficiency [8–12]. AF relaying is more popular in practice because the relays only amplify and process the signal linearly before they retransmit it again and thus AF leads to low-complexity relay transceivers. A lot of research therein assumes perfect channel knowledge at the nodes, which is unpractical in wireless communications, so it is important to develop efficient channel estimation algorithms. Channel estimation for this communication scene has been studied in [13–18]. Specifically, in [13] the maximum-likelihood-based estimator, flat-fading channels were reported, and in [14, 15], the cascaded source-relay-source channels were estimated using block-based training under the assumption of time-invariant frequency-selective fading channels although the individual channels were also

estimated using pilot-tone-based training in [14]. Different from [14, 15], where the relay only amplifies and forwards the received signal, the work in [16] allowed the relay to estimate the channel parameters and allocated the powers for these parameters. The channel estimation problem was extended to the multiple antennas scene at all the three nodes in [17], and an optimal method was proposed to design the training signals based on the mean-square-error (MSE) criterion. A blind channel estimation algorithm by applying a nonredundant linear precoding at both source nodes was proposed for the frequency-selective channels in [18].

The expectation-maximization (EM) algorithm is a general method for solving maximum likelihood estimation problems given incomplete data, and it consists of two iterative steps: the expectation step and the maximization step. These two steps are iterated until the estimated values converge [19, 20]. The EM-based maximum likelihood estimation for one-way relay channel has been investigated in [21–24], where different variables are treated as the missing data and iterative process is provided for the channel estimation. To the best of our knowledge, the EM estimation algorithm in two-way relay channels has not been investigated yet.

Since the received signal at the source node in the relay-based M2M two-way communications is a superimposed signal, which can be decomposed into two signal components with two unknown channel parameters, one is the self-interference signal, and the other is the desired signal from the other source node. Take the channel estimation at source node 1, for example, the two unknown parameters are source1-relay-source1 cascaded channel and source2-relay-source1 cascaded channel. The source1-relay-source1 channel is the unobserved data when estimating source2-relay-source1 channel, while the source2-relay-source1 channel is also the unobserved data when estimating source1-relay-source1 channel. Since the superimposed signal structure at the receiver is conducive to the application of the EM algorithm and the closed-form solution of maximum likelihood channel estimation does not exist for this relay-based M2M two-way communication, the EM algorithm is utilized to provide the maximum likelihood solution in the presence of unobserved data through stable iterations.

In this paper, we propose an EM-based channel estimation algorithm for relay-based M2M two-way communication to jointly estimate the two cascaded channels. The idea is to decompose the observed data into its signal components and then estimate the parameters of each signal component separately. In the EM algorithm, the received signal at the source node is treated as the incomplete data, and the set of the two signal components from the two source nodes is modeled as the complete data. Conditioned upon the incomplete observations, the channel estimation algorithm maximizes the expectation of log likelihood function defined over the complete data, by averaging over the unknown underlying parameters and using the current estimates of the cascaded channels without channel statistical information. The algorithm iterates back and forth, using the current channel parameter estimates to decompose the received signal better and thus increase the likelihood of the next channel parameter estimates. Even in the absence of training sequence, the cascaded channel estimates can still be obtained through the iterations between the E-step and M-step. The Bayesian Cramér-Rao lower bounds are derived under random parameters for cascaded channel estimation, and the validity of the proposed studies is verified by Monte Carlo simulations.

The rest of the paper is organized as follows. Section 2 describes the three-node model in relay-based M2M networks. Section 3 presents the EM-based cascaded channel estimation algorithms. Section 4 provides the Bayesian Cramér-Rao lower bounds of the cascaded channel estimation. Simulation results and conclusion are given in Sections 5 and 6, respectively.

2. System Model

Consider a three-node cooperation model in wireless M2M networks where two source nodes, T_1 and T_2, exchange information through one relay node R, as shown in Figure 1, which is also the model of two-way relay network. Each node has a single omnidirectional antenna and operates in half-duplex mode. The direct link between T_1 and T_2 is very weak or even nonexistent, and thus the source-destination communication is performed only via the relay. The transmission is divided into two phases. During Phase I, both T_1 and T_2 send a signal frame to R via an uplink manner, whereas during Phase II, R processes the received signals and broadcasts them to T_1 and T_2.

The baseband channel between T_1 and R is denoted by $\mathbf{h} = [h_0, h_1, \ldots, h_{L_h-1}]^T$, and the one between T_2 and R is denoted by $\mathbf{g} = [g_0, g_1, \ldots, g_{L_g-1}]^T$, where L_h and L_g are the numbers of the taps of the corresponding channel delay spread. $h_l \in \mathcal{CN}(0, \sigma_{h,l}^2)$ and $g_l \in \mathcal{CN}(0, \sigma_{g,l}^2)$ are independent from each other. The frequency-domain representations are $H_k = \sum_{l=0}^{L_h-1} h(l) e^{-j(2\pi/N)kl}$, $k = 0, \ldots, N-1$ and $G_k = \sum_{l=0}^{L_g-1} g(l) e^{-j(2\pi/N)kl}$, $k = 0, \ldots, N-1$, respectively. Both $\mathbf{h}$ and $\mathbf{g}$ are assumed to remain unchanged at least for one round of data exchange. As for time-division duplexing (TDD), $\mathbf{h}$ and $\mathbf{g}$ can be considered as reciprocal. The average transmission powers of T_1, T_2, and R are denoted as P_s, P_s, and P_r, respectively.

Suppose that each OFDM symbol contains N information symbols. The time-domain OFDM signals transmitted from T_1 and T_2 after inverse discrete Fourier transform (IDFT) are, respectively, given by

$$\begin{aligned} s_{1n} &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} d_{1k} e^{j2\pi kn/N}, \quad n = 0, \ldots, N-1, \\ s_{2n} &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} d_{2k} e^{j2\pi kn/N}, \quad n = 0, \ldots, N-1, \end{aligned} \tag{1}$$

where $\{d_{1k}, k = 0, \ldots, N-1\}$ and $\{d_{2k}, k = 0, \ldots, N-1\}$ are zero-mean independent and identically distributed (i.i.d.) data with average power of P_s. Denoting that frequency-domain $\mathbf{d}_1 = [d_{10}, d_{11}, \ldots, d_{1(N-1)}]$, $\mathbf{d}_2 = [d_{20}, d_{21}, \ldots, d_{2(N-1)}]$ and time-domain signals

Figure 1: Three-node cooperation model.

$\mathbf{s}_1 = [s_{10}, s_{11}, \ldots, s_{1(N-1)}]$, $\mathbf{s}_2 = [s_{20}, s_{21}, \ldots, s_{2(N-1)}]$, the time-domain OFDM signals transmitted from T_1 and T_2 can be rewritten in vector-matrix form as

$$\mathbf{s}_1 = \mathbf{F}^H \mathbf{d}_1, \qquad \mathbf{s}_2 = \mathbf{F}^H \mathbf{d}_2, \tag{2}$$

where $\mathbf{F}$ is the normalized DFT matrix whose (p, q)th entry is $\mathbf{F}(i, j) = (1/\sqrt{N})e^{-j2\pi ij/N}$.

The time-domain symbols $\{s_{1n}, n = 0, \ldots, N-1\}$ and $\{s_{2n}, n = 0, \ldots, N-1\}$ go through regular OFDM transmission steps. Both T_1 and T_2 insert the cyclic prefix (CP) of length $L_{cp,T_1,T_2} \geq \max\{L_h - 1, L_g - 1\}$ in front of the OFDM block to avoid intersymbol interference. The time-domain received signal at the relay after removal of CP and DFT is expressed as

$$r_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} H_k d_{1k} e^{j2\pi kn/N} + \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} G_k d_{2k} e^{j2\pi kn/N} + v_n, \quad n = 0, \ldots, N-1, \tag{3}$$

where H_k is the complex gain of the kth subcarrier of channel $\mathbf{h}$, G_k is the complex gain of the kth subcarrier of channel $\mathbf{g}$, and v_n is zero-mean circular complex Gaussian with the variance σ_v^2 at the relay. The time-domain received signal at the relay can be rewritten in vector-matrix form as

$$\mathbf{r} = \mathbf{H}\mathbf{s}_1 + \mathbf{G}\mathbf{s}_2 + \mathbf{v}, \tag{4}$$

where $\mathbf{r} = [r_0, r_1, \ldots, r_{(N-1)}]$, $\mathbf{v} = [v_0, v_1, \ldots, v_{(N-1)}]$, $\mathbf{H}$ is the $N \times N$ circulant matrix whose first column is $[\mathbf{h}^T, \mathbf{0}_{1\times(N-L_h)}]^T$, and $\mathbf{G}$ is the $N \times N$ circulant matrix whose first column is $[\mathbf{g}^T, \mathbf{0}_{1\times(N-L_g)}]^T$. Define $\Psi_m(\mathbf{x})$ as the $N \times m$ column-wise circulant matrix with the first column $\mathbf{x}$; then we have $\mathbf{H}\mathbf{s}_1 = \Psi_{L_h}(\mathbf{s}_1)\mathbf{h}$ and $\mathbf{G}\mathbf{s}_2 = \Psi_{L_g}(\mathbf{s}_2)\mathbf{g}$.

After DFT, the received signal at the relay of the kth subcarrier is written as

$$R_k = H_k d_{1k} + G_k d_{2k} + V_k, \quad k = 0, \ldots, N-1, \tag{5}$$

where V_k is the corresponding frequency response of the relay noise over the kth subcarrier with the variance σ_V^2.

The relay amplifies the received signal r_n, adds a new CP, and forwards the signal $\mathbf{r}_t = \alpha\mathbf{r}$ to T_1 and T_2. The relay amplification factor α is chosen as $\sqrt{P_r/(\sigma_h^2 P_s + \sigma_g^2 P_s + \sigma_n^2)}$.

For symmetry, only the process at T_1 is discussed. The time-domain received signal at T_1, after CP removal, is

$$y_n = \frac{\alpha}{\sqrt{N}} \sum_{k=0}^{N-1} H_k^2 d_{1k} e^{j2\pi kn/N} + \frac{\alpha}{\sqrt{N}} \sum_{k=0}^{N-1} H_k G_k d_{2k} e^{j2\pi kn/N} + \frac{\alpha}{\sqrt{N}} \sum_{k=0}^{N-1} H_k V_k e^{j2\pi kn/N} + n_n, \quad n = 0, \ldots, N-1, \tag{6}$$

where $\mathbf{h}_2 = \mathbf{h} \otimes \mathbf{g}$ is zero-mean circular complex Gaussian with variance σ^2 at the receiver of T_1. Defining the time-domain cascaded channels $\mathbf{h}_1 = \mathbf{h} \otimes \mathbf{h}$ and $\mathbf{h}_2 = \mathbf{h} \otimes \mathbf{g}$, $\mathbf{y} = [y_0, y_1, \ldots, y_{(N-1)}]$, $\mathbf{n} = [n_0, n_1, \ldots, n_{(N-1)}]$, the time-domain received signal at T_1 can be rewritten in vector-matrix form as

$$\begin{aligned} \mathbf{y} &= \alpha\mathbf{H}\mathbf{H}\mathbf{s}_1 + \alpha\mathbf{H}\mathbf{G}\mathbf{s}_2 + \alpha\mathbf{H}\mathbf{v} + \mathbf{n} \\ &= \alpha\Psi_{2L_h-1}(\mathbf{s}_1)\mathbf{h}_1 + \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\mathbf{h}_2 + \mathbf{n}_a, \end{aligned} \tag{7}$$

where $\mathbf{n}_a = \alpha\mathbf{H}\mathbf{v} + \mathbf{n}$ is the total noise of the receiver at T_1.

After DFT, the received signal at T_1 of the kth subcarrier is written as

$$Y_k = \alpha H_k^2 d_{1k} + \alpha H_k G_k d_{2k} + \alpha H_k V_k + N_k, \quad k = 0, \ldots, N-1, \tag{8}$$

where N_k is the corresponding frequency response of the relay noise over the kth subcarrier. The frequency-domain received signal at T_1 can be rewritten in vector-matrix form as

$$\begin{aligned} \mathbf{Y} &= \alpha\,\mathrm{diag}\{\mathbf{d}_1\}\mathbf{H}_1 + \alpha\,\mathrm{diag}\{\mathbf{d}_2\}\mathbf{H}_2 + \mathbf{N}_a \\ &= \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\mathbf{h}_1 + \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\mathbf{h}_2 + \mathbf{N}_a, \end{aligned} \tag{9}$$

where $\mathbf{Y}$ is the corresponding frequency response of the time-domain received signal $\mathbf{y}$, $\mathbf{D}_1 = \mathrm{diag}\{\mathbf{d}_1\}$, $\mathbf{D}_2 = \mathrm{diag}\{\mathbf{d}_2\}$, $\mathbf{N}_a = \mathbf{F}\mathbf{n}_a$ is the corresponding frequency response of the noise $\mathbf{n}_a$, and $\mathbf{N}_a(k) = \alpha H_k V_k + N_k$. $\mathbf{H}_1 = \mathbf{F}_{2L_h-1}\mathbf{h}_1$ and $\mathbf{H}_2 = \mathbf{F}_{L_h+L_g-1}\mathbf{h}_2$ are the corresponding frequency responses of the time-domain cascaded channels $\mathbf{h}_1$ and $\mathbf{h}_2$, $\mathbf{H}_1(k) = H_k^2$, and $\mathbf{H}_2(k) = H_k G_k$. $\mathbf{F}_{2L_h-1}$ is the first $(2L_h - 1)$ column of the DFT matrix $\mathbf{F}$, and $\mathbf{F}_{L_h+L_g-1}$ is the first $(L_h + L_g - 1)$ column of $\mathbf{F}$. The task of cascaded channel estimation is to find $\mathbf{h}_1$ and $\mathbf{h}_2$ from $\mathbf{Y}$.

3. The EM Algorithm Based Channel Estimation

The EM algorithm is a technique for finding maximum likelihood estimates of system parameters in a broad range of problems where observed data are incomplete [19]. The expectation step is performed with respect to unknown underlying parameters, using the current estimate of the parameters, conditioned upon the incomplete observations. The maximization step then provides a new estimate of

the parameters that maximizes the expectation of log likelihood function defined over complete data, conditioned on the most recent observation and the last estimate. The EM algorithm consists of two iterative steps: the expectation step and the maximization step. These two steps are iterated until the estimated values converge [20]. When the underlying complete data come from an exponential family whose maximum-likelihood estimates are easily computed, then each maximization step of an EM algorithm is likewise easily computed.

The received signal at the source node in the relay-based M2M two-way communications is a superimposed signal, which can be decomposed into two signal components with two unknown channel parameters, one is the self-interference signal, and the other is the desired signal from the other source node. Take the channel estimation at source node 1, for example, the two unknown parameters are source1-relay-source1 cascaded channel and source2-relay-source1 cascaded channel. The source1-relay-source1 channel is the unobserved data when estimating source2-relay-source1 channel, while the source2-relay-source1 channel is also the unobserved data when estimating source1-relay-source1 channel. Since the superimposed signal structure at the receiver is conducive to the application of the EM algorithm and the closed-form solution of maximum likelihood channel estimation does not exist, the EM algorithm is utilized to provide the maximum likelihood solution in the presence of unobserved data through stable iterations.

In the EM channel estimation, the set of unknown parameters is $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$, the received signal $\mathbf{Y}/\mathbf{y}$ at T_1 is treated as the observed (incomplete) data, and the set of the two signal components from the two source nodes is modeled as the complete data. Conditioned upon the incomplete observations, the algorithm maximizes the expectation of log likelihood function defined over the complete data, by averaging over the unknown underlying parameters and using the current estimate of the parameters.

3.1. Cascaded Channel Estimation Based on Frequency Domain Processing. The frequency-domain received signal $\mathbf{Y}$ at T_1, also the incomplete data, can be written as

$$\mathbf{Y} = \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\mathbf{h}_1 + \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\mathbf{h}_2 + \mathbf{N}_a. \tag{10}$$

A natural choice for the complete data $\mathbf{Y}_c$ is obtained by decomposing the observed data $\mathbf{Y}$ into its signal components:

$$\mathbf{Y}_c = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\mathbf{h}_1 + \mathbf{N}_1 \\ \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\mathbf{h}_2 + \mathbf{N}_2 \end{bmatrix}, \tag{11}$$

where $\mathbf{Y}_1$ is the component of the received signal $\mathbf{Y}$ transmitted by the source node T_1 through the channel with impulse response $\mathbf{h}_1$, $\mathbf{Y}_2$ is the component of the received signal $\mathbf{Y}$ transmitted by the source node T_2 through the channel with impulse response $\mathbf{h}_2$. $\mathbf{N}_1$ and $\mathbf{N}_2$ are obtained by arbitrarily decomposing the total noise $\mathbf{N}_a$ into two components:

$$\mathbf{N}_1 + \mathbf{N}_2 = \mathbf{N}_a. \tag{12}$$

It is found to be most convenient to let $\mathbf{N}_1$ and $\mathbf{N}_2$ be statistically independent, zero-mean, and Gaussian with the covariance matrix $\mathbf{R}_1$ and $\mathbf{R}_2$, respectively. Denote the covariance matrix of $\mathbf{N}_a$ as $\mathbf{R}_a$; then $\mathbf{R}_1 + \mathbf{R}_2 = \mathbf{R}_a$.

The log-likelihood of the complete data $\mathbf{Y}_c$ is

$$\log p\left(\mathbf{Y}_c \mid \theta\right) = c - \left(\mathbf{Y}_c - \mathbf{D}_c\right)^H \Lambda^{-1}\left(\mathbf{Y}_c - \mathbf{D}_c\right), \tag{13}$$

where c contains all the terms that are independent of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ and the mean vector $\mathbf{D}_c$ and the covariance matrix Λ are given, respectively, by

$$\begin{gathered} \mathbf{D}_c = \begin{bmatrix} \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\mathbf{h}_1 \\ \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\mathbf{h}_2 \end{bmatrix}, \\ \Lambda = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 \end{bmatrix} = \begin{bmatrix} \beta_1\mathbf{R}_{\mathbf{n}\mid\theta} & \mathbf{0} \\ \mathbf{0} & \beta_2\mathbf{R}_{\mathbf{n}\mid\theta} \end{bmatrix}, \end{gathered} \tag{14}$$

where β_1 and β_2 are both real numbers, $0 < \beta_1 < 1$, $0 < \beta_2 < 1$, and $\beta_1 + \beta_2 = 1$.

Suppose that $\widehat{\theta}^{(i)} = \{\widehat{\mathbf{h}}_1^{(i)}, \widehat{\mathbf{h}}_2^{(i)}\}$ denotes the current value of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ after i iterations of the EM algorithm. The $(i+1)$th iteration can be described in two steps as follows.

(1) E-step: given the observations $\mathbf{Y}$ and the last estimates $\widehat{\theta}^{(i)} = \{\widehat{\mathbf{h}}_1^{(i)}, \widehat{\mathbf{h}}_2^{(i)}\}$, compute the expectation $Q(\theta \mid \widehat{\theta}^{(i)})$ of the log-likelihood of the complete data $\mathbf{Y}_c$, which is also the conditional expectation of (13),

$$\begin{aligned} Q\left(\theta \mid \widehat{\theta}^{(i)}\right) &= \mathrm{E}\left[\log p\left(\mathbf{Y}_c \mid \theta\right) \mid \mathbf{Y}, \widehat{\theta}^{(i)}\right] \\ &= d - \left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right)^H \Lambda^{-1}\left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right), \end{aligned} \tag{15}$$

where d contains all the terms that are independent of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$, and the ith estimate of the conditional expectation of the complete data $\mathbf{Y}_c$ is [25]

$$\begin{aligned} \widehat{\mathbf{Y}}_c^{(i)} &= \mathrm{E}\left\{\mathbf{Y}_c \mid \mathbf{Y}, \widehat{\theta}^{(i)}\right\} \\ &= \left\{\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \mid \widehat{\theta}^{(i)}\right\} + \begin{bmatrix} \beta_1\mathbf{R}_{\mathbf{n}\mid\theta} \\ \beta_2\mathbf{R}_{\mathbf{n}\mid\theta} \end{bmatrix} \mathbf{R}_{\mathbf{n}\mid\theta}^{-1} \\ &\quad \times \left[\mathbf{Y} - \mathbf{A}\mathrm{E}\left\{\mathbf{Y}_c \mid \widehat{\theta}^{(i)}\right\}\right]. \end{aligned} \tag{16}$$

$Q(\theta \mid \widehat{\theta}^{(i)})$ contains both the unknown variable $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ and the known constant $\widehat{\theta}^{(i)} = \{\widehat{\mathbf{h}}_1^{(i)}, \widehat{\mathbf{h}}_2^{(i)}\}$. The conditional expectations of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are given, respectively, by

$$\begin{aligned} \widehat{\mathbf{Y}}_1^{(i)} &= \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\widehat{\mathbf{h}}_1^{(i)} + \beta_1 \\ &\quad \times \left[\mathbf{Y} - \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\widehat{\mathbf{h}}_1^{(i)} - \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\widehat{\mathbf{h}}_2^{(i)}\right], \\ \widehat{\mathbf{Y}}_2^{(i)} &= \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\widehat{\mathbf{h}}_2^{(i)} + \beta_2 \\ &\quad \times \left[\mathbf{Y} - \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\widehat{\mathbf{h}}_1^{(i)} - \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\widehat{\mathbf{h}}_2^{(i)}\right]. \end{aligned} \tag{17}$$

(2) M-step: maximize the conditional expectation $Q(\theta \mid \widehat{\theta}^{(i)})$ of the log-likelihood of the complete data $\mathbf{Y}_c$,

$$\begin{aligned} \widehat{\theta}^{(i+1)} &= \arg\max_{\theta} Q\left(\theta \mid \widehat{\theta}^{(i)}\right) \\ &= \arg\max_{\theta}\left[-\left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right)^H \Lambda^{-1}\left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right)\right], \end{aligned} \tag{18}$$

which is equivalent to solve

$$\begin{aligned}\widehat{\mathbf{h}}_1^{(i+1)} &= \arg\min_{\mathbf{h}_1}\left\{\left\|\widehat{\mathbf{Y}}_1^{(i)} - \alpha\mathbf{D}_1\mathbf{F}_{2L_h-1}\mathbf{h}_1\right\|^2\right\},\\ \widehat{\mathbf{h}}_2^{(i+1)} &= \arg\min_{\mathbf{h}_2}\left\{\left\|\widehat{\mathbf{Y}}_2^{(i)} - \alpha\mathbf{D}_2\mathbf{F}_{L_h+L_g-1}\mathbf{h}_2\right\|^2\right\}.\end{aligned} \tag{19}$$

We obtain

$$\begin{aligned}\widehat{\mathbf{h}}_1^{(i+1)} &= \frac{1}{\alpha}\mathbf{F}_{2L_h-1}^H\mathbf{D}_1^{-1}\widehat{\mathbf{Y}}_1^{(i)},\\ \widehat{\mathbf{h}}_2^{(i+1)} &= \frac{1}{\alpha}\mathbf{F}_{L_h+L_g-1}^H\mathbf{D}_2^{-1}\widehat{\mathbf{Y}}_2^{(i)}.\end{aligned} \tag{20}$$

The new estimates $\widehat{\theta}^{(i+1)} = \{\widehat{\mathbf{h}}_1^{(i+1)}, \widehat{\mathbf{h}}_2^{(i+1)}\}$ are utilized in the E-step of the next iteration to update $\widehat{\mathbf{Y}}_1^{(i+1)}$ and $\widehat{\mathbf{Y}}_2^{(i+1)}$. Each iteration cycle increases the likelihood until convergence is accomplished.

The most striking feature of the algorithm is that it decouples the multiple-input channel estimation problem into two separate single-input channel estimation problems, a much more palatable problem. Therefore, the complexity of the channel estimation algorithm is essentially unaffected by the assumed number of signal components. Since $\mathbf{F}$ and $\mathbf{F}^H$ are DFT matrices and $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrices, the matrix inversion operation is relatively simple.

3.2. Cascaded Channel Estimation Based on Time Domain Processing. The time-domain received signal $\mathbf{y}$ at T_1 can be written as

$$\mathbf{y} = \alpha\Psi_{2L_h-1}(\mathbf{s}_1)\mathbf{h}_1 + \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\mathbf{h}_2 + \mathbf{n}_a. \tag{21}$$

A natural choice for the complete data $\mathbf{y}_c$ is obtained by decomposing the observed data $\mathbf{y}$ into its signal components:

$$\mathbf{y}_c = \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\alpha\Psi_{2L_h-1}(\mathbf{s}_1)\mathbf{h}_1 + \mathbf{n}_1\\ \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\mathbf{h}_2 + \mathbf{n}_2\end{bmatrix}, \tag{22}$$

where $\mathbf{y}_1$ is the component of the received signal $\mathbf{y}$ transmitted by the source node T_1 through the channel with impulse response $\mathbf{h}_1$, $\mathbf{y}_2$ is the component of the received signal $\mathbf{y}$ transmitted by the source node T_2 through the channel with impulse response $\mathbf{h}_2$, and $\mathbf{n}_1$ and $\mathbf{n}_2$ are obtained by arbitrarily decomposing the total noise $\mathbf{n}_a$ into two components:

$$\mathbf{n}_1 + \mathbf{n}_2 = \mathbf{n}_a. \tag{23}$$

Let $\mathbf{n}_1$ and $\mathbf{n}_2$ be statistically independent, zero-mean, and Gaussian with the covariance matrix $\mathbf{R}_1$ and $\mathbf{R}_2$, respectively. Denote the covariance matrix of $\mathbf{n}_a$ as $\mathbf{R}_a$; then $\mathbf{R}_1 + \mathbf{R}_2 = \mathbf{R}_a$.

The log-likelihood of the complete data $\mathbf{y}_c$ is

$$\log p(\mathbf{y}_c \mid \theta) = c - (\mathbf{y}_c - \mathbf{s}_c)^H\Lambda^{-1}(\mathbf{y}_c - \mathbf{s}_c), \tag{24}$$

where c contains all the terms that are independent of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ and the mean vector $\mathbf{s}_c$ and the covariance matrix Λ are given, respectively, by

$$\begin{aligned}\mathbf{s}_c &= \begin{bmatrix}\alpha\Psi_{2L_h-1}(\mathbf{s}_1)\mathbf{h}_1\\ \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\mathbf{h}_2\end{bmatrix},\\ \Lambda &= \begin{bmatrix}\mathbf{R}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_2\end{bmatrix} = \begin{bmatrix}\beta_1\mathbf{R}_{\mathbf{n}|\theta} & \mathbf{0}\\ \mathbf{0} & \beta_2\mathbf{R}_{\mathbf{n}|\theta}\end{bmatrix},\end{aligned} \tag{25}$$

where β_1 and β_2 are both real numbers, $0 < \beta_1 < 1$, $0 < \beta_2 < 1$, $\beta_1 + \beta_2 = 1$.

Suppose that $\widehat{\theta}^{(i)} = \{\widehat{\mathbf{h}}_1^{(i)}, \widehat{\mathbf{h}}_2^{(i)}\}$ denotes the current value of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ after i iterations of the EM algorithm. The $(i+1)$th iteration can be described in two steps as follows.

(1) E-step: given the observations $\mathbf{y}$ and the last estimates $\widehat{\theta}^{(i)} = \{\widehat{\mathbf{h}}_1^{(i)}, \widehat{\mathbf{h}}_2^{(i)}\}$, compute the expectation $Q(\theta \mid \widehat{\theta}^{(i)})$ of the log-likelihood of the complete data $\mathbf{y}_c$, which is also the conditional expectation of (24),

$$Q\left(\theta \mid \widehat{\theta}^{(i)} = d - \left(\widehat{\mathbf{y}}_c^{(i)} - \mathbf{s}_c\right)^H\Lambda^{-1}\left(\widehat{\mathbf{y}}_c^{(i)} - \mathbf{s}_c\right), \tag{26}$$

where d contains all the terms that are independent of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ and the ith estimate of the conditional expectation of the complete data $\mathbf{y}_c$ is [23]

$$\begin{aligned}\widehat{\mathbf{y}}_c^{(i)} &= \mathrm{E}\left\{\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} \mid \widehat{\theta}^{(i)}\right\} + \begin{bmatrix}\beta_1\mathbf{R}_{\mathbf{n}|\theta}\\ \beta_2\mathbf{R}_{\mathbf{n}|\theta}\end{bmatrix}\mathbf{R}_{\mathbf{n}|\theta}^{-1}\\ &\quad\times\left[\mathbf{y} - \mathbf{A}\mathrm{E}\left\{\mathbf{y}_c \mid \widehat{\theta}^{(i)}\right\}\right].\end{aligned} \tag{27}$$

The conditional expectations of $\mathbf{y}_1$ and $\mathbf{y}_2$ are given, respectively, by

$$\begin{aligned}\widehat{\mathbf{y}}_1^{(i)} &= \alpha\Psi_{2L_h-1}(\mathbf{s}_1)\widehat{\mathbf{h}}_1^{(i)} + \beta_1\\ &\quad\times\left[\mathbf{y} - \alpha\Psi_{2L_h-1}(\mathbf{s}_1)\widehat{\mathbf{h}}_1^{(i)} - \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\widehat{\mathbf{h}}_2^{(i)}\right],\\ \widehat{\mathbf{y}}_2^{(i)} &= \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\widehat{\mathbf{h}}_2^{(i)} + \beta_2\\ &\quad\times\left[\mathbf{y} - \alpha\Psi_{2L_h-1}(\mathbf{s}_1)\widehat{\mathbf{h}}_1^{(i)} - \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\widehat{\mathbf{h}}_2^{(i)}\right].\end{aligned} \tag{28}$$

(2) M-step: maximize the conditional expectation $Q(\theta \mid \widehat{\theta}^{(i)})$,

$$\widehat{\theta}^{(i+1)} = \arg\max_{\theta}\left[-\left(\widehat{\mathbf{y}}_c^{(i)} - \mathbf{s}_c\right)^H\Lambda^{-1}\left(\widehat{\mathbf{y}}_c^{(i)} - \mathbf{s}_c\right)\right]. \tag{29}$$

We obtain

$$\begin{aligned}\widehat{\mathbf{h}}_1^{(i+1)} &= \left(\alpha\Psi_{2L_h-1}^H(\mathbf{s}_1)\Psi_{2L_h-1}(\mathbf{s}_1)\right)^{-1}\\ &\quad\times\Psi_{2L_h-1}^H(\mathbf{s}_1)\widehat{\mathbf{y}}_1^{(i)},\\ \widehat{\mathbf{h}}_2^{(i+1)} &= \left(\alpha\Psi_{L_h+L_g-1}^H(\mathbf{s}_2)\Psi_{L_h+L_g-1}(\mathbf{s}_2)\right)^{-1}\\ &\quad\times\Psi_{L_h+L_g-1}^H(\mathbf{s}_2)\widehat{\mathbf{y}}_2^{(i)}.\end{aligned} \tag{30}$$

The new estimates $\widehat{\theta}^{(i+1)} = \{\widehat{\mathbf{h}}_1^{(i+1)}, \widehat{\mathbf{h}}_2^{(i+1)}\}$ are utilized in the E-step of the next iteration to update $\widehat{\mathbf{y}}_1^{(i+1)}$ and $\widehat{\mathbf{y}}_2^{(i+1)}$. Each iteration cycle increases the likelihood until convergence is accomplished.

3.3. Cascaded Channel Estimation without Training Sequence. In Sections 3.1 and 3.2, the channel estimate is continuously updated by transmitting pilot symbols using specified time-frequency lattices. When no training sequence is available

and signals are still to be detected from the observations, the EM algorithm is applied to take an average over the unknown channel impulse response for reducing bit errors caused by uncertainty in the channel. Here no training sequence means that the source node T_1 only knows its own signals $\mathbf{d}_1$, while $\mathbf{d}_2$ is completely unknown to T_1.

The selection of the complete data $\mathbf{Y}_c$ is the same as Section 3.1, and the log-likelihood of the complete data $\mathbf{Y}_c$ is the same as (13).

Suppose that $\widehat{\theta}^{(i)} = \{\widehat{\mathbf{h}}_1^{(i)}, \widehat{\mathbf{h}}_2^{(i)}\}$ denotes the current value of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ after i iterations of the EM algorithm. The $(i+1)$th iteration can be described in two steps as follows.

(1) E-step: given the observations $\mathbf{Y}$ and the last estimates $\widehat{\theta}^{(i)} = \{\widehat{\mathbf{h}}_1^{(i)}, \widehat{\mathbf{h}}_2^{(i)}\}$, compute the expectation $Q(\theta \mid \widehat{\theta}^{(i)})$ of the log-likelihood of the complete data $\mathbf{Y}_c$, which is the conditional expectation of (13)

$$Q\left(\theta \mid \widehat{\theta}^{(i)}\right) = d - \left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right)^H \Lambda^{-1} \left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right), \tag{31}$$

where d contains all the terms that are independent of $\theta = \{\mathbf{h}_1, \mathbf{h}_2\}$ and the ith estimate of the conditional expectations of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are given, respectively, by [23]

$$\begin{aligned}
\widehat{\mathbf{Y}}_1^{(i)} &= \alpha \mathbf{D}_1 \mathbf{F}_{2L_h-1} \widehat{\mathbf{h}}_1^{(i)} + \beta_1 \\
&\quad \times \left[\mathbf{Y} - \alpha \mathbf{D}_1 \mathbf{F}_{2L_h-1} \widehat{\mathbf{h}}_1^{(i)} - \alpha \widehat{\mathbf{D}}_2^{(i)} \mathbf{F}_{L_h+L_g-1} \widehat{\mathbf{h}}_2^{(i)}\right], \\
\widehat{\mathbf{Y}}_2^{(i)} &= \alpha \widehat{\mathbf{D}}_2^{(i)} \mathbf{F}_{L_h+L_g-1} \widehat{\mathbf{h}}_2^{(i)} + \beta_2 \\
&\quad \times \left[\mathbf{Y} - \alpha \mathbf{D}_1 \mathbf{F}_{2L_h-1} \widehat{\mathbf{h}}_1^{(i)} - \alpha \widehat{\mathbf{D}}_2^{(i)} \mathbf{F}_{L_h+L_g-1} \widehat{\mathbf{h}}_2^{(i)}\right],
\end{aligned} \tag{32}$$

where $\widehat{\mathbf{D}}_2^{(i)}$ is the estimate of $\mathbf{D}_2$ in the ith iteration,

$$\begin{gathered}
\widehat{\mathbf{D}}_2^{(i)} = \mathrm{E}\left\{\mathbf{D}_2 \mid \mathbf{y}, \widehat{\theta}^{(i)}\right\} = \mathrm{diag}\left\{\widehat{\mathbf{d}}_2^{(i)}\right\}, \\
\widehat{\mathbf{d}}_2^{(i)} = \left[\widehat{d}_2^{(i)}(0), \widehat{d}_2^{(i)}(1), \ldots, \widehat{d}_2^{(i)}(N-1)\right]^T, \\
\widehat{d}_2^{(i)}(n) = \sum_{m=1}^{M} \rho_m^{(i)}(n) s_m, \qquad \widehat{d}_{2m}^{(i)}(n) = s_m, \qquad \widehat{d}_{2j}^{(i)}(n) = s_j, \\
\rho_m^{(i)}(n) = \left(\exp\left\{-\left(Y(n) - \mu^{(i)}(n)\right)^H \mathbf{R}_{\mathbf{n}|\theta}^{-1} \left(Y(n) - \mu^{(i)}(n)\right)\right\}\right) \\
\times \left(\sum_{j=1}^{M} \exp\left\{-\left(Y(n) - \mu_j^{(i)}(n)\right)^H \right.\right. \\
\left.\left. \times \mathbf{R}_{\mathbf{n}|\theta}^{-1} \left(Y(n) - \mu_j^{(i)}(n)\right)\right\}\right)^{-1}, \\
\mu^{(i)}(n) = \widehat{Y}_1^{(i)}(n) + \widehat{Y}_{2m}^{(i)}(n), \\
\mu_j^{(i)}(n) = \widehat{Y}_1^{(i)}(n) + \widehat{Y}_{2j}^{(i)}(n), \\
\widehat{\mathbf{Y}}_{2m}^{(i)} = \mathrm{diag}\left\{\alpha\left(\widehat{\mathbf{d}}_{2m}^{(i)}\right)\right\} \mathbf{F}_{L_h+L_g-1} \widehat{\mathbf{h}}_2^{(i)}, \\
\widehat{\mathbf{Y}}_{2j}^{(i)} = \mathrm{diag}\left\{\alpha\left(\widehat{\mathbf{d}}_{2j}^{(i)}\right)\right\} \mathbf{F}_{L_h+L_g-1} \widehat{\mathbf{h}}_2^{(i)},
\end{gathered} \tag{33}$$

where $\widehat{\mathbf{d}}_2^{(i)}$ is the estimate of $\mathbf{d}_2$ in the ith iteration and $\widehat{d}_2^{(i)}(n)$ is the estimate of $d_2(n)$. s_m and s_j are different points in the constellation of $d_2(n)$, and $\rho_m^{(i)}(n)$ is the probability of $d_2(n) = s_m$.

(2) M-step: maximize the conditional expectation $Q(\theta \mid \widehat{\theta}^{(i)})$,

$$\widehat{\theta}^{(i+1)} = \arg\max_{\theta} \left[-\left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right)^H \Lambda^{-1} \left(\widehat{\mathbf{Y}}_c^{(i)} - \mathbf{D}_c\right)\right]. \tag{34}$$

We obtain

$$\begin{aligned}
\widehat{\mathbf{h}}_1^{(i+1)} &= \frac{1}{\alpha} \mathbf{F}_{2L_h-1}^H \mathbf{D}_1^{-1} \widehat{\mathbf{Y}}_1^{(i)}, \\
\widehat{\mathbf{h}}_2^{(i+1)} &= \frac{1}{\alpha} \mathbf{F}_{L_h+L_g-1}^H \left(\widehat{\mathbf{D}}_2^{(i)}\right)^{-1} \widehat{\mathbf{Y}}_2^{(i)}.
\end{aligned} \tag{35}$$

The new estimates $\widehat{\theta}^{(i+1)} = \{\widehat{\mathbf{h}}_1^{(i+1)}, \widehat{\mathbf{h}}_2^{(i+1)}\}$ are utilized in the E-step of the next iteration to update $\widehat{\mathbf{Y}}_1^{(i+1)}$ and $\widehat{\mathbf{Y}}_2^{(i+1)}$. Each iteration cycle increases the likelihood until convergence is accomplished.

4. The Bayesian Cramér-Rao Lower Bound

For many practical estimation problems, optimal estimators such as the maximum-likelihood estimator, maximum a posteriori estimator or minimum mean square error estimator are infeasible, so suboptimal estimators are needed, which are typically evaluated by determining MSE through simulations and comparing this error to theoretical performance bounds. In particular, the family of Cramér-Rao lower bounds (CRLBs) has been shown to give tight estimation lower bounds in a number of practical scenarios [26–29].

The CRLB for the estimation of deterministic parameters is given by the inverse of the Fisher information matrix (FIM), and Van Trees derived an analogous bound to the CRLB for random variables, referred to as "Bayesian CRLB" [30]. Unlike the standard and modified CRLBs, the statistical dependence is naturally considered within the Bayesian CRLB framework. With the assistance of Bayesian CRLB, the performance of the suboptimal estimators can be assessed, and the optimal training design could be obtained.

Denoting $\theta = [\mathbf{h}_1^T, \mathbf{h}_2^T]^T$, the corresponding FIM is defined as

$$\mathcal{J} = \mathrm{E}\left\{\frac{\partial \ln p(\mathbf{y}, \theta)}{\partial \theta^*} \left(\frac{\partial \ln p(\mathbf{y}, \theta)}{\partial \theta^*}\right)^H\right\} = \begin{bmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \mathcal{J}_{12}^H & \mathcal{J}_{22} \end{bmatrix}, \tag{36}$$

where the expectation is taken over the joint probability $p(\mathbf{y}, \theta)$. The Bayesian CRLB is the lower bound of any error covariance matrix $\mathrm{E}\{(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^H\}$,

$$\text{Bayesian CRLB} = \mathcal{J}^{-1}. \tag{37}$$

Figure 2: Channel estimation NMSE versus SNR.

Lemma 1. *The FIM can be calculated as*

$$\begin{aligned}
\mathcal{J}_{11} &= \frac{1}{\sigma_n^2\left(\alpha^2\sigma_h^2+1\right)}\left[\alpha\Psi_{2L_h-1}^H\left(\mathbf{s}_1\right)\right] \\
&\quad\times\left[\alpha\Psi_{2L_h-1}\left(\mathbf{s}_1\right)\right]+\mathbf{R}_{\mathbf{h}_1}^{-1}, \\
\mathcal{J}_{12} &= \frac{1}{\sigma_n^2\left(\alpha^2\sigma_h^2+1\right)}\left[\alpha\Psi_{2L_h-1}^H\left(\mathbf{s}_1\right)\right]\left[\alpha\Psi_{L_h+L_g-1}\left(\mathbf{s}_2\right)\right], \\
\mathcal{J}_{21} &= \frac{1}{\sigma_n^2\left(\alpha^2\sigma_h^2+1\right)}\left[\alpha\Psi_{L_h+L_g-1}^H\left(\mathbf{s}_2\right)\right]\left[\alpha\Psi_{2L_h-1}\left(\mathbf{s}_1\right)\right], \\
\mathcal{J}_{22} &= \frac{1}{\sigma_n^2\left(\alpha^2\sigma_h^2+1\right)}\left[\alpha\Psi_{L_h+L_g-1}^H\left(\mathbf{s}_2\right)\right] \\
&\quad\times\left[\alpha\Psi_{L_h+L_g-1}\left(\mathbf{s}_2\right)\right]+\mathbf{R}_{\mathbf{h}_2}^{-1},
\end{aligned} \tag{38}$$

where $\mathbf{R}_{\mathbf{h}_1}$ *and* $\mathbf{R}_{\mathbf{h}_2}$ *are the covariance matrices of* $\mathbf{h}_1$ *and* $\mathbf{h}_2$, *respectively,* $\mathbf{R}_{\mathbf{h}_1}=E_\theta\{\mathbf{h}_1\mathbf{h}_1^H\}$, $\mathbf{R}_{\mathbf{h}_2}=E_\theta\{\mathbf{h}_2\mathbf{h}_2^H\}$.

Proof. See Appendix. □

It indicates that the Bayesian CRLB is the inverse of the FIM $\mathcal{J}$, and $\mathcal{J}$ is a block matrix. If $\mathcal{J}$ can be simplified to a diagonal matrix, then the inverse operation becomes simple. Therefore, when calculating the Bayesian CRLB, we may assume that $\mathbf{s}_1$ and $\mathbf{s}_2$ are orthogonal to each other, which means $\mathcal{J}_{12}=\mathbf{0}$ and $\mathcal{J}_{21}=\mathbf{0}$, and $\mathcal{J}$ is simplified to a diagonal matrix. The Bayesian CRLBs for $\mathbf{h}_1$ and $\mathbf{h}_2$ can be, respectively, expressed as

$$\text{Bayesian CRLB}_{\mathbf{h}_1}=\mathcal{J}_{11}^{-1},\qquad \text{Bayesian CRLB}_{\mathbf{h}_2}=\mathcal{J}_{22}^{-1}. \tag{39}$$

The channel error covariance matrices are lower bounded by

$$\begin{aligned}
\text{Cov}_{\mathbf{h}_1} &= \text{E}\left\{\Delta\mathbf{h}_1\Delta\mathbf{h}_1^H\right\}\geq\text{Bayesian CRLB}_{\mathbf{h}_1}, \\
\text{Cov}_{\mathbf{h}_2} &= \text{E}\left\{\Delta\mathbf{h}_2\Delta\mathbf{h}_2^H\right\}\geq\text{Bayesian CRLB}_{\mathbf{h}_2}.
\end{aligned} \tag{40}$$

Furthermore, the channel estimation MSEs are lower bounded by

$$\begin{aligned}
\text{MSE}_{\mathbf{h}_1} &= \text{tr}\left(\text{Cov}_{\mathbf{h}_1}\right)\geq\text{tr}\left(\text{Bayesian CRLB}_{\mathbf{h}_1}\right), \\
\text{MSE}_{\mathbf{h}_2} &= \text{tr}\left(\text{Cov}_{\mathbf{h}_2}\right)\geq\text{tr}\left(\text{Bayesian CRLB}_{\mathbf{h}_2}\right).
\end{aligned} \tag{41}$$

5. Simulations

In this section, we provide numerical results to verify our studies in the relay-based M2M two-way communications, and only the channel estimation at T_1 is considered because of symmetry. For simplicity, $P_s=P_r$ is assumed, the noise variance is set as 1, and SNR is defined as $P_s/\sigma_n^2=P_s$. In all simulations, the QPSK modulation is applied, and the OFDM symbol length is taken as $N=64$. The channel delay spread taps are set as $L_h=L_g=6$, and all channel taps have unit variances. All the simulation results are averaged over $M_c=1000$ Monte-Carlo runs, and between Monte-Carlo runs, independent realizations of noise, input, and channels are used. The random initialization of $\mathbf{h}_1$ and $\mathbf{h}_2$ is utilized as the initial value of the EM algorithm. The normalized estimation mean square errors (NMSE) is adopted as the figure of merit to evaluate the channel estimation accuracy:

$$\begin{aligned}
\text{NMSE}_1 &\triangleq \frac{1}{M_c}\sum_{m=1}^{M_c}\frac{\left\|\hat{\mathbf{h}}_1-\mathbf{h}_1\right\|^2}{\left\|\mathbf{h}_1\right\|^2}, \\
\text{NMSE}_2 &\triangleq \frac{1}{M_c}\sum_{m=1}^{M_c}\frac{\left\|\hat{\mathbf{h}}_2-\mathbf{h}_2\right\|^2}{\left\|\mathbf{h}_2\right\|^2}.
\end{aligned} \tag{42}$$

For cascaded channel estimation based on frequency domain processing at T_1, the NMSE performance against SNR and number of iterations are shown in Figures 2 and 3, respectively. We can see that (1) the EM-based algorithm improves the estimation accuracy by iteratively refining the cascaded channel estimates and finally approaches the corresponding maximum likelihood solutions. After 4 iterations, the NMSEs of both $\mathbf{h}_1$ and $\mathbf{h}_2$ improve a lot from the initial estimates. The improvement in the first iteration is the most significant, and the improvement in the following iterations slows down. (2) The NMSE curve of $\mathbf{h}_1$ estimation almost coincides with the NMSE curve of $\mathbf{h}_2$ estimation completely. This is because in the cascaded channel estimation, $\mathbf{h}_1$ and $\mathbf{h}_2$ are symmetrical, so the estimation performance of $\mathbf{h}_1$ and $\mathbf{h}_2$ are the same. (3) The impact of iteration number on the NMSE performance is shown in Figure 3. Here we take the estimation of $\mathbf{h}_1$ for example. In different SNR conditions, EM algorithm converges to the maximum likelihood solution within 10 iterations. When SNR is relatively low, the convergence of the EM algorithm is relatively fast. As

Figure 3: Channel estimation NMSE versus number of iterations.

Figure 4: Channel estimation NMSE versus SNR.

Figure 5: Channel estimation NMSE versus number of iterations.

Figure 6: Channel estimation NMSE versus SNR.

SNR increases, the required iterations for convergence to the maximum likelihood solution increase accordingly.

For cascaded channel estimation based on time domain processing at T_1, the NMSE performance against SNR and number of iterations are shown in Figures 4 and 5, respectively. The NMSE curves demonstrate the same trend with Figures 2 and 3.

Figure 6 shows the NMSE performance comparison between the frequency domain processed estimation and the time domain processed estimation. The Bayesian CRLBs are exactly the same for the estimation of $\mathbf{h}_1$ and $\mathbf{h}_2$, which are also plotted to measure the NMSE performance. We can see that (1) the frequency domain processed estimation converges to the maximum likelihood solution after 4 iterations, while the time domain processed estimation converges to the maximum likelihood solution after 8 iterations. When the number of iterations is the same, the estimation results of the time domain process are more accurate than the results of the frequency domain process and more close to the corresponding Bayesian CRLB.

Figure 7 shows the NMSE performance comparison of the time domain processed estimation, when the power allocation between the subcarriers is not equal. The training

Figure 7: Channel estimation NMSE versus SNR.

Figure 8: Channel estimation NMSE versus SNR.

sequences $\mathbf{d}_1$ and $\mathbf{d}_2$ are subject to the following constraints. For $\mathbf{d}_1$, the power for subcarriers from 1 to 32 is halved, while the power for subcarriers from 33 to 64 increases to 1.5 times the original value. The power allocation between subcarriers is just the opposite for $\mathbf{d}_2$. We can see that, compared with Figure 4, the unequal power allocation almost causes no change in the NMSE performance.

Figure 8 shows the NMSE performance of cascaded channel estimation without training sequence. We can see that the EM-based algorithm improves the estimation accuracy by iteratively refining the cascaded channel estimates. However, since $\mathbf{d}_2$ is unknown to the receiver at T_1, the convergence is slower than the counterpart in Figure 2. Moreover, the NMSE performance of $\mathbf{h}_2$ estimation is worse than the NMSE performance of $\mathbf{h}_1$ estimation, due to the inaccurate estimates of $\mathbf{d}_2$. With the increase of the number of iterations, the channel estimation results become more and more accurate, the estimates of $\mathbf{d}_2$ becomes more and more accurate accordingly, which further improves the accuracy of channel estimation in turn.

If it is possible to select a more accurate initial value to enable the EM iterations in the relay-based M2M two-way communications, the speed of convergence can be accelerated, and the performance of channel estimation can be further improved.

6. Conclusion

In this paper, we have investigated the EM-based cascaded channel estimation algorithm in relay-based M2M two-way communications, which converts the multiple-input channel estimation problem into two separate single-input channel estimation problems, leading to a considerable simplification in the computations involved and increasing the possibility of the realization of M2M communications. In the EM algorithm, the received signal at the source node is treated as the incomplete data, and the set of the two signal components from the two source nodes is modeled as the complete data. Conditioned upon the incomplete observations, the algorithm maximizes the expectation of log likelihood function defined over the complete data, by averaging over the unknown underlying parameters and using the current estimates of the cascaded channels. The algorithm iterates back and forth, using the current channel parameter estimates to decompose the received signal better and thus increase the likelihood of the next channel parameter estimates. Even in the absence of training sequence, the cascaded channel estimates can still be obtained through the iterations between the E-step and M-step. The Bayesian Cramér-Rao lower bounds have been derived under random parameters for cascaded channel estimation. The proposed method works well without channel statistical information, and the simulation demonstrates good agreement of the theoretical lower bound and the practical estimation performance.

Appendix

Proof of Lemma 1

First calculate the likelihood probability $p(\mathbf{y} \mid \theta)$. Denote $\mathbf{W}_1$ as $N \times N$ circulant matrix whose first column is $[\mathbf{h}_1^T, \mathbf{0}_{1\times(N-2L_h+1)}]^T$ and $\mathbf{W}_2$ as $N \times N$ circulant matrix whose first column is $[\mathbf{h}_2^T, \mathbf{0}_{1\times(N-L_h-L_g+1)}]^T$. Suppose $N \geq \max(2L_h - 1, L_h + L_g - 1)$; then we have $\mathbf{HHs}_1 = \mathbf{W}_1\mathbf{s}_1 = \Psi_{2L_h-1}(\mathbf{s}_1)\mathbf{h}_1, \mathbf{HGs}_2 = \mathbf{W}_2\mathbf{s}_2 = \Psi_{L_h+L_g-1}(\mathbf{s}_2)\mathbf{h}_2$. Conditioned on a specific channel realization, the noise term in $\mathbf{y}$ is AWGN with variance $\mathbf{R}_{\mathbf{n}|\theta} = \sigma_n^2(\alpha^2\mathbf{HH}^H + \mathbf{I})$. The estimation of individual channel $\mathbf{h}$ is not considered in this paper, so the instantaneous value of $\mathbf{HH}^H$ is substituted by its expectation

$\mathrm{E}_\theta\{\mathbf{H}\mathbf{H}^H\} = \sigma_h^2\mathbf{I}$, which is to say, $\mathbf{R}_{\mathbf{n}|\theta} = \sigma_n^2(\alpha^2\sigma_h^2+1)\mathbf{I}$. Denote the mean value of $\mathbf{y}$ conditioned on θ as

$$\mu = \alpha\Psi_{2L_h-1}(\mathbf{s}_1)\mathbf{h}_1 + \alpha\Psi_{L_h+L_g-1}(\mathbf{s}_2)\mathbf{h}_2, \tag{A.1}$$

and the log-likelihood function is

$$\ln p(\mathbf{y} \mid \theta) = c - (\mathbf{y}-\mu)^H\mathbf{R}_{\mathbf{n}|\theta}^{-1}(\mathbf{y}-\mu), \tag{A.2}$$

where c contains all the terms that are independent of $\theta = [\mathbf{h}_1^T, \mathbf{h}_2^T]^T$. The related partial derivatives are listed as follows:

$$\begin{aligned}\frac{\partial\mu}{\partial h_{1j}} &= \alpha\Psi_{2L_h-1,[:,j]}(\mathbf{s}_1),\\ \frac{\partial\mu}{\partial h_{2j}} &= \alpha\Psi_{L_h+L_g-1,[:,j]}(\mathbf{s}_2),\\ \frac{\partial\mathbf{R}_{\mathbf{n}|\theta}}{\partial h_{1i}^*} &= \frac{\partial\mathbf{R}_{\mathbf{n}|\theta}}{\partial h_{2i}^*} = \mathbf{0}.\end{aligned} \tag{A.3}$$

Then calculate the Fisher information matrix (FIM). Since $\theta = [\mathbf{h}_1^T, \mathbf{h}_2^T]^T$, the FIM can be calculated as follows.

(1) *Calculate block submatrix* $\mathcal{J}_{11}$. Since $p(\mathbf{y} \mid \theta)$ is independent of $p(\theta)$, $\mathbf{h}_1$ is independent of $\mathbf{h}_2$; then $p(\mathbf{y},\theta) = p(\mathbf{y} \mid \theta)p(\theta)$, $p(\theta) = p(\mathbf{h}_1)p(\mathbf{h}_2)$, and $\mathcal{J}_{11}$ is calculated as

$$\begin{aligned}\mathcal{J}_{11} &= \mathrm{E}\left\{\frac{\partial\ln p(\mathbf{y},\theta)}{\partial\mathbf{h}_1^*}\left(\frac{\partial\ln p(\mathbf{y},\theta)}{\partial\mathbf{h}_1^*}\right)^H\right\}\\ &= \mathrm{E}\left\{\frac{\partial\ln(p(\mathbf{y} \mid \theta)p(\theta))}{\partial\mathbf{h}_1^*}\left(\frac{\partial\ln(p(\mathbf{y} \mid \theta)p(\theta))}{\partial\mathbf{h}_1^*}\right)^H\right\}\\ &= \mathrm{E}_\theta\{\mathcal{F}_{11}\} + \mathbf{R}_{\mathbf{h}_1}^{-1},\end{aligned} \tag{A.4}$$

where the first and the second expectations are over $p(\theta)$ and $p(\mathbf{y} \mid \theta)$, respectively. $\mathcal{F}_{11}$ is the FIM for the deterministic θ,

$$\begin{aligned}\mathcal{F}_{11} &= \mathrm{E}_{\mathbf{y}|\theta}\left\{\frac{\partial\ln p(\mathbf{y} \mid \theta)}{\partial\mathbf{h}_1^*}\left(\frac{\partial\ln p(\mathbf{y} \mid \theta)}{\partial\mathbf{h}_1^*}\right)^H\right\}\\ &= \alpha^2\Psi_{2L_h-1}^H(\mathbf{s}_1)\mathbf{R}_{\mathbf{n}|\theta}^{-1}\Psi_{2L_h-1}(\mathbf{s}_1).\end{aligned} \tag{A.5}$$

Therefore,

$$\begin{aligned}\mathcal{J}_{11} &= \mathrm{E}_\theta\{\mathcal{F}_{11}\} + \mathbf{R}_{\mathbf{h}_1}^{-1}\\ &= \frac{\alpha^2}{\sigma_n^2(\alpha^2\sigma_h^2+1)}\Psi_{2L_h-1}^H(\mathbf{s}_1)\Psi_{2L_h-1}(\mathbf{s}_1) + \mathbf{R}_{\mathbf{h}_1}^{-1}.\end{aligned} \tag{A.6}$$

(2) *Calculate block submatrices* $\mathcal{J}_{12}$, $\mathcal{J}_{21}$

$$\begin{aligned}\mathcal{J}_{12} &= \mathrm{E}\left\{\frac{\partial\ln p(\mathbf{y},\theta)}{\partial\mathbf{h}_1^*}\left(\frac{\partial\ln p(\mathbf{y},\theta)}{\partial\mathbf{h}_2^*}\right)^H\right\}\\ &= \mathrm{E}\left\{\frac{\partial\ln(p(\mathbf{y} \mid \theta)p(\theta))}{\partial\mathbf{h}_1^*}\left(\frac{\partial\ln(p(\mathbf{y} \mid \theta)p(\theta))}{\partial\mathbf{h}_2^*}\right)^H\right\}\\ &= \mathrm{E}_\theta\{\mathcal{F}_{12}\},\end{aligned} \tag{A.7}$$

where $\mathcal{F}_{12}$ is the FIM for the deterministic θ,

$$\begin{aligned}\mathcal{F}_{12} &= \mathrm{E}_{\mathbf{y}|\theta}\left\{\frac{\partial\ln p(\mathbf{y} \mid \theta)}{\partial\mathbf{h}_1^*}\left(\frac{\partial\ln p(\mathbf{y} \mid \theta)}{\partial\mathbf{h}_2^*}\right)^H\right\}\\ &= \alpha^2\Psi_{2L_h-1}^H(\mathbf{s}_1)\mathbf{R}_{\mathbf{n}|\theta}^{-1}\Psi_{L_h+L_g-1}(\mathbf{s}_2).\end{aligned} \tag{A.8}$$

Therefore,

$$\begin{aligned}\mathcal{J}_{12} &= \mathrm{E}_\theta\{\mathcal{F}_{12}\}\\ &= \frac{\alpha^2}{\sigma_n^2(\alpha^2\sigma_h^2+1)}\Psi_{2L_h-1}^H(\mathbf{s}_1)\Psi_{L_h+L_g-1}(\mathbf{s}_2),\\ \mathcal{J}_{21} &= \mathcal{J}_{12}^H = \frac{\alpha^2}{\sigma_n^2(\alpha^2\sigma_h^2+1)}\Psi_{L_h+L_g-1}^H(\mathbf{s}_2)\Psi_{2L_h-1}(\mathbf{s}_1).\end{aligned} \tag{A.9}$$

(3) *Calculate block submatrices* $\mathcal{J}_{22}$

$$\begin{aligned}\mathcal{J}_{22} &= \mathrm{E}\left\{\frac{\partial\ln p(\mathbf{y},\theta)}{\partial\mathbf{h}_2^*}\left(\frac{\partial\ln p(\mathbf{y},\theta)}{\partial\mathbf{h}_2^*}\right)^H\right\}\\ &= \mathrm{E}\left\{\frac{\partial\ln(p(\mathbf{y} \mid \theta)p(\theta))}{\partial\mathbf{h}_2^*}\left(\frac{\partial\ln(p(\mathbf{y} \mid \theta)p(\theta))}{\partial\mathbf{h}_2^*}\right)^H\right\}\\ &= \mathrm{E}_\theta\{\mathcal{F}_{22}\} + \mathbf{R}_{\mathbf{h}_2}^{-1},\end{aligned} \tag{A.10}$$

where $\mathcal{F}_{22}$ is the FIM for the deterministic θ,

$$\begin{aligned}\mathcal{F}_{22} &= \mathrm{E}_{\mathbf{y}|\theta}\left\{\frac{\partial\ln p(\mathbf{y} \mid \theta)}{\partial\mathbf{h}_2^*}\left(\frac{\partial\ln p(\mathbf{y} \mid \theta)}{\partial\mathbf{h}_2^*}\right)^H\right\}\\ &= \alpha^2\Psi_{L_h+L_g-1}^H(\mathbf{s}_2)\mathbf{R}_{\mathbf{n}|\theta}^{-1}\Psi_{L_h+L_g-1}(\mathbf{s}_2).\end{aligned} \tag{A.11}$$

Therefore,

$$\begin{aligned}\mathcal{J}_{22} &= \mathrm{E}_\theta\{\mathcal{F}_{22}\} + \mathbf{R}_{\mathbf{h}_2}^{-1}\\ &= \frac{\alpha^2}{\sigma_n^2(\alpha^2\sigma_h^2+1)}\\ &\quad\times\Psi_{L_h+L_g-1}^H(\mathbf{s}_2)\Psi_{L_h+L_g-1}(\mathbf{s}_2) + \mathbf{R}_{\mathbf{h}_2}^{-1}.\end{aligned} \tag{A.12}$$

Acknowledgments

This research was partly supported by the National Science Foundation of China (Grant no. NFSC 61071083, and 61371073) and the National High Technology Research and Development Program of China (863 Program no. 2012AA01A506).

References

[1] H. Cho and J. Puthenkulam, "Machine to Machine (M2M) Communication Study Report," IEEE 802.16ppc-10/0002r6, May 2010.

[2] Harbor Research Report, Machine-To-Machine (M2M) & Smart Systems Forecast 2010-2014, 2009.

[3] 3GPP Technical Report, "System improvements for machine-type communications (release 10)," TR 23.888, July 2010.

[4] S. Andreev, O. Galinina, and Y. Koucheryavy, "Energy-efficient client relay scheme for machine-to-machine communication," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference (GLOBECOM '11)*, December 2011.

[5] M. Gerasimenko, V. Petrov, and O. Galinina, "Impact of machine-type communications on energy and delay performance of random access channel in LTE-advanced," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 4, pp. 366–377, 2013.

[6] T. Liu, L. Song, Y. Li, Q. Huo, and B. Jiao, "Performance analysis of hybrid relay selection in cooperative wireless systems," *IEEE Transactions on Communications*, vol. 60, no. 3, pp. 779–788, 2012.

[7] Q. Huo, L. Song, Y. Li, and B. Jiao, "A distributed differential space-time coding scheme with analog network coding in two-way relay networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4998–5004, 2012.

[8] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 1668–1672, July 2006.

[9] P. Popovski and H. Yomo, "Physical network coding in two-way wireless relay channels," in *Proceedings of the IEEE International Conference on Communications (ICC '07)*, pp. 707–712, June 2007.

[10] T. Cui, T. Ho, and J. Kliewer, "Memoryless relay strategies for two-way relay channels: Performance analysis and optimization," in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 1139–1143, May 2008.

[11] L. Song, Y. Li, A. Huang, B. Jiao, and A. V. Vasilakos, "Differential modulation for bidirectional relaying with analog network coding," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3933–3938, 2010.

[12] L. Song, G. Hong, B. Jiao, and M. Debbah, "Joint relay selection and analog network coding using differential modulation in two-way relay channels," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2932–2939, 2010.

[13] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Transactions on Communications*, vol. 57, no. 10, pp. 3024–3033, 2009.

[14] F. Gao, R. Zhang, and Y.-C. Liang, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4443–4455, 2009.

[15] W. Yang, Y. Cai, J. Hu, and W. Yang, "Channel estimation for two-way relay OFDM networks," *Eurasip Journal on Wireless Communications and Networking*, vol. 2010, Article ID 186182, 2010.

[16] B. Jiang, F. Gao, X. Gao, and A. Nallanathan, "Channel estimation and training design for two-way relay networks with power allocation," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 2022–2032, 2010.

[17] T.-H. Pham, Y.-C. Liang, A. Nallanathan, and H. K. Garg, "Optimal training sequences for channel estimation in bi-directional relay networks with multiple antennas," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 474–479, 2010.

[18] X. Liao, L. Fan, and F. Gao, "Blind channel estimation for OFDM modulated two-way relay network," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '10)*, April 2010.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximun likelihood estimation from incomplete data," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[20] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[21] J.-S. Sheu and W.-H. Sheen, "An EM algorithm-based channel estimation for OFDM amplify-and-forward relaying systems," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, May 2010.

[22] P. Lioliou, M. Viberg, and M. Matthaiou, "Bayesian channel estimation techniques for AF MIMO relaying systems," in *Proceedings of the IEEE 74th Vehicular Technology Conference (VTC Fall '11)*, September 2011.

[23] P. Lioliou, M. Viberg, and M. Matthaiou, "Bayesian approach to channel estimation for AF MIMO relaying systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 8, pp. 1440–1451, 2012.

[24] C. Zhang, S. Tang, and P. Ren, "EM algorithm based channel estimation for amplify-and-forward relay networks with unknown noise correlation," in *Proceedings of the IEEE Vehicular Technology Conference (VTC Fall '12)*, pp. 1–5, September 2012.

[25] G. Geoffrey and S. David, *Probability and Random Processes*, ISBN 0-19-857222-0, Oxford University Press, 3rd edition, 2001.

[26] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley, New York, NY, USA, 1968.

[27] L. P. Seidman, "Performance limitations and error calculations for parameter estimation," *Proceedings of the IEEE*, vol. 58, no. 5, pp. 644–652, 1970.

[28] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Transactions on Information Theory*, vol. 15, no. 3, pp. 386–391, 1969.

[29] R. W. Miller and C. B. Chang, "A modified Cramr-Rao bound and its applications," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 398–400, 1978.

[30] Gill and Levit, *Applications of the Van Trees Inequality: The Bayesian Cramr-Rao Bound*, Bernoulli, 1995.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 313797, 8 pages
http://dx.doi.org/10.1155/2013/313797

Research Article

Random Access Control for M2M in LTE System

Zhide Chen and Yali Zeng

Key Laboratory of Network Security and Cryptology, Fujian Normal University, Fuzhou 350007, China

Correspondence should be addressed to Zhide Chen; zhidechen@fjnu.edu.cn

Received 19 July 2013; Revised 14 October 2013; Accepted 22 October 2013

Academic Editor: Jianhua He

Copyright © 2013 Z. Chen and Y. Zeng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the implementation of 4G mobile networks, the number of M2M users will be much large in a cell. When massive M2M users try to randomly access simultaneously, this will impose a great influence on H2H communication. Without effective control, these M2M users will lead to a system breakdown. In this paper, we put forward a model, which allows low delay tolerance of users to have a higher priority for random access, when contending with high delay tolerance of users for limited random access resources. In this model, by controlling access rates of different delay tolerance of users, the random access success probability will be improved efficiently. In the meantime, it will achieve the highest success probability and optimize the system performance by setting optimal access rates for different delay tolerance of users. This will be a practical guide to M2M communication in LTE system.

1. Introduction

3G mobile communication is moving towards 4G with the rapid development of mobile communication technology. An increasing number of mobile and nonmobile users will be a great challenge for 4G communication. It means that the wireless network system should have better performance to support more and more communications. In 2004, the 3GPP and 3GPP2 launched Long Term Evolution (LTE) project, aiming to introduce some key technologies to achieve higher and better system performance. However, LTE system is mainly designed for traditional H2H users. With the tremendous increase of M2M users, it is bound to have a great impact on H2H communications. It can improve the system performance effectively through appropriately controlling M2M users to access. Higher success probability and better system performance can be achieved through setting different access rates for different levels of users. This paper proves that hierarchical control of different users can efficiently improve random access success probability and optimize system performance.

The full paper is organized as follows. In Section 2, we briefly introduce research background and related works. Section 3 analyzes the influence of a large number of M2M users for simultaneous random access in LTE network after studying the LTE random access process. A system model is proposed in Section 4 to improve random access success probability when numerous users access simultaneously. In the next section, we analyze the model and come to a conclusion that setting optimal access rates for different levels of users will achieve the highest success probability. Section 6 builds a simulation to verify the system model. Finally, a brief summary of this paper is presented in Section 7.

2. Research Background and Related Works

2.1. Research Background

2.1.1. LTE. Since traditional 2G/3G network cannot meet the needs of multiple terminals of wireless bandwidth resources, the 3GPP launched LTE/LTE-A. LTE is short for Long Term Evolution (Figure 1). It is not really 4G, and it can be called 3.9G. The real 4G is LTE-advanced. In the fourth-generation broadband wireless multiple access systems data rate will reach probably hundreds of megabits per second [1]. Moreover, downlink data rate target will peak at 100 Mbps and uplink target data rate will peak at 50 Mbps for a bandwidth of 20 MHz in LTE system [2]. However, [3] analyses that downlink peak data rate over 1 Gb/s and uplink peak data rate over 500 Mb/s becomes possible. To meet the need of high data rate, downlink multiple access mechanism

Figure 1: LTE system architecture diagram.

Figure 2: Pure ALOHA.

Figure 3: Slot ALOHA.

utilizes orthogonal frequency division multiplexing (OFDM) and uplink multiple access mechanism utilizes single carrier frequency division multiple access (SC-FDMA) [4].

LTE supports a higher data rate and cell capacity; such a system can be better for meeting the need of IOT for broadband resources. Time delay is 10 times lower than 2G/3G in LTE system, which is better for meeting the high requirements on the system delay of M2M equipment, such as some M2M devices used in monitoring and alarm systems. The number and traffic of wireless network users will grow exponentially [5]. Furthermore, LTE is increasingly seen as a service system for real-time M2M communication [6]. Asynchronous M2M users' accesses utilize access reservation protocol in LTE system. Access reservation protocol is helpful for asynchronous communication or intermittent transmission [7]. The LTE access preambles are selected from the orthogonal preamble set which are obtained from the Zadoff-Chu sequence [8]. LTE access process includes contention phase and data phase; LTE contention phase utilizes time-sharing ALOHA [9]. This ALOHA mechanism proposes a combination model [10]. More information about this combination model is in [11]; earlier research on access reservation protocol is in [12].

2.1.2. M2M. M2M, in a narrow sense, is short for machine to machine. M2M communication is defined as the machine type communication in 3GPP [13]. M2M applies in a very wide range of areas such as industry and agriculture. There are many examples of M2M applications, such as fleet management, smart meters, and goods tracking. We can see that the number of smart meters is far more than 35,000 in New York City with 2 km cell radii [13]. According to the prediction in [14], benefits from M2M communication will greatly improve in the next decade.

2.1.3. ALOHA. The LTE system utilizes slotted ALOHA. ALOHA protocol is divided into two types, one is Pure ALOHA (Figure 2) and the other is Slot ALOHA (Figure 3). Slot ALOHA requires clock synchronization in the whole network. Users can arrive at any time, but they can send data packets until the beginning of the next slot. It may cause part conflict or complete conflict in Pure ALOHA; however, it only causes complete conflict in Slot ALOHA.

2.2. Related Works. It will have a huge impact on the LTE system when multiple M2M users access at the same time. Therefore, aiming to solve the problem of congestion, the 3GPP proposed the following schemes:

(1) introducing new barring factors for M2M equipments to avoid network congestion and control access,

(2) classifying RACH resources,

(3) setting specific back-off time for M2M equipments,

(4) allocating specific slot for M2M equipments to access, and so on.

A fast retrial and dynamic control of random access algorithm in LTE system was proposed [15], and then it got the possible optimal rate of random access when there were some M2M users' packets and retrial packets. However, it just considered M2M users; that is to say, all M2M users were regarded as the same level in [15]. Obviously, it is impractical. In [16], two new dynamic spectrum access schemes for secondary users were proposed which were based on continuous-time Markov chains (CTMC), and then it got the optimal probability of secondary users for random access under the premise of the collision threshold value. In [16], users were divided into two levels. Based on [16, 17], access channel got the optimal probability of secondary users for random access based on Markov in two kinds of channel states.

3. LTE Random Access

3.1. Access Pattern. In a mobile communication network, users need to access if they want to communicate with others. When users synchronized with uplink, they can transmit data packets. In LTE system, there are two kinds of random access patterns, one is based on contention (Figure 4) and the other is based on the noncontention (Figure 5). eNB is

Figure 4: Contention-based.

Figure 5: Noncontention-based.

evolved node and UE is user equipment in the figure. The main random access pattern is contention-based pattern in LTE system, so we focus on this pattern in this paper. The process of contention-based pattern is as follows:

(1) random access initialization,

(2) random access resource selection,

(3) random access preambles transmission,

(4) random access response and Msg3,

(5) contention resolution.

3.2. Problem Arises. It will lead to a larger chance of choosing the same preamble in the stage of choosing access resources, if numerous users access simultaneously and preambles and time-frequency resources are limited. If many users send preambles with the same time-frequency resources, this will cause the eNB to get the same RA-RNTIs according to the received time-frequency resources and to get the same PDCCHs scrambled with the same RA-RNTIs. The UL grants, preamble indexes, and TC-RNTIs also will be the same. If all of the M2M users can receive PDCCHs, which are scrambled with their own RA-RNTIs, they will send Msg3 according to UL grant's instructions. Moreover, M2M users will send Msg3 in the same position because of the same UL grants. However, the eNB just decodes the first received Msg3. It means other users will fail to access. If none of the users receive RAR or if decoded preamble index is different from their own, they will choose a random value in [0, back-off time]. The number of optional values is far less than the number of users who are choosing the back-off values at the same time, so many users will choose the same values, and these users will collide once again in the next retrial.

4. System Model

In LTE system, due to the high density of M2M users, there may be massive M2M users in a cell. These users may access randomly in a very short period of time in the contention-based pattern. As we know, random access resources are limited, which are less than 64 preambles used in the competition in a cell. When a cell radius is larger, these preambles are more possible to be pairwise orthogonal [18].

Now, we suppose that there are N ($N < 64$) preamble sequences, and they are orthogonal to each other. We regard the N preamble sequences as N parallel logical channels (hereinafter to be referred to as channels) [1]; that is to say, there are N channels that can be used in a slot. When there is no user in a channel for random access in the current time slot, this indicates that the channel is idle. When there is one and only one user in a channel for random access in the current time slot, this indicates that the user accesses successfully the channel and the channel is successful. When there are more than two users in the channel for random access, the channel is colliding. Figure 6 shows that four preamble sequences are available in a slot; that is, there are four logical channels available. User 1, user 2, and user 3 are accessing the first slot. User 1 chooses channel 1. User 1 accesses successfully as there is only one user in this channel in this slot. User 2 and user 3 choose channel 4 in the meantime, so it causes collision. None of the users choose channel 2 or channel 3, so they are idle.

Now, we define the following notations:

n_H: number of H2H users for random access in this channel in this slot,

n_M: number of M2M users for random access in this channel in this slot,

n_1: number of the primary users for random access in this channel in this slot,

n_2: number of the secondary users for random access in this channel in this slot,

n_3: number of the third users for random access in this channel in this slot,

FIGURE 6: Contention-based random access with 4 logical channels.

FIGURE 7: Contention-based random access flowchart of two levels of users.

λ_H: arrivals of H2H users,

λ_M: arrivals of M2M users,

λ_1: arrivals of the primary users,

λ_2: arrivals of the secondary users,

λ_3: arrivals of the third users,

α: the rate of M2M users' access,

α_2: the rate of the secondary users' access,

α_3: the rate of the third users' access.

In LTE system, when users are divided into two levels, we can draw the contention-based random access flowchart as follows (Figure 7).

When packets arrive in the same channel in the same time slot, the eNB determines the following.

(1) If there are only H2H user's packets, go to step 2; otherwise, go to step 3.

(2) If there is only one H2H user's packet, let the H2H user access, and then go to step 6; otherwise, go to step 7.

(3) If there are only M2M users' packets, go to step 4; otherwise, go to step 5.

(4) If there is only one M2M user's packet, let the M2M user access with rate α, and then go to step 6; otherwise, go to step 7.

(5) If there are multiple H2H users and M2M users packets, go to step 2.

(6) The channel is successful.

(7) The channel is colliding.

All of the colliding packets have to retry until a random time.

5. Analysis

5.1. Two Levels. From the system model in Section 4, we can know that users will choose a channel to access from N logical channels when users want to communicate with others. By controlling the access rate α of M2M users, we want to achieve the highest success probability of random access in a channel. We come to a conclusion that the success probability of random access will achieve the highest value when $\alpha = 1/\lambda_M$. The proof is as follows.

The probability of i H2H users for random access in a channel in a current time slot is

$$p(n_H = i) = \frac{\lambda_H^i}{i!} e^{-\lambda_H}. \tag{1}$$

The probability of j M2M users for random access in a channel in a current time slot is

$$p(n_M = j) = \frac{(\alpha\lambda_M)^j}{j!} e^{-\alpha\lambda_M}. \tag{2}$$

Because H2H users' delay tolerance is lower than M2M users', H2H users can access if H2H users' data packets arrive, regardless of whether the M2M users' data packets arrive or not. The premise of access success for H2H users is that there is one and only one H2H user to access the channel in the slot, so the access success probability of H2H users is

$$p_{H,\text{succ}} = p(n_H = 1) = \lambda_H e^{-\lambda_H}. \tag{3}$$

M2M users are allowed to access when there are not any H2H users' data packets. The premise of access success for M2M users is that there is one and only one M2M user to access the channel in the slot, so the access success probability of M2M users is

$$p_{M,\text{succ}} = p(n_H = 0) * p(n_M = 1) = e^{-\lambda_H} * \alpha\lambda_M e^{-\alpha\lambda_M}. \tag{4}$$

The access success probability of a channel in the slot is

$$P_{\text{success}} = p_{H,\text{succ}} + p_{M,\text{succ}} = \lambda_H e^{-\lambda_H} + e^{-\lambda_H} * \alpha\lambda_M e^{-\alpha\lambda_M}. \tag{5}$$

The probability of idle state of a channel in the slot is

$$P_{\text{idle}} = p(n_H = 0) * p(n_M = 0) = e^{-\lambda_H} e^{-\alpha\lambda_M}. \quad (6)$$

The probability of colliding state of a channel in the slot is

$$\begin{aligned} P_{\text{collision}} &= 1 - P_{\text{success}} - P_{\text{idle}} \\ &= 1 - \left(\lambda_H e^{-\lambda_H} + e^{-\lambda_H} * \alpha\lambda_M e^{-\alpha\lambda_M}\right) \\ &\quad - e^{-\lambda_H} e^{-\alpha\lambda_M} e; \end{aligned} \quad (7)$$

to gain the value of α, we need to get the derivation of (5); its first derivative is

$$P'_{\text{success}} = \lambda_M e^{-\lambda_H - \alpha\lambda_M} (1 - \alpha\lambda_M). \quad (8)$$

Hypothesis (8) is equal to 0, so we can get $\alpha = 1/\lambda_M$, and the second derivative of (8) is

$$P''_{\text{success}} = \lambda_M e^{-\lambda_H - \alpha\lambda_M} \left(\alpha\lambda_M^2 - 2\lambda_M\right). \quad (9)$$

Substitute $\alpha = 1/\lambda_M$ into (9); we can get

$$P''_{\text{success}} = -\lambda_M^2 e^{-\lambda_H - 1} < 0, \quad (10)$$

so P_{success} can be the maximum value when $\alpha = (1/\lambda_M)$; that is to say, when $\alpha = 1/\lambda_M$, the random access success probability will achieve the highest value; that is,

$$P_{2,\max} = \lambda_H e^{-\lambda_H} + e^{-\lambda_H - 1}. \quad (11)$$

5.2. Three Levels. We divide random access users into three levels here. The primary users have the lowest delay tolerance, the secondary users have the middle delay tolerance, and the third users have the highest delay tolerance. Three levels of users arrival rates obey the Poisson random processes with parameters λ_1, λ_2, and λ_3, respectively. We allow the primary users to access when the primary users' data packets arrive, no matter whether the secondary and third users' data packets arrive or not. We allow the secondary users to access with rate α_2, when the secondary users' data packets arrive and there are no primary users' data packets, no matter whether the third users arrive or not. We allow the third users to access with rate α_3 when there are no primary or secondary users' data packets. By controlling the access rate α_2 of the secondary users and the access rate α_3 of the third users, we want to achieve the highest random access success probability in a channel. We come to a conclusion that the random access success probability will achieve the highest value when $\alpha_2 = (1 - e^{-1})/\lambda_2$ and $\alpha_3 = 1/\lambda_3$. The proof is as follows.

The random access probability of i primary users in a channel in a current time slot is

$$p(n_1 = i) = \frac{\lambda_1^i}{i!} e^{-\lambda_1}. \quad (12)$$

The random access probability of j secondary users in a channel in a current time slot is

$$p(n_2 = j) = \frac{(\alpha_2\lambda_2)^j}{j!} e^{-\alpha_2\lambda_2}. \quad (13)$$

The random access probability of k third users in a channel in a current time slot is

$$p(n_3 = k) = \frac{(\alpha_3\lambda_3)^k}{k!} e^{-\alpha_3\lambda_3}. \quad (14)$$

The access success probability of the primary users is

$$P_{1,\text{su}} = p(n_1 = 1) = \lambda_1 e^{-\lambda_1}. \quad (15)$$

The access success probability of the secondary users is

$$P_{2,\text{su}} = p(n_1 = 0)\, p(n_2 = 1) = e^{-\lambda_1} * \alpha_2\lambda_2 e^{-\alpha_2\lambda_2}. \quad (16)$$

The access success probability of the third-level users is

$$\begin{aligned} P_{3,\text{su}} &= p(n_1 = 0)\, p(n_2 = 0)\, p(n_3 = 1) \\ &= e^{-\lambda_1} * e^{-\alpha_2\lambda_2} * \alpha_3\lambda_3 e^{-\alpha_3\lambda_3}. \end{aligned} \quad (17)$$

The access success probability in a channel in the slot is

$$\begin{aligned} P_{\text{su}} &= P_{1,\text{su}} + P_{2,\text{su}} + P_{3,\text{su}} \\ &= \lambda_1 e^{-\lambda_1} + e^{-\lambda_1} * \alpha_2\lambda_2 e^{-\alpha_2\lambda_2} \\ &\quad + e^{-\lambda_1} * e^{-\alpha_2\lambda_2} * \alpha_3\lambda_3 e^{-\alpha_3\lambda_3}. \end{aligned} \quad (18)$$

To gain the maximum value of P_{su}, we need to get the first and second derivatives of (18); he first derivative of P_{su} is

$$\begin{aligned} P_{11} &= \frac{\partial P_{\text{su}}}{\partial \alpha_2} = \lambda_2 e^{-\lambda_1 - \alpha_2\lambda_2} - \alpha_2\lambda_2^2 e^{-\lambda_1 - \alpha_2\lambda_2} - \alpha_3\lambda_3\lambda_2 e^{-\alpha_3\lambda_3} \\ &= \lambda_2 e^{-\lambda_1 - \alpha_2\lambda_2} \left(1 - \alpha_2\lambda_2 - \alpha_3\lambda_3 e^{-\alpha_3\lambda_3}\right). \end{aligned} \quad (19)$$

The first derivative of P_{su} is

$$\begin{aligned} P_{12} &= \frac{\partial P_{\text{su}}}{\partial \alpha_3} = \lambda_3 e^{-\lambda_1 - \alpha_2\lambda_2 - \alpha_3\lambda_3} - \alpha_3\lambda_3^2 e^{-\lambda_1 - \alpha_2\lambda_2 - \alpha_3\lambda_3} \\ &= \lambda_3 e^{-\lambda_1 - \alpha_2\lambda_2 - \alpha_3\lambda_3} (1 - \alpha_3\lambda_3). \end{aligned} \quad (20)$$

The second derivative of P_{su} is

$$\begin{aligned} P_{21} &= \frac{\partial^2 P_{\text{su}}}{\partial \alpha_2^2} = \frac{\partial P_{11}}{\partial \alpha_2} \\ &= -\lambda_2^2 e^{-\lambda_1 - \alpha_2\lambda_2} \left(1 - \alpha_2\lambda_2 - \alpha_3\lambda_3 e^{-\alpha_3\lambda_3}\right) - \lambda_2^2 e^{-\lambda_1 - \alpha_2\lambda_2} \\ &= -\lambda_2^2 e^{-\lambda_1 - \alpha_2\lambda_2} \left(2 - \alpha_2\lambda_2 - \alpha_3\lambda_3 e^{-\alpha_3\lambda_3}\right). \end{aligned} \quad (21)$$

The second derivative of P_{su} is

$$\begin{aligned} P_{22} &= \frac{\partial^2 P_{\text{su}}}{\partial \alpha_3^2} = \frac{\partial P_{12}}{\partial \alpha_3} \\ &= -\lambda_3^2 e^{-\lambda_1 - \alpha_2\lambda_2 - \alpha_3\lambda_3} (1 - \alpha_3\lambda_3) - \lambda_3^2 e^{-\lambda_1 - \alpha_2\lambda_2 - \alpha_3\lambda_3} \\ &= -\lambda_3^2 e^{-\lambda_1 - \alpha_2\lambda_2 - \alpha_3\lambda_3} (2 - \alpha_3\lambda_3). \end{aligned} \quad (22)$$

The second mixed derivative of P_{su} is

$$\begin{aligned} P_{23} &= \frac{\partial^2 P_{su}}{\partial\alpha_2\partial\alpha_3} = \frac{\partial P_{11}}{\partial\alpha_3} = \lambda_2 e^{-\lambda_1-\alpha_2\lambda_2}\left(\lambda_3 e^{-\alpha_3\lambda_3} - \lambda_3^2\alpha_3\lambda_3\right) \\ &= -\lambda_2 e^{-\lambda_1-\alpha_2\lambda_2}\lambda_3 e^{-\alpha_3\lambda_3}\left(1-\lambda_3\alpha_3\right) \\ &= -\lambda_2\lambda_3 e^{-\lambda_1-\alpha_2\lambda_2-\alpha_3\lambda_3}\left(1-\alpha_3\lambda_3\right). \end{aligned} \tag{23}$$

Hypothesis (19) is equal to 0; we can get

$$\alpha_2 = \frac{1-\alpha_3\lambda_3 e^{-\alpha_3\lambda_3}}{\lambda_2}. \tag{24}$$

Hypothesis (20) is equal to 0; we can get

$$\alpha_3 = \frac{1}{\lambda_3}. \tag{25}$$

Substitute (25) into (24); we can get:

$$\alpha_2 = \frac{1-e^{-1}}{\lambda_2}. \tag{26}$$

Substitute (25) and (26) into (21), (22), and (23), respectively; we can get

$$\begin{aligned} P_{21} &= -\lambda_2^2 e^{-\lambda_1-1+e^{-1}}\left(2-1+e^{-1}-e^{-1}\right) = -\lambda_2^2 e^{-\lambda_1-1+e^{-1}}, \\ P_{22} &= -\lambda_3^2 e^{-\lambda_1-1+e^{-1}-1} = -\lambda_3^2 e^{-\lambda_1+e^{-1}-2}, \\ P_{23} &= -\lambda_2\lambda_3 e^{-\lambda_1-1+e^{-1}-1}\left(1-\alpha_3\lambda_3\right) = 0. \end{aligned} \tag{27}$$

According to (27), we can get

$$P_{21} * P_{22} - P_{23}^2 > 0, \quad P_{21} < 0, \tag{28}$$

so P_{su} can be the maximum value when $\alpha_2 = (1-e^{-1})/\lambda_2$ and $\alpha_3 = 1/\lambda_3$; that is,

$$\begin{aligned} P_{3,\max} &= \lambda_1 e^{-\lambda_1} + e^{-\lambda_1} * \left(1-\frac{1}{e}\right) * e^{-1+1/e} \\ &\quad + e^{-\lambda_1} * e^{-1+1/e} * e^{-1}. \end{aligned} \tag{29}$$

6. Simulation

6.1. Two Levels. If there are only H2H users, we set λ_H range from 0 to 200 and suppose that there are 50 preambles (logical channels), and then we can get Figure 8.

Figure 8 shows that the random access success probability increases as the number of users increases when the number of H2H users is within a certain range. However, as the number of H2H users increases, the random access success probability will reduce due to collision or other reasons, when the number of users reaches a certain number (the average arrival number of H2H users per second is more than 50 in Figure 8).

Figure 8: The average arrival number of H2H versus access probability.

Figure 9: The average arrival number of M2M versus access probability.

When multiple M2M users and H2H users are accessing we meanwhile set the average arrival number of H2H users at 50 per second and set the average arrival number of M2M users at a range from 0 to 5000 per second; that is to say, the average of arrival number of H2H users is 1 per channel per second. When setting different access rates of M2M users, we can get Figure 9.

From Figure 9, we can know, setting different access rates for M2M users, that the lower the access rate is, the higher the random access success probability is, when the number of users reaches a certain number (the average arrival number of M2M users is 50 per second in Figure 9); that is, controlling access rate can improve the access success probability in the same number of users.

Figure 10: The average arrival number of M2M versus collision probability.

According to $P_{2,\max}$ in Section 5.1, we can get the theoretical maximum probability of random access success: $P_{2,\max} = 1 * e^{-1} + e^{-2} \approx 0.5032147244 \approx 0.5032$. We can also know from Figure 9, when the average arrival number of M2M is 1000/s, that the maximum value of simulation is 0.5032, which is approximately equal to the theoretical value 0.5032147244. Since we assume that there are 50 logical channels, the arrival number of users should be divided by 50 per channel; that is, the arrival number of users is 1000/50 = 20. That is, it can get the maximum value when access rate is 1/20 = 0.05. The simulation result matches the theoretical result. Hence, we can come to a conclusion that we can get the maximum success probability of random access when $\alpha = 1/\lambda_M$.

When multiple M2M users and H2H users are accessing, setting the average arrival number of H2H users at 50 per second, allowing M2M users to access with a certain rate, we can meanwhile get Figure 10.

From Figure 10, we can know that the lower the access rate is, the higher the access success probability is, when the number of M2M users reaches a certain number (the average arrival number of M2M users is 1500 per second in Figure 10); that is to say, controlling the access rate can obviously reduce the collision probability of random access in the same number of users.

6.2. Three Levels. We also suppose that there are 50 logical channels here, the average arrival number of primary users is 50 per second, the average arrival number of secondary users is $0 \le \lambda_2 \le 600$, and the average arrival number of third users is $0 \le \lambda_3 \le 5000$.

When we do not control the access rate of the secondary and third users, we can get Figure 11. When we properly control the access rate of the secondary and third users (here values of 0.5 and 0.1, resp.), we can get Figure 12. When we control the access rate of the secondary and third users with a smaller value (here values of 0.1 and 0.05, resp.), we can get Figure 13.

Figure 11: Access rate 1.

Figure 12: Access rate of 0.5 and 0.1, respectively.

From Figure 11, we can know that, when we do not control the access rate of the secondary and third users, the success probability of random access increases with the number of users increasing when the number of them is small. When the number of two-level users is larger, channels become colliding due to multiple users accessing simultaneously; moreover, the success probability of random access declines sharply. When the number of the third users is constant (we assume it is 2000), the success probability of random access increases with the number of the secondary users increasing in a small range, and then it is down to a constant value.

The contrast between Figures 11 and 12 tells us, when we control the secondary and third users' access properly, that the success probability of random access can improve obviously.

According to Section 5.2, we can get the theoretical maximum success probability of random access: $P_{3,\max} = e^{-1} + e^{-1-1+1/e} * (1 - 1/e) + e^{-3+1/e} \approx 0.56339397532 \approx 0.5634$. In Figure 12, we set the theoretical access rate of the secondary and third users at 0.5 and 0.1, respectively. When

Figure 13: Access rate of 0.1 and 0.05, respectively.

the number of secondary and third users is 51.26/50 = 1.0252 and 633.2/50 = 12.664, respectively, the access rate of the secondary users is about 0.6 and that of the third users is about 0.08 according to (25) and (26) in Section 5.2, which are approximately equal to the simulation values 0.5 and 0.1. Hence, we can get the maximum probability 0.5634 when $\alpha_2 = (1 - e^{-1})/\lambda_2$ and $\alpha_3 = 1/\lambda_3$.

The contrast between Figures 11, 12, and 13 tells us that the smaller the access rate of the secondary and third users is, the larger the success probability of random access is in the same number of users.

7. Conclusion

It has become an inevitable trend for M2M accessing the LTE system with the development of the Internet of things technology. This paper first simply introduces the LTE and M2M technologies and the LTE random access process and then analyses the impact of the LTE network when a great number of users are accessing simultaneously. The proposed model can improve the random access success probability efficiently by controlling the access rate of different levels of users. Furthermore, it can optimize the system performance by setting optimal access rates for different levels of users. This will have guiding and practical significance for M2M communication in LTE system.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Fujian Province (2012J01252), Fujian Province University-Industry Cooperation of Major Science and Technology Project (2011H6008), National Natural Science Foundation (61072080), and the Development Project of Fujian Provincial Strategic Emerging Industries Technologies: Key Technologies in development of Next generation Integrated High Performance Gateway, and Fujian development and reform commission high-technical [2013]266.

References

[1] T. V. K. Chaitanya and E. G. Larsson, "Improving 3GPP-LTE uplink control signaling performance using complex-field coding," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 161–171, 2013.

[2] E. Dahlman, S. Parkvall, and J. Sköld, *3G Evolution- HSPA and LTE for Mobile Broadband*, Academic Press, 2nd edition, 2008.

[3] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 10–22, 2010.

[4] D. Astély, E. Dahlman, A. Furuskar et al., "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[5] 3rd Generation Partnership Project. 3GPP LTE-Advanced, http://www.3gpp.org/lte-advanced.

[6] H. Thomsen, N. K. Pratas, and C. Stefanovic, "Petar popovski: analysis of the LTE access reservation protocol for real-time traffic," CoRR Abs 1301.2427, 2013.

[7] D. P. Bertsekas and R. G. Gallager, *Data Networks*, Prentice-Hall, 1987.

[8] Physical Layer Procedures, 3GPP TS 36.213.

[9] J. Research, "Embedded Mobile & M2M Strategies," Tech. Rep., 2010.

[10] J. E. Wieselthier, A. Ephremides, and L. A. Michaels, "Exact analysis and performance evaluation of framed ALOHA with capture," *IEEE Transactions on Communications*, vol. 37, no. 2, pp. 125–137, 1989.

[11] C.-H. Wei, R.-G. Cheng, and S.-L. Tsao, "Modeling and estimation of one-shot random access for finite-user multichannel slotted aloha systems"," *IEEE Communications Letters*, vol. 16, no. 8, pp. 1196–1199,, 2012.

[12] L. Roberts, "Dynamic allocation of satellite capacity through packet reservation," in *Proceedings of the National Computer Conference and Exposition (AFIPS '73)*, pp. 711–716, June 1973.

[13] A. Meader and P. Rost, "The Challenge of M2M Communications for the Cellular Radio Access Network," EuroView, 2011.

[14] A. B. I. Research, "Cellular Machine-to-Machine (M2M) Markets," Tech. Rep., 2010.

[15] Z. Feng and X. F. Zhong, "Fast Retrial and Dynamic Access Control Algorithm for LTE-Advanced Based M2M Network," AICT, 2012.

[16] Y. J. Yao, Z. Y. Feng, and D. Miao, "Markov-based optimal access probability for dynamic spectrum access in cognitive radio networks," in *Proceedings of the IEEE 71st Vehicular Technology Conference (VTC '10)*, May 2010.

[17] Y. J. Yao, Z. Y. Feng, W. Li, and Y. Qian, "Dynamic spectrum access with QoS guarantee for wireless networks: a Markov approach," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, December 2010.

[18] M. Amirijoo, P. Frenger, F. Gunnarsson, J. Moe, and K. Zetterberg, "On self-optimization of the random access procedure in 3G long term evolution," in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management-Workshops (IM '09)*, pp. 177–184, June 2009.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 902574, 11 pages
http://dx.doi.org/10.1155/2013/902574

Research Article

Joint Access Control and Subchannel Allocation Scheme for Femtocell-Based M2M Network Using a Truthful Mechanism

Chengmei Li, Jianjun Wu, Ziqiang Feng, Xi Luan, and Haige Xiang

State Key Laboratory of Advanced Optical Communication Systems and Networks, Peking University, Beijing 100871, China

Correspondence should be addressed to Jianjun Wu; just@pku.edu.cn

Received 27 August 2013; Accepted 22 October 2013

Academic Editor: Yan Zhang

Copyright © 2013 Chengmei Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A joint access control and subchannel allocation (JACSA) scheme is proposed in this paper for femtocell-based machine-to-machine (M2M) network to provide better communication services. As short-range and cost-beneficial eNodeBs, femtocells can improve indoor coverage and data transmission, and can further be used for M2M communication. There are two challenges for femtocell network, access control and truth-telling. Femtocell machine-type communication devices (FMD) select femtocell access points (FAPs) according to the reported channel capacities, while the true values are private information of each femtocell. Therefore, selfish FAPs have incentive to report larger capacities to win greater opportunity to be selected. To solve the aforementioned two problems, a JACSA scheme based on the Arrow-d'Aspremont-Gerard-Varet (AGV) is proposed both in open access and hybrid access scenarios, and formula derivations are given. We prove that, compared with the optimal subchannel allocation (OSA) scheme, JACSA scheme has the feature of achieving near optimal performances with much lower computational complexity. Furthermore, we compare the allocation results for open access and hybrid access in the proposed JACSA scheme. Finally, simulations are performed, and the results verify the availability of our proposed scheme.

1. Introduction

Machine-to-machine (M2M) or machine-type communication (MTC) is an emerging technology that allows devices to communicate with MTC server or each other in absence of human intervention or interaction [1]. In the M2M networks, which have the characteristics of large device number and small data transmission, machine-type devices communicate with each other through wired or wireless connections. As smart home and smart grid become popular, such as home security sensing, lighting control, and the sensors of actuators, M2M communications have a widespread and promising coverage [2, 3].

Due to their commercial prospects and properties, M2M networks have been widely studied [4–7]. In [4], the authors briefly review the features of M2M services in the third generation (3G) long-term evolution (LTE) and the advanced (LTE-advanced) networks, and a variety of radio resource allocation schemes are proposed. In [5], focusing on the problem of heavy random access (RA) load caused by accommodating the huge population of M2M or MTC customers/devices in LTE networks, two methods are proposed to solve the RA congestion. Different from [4, 5], where the M2M communications exist in LTE networks, the work in [6] promotes the new paradigm of cognitive machine-to-machine (CM2M) communication, by exploiting cognitive radio technology in M2M communications, and proposes a coordination-based energy-efficient spectrum discovery scheme that can be used in smart grid neighborhood area networks, which is shown to significantly save energy consumption. For the M2M communications of smart meters, cognitive radio (CR) functions based on different multiobjective genetic algorithms are proposed to find out the tradeoff between power efficiency and spectrum efficiency in different operating environments [7].

Recently, there are a few emerging research interests focusing on the femtocell-based M2M networks [8, 9]. Femtocells are small, inexpensive, low-power base stations (BSs) for better indoor voice coverage and data reception. They connect to their own wired backhaul connections,

and thus can efficiently offload data traffic from macrocell networks, which appeals to operators [10, 11]. Due to the short transmission distance, femtocells require very low transmission power and thus extend battery life of machines. In addition, femtocells can ease the RA congestion problem if M2M traffic is shared among the femtocell base stations [8]. In [8], the authors propose the group-based time control mechanism to improve the network overload and delay performance in the femtocell-based MTC networks. In [9], the potential added values are analyzed, as well as challenges, in augmenting personal telehealth systems with MTC personal health devices with integrated 3GPP interfaces (GPRS, HSPA, LTE, etc.) and operating in a femtocell network environment.

Femtocells have many benefits both for users (e.g., M2M devices) and operators. However, these benefits are not easy to accomplish simultaneously due to many challenges. One of the major challenges in the femtocell network is the access control problem because the choice of access control mechanism directly influences the interference and performance of the network.

Since femtocells provide better coverage and higher data rates for indoor machines, owners of femtocells prefer the closed access mechanism in order to fully use the femtocell resources, while operators are more interested in the open access mechanism, because open access femtocells can help operators offload the data traffic from the macrocell network and thus assist in providing higher network throughput and outdoor capacity. Nevertheless, hybrid access, a tradeoff of the open access and closed access, is the most promising access mechanism which is likely to benefit both users and operators [12, 13].

Besides, there are some studies on resource allocation using game theory [14–16] and auction theory [17, 18]. Both theories are important mathematical tools that can effectively solve the complex interactions and cheating problems among rational players and guarantee fairness and truth-telling characters. They are widely used in wireless communication [19] and show great advantages in resource allocation problems [20, 21] and access control problems [22–24].

In [22], the paper considers the selfishness of femtocell owners and proposes a utility-aware refunding framework to motivate femtocell owners to adopt hybrid access through refunding. The framework is formulated and analyzed as a Stackelberg game, and the authors prove that maximum utility can be achieved at the unique Nash Equilibrium. In [23], the authors considered the scenario that femtocell owners rent spectrum from the macrocell and proposed a spectrum leasing framework based on game theory. They model the process as a three-stage Stackelberg game and obtain the optimal spectrum leasing price, spectrum leasing ratio, and open access ratio from the Nash equilibrium. The authors in [24] propose a reverse auction in access control problem, and they use Vickery-Clarke-Grove (VCG) mechanism [17] to motivate femtocell owners participate in the auction process and maximize social welfare. Nevertheless, the VCG mechanism cannot implement the budget balance of the network, which means extra cost needs to be paid by the players in the game, and thus, the total welfare is impaired. Compared with the VCG mechanism, the AGV mechanism [17, 25] is an incentive efficient mechanism that can maximize the expected total payment of all the players in the game and achieve budget balance.

In this paper, we consider a femtocell based on M2M network with OFDMA techniques and assume all the subchannels are orthogonal and using the same spectrum. FAPs need to report their channel capacity to MTC devices (FMDs), and these devices select FAPs according to the reported information and pay for the femtocell resource they use. The channel capacity is private information, so femtocells do not know others' channel information. Since FAP cannot get payment from FMDs unless it is selected, each FAP has an incentive to report exaggerated channel capacity to get more chance to be selected, which may lead to cheating and vicious competition in the access control process and thus decrease the throughput of the network. Under these conditions, we propose a JACSA scheme to solve the problems. By adding a transfer payment to the total payment of each FAP, we prove that there is no incentive for FAPs to report exaggerated channel information. Each femtocell can get its maximum expected total payment when it reports the true information and any cheating behavior leads to a decrease in its expected total payment. Moreover, the subchannel allocation solution based on the truthful mechanism contains two steps. Firstly, FMDs select FAPs according to their reported information. Then all FAPs allocate their subchannels to the linked FMDs aiming at maximizing their own payment. The truth-telling mechanism results in that each FAP reports its real subchannel capacity, which ensures one feature that near optimal performance can be achieved. The subchannel allocation method, with which each FAP (FUE) only chooses FUEs (FAPs) with the largest capacities, decides another feature of low computational complexity of this scheme. Further analysis proves effectiveness of the truthful mechanism and shows the near optimal performance of the subchannel allocation solution with low computational complexity.

This paper is organized as follows. In Section 2, the system model of open access and hybrid access femtocell-based M2M network is described. Then the access control and resource allocation problem are formulated. In Section 3, we propose the JACSA scheme and prove the verity of AGV mechanism in both open and hybrid access scenarios. The simulation results of open access scenario in JACSA scheme versus OSA scheme, hybrid access scenario in JACSA scheme versus OSA scheme, and open access scenario versus hybrid scenario in JACSA scheme are shown in Section 4. Finally, conclusions are drawn in Section 5.

2. System Model

Access control mechanisms in femtocell networks could be classed into three categories [12, 26]. (1) *Closed access*: in this scenario, only registered user equipments (UEs), which are defined as closed subscriber group (CSG) in the third generation partnership project (3GPP), are able to connect to femtocells. (2) *Open access*: in this scenario, all public access to femtocells is allowed. (3) *Hybrid access*: in this scenario,

Figure 1: System model of the OFDMA femtocell network.

there is a mixing of open access and closed access modes. In this paper, we consider a two-tier femtocell-based M2M network with the open access and hybrid access mechanism and compare these two situations.

(1) *Open access*: the system consists of one macrocell base station (MBS), N_O FAPs denoted by $\{F_1, F_2, \ldots, F_{N_O}\}$ which provide services to all M_O FMDs denoted by $\{U_1, U_2, \ldots, U_{M_O}\}$.

(2) *Hybrid access*: besides the FAPs and FMDs described in the open access scenario (which we call OFAP and OFMD for short), there are N_C additional FAPs denoted by $\{F_{N_O+1}, F_{N_O+2}, \ldots, F_{N_O+N_C}\}$ and M_C FMDs in CSGs denoted by $\{M_{M_O+1}, M_{M_O+2}, \ldots, M_{M_O+M_C}\}$ (which we name as CFAP and CFMD), with each FAP F_{N_O+n} providing service only to M_{C_n} FMDs denoted by $\{U_{M^{n-1}+1}, U_{M^{n-1}+2}, \ldots, U_{M^n}\}$ with $M^n = M_O + \sum_{i=1}^{n} M_{C_i}$. This means that only FMDs belonging to the CSG of corresponding FAP have authorization for this access point. The total numbers of FAPs and FMDs are $N = N_O + N_C$ and $M = M_O + M_C$, respectively.

For hybrid access scenario, there exist three cases:

(1) *OO case*: OFAPs F_n $(n = 1, 2, \ldots, N_O)$ report capacities to OFMDs U_m $(m = 1, 2, \ldots, M_O)$, and each OFMD may choose one OFAP to access.

(2) *OC case*: OFAPs F_n $(n = 1, 2, \ldots, N_O)$ report capacities to CFMDs U_m $(m = M_O + 1, M_O + 2, \ldots, M)$.

(3) *CC case*: Each CFAP F_n $(n = N_O + 1, N_O + 2, \ldots, N)$ reports capacities to CFMDs U_m $(m = M^{n-1} + 1, M^{n-1}+2, \ldots, M^n)$ in its own CSG. Each CFMD may choose either one OFAP or one CFAP to access.

The open access scenario can be regarded as a special case of the hybrid access scenario with $N_C = 0$ and $M_C = 0$. For simplicity, we only introduce the hybrid access scenario below and consider the open access scenario as its special case.

The interference analysis model is shown in Figure 1. M2M communications have the characteristics of dealing with a huge number of devices. A large amount of MTC device will increase congestion in the network. As a result, only part of MTC devices can access the network. Here we only discusses the truth-telling problem of MTC devices which are able to access the network. The interference exists when the MBS and all FAPs share the same K subchannels with equal bandwidth of W. Here we consider the worst condition that interference exists in all the subchannels. The transmission power of MBS and FAP (e.g., F_n) over the K subchannels can be denoted by vectors $P = \{P_1, P_2, \ldots, P_K\}$ and $P_n = \{P_{n,1}, P_{n,2}, \ldots, P_{n,K}\}$ [27]. For additive white gaussian noise (AWGN) with variance of σ^2, the signal to interference plus noise ratio (SINR) of F_n to U_m in subchannel k is

$$\mathrm{SINR}_{n,m,k} = \frac{P_{n,k}\left|h_{n,m,k}\right|^2}{\sum_{i=1, i\neq n}^{N} P_{i,k}\left|h_{i,m,k}\right|^2 + P_k\left|h_{m,k}\right|^2 + \sigma^2}, \quad (1)$$

with $h_{n,m,k}$ and $h_{m,k}$ the channel frequency responses of the kth subchannel from F_n to U_m and from MBS to U_m, respectively, where the path loss and Rayleigh fading factor are contained. The channel capacity of F_n to U_m in subchannel k is

$$C_{n,m,k} = W \log\left(1 + \mathrm{SINR}_{n,m,k}\right). \quad (2)$$

During the access control and subchannel allocation processes, the FAPs first report to FMDs their channel state information (CSI), which may be false information. According to the quality of CSIs, FMDs select FAPs by comparing the capacities they report, where we assume that each FMD can only access one of the FAPs to enjoy the service. Based on the true subchannel information, FAPs allocate the capacities of all subchannels to one or more applying FMD(s) to maximum their profits, and each FMD pays for the sum of the subchannel capacities provided by the FAP. Therefore, the sum of the subchannel capacities of F_n used by U_m can be written as

$$C_{n,m} = \sum_{k\in\kappa_{n,m}} C_{n,m,k}, \quad (3)$$

where $\kappa_{n,m}$ is the set of subchannels of F_n allocated to U_m. The throughput of the network is given by

$$C = \sum_{n=1}^{N_O}\sum_{m=1}^{M_O} C_{n,m} + \sum_{n=1}^{N_O}\sum_{m=M_O+1}^{M} C_{n,m} + \sum_{n=N_O+1}^{N}\sum_{m=M^{n-1}+1}^{M^n} C_{n,m}. \quad (4)$$

The three terms on the right side of (4) represent the throughput of OO, OC, and CC cases, respectively. The main problem here is to maximize the throughput of the network.

In order to solve the access control and subchannel allocation problem, we need to construct a mathematical model to describe it. During the accessing process of femtocell

network, each FAP aims at increasing its own revenues by providing larger capacities, while the FMDs prefer better communication quality by selecting FAPs with larger capacities. However, the information shared by FMDs and FAPs is unbalanced, and the FMDs have incomplete information: they do not know either the real channel capacity of the each FAP or the real capacity of each subchannel. They have no other choice but to trust the FAPs of their reported information, and can only select the FAPs according to their reported capacities. For the FMDs, to select the FAPs with largest reported average capacities is always a good idea. After all, FAPs with larger average capacities are more probably to provide subchannels with larger capacities. Furthermore, the FMD can guarantee that it has larger opportunity to attain a subchannel when it select to attach the FAP with largest capacity rather than a smaller one, since the FAP always grants its subchannels to those FMDs with largest capacities. In a simple scheme, the FAPs are paid by FMDs only when they are selected. If no FMD selects the FAP, it will get nothing. This result may drive the FAPs to report better channel information to FMDs in order to win greater opportunity to be selected, causing unfairness in subchannel allocation and reduction of the throughput of the network

$$\widehat{C} \leq \widetilde{C}, \tag{5}$$

with $\widehat{C}$ and $\widetilde{C}$ representing the throughput of the network calculated according to the reported and real information, respectively. Considering the maximization of the throughput, the FAPs should all report true information.

When FMDs select FAPs according to their reported information, the truth-telling problem needs to be solved effectively to maximize the throughput of the network.

3. Joint Access Control and Subchannel Allocation Scheme

To formulate the problem above, we denote the real and the reported subchannel capacities of F_n to U_m as $(\widetilde{C}_{n,m,1}, \widetilde{C}_{n,m,2}, \ldots, \widetilde{C}_{n,m,K})$ and $(\widehat{C}_{n,m,1}, \widehat{C}_{n,m,2}, \ldots, \widehat{C}_{n,m,K})$, respectively. Similarly, the averaged real and reported subchannel capacities of F_n to U_m can be expressed as $\widetilde{\mathscr{C}}_{n,m}$ and $\widehat{\mathscr{C}}_{n,m}$, respectively. The FMD pays ξ to FAP per unit channel capacity, and the revenue of F_n from U_m can be expressed as

$$R_{n,m} = \begin{cases} \sum_{k \in \kappa_{n,m}} \xi \widetilde{C}_{n,m,k}, & F_n \text{ is selected by } U_m, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The total revenue of F_n can be expressed as

$$R_n = \begin{cases} \sum_{m=1}^{M} R_{n,m}, & \text{for OO, OC cases,} \\ \sum_{m=M^{n-1}+1}^{M^n} R_{n,m}, & \text{for CC case.} \end{cases} \tag{7}$$

Here we consider F_n and U_m as an example and neglect whether they are OFAP (OFMD) or CFAP (CFMD), which does not influence the following discussion. The real capacity of each subchannel for any FAP obeys a certain probability density function (PDF) defined as $f(\widetilde{C}_{n,m,k})$, and the real average subchannel capacity obeys the PDF which is expressed as $p(\widetilde{\mathscr{C}}_{n,m})$, which is known by all FAPs. The information enjoyed by F_n, $\overline{F}_n$, U_m, and $\overline{U}_m$ (where $\overline{F}_n$ represents any one in the sets of all FAPs except F_n, and $\overline{U}_m$ in the sets of all FMDs except U_m) is

(1) F_n enjoys its real channel capacities to U_m $(\widetilde{C}_{n,m,1}, \widetilde{C}_{n,m,2}, \ldots, \widetilde{C}_{n,m,K})$ as private information belonging to its own, just the same situation as $\overline{F}_n$. F_n reports its channel capacities to U_m as $(\widehat{C}_{n,m,1}, \widehat{C}_{n,m,2}, \ldots, \widehat{C}_{n,m,K})$, which is not known to $\overline{F}_n$. Besides, F_n knows well about the PDF of the real capacities of all FAPs as $\overline{F}_n$.

(2) F_n does not know either real or reported channel capacities of $\overline{F}_n$, which is also private information belonging to $\overline{F}_n$ its own.

(3) U_m only knows the capacities all FAPs have reported to it and treats them as real capacities, according to which it selects one FAP to access. It does not know either reported or real capacities of FAPs to $\overline{U}_m$.

(4) For $\overline{U}_m$, the situation is the same as U_m.

Based on the PDF of capacities, F_n can calculate its expected revenue from U_m as

$$\mathscr{R}_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right) = \xi \widetilde{\mathscr{C}}_{n,m} P(n,m), \tag{8}$$

where

$$P(n,m) = \begin{cases} \left(\int_0^{\widehat{\mathscr{C}}_{n,m}} p(x)\mathrm{d}x\right)^{N_O - 1}, & \text{for OO case,} \\ \left(\int_0^{\widehat{\mathscr{C}}_{n,m}} p(x)\mathrm{d}x\right)^{N_O}, & \text{for OC, CC cases} \end{cases} \tag{9}$$

is the probability of F_n selected by U_m. We can easily know that when $\widehat{\mathscr{C}}_{n,m} \to \infty$, $P(n,m) \to 1$. As the expected revenue is monotone increasing function of the reported capacities, F_n has an incentive to report larger average subchannel capacity. More important, if F_n thinks that all FAPs have the same motivation, it will get nothing if it does not report ∞. In this situation all information received by FMDs is false, making all FMDs blind on selecting the FAPs.

In order to solve the above problem, a truthful mechanism AGV in JACSA scheme can be employed to prevent FAPs reporting false information, when each FAP gets its maximum expected total revenue when it reports the real capacities. We will introduce the AGV mechanism in the next chapter.

Here we assume that all FAPs would report their real capacities to FMDs, and the next problem to be solved is how the FAPs allocate their subchannels among linked FMDs to maximize the throughput. Global optimum can not be

reached, since each FMD only knows its own choices of FAPs, and it can not learn other FMDs' choices, and those FMDs who select the same FAP may compete for subchannel resources. For the optimal subchannel allocation (OSA) scheme which maximizing the throughput of the whole network, all subchannel information (which, of course, should be the true values) and allocation probabilities should be concerned, and the complexity of the algorithm is $O(N^M K)$. In a rapidly changing wireless communication environment, this process is too time-consuming and cannot be achieved. Besides, as all FAPs may be so selfish that they only care about maximize their own throughput, an additional entity must be included to allocate the subchannels globally, which raises the total cost of the network. With the JACSA scheme, however, the complexity is only $O(N^M K)$, which is much lower than the OSA scheme, especially when N and M are very large. From the following simulation results, we can see that, in the current JACSA scheme, as long as each FAP maximizs its own utility, the maximum throughput of the whole network can be well approached with a percentage of about 90%. The whole process of JACSA scheme is described as follows.

(i) Based on the truth-telling mechanism, all FAPs report real average capacities to FMDs. OFAP F_n reports $\widetilde{\mathscr{C}}_{n,m}$ to any FMD U_m ($n = 1, 2, \ldots, N_O$ and $m = 1, 2, \ldots, M$), and CFAP $F_{n'}$ reports $\widetilde{\mathscr{C}}_{n',m'}$ to CFMD $U_{m'}$ ($n' = N_O + 1, N_O + 2, \ldots, N$ and $m' = M^{n'-1} + 1, M^{n'-1} + 2, \ldots, M^{n'}$) in its CSG.

(ii) According to the reported information, FMDs select one FAP with the largest average capacity to access. OFMD can only access N_O OFAPs, while CFMD can access N_O OFAPs and one additional CFAP.

(iii) According to the accessing information, FAPs allocate their subchannel resources among FMDs who select them. For each subchannel, FAP will allocate it to the FMD with the most capacity to maximize its throughput. OFAPs will allocate their subchannels among all FMDs, while CFAPs only allocate among CFMDs in their CSGs.

(iv) Subchannel allocation finished. FMDs will access the FAPs who have allocated resources to them.

4. AGV Mechanism

Truth-telling is achieved by adding a transfer payment to the expected revenue $\mathscr{R}_{n,m}$, which is different for the OO, OC, and CC case.

(1) For OO case, the transfer payment is

$$\begin{aligned}\mathscr{T}_{n,m}&\left(\widehat{\mathscr{C}}_{1,m}, \ldots, \widehat{\mathscr{C}}_{N_O,m}\right)\\ &= \Gamma_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right) - \frac{1}{N_O - 1}\sum_{i=1, i\neq n}^{N_O} \Gamma_{i,m}\left(\widehat{\mathscr{C}}_{i,m}\right),\end{aligned} \tag{10}$$

where the externality

$$\Gamma_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right) = \sum_{i=1, i\neq n}^{N_O} E\left[\mathscr{R}_{i,m}\left(\widehat{\mathscr{C}}_{n,m}\right)\right] \tag{11}$$

represents the sum of other FAPs' expected payment from U_m when F_n has reported $\widehat{\mathscr{C}}_{n,m}$.

(2) For OC case, the transfer payment is

$$\begin{aligned}\mathscr{T}_{n,m}&\left(\widehat{\mathscr{C}}_{1,m}, \ldots, \widehat{\mathscr{C}}_{N_O,m}, \widehat{\mathscr{C}}_{N_O+q,m}\right)\\ &= \Gamma_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right)\\ &\quad - \frac{1}{N_O}\left[\sum_{i=1, i\neq n}^{N_O} \Gamma_{i,m}\left(\widehat{\mathscr{C}}_{i,m}\right) + \Gamma_{N_O+q,m}\left(\widehat{\mathscr{C}}_{N_O+q,m}\right)\right],\end{aligned} \tag{12}$$

where the externality is

$$\begin{aligned}\Gamma_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right) &= \sum_{i=1, i\neq n}^{N_O} E\left[\mathscr{R}_{i,m}\left(\widehat{\mathscr{C}}_{n,m}\right)\right]\\ &\quad + E\left[\mathscr{R}_{N_O+q,m}\left(\widehat{\mathscr{C}}_{n,m}\right)\right],\\ \Gamma_{N_O+q,m}\left(\widehat{\mathscr{C}}_{N_O+q,m}\right) &= \sum_{i=1}^{N_O} E\left[\mathscr{R}_{i,m}\left(\widehat{\mathscr{C}}_{N_O+q,m}\right)\right].\end{aligned} \tag{13}$$

(3) For CC case, the transfer payment is

$$\begin{aligned}\mathscr{T}_{N_O+q,m}&\left(\widehat{\mathscr{C}}_{1,m}, \ldots, \widehat{\mathscr{C}}_{N_O,m}, \widehat{\mathscr{C}}_{N_O+q,m}\right)\\ &= \Gamma_{N_O+q,m}\left(\widehat{\mathscr{C}}_{N_O+q,m}\right) - \frac{1}{N_O}\sum_{i=1}^{N_O} \Gamma_{i,m}\left(\widehat{\mathscr{C}}_{i,m}\right),\end{aligned} \tag{14}$$

where the externality is the same as OC case.

The expected total payment or the expected utility of F_n from U_m is

$$\mathscr{U}_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right) = \mathscr{R}_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right) + \mathscr{T}_{n,m}. \tag{15}$$

The existence of transfer payment can effectively prevent the FAPs from reporting false information. The FAPs who report larger capacities than their real values can be punished, and the loss of any FAP's revenue caused by exaggeratedly reported capacities of other FAPs can be compensated. If F_n reports its subchannel capacities with larger values ($\widehat{\mathscr{C}}_{n,m} > \widetilde{\mathscr{C}}_{n,m}$), it has more chance to be selected by U_m, but the transfer payment is even larger, decreasing its expected total revenue. On the contrary, if F_n reports its true values but one of the other FAPs reports larger capacities than its real values and is selected by FMDs instead of F_n, it will have to pay F_n some transfer payment. On the other hand, if F_n reports smaller average capacities ($\widehat{\mathscr{C}}_{n,m} < \widetilde{\mathscr{C}}_{n,m}$), it has smaller probability to be selected by U_m, and the transfer payment from other FAPs does not compensate the reduction, making its expected utility decreasing.

Proposition 1. *Each FAP maximizes its expected utility from FMDs only when it reports its real information*

$$\widehat{\mathscr{C}}_{n,m} = \widetilde{\mathscr{C}}_{n,m}. \tag{16}$$

Proof. Without loss of generality, we consider the expected utility of F_n paid by U_m, which can be expressed as follows.

(1) For OO case

$$\begin{aligned} &E\left[\mathscr{U}_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] \\ &\quad = E\left[\mathscr{R}_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right) + \mathscr{T}_{n,m}\left(\widehat{\mathscr{C}}_{1,m},\ldots,\widehat{\mathscr{C}}_{N_O,m}\right)\right]; \end{aligned} \tag{17}$$

inserting (10) into the above equation, we get

$$\begin{aligned} E\left[\mathscr{U}_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] &= E\left[\mathscr{R}_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] \\ &\quad + E\left[\sum_{i=1,i\neq n}^{N_O} \mathscr{R}_{i,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] \\ &\quad - \frac{1}{N_O-1}\sum_{j=1,j\neq n}^{N_O} \Gamma_{j,m}\left(\widehat{\mathscr{C}}_{j,m}\right) \\ &= E\left[\sum_{i=1}^{N_O} \mathscr{R}_{i,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] \\ &\quad - \frac{1}{N_O-1}\sum_{j=1,j\neq n}^{N} \Gamma_{j,m}\left(\widehat{\mathscr{C}}_{j,m}\right). \end{aligned} \tag{18}$$

(2) For OC case

$$\begin{aligned} &E\left[\mathscr{U}_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] \\ &\quad = E\left[\mathscr{R}_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right) + \mathscr{T}_{n,m}\left(\widehat{\mathscr{C}}_{1,m},\ldots,\widehat{\mathscr{C}}_{N_O,m},\widehat{\mathscr{C}}_{N_O+q,m}\right)\right]; \end{aligned} \tag{19}$$

inserting (12) into the above equation, we get

$$\begin{aligned} &E\left[\mathscr{U}_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] \\ &\quad = E\left[\sum_{i=1}^{N_O} \mathscr{R}_{i,m}\left(\widetilde{\mathscr{C}}_{n,m}\right) + \mathscr{R}_{N_O+q,m}\left(\widetilde{\mathscr{C}}_{n,m}\right)\right] \\ &\quad - \frac{1}{N_O}\left[\sum_{j=1,j\neq n}^{N} \Gamma_{j,m}\left(\widehat{\mathscr{C}}_{j,m}\right) + \Gamma_{N_O+q,m}\left(\widehat{\mathscr{C}}_{N_O+q,m}\right)\right]. \end{aligned} \tag{20}$$

(3) For CC case

$$\begin{aligned} &E\left[\mathscr{U}_{N_O+q,m}\left(\widetilde{\mathscr{C}}_{N_O+q,m}\right)\right] \\ &\quad = E\left[\mathscr{R}_{N_O+q,m}\left(\widetilde{\mathscr{C}}_{N_O+q,m}\right)\right. \\ &\qquad \left. + \mathscr{T}_{N_O+q,m}\left(\widehat{\mathscr{C}}_{1,m},\ldots,\widehat{\mathscr{C}}_{N_O,m},\widetilde{\mathscr{C}}_{N_O+q,m}\right)\right], \end{aligned} \tag{21}$$

inserting (14) into the above equation, we get

$$\begin{aligned} &E\left[\mathscr{U}_{N_O+q,m}\left(\widetilde{\mathscr{C}}_{N_O+q,m}\right)\right] \\ &\quad = E\left[\sum_{i=1}^{N_O} \mathscr{R}_{i,m}\left(\widetilde{\mathscr{C}}_{N_O+q,m}\right) + \mathscr{R}_{N_O+q,m}\left(\widetilde{\mathscr{C}}_{N_O+q,m}\right)\right] \\ &\quad - \frac{1}{N_O}\left[\sum_{j=1,j\neq n}^{N} \Gamma_{j,m}\left(\widehat{\mathscr{C}}_{j,m}\right)\right]. \end{aligned} \tag{22}$$

The terms on the right side of (18), (20), and (22) represent the expected revenue of all FAPs paid by U_m when F_n reports $\widetilde{\mathscr{C}}_{n,m}$ to U_m and the average externality of other FAPs except F_n, which is independent on $\widetilde{\mathscr{C}}_{n,m}$ to U_m. Therefore, the first term describing the expected total revenue of all FAPs determines the expected utility of F_n from U_m. Note that according to (8), the expected total revenue is decided by the reported capacities which may not be true. Besides, we easily know that if there exist FAPs who cheat about their capacities, the selections of FMDs may be misled, leading to decreasing of the expected total revenue. Therefore, only when all FMDs select FAPs according to the real capacities, the whole network can get the maximum total revenue. The maximum of the whole network also means that the utility of F_n from U_m reaches its maximum value. The coincidence between the global maximum and the individual optimum ensures that F_n has no incentive to report false information and the equilibrium can be achieved.

□

Proposition 2. *The network pays no extra costs for the truthful mechanism.*

Proof. From the transfer function equations (10), (12), and (5) we can obtain the sum of the transfer payment of all FAPs with U_m ($\forall m$) as follows:

(1) for open FMDs (OO case)

$$\begin{aligned} &\sum_{n=1}^{N_O} \mathscr{T}_{n,m}\left(\widehat{\mathscr{C}}_{1,m},\ldots,\widehat{\mathscr{C}}_{N_O,m}\right) \\ &\quad = \sum_{n=1}^{N_O} \Gamma_{n,m}\left(\widetilde{\mathscr{C}}_{n,m}\right) - \frac{1}{N_O-1}\sum_{n=1}^{N_O}\sum_{i=1}^{N_O} \Gamma_{i,m}\left(\widehat{\mathscr{C}}_{i,m}\right) \\ &\qquad + \frac{1}{N_O-1}\sum_{j=1}^{N_O} \Gamma_{j,m}\left(\widehat{\mathscr{C}}_{j,m}\right) \\ &\quad = \frac{N_O}{N_O-1}\sum_{j=1}^{N_O} \Gamma_{j,m}\left(\widehat{\mathscr{C}}_{j,m}\right) - \frac{N_O}{N_O-1}\sum_{i=1}^{N} \Gamma_{i,m}\left(\widehat{C}_{i,m}\right) \\ &\quad = 0. \end{aligned} \tag{23}$$

(2) For closed FMDs (OC and CC case)

$$\begin{aligned}
&\sum_{n=1}^{N_O} \mathcal{T}_{n,m}\left(\widehat{\mathscr{C}}_{1,m},\ldots,\widehat{\mathscr{C}}_{N_O,m},\widehat{\mathscr{C}}_{N_O+q,m}\right)+\mathcal{T}_{N_O+q,m}\\
&\quad=\sum_{n=1}^{N_O}\Gamma_{n,m}\left(\widehat{\mathscr{C}}_{n,m}\right)-\frac{1}{N_O}\sum_{n=1}^{N_O}\sum_{i=1}^{N_O}\Gamma_{i,m}\left(\widehat{\mathscr{C}}_{i,m}\right)\\
&\qquad+\frac{1}{N_O}\sum_{j=1}^{N_O}\Gamma_{j,m}\left(\widehat{\mathscr{C}}_{j,m}\right)-\frac{1}{N_O}\sum_{n=1}^{N_O}\mathcal{T}_{N_O+q,m}\left(\widehat{\mathscr{C}}_{N_O+q,m}\right)\\
&\qquad+\mathcal{T}_{N_O+q,m}\\
&\quad=0.
\end{aligned} \tag{24}$$

The total transfer payment of all FAPs with U_m equals zero. Therefore, the total transfer payment of the whole network also equals zero. □

5. Simulation Results

Based on the analysis in previous sections, simulation results are given to evaluate the performance of the proposed JACSA scheme. Generally, the reported subchannel capacity of each FAP is assumed to obey the exponential distribution described by PDF $f(x) = e^{-x}$, and thus the average subchannel capacity of each FAP follows Erlang distribution with PDF $p(x) = K^K x^{K-1} e^{-Kx}/(K-1)!$. As in our scheme, the FMDs deal with the information got from the FAPs in an easy way: they only care about the average capacities reported by FAPs instead of detailed subchannel capacities. For simplicity, we assume that each FAP calculates its own average subchannel capacity and reports a value to each FMD, from which FMD selects FAP. We consider two scenarios here, the open access and the hybrid access. Now we will give the simulation results of verification of the truth-telling mechanism, the open access scenario in JACSA scheme versus OSA scheme, the hybrid access scenario in JACSA scheme versus OSA scheme, and the open access scenario versus hybrid access scenario in JACSA scheme, respectively.

Firstly, we will focus on verification of truth-telling mechanism. For the open scenario, we investigate the performance of the truth-telling mechanism and consider a femtocell network with the settings of $M_O = 5$, $N_O = 2$, and $K = 8$. We also assume that the price per unit of channel capacity $\xi = 1$ and a random sample of the real average channel capacity is obtained as

$$\mathscr{C}_{\text{open}} = \begin{pmatrix} \widetilde{\mathscr{C}}_{1,1} & \widetilde{\mathscr{C}}_{2,1} \\ \widetilde{\mathscr{C}}_{1,2} & \widetilde{\mathscr{C}}_{2,2} \\ \widetilde{\mathscr{C}}_{1,3} & \widetilde{\mathscr{C}}_{2,3} \\ \widetilde{\mathscr{C}}_{1,4} & \widetilde{\mathscr{C}}_{2,4} \\ \widetilde{\mathscr{C}}_{1,5} & \widetilde{\mathscr{C}}_{2,5} \end{pmatrix} = \begin{pmatrix} 1.00 & 0.85 \\ 0.82 & 0.84 \\ 1.26 & 0.58 \\ 1.07 & 0.71 \\ 0.97 & 1.04 \end{pmatrix}. \tag{25}$$

For the hybrid access scenario, besides the channel capacities above, the open and closed FAPs also report to the closed FMDs. We consider that $N_C = 1$, $M_C = M_{C_1} = 2$, and $K = 8$. The channel capacity reported to closed FMDs is then

$$\mathscr{C}_{\text{closed}} = \begin{pmatrix} \widetilde{\mathscr{C}}_{1,6} & \widetilde{\mathscr{C}}_{2,6} & \widetilde{\mathscr{C}}_{3,6} \\ \widetilde{\mathscr{C}}_{1,7} & \widetilde{\mathscr{C}}_{2,7} & \widetilde{\mathscr{C}}_{3,7} \end{pmatrix} = \begin{pmatrix} 0.68 & 0.95 & 1.11 \\ 1.22 & 1.49 & 0.54 \end{pmatrix}. \tag{26}$$

Figure 2: Expected total payment when different average subchannel capacities are reported to U_1.

Without loss of generality, we consider U_4 as an example for the open access scenario and U_6 for the hybrid access scenario.

In Figure 2, we demonstrate the variation of F_n's expected total payment paid by the open FMD U_4 ($n = 1, 2$) and closed FMD U_6 ($n = 1, 2, 3$), when the reported information changes in the proposed mechanism. Given that the other nodes are honest, we find that F_n gets its maximum expected total payment only when it reports the true information, which is proved in Proposition 1. There is no incentive for FAPs to report exaggerated information.

In Figure 2, in the hybrid access scenario, the open FAP F_1, F_2 and closed FAP F_3 get their largest expected total payments, whose values are 0.3008, 0.2471, and 0.1977, respectively, at subchannel capacities at $\widetilde{\mathscr{C}}_{1,6} = 0.55$, $\widetilde{\mathscr{C}}_{2,6} = 0.96$, and $\mathscr{C}_{3,6} = 1.13$ for FMD U_6, respectively. These capacities are very close to their true capacities, whose values are 0.68, 0.95, and 1.11, respectively. The expected total payments are 0.2993, 0.2429, and 0.1956, respectively, which differ from their maximum values of less than 0.5%. These derivations from their true values are due to the fluctuations of the calculated expected total payments for limited simulation times. Among all FAPs, F_3, which has the largest true capacity, will be selected by U_6. The expected total payment is less than $\xi\widetilde{\mathscr{C}}_{1,6} = 1.13$ because it needs to pay transfer payment to other FAPs.

Figure 3 shows the expected transfer payment of F_n with open FMD U_4 ($n = 1, 2$) and closed FMD U_6

Figure 3: Expected transfer payment when different average subchannel capacities are reported to U_1.

Figure 4: Throughput comparison at different K.

Figure 5: Throughput comparison at different M_O.

($n = 1, 2, 3$). We can see that the transfer payment of each FAP monotonically decreases because the larger the reported average subchannel capacity is, the more transfer payment should be paid to others. Therefore, FAPs will not report exaggerate information if they take the transfer payment into consideration. We also find that the transfer payment curve of F_n with the largest real average subchannel capacity lies above other curves. This is because the expected payment is proportional to the real channel capacity (25) and (26). So the FAP with larger real channel capacity needs larger transfer payment to guarantee truth-telling. Furthermore, the transfer payment can be seen as a kind of tax.

When all FAPs report true information, the transfer payments of all FAPs F_n with U_6 are $\mathcal{T}_{1,6} = 0.2765$, $\mathcal{T}_{2,6} = 0.0148$, and $\mathcal{T}_{3,6} = -0.2906$, respectively. We can see that $\mathcal{T}_{1,6} + \mathcal{T}_{2,6} + \mathcal{T}_{3,6} = 6.5 \times 10^{-4} \approx 0$, an example of Proposition 2. Even after we consider the fluctuations in the limited simulation times, the transfer payments at the points with maximum expected transfer payments are 0.2966, 0.0081, and −0.3167, and their sum is −0.0120, which is of the order of 1% and is within a reasonable error range.

Second, we consider the performance of the JACSA scheme and OSA scheme for an open access scenario. We primarily focus on the allocation result and throughput of the network at different K and M_O. Primarily, we set $M_O = 5$ and $N_O = 3$ and change the value of K to investigate the influence of the number of subchannels. From Figure 4 we can see that when the number of subchannel (K) increases, the throughput of the network will increase at the same time. Because each FAP has more subchannels for allocation when K increases and more FMDs get the chance to obtain more channel resource than the scenario with lower K, which increase the throughput of the network. Besides, as K increases, the relative difference between the joint and optimal subchannel allocation is almost the same, and the JACSA scheme can give a throughput about 90% of the OSA scheme with a much lower complexity.

Then we set $N_O = 3$ and $K = 8$ and change the value of M_O to investigate the influence of FMDs. Figure 5 shows the throughput of the JACSA scheme and OSA scheme under the condition of different number of FMDs. As the number of open FMDs increases, the subchannel can be used more sufficiently, and thus the throughput of the system is larger.

Third, comparing the JACSA scheme with the OSA scheme for the hybrid access scenario. We set $N_O = 2$, $M_O = 5$, $N_C = 1$, and $M_C = M_{C_1} = 2$ and change the value of

Figure 6: Throughput comparison at different K.

Figure 7: Throughput comparison at different M_O.

K to investigate the influence of the number of subchannels. Figure 6 shows the throughput of all open and closed FAPs in JACSA and OSA scheme. Similar with the open access scenario, the total throughput does not change much with increasing of K. However, the throughput of all open FAPs is smaller for the OSA scheme than that for the JACSA scheme, while the throughput of all closed FAPs is larger. This means that the current JACSA scheme does not sufficiently use the closed FAPs.

Similarly, we consider the influence of the number of closed FMDs in the hybrid access scenario. Figure 7 shows the throughput of all open and closed FAPs in the JACSA and OSA schemes. Increasing of the number of closed FMDs means that both subchannels of the open and closed FAPs are used sufficiently and the throughput of both kinds of FAPs increases.

Finally, in the hybrid access scenario, due to the existence of the closed FAP, which is exclusive to closed FMDs, these closed FMDs have more access choices, and thus they probably have larger per user capacity. Compared with the situation where this FAP is open, however, there is less chance for access, thus total throughput decreases. First, we look from the perspective of the FAPs and calculate the total throughput of the system. Figure 8 shows the throughput of open and closed FAPs with $N_O = 2$ and $N_C = 1$, respectively, and a total number of FMDs $M_O + M_C = 8$ and transfers one FMD from open status to closed. In order to compare with the open access scenario, we also plot the $N_O = 3$ and $M_O = 8$ case. The horizontal axis represents the combinations of FAPs and FMDs. We can see that compared with open and hybrid access scenarios, the throughput is less in the latter, when the maximum value is equal to the throughput in the former. When there is no closed FMD in the hybrid, as there is one empty FAP which has no access, the total throughput is minimum. With the increasing of the closed FMDs, the total throughput and that for all closed FAPs

Figure 8: Throughput comparison with different combinations of FAPs and FMDs. The figure shows the number of open FAPs and open FMDs, and the total number of FAPs and FMDs satisfies $N_O + N_C = 3$ and $M_O + M_C = 8$.

increases, and the throughput for all open FAPs decreases slightly, as more FMDs may access closed FAPs. Therefore, no matter how many closed FMDs there are, total throughput is smaller in the hybrid access scenario, compared with open access.

From the perspective of FMDs, the average capacity is enjoyed by each FMD, as shown in Figure 9. When there is no closed FMD, as there are less open FAPs, the average capacity for open FMDs is smaller. If one FMD decides to join the CSG, the average capacity to be obtained is larger than that of an open FMD, and is also larger than that received by other

Figure 9: Average capacity per FMD enjoys with different combinations of FAPs and FMDs. The figure shows the number of open FAPs and open FMDs, and the total number of FAPs and FMDs satisfies $N_O + N_C = 3$ and $M_O + M_C = 8$.

open FMDs. As the closed FMDs probably access the closed FAP, the open FMDs have less competition and enjoy larger average capacity. As more FMDs join the CSG, as there are more competitions for the closed FAP, the average capacity decreases for each closed FMD but is still larger than that received by open FMDs and received with an open FMD. If all FMDs become closed, the situation is the same as in the open access scenario, and the capacity of each closed FMD is minimum. This indicates that, when closed FAPs exist, by becoming a closed FMD, the average capacity each FMD enjoys grows.

6. Conclusion

In this paper, we focus on the OFDMA femtocell-based M2M network with several femtocells and MTC devices distributed randomly in open access scenario and hybrid access scenario. In order to solve the truth-telling and subchannel allocation problems under these two scenarios, we propose a joint access control and subchannel allocation scheme based on the AGV mechanism. We use the transfer payment to balance the total payment and prove that there is only one equilibrium to be reached at the point when all FAPs reported their true capacities. By comparing the JACSA scheme with the OSA scheme in the two scenarios, we show that the JACSA scheme can obtain near optimal results with lower complexity. Finally, we compare the open access scenario with hybrid access scenario in the JACSA scheme, and the simulation results show that from the view of FAPs, an open access scenario is better to get more payments, but from the view of FMDs, when there are closed FAPs, it is better to join the CSG for better services.

Acknowledgments

This research was partly supported by the National Science Foundation of China (Grant no. NFSC #61071083, no. NFSC #61371073) and the National High Technology Research and Development Program of China (863 program, no. 2012AA01A506).

References

[1] 3GPP, "Service requirements for Machine-Type Communications (MTC)," Tech. Rep. TS 22.368 V11.2.0, 2011.

[2] M. Starsinic, "System architecture challenges in the home M2M network," in *Proceedings of the Long Island Systems, Applications and Technology Conference (LISAT '10)*, May 2010.

[3] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: architectures, standards, and QoS improvement," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, 2011.

[4] K. Zheng, F. Hu, and W. Wang, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184–192, 2012.

[5] K.-D. Lee, S. Kim, and B. Yi, "Throughput comparison of random access methods for M2M service over LTE networks," in *Proceedings of the IEEE GLOBECOM Workshops (GC Wkshps '11)*, pp. 373–377, December 2011.

[6] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive machine-to-machine communications: visions and potentials for the smart grid," *IEEE Network Magazine*, vol. 26, no. 3, pp. 6–13, 2012.

[7] Q. D. Vo, J.-P. Choi, H. M. Chang, and W. C. Lee, "Green perspective cognitive radio-based M2M communications for smart meters," in *Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC '10)*, pp. 382–383, November 2010.

[8] A.-H. Tsai and J.-H. Huang, "Overload control for machine type communications with femtocells," in *Proceedings of the IEEE Vehicular Technology Conference (VTC Fall '12)*, pp. 1–5, September 2012.

[9] E. Mutafungwa, "Applying MTC and femtocell technologies to the continua health reference architecture," in *Proceedings of the Grid and Pervasive Computing Workshops*, January 2012.

[10] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: past, present, and future," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 497–508, 2012.

[11] H. Claussen, L. T. W. Ho, and L. G. Samuel, "An overview of the femtocell concept," *Bell Labs Technical Journal*, vol. 13, no. 1, pp. 221–246, 2008.

[12] G. de la Roche, A. Valcarce, D. López-Pérez, and J. Zhang, "Access control mechanisms for femtocells," *IEEE Communications Magazine*, vol. 48, no. 1, pp. 33–39, 2010.

[13] J. Xiang, Y. Zhang, T. Skeie, and L. Xie, "Downlink spectrum sharing for cognitive radio femtocell networks," *IEEE Systems Journal*, vol. 4, no. 4, pp. 524–534, 2010.

[14] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, Cambridge, Mass, USA, 1991.

[15] J. Chen, R. Zhang, L. Song, Z. Han, and B. Jiao, "Joint relay and jammer selection for secure two-way relay networks," *IEEE*

Transactions on Information Forensics and Security, vol. 7, no. 1, pp. 310–320, 2012.

[16] C. Xu, L. Song, and Z. Han, *Resource Management for Device-to-Device Underlay Communication*, Springer Briefs in Computer Science, 2014.

[17] V. Krishna and Auction Theory Academic Press, 2002.

[18] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, and B. Jiao, "Efficiency resource allocation for device-to-device underlay communication systems: a reverse iterative combinatorial auction based approach," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 348–358, 2013.

[19] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models and Applications*, Cambridge University Press, 2011.

[20] Z. Han, Z. Ji, and K. J. R. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1366–1376, 2005.

[21] J. Deng, R. Q. Zhang, L. Y. Song, Z. Han, G. Yang, and B. L. Jiao, "Joint power control and subchannel allocation for OFDMA femtocell networks using distributed auction game," in *Proceedings of the International Conference on Wireless Communications & Signal Processing (WCSP '12)*, pp. 1–6, October 2012.

[22] Y. Chen, J. Zhang, and Q. Zhang, "Utility-aware refunding framework for hybrid access femtocell network," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1688–1697, 2012.

[23] Y. Yi, J. Zhang, Q. Zhang, and T. Jiang, "Spectrum leasing to femto service provider with hybrid access," in *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (IEEE INFOCOM '12)*, pp. 1215–1223, March 2012.

[24] Y. Chen, J. Zhang, Q. Zhang, and J. Jia, "A reverse auction framework for access permission transaction to promote hybrid access in femtocell network," in *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (IEEE INFOCOM '12)*, pp. 2761–2765, March 2012.

[25] J. Deng, R. Zhang, L. Song, Z. Han, and B. Jiao, "Truthful mechanisms for secure communication in wireless cooperative system," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4236–4245, 2013.

[26] A. Golaup, M. Mustapha, and L. B. Patanapongpibul, "Femtocell access control strategy in UMTS and LTE," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 117–123, 2009.

[27] P. Lee, T. Lee, J. Jeong, and J. Shin, "Interference management in LTE femtocell systems using fractional frequency reuse," in *Proceedings of the 12th International Conference on Advanced Communication Technology (ICACT '10)*, pp. 1047–1051, February 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 540849, 11 pages
http://dx.doi.org/10.1155/2013/540849

Research Article

Multiple Beam Selection for Combing M2M Communication Networks and Cellular Networks with Limited Feedback

Xiaoning Zhang,[1] Lin Bai,[2] and Wenyang Guan[1]

[1] *State Key Laboratory of Advanced Optical Communication Systems and Networks, Peking University, Beijing 100871, China*
[2] *School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Lin Bai; l.bai@buaa.edu.cn and Wenyang Guan; guanwenyang@pku.edu.cn

Received 18 August 2013; Accepted 19 October 2013

Academic Editor: Jianhua He

Copyright © 2013 Xiaoning Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study the scenario in which a large number of machine-type communication devices (MTCDs) communicate with each other by utilizing the help of the base station (BS) through some MTCD gateways. We consider an overlay mode of an orthogonal frequency division multiple access (OFDMA) based cellular system using orthogonal beamforming to provide broadband wireless access for the MTCD gateways. In order to avoid the interference with mobile users, the beamforming vectors to the MTCD gateways have to be orthogonal to the channel vectors of mobile users, which become a beamforming constraint for the MTCD gateways. However, with limited feedback of channel state information (CSI) at a BS, the orthogonal beamforming constraints may not be achieved. In such a practical case, conventional feedback schemes are feasible but not efficient due to the orthogonality constraints. In this paper, we propose a novel multiple beam selection (MBS) approach with limited feedback for MTCD gateways by taking into account the previous orthogonality constraints. Simulation results show that the performance improvement of the proposed approach over the conventional ones is generally about 10% when a BS is equipped with an array of 6 elements.

1. Introduction

Machine-to-machine (M2M) communication networks have been proposed to connect machines without the intervention from human beings. In recent years, the research of combing M2M networks and mobile cellular networks draws much attention [1, 2]. In an M2M network, there are lots of machine-type communication devices (MTCDs). These MTCDs access a cellular network by some MTCD gateways [3], which are assumed as secondary users in the cellular network. There are hundreds of MTCDs connecting to one MTCD gateway. In the combined M2M cellular networks, the MTCD gateway may need to communicate with the BS at a high data rate.

As the mobile users in a cellular system have a higher priority than the MTCD gateways, the beamforming vectors for the mobile users are designed to maximize the beamforming gains. In the overlay mode, since the signals transmitted to the MTCD gateways should not interfere with the mobile users [4], the beamforming vectors have to be orthogonal to the channel vectors to the mobile users, which become a constraint in finding the beamforming vectors for the MTCD gateways. The throughput performance can be improved when the orthogonal beamforming is combined with multiuser scheduling over multiple subcarriers, where the users with good channel conditions are selected for each transmission. Channel state information (CSI) is required in both beamforming and multiuser scheduling. However, unfortunately, in a practical system, the CSI available at the BS is imperfect due to specific system configurations. Therefore, it is worthwhile to design orthogonal beamforming using imperfect CSI for this kind of hybrid network.

1.1. Related Work. There are mainly two ways for the BS to get information from each user, that is, channel reciprocity and feedback channel. In a time division duplexing (TDD) system, it is possible to use the channel reciprocity at the BS, which refers to estimating the downlink channel from the uplink signal. However, the channel reciprocity requires calibration of the radio devices of both the receiver and the transmitter, for example, A/D converters, mixers, filters, and antennas. This requires extra special hardware as well as special protocols to facilitate the channel measurement and calibration [5, 6]. Therefore, most of the research papers

assume that a feedback channel exists from each user to the BS for carrying feedback information as those in [7–9].

It is generally assumed that the feedback channel from each user to the BS is not perfect. In [10–12], the delayed feedback channel is considered. The effects of outdated CSI at the BS are analyzed by means of Markov model of the channel temporal correlation in [11, 12]. The effects of the noisy feedback channel are considered in [13–15], where the feedback channel is assumed as a discrete symmetric channel. Plenty of the pieces of literature focus on the feedback channel with finite rate (see, e.g., [16] and the references therein), which is referred to as limited feedback channel.

From the limited feedback channel, each user is allowed to send a small number of information bits back to the BS. Most of the conventional feedback schemes focus on the quantization of the channel direction information (CDI) [9] using some vector quantization algorithms [17]. In [8], a codebook of orthogonal beamforming matrices is used for the quantization. Each user quantizes the channel directions according to a beamforming codebook with a priori knowledge by both the BS and the users. Possible ways of reducing the number of feedback bits have been proposed in [18, 19], where only the mean or covariance of CSI is sent back to the BS.

As for the overlay mode based on the orthogonal beamforming that we consider in this paper, the conventional feedback strategies are no longer suitable; using conventional feedback schemes, the BS acquires an approximation of each users' channel directions the resulting beamforming vectors designed according to this approximation will not be best aligned with the MTCD gateways' channel direction. Besides, due to the orthogonality constraint, the conventional feedback strategies become inefficient. That is, the codebook used and designed in [8] is uniformly distributed in the whole space, while the actual beamforming vector lies in a subspace due to the orthogonality constraint. Thus, the vectors in the codebook that lie out of the subspace provide redundant information and result in excessive feedback. Therefore, a new feedback strategy is needed in the overlay mode for MTCD gateways.

In this study, we propose a new multiple beam selection (MBS) strategy with limited feedback of CSI. A (downlink) broadcasting channel is used to inform MTCD gateways with all the available beams. Each user chooses the best one out of the available beams and sends the index of the chosen beam back to the BS. As the beams are dynamically generated at the BS according to the interference pattern of the users, the orthogonality is guaranteed and the CSI can be fed back through the feedback channel to the BS more efficiently.

This paper addresses joint beamforming, scheduling, and feedback for broadband wireless access to provide high throughput for MTCD gateways. The design of beamforming vectors and limited feedback from each user are closely connected due to the orthogonal constraint for MTCD gateways. In conventional feedback schemes, the beamforming vectors are calculated after receiving the feedback information from each user. However, in the proposed MBS scheme, the beamforming vectors are calculated in advance, awaiting for each user's choice. A similar thought has been applied to opportunistic space division multiple access (OSDMA) in [20, 21].

1.2. Organization and Notations. The remainder of this paper is organized as follows. The system model is given in Section 2. In Section 3, a detailed study on the limited feedback for MTCD gateways is presented and the MBS strategy is proposed. The performance analysis is given in Section 4 followed by a comprehensive evaluation of the proposed MBS strategy using Monte Carlo simulation in Section 5. We conclude the paper with some remarks in Section 6.

The superscripts T and H stand for the transpose and Hermitian transpose, respectively. Upper and lower bold-faced letters are used for matrices and column vectors, respectively. Denote by $|x|$ the absolute value of a scalar x or cardinality of x if x is a set, $\|\mathbf{x}\|$ the 2-norm of vector $\mathbf{x}$, $\perp$ perpendicularity, and $\mathbf{I}$ the identity matrix of a certain size implicitly given by the context. We denote by $[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N]$ the concatenation of N column vectors and by $[\mathbf{W}, \mathbf{c}]$ the concatenation of a matrix $\mathbf{W}$ and a vector $\mathbf{c}$ of the same column size. Elements of a set are enumerated as $\{a_1, \ldots, a_N\}$. $\mathbb{E}[\cdot]$ denotes the statistical expectation. $\mathcal{CN}(\mathbf{a}, \mathbf{R})$ represents the distribution of circularly symmetric complex Gaussian (CSCG) random vectors with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{R}$, and $\mathcal{C}^n$ denotes n-dimensional complex space. We denote the probability of an event X by $\Pr(X)$. log and $\log_2$ represent the natural logarithm and binary logarithm, respectively.

2. System Model

As in [22], in this paper, we consider downlink beamforming for MTCD gateways. Suppose that a BS in a cellular network is equipped with an antenna array of L elements. For downlink channels, there are M orthogonal subcarriers. The BS can support two different groups of users. One group has ordinary mobile users, which is called Group I. The other group consists of MTCD gateways, which is called Group II. The users in Group II want to have high data rate services. It is assumed that the users in Group I have higher priority than those in Group II. In the overlay mode, the downlink transmissions to the users in Group II should not affect the users in Group I.

It is assumed that there are K users in Group I, and Q users out of K are chosen for data transmissions in each subcarrier. These are called active users in Group I. The number of users in Group II is denoted by U. Only one of these users is selected for transmission through each subcarrier. In general, Q is small (usually $Q = 1$ if no spatial multiplexing is considered and U is also small as the number of active fixed subscribers at a time would be limited, especially in a rural area where there are not so many mobile users).

The transmitted signal vectors from the antenna array to users in Groups I and II through subcarrier m, denoted by $\mathbf{s}_{\mathrm{I},m}$ and $\mathbf{s}_{\mathrm{II},m}$, are given by

$$\mathbf{s}_{\mathrm{I},m} = \sum_{q=1}^{Q} \mathbf{w}_{\mathrm{I},m,q} a_{m,q}; \qquad \mathbf{s}_{\mathrm{II},m} = \mathbf{w}_{\mathrm{II},m} b_m, \tag{1}$$

where $a_{m,q}$ and b_m are independent data symbols for users in Groups I and II, respectively, and $\mathbf{w}_{\mathrm{I},m,q}$ and $\mathbf{w}_{\mathrm{II},m}$ are the corresponding beamforming weight vectors. Throughout the paper, we assume that the beamforming vectors are normalized; that is, $\|\mathbf{w}_{\mathrm{I},m,q}\| = \|\mathbf{w}_{\mathrm{II},m}\| = 1$ for all m. Denote $P_{\mathrm{I}} = \mathbb{E}[|a_{m,q}|^2]$ and $P_{\mathrm{II}} = \mathbb{E}[|b_m|^2]$ as the transmitted powers for Groups I and II, respectively.

We denote by k_m the index of a user in Group I who receives the signal through subcarrier m. The received signal at this user is given by

$$x_{m,k_m} = \mathbf{h}_{m,k_m}^H \left(\mathbf{s}_{\mathrm{I},m} + \mathbf{s}_{\mathrm{II},m}\right) + v_{m,k_m}, \tag{2}$$

where $\mathbf{h}_{m,k_m}^H$ is the $1\times L$ downlink channel vector of subcarrier m from the transmit antenna array to the k_mth user in Group I and $v_{m,k_m} \sim \mathcal{CN}(0,\sigma^2)$ is an independent background noise.

If a user in Group II also receives a signal through subcarrier m whose index is denoted by u_m, the received signal at this user is given by

$$y_{m,u_m} = \mathbf{g}_{m,u_m}^H \left(\mathbf{s}_{\mathrm{I},m} + \mathbf{s}_{\mathrm{II},m}\right) + v_{m,u_m}, \tag{3}$$

where $\mathbf{g}_{m,u_m}^H$ is the $1\times L$ downlink channel vector of subcarrier m from the transmit antenna array to the u_mth user in Group II and $v_{m,u_m} \sim \mathcal{CN}(0,\sigma^2)$ is also an independent background noise.

Note that the characteristics of the channel vectors to users in Groups I and II are different. In general, $\mathbf{g}_{m,u_m}$ is not varying rapidly, while $\mathbf{h}_{m,k_m}$ is time varying as Group I users are mobile. The beamforming weight vectors $\mathbf{w}_{\mathrm{I},m,q}$ and $\mathbf{w}_{\mathrm{II},m}$ are decided by the BS according to the information sent back by each user. If the feedback channel is perfect, the CSI of each user, that is, $\mathbf{h}_{m,k_m}$ and $\mathbf{g}_{m,u_m}$, is known by the BS.

Throughout the paper, we focus on the data transmissions of the MTCD gateways. Generally we assume that the $\mathbf{h}_{m,k_m}$'s are available at the BS for beamforming and discuss the impact of limited feedback with $\mathbf{g}_{m,u_m}$. Note that the impact of imperfect $\mathbf{h}_{m,k_m}$ is considered in [8, 9]. We do not consider this issue in detail.

3. Beamforming with Limited Feedback

In this section, the beamforming algorithms for MTCD gateways are proposed. First, the algorithms with perfect CSI of the Group II users are briefly described in Section 3.1. The motivations for the new feedback scheme are given in Section 3.2. In Section 3.3, the beamforming algorithms with limited feedback CSI of the MTCD gateways are proposed.

3.1. Orthogonal Beamforming with Perfect CSI. As users in Group I have a higher priority than those in Group II, the signals to users in Group II should not interfere with the signals to users in Group I. For broadband access, it is required that the data rate to users in Group II from the BS is sufficiently high, while the number of users in Group II would not be large. In order to deal with these issues, we consider orthogonal beamforming. Using orthogonal beamforming, space division multiple access (SDMA) [23] can be implemented for downlink transmission to users in Groups I and II. The main difference of the orthogonal beamforming approach in this paper from that for the conventional SDMA is that the overlay mode based beamforming approach should take into account the priority between users in Groups I and II and maximize the data rate to users in Group II.

Note that a user scheduling problem exists in Group I as $Q < K$. When $Q > 1$, this selection problem cannot be solved by simply choosing the users with the best signal-to-interference-plus-noise ratio (SINR) or signal-to-noise ratio (SNR). Fortunately, there is an existing approach in [24].

In [24], the semiorthogonal user selection (SUS) algorithm is used for scheduling Q active users out of K users through each subcarrier. The SUS algorithm results in a group of users with channel vectors $\mathbf{h}_{m,k_m^q}$, $q = 1,\ldots,Q$, and

$$1 - \frac{\left|\mathbf{h}_{m,k_m^i}^H \mathbf{h}_{m,k_m^j}\right|^2}{\left\|\mathbf{h}_{m,k_m^i}\right\|^2 \left\|\mathbf{h}_{m,k_m^j}\right\|^2} > 1-\epsilon, \quad i,j = 1,\ldots,Q, i\neq j, \tag{4}$$

where ϵ is a small positive constant. Note that ϵ is a parameter related to the number of users K similar to α in [24], which is used to find semiorthogonal users. The SUS algorithm chooses the users with semiorthogonal channel directions. Thus, hereafter, we assume that the chosen users in Group I have orthogonal channel directions; that is,

$$\mathbf{h}_{m,k_m^i} \perp \mathbf{h}_{m,k_m^j}, \quad \forall m = 1,\ldots,M, \quad i,j = 1,\ldots,Q,\ i\neq j. \tag{5}$$

This becomes true when the number of users K is sufficiently large [24]. We denote by $\mathcal{Q}_m$ the set of the Group I active users in each subcarrier m. From (5), the beamforming vectors for the Q selected users are simply represented by

$$\mathbf{w}_{\mathrm{I},m,i} = \frac{\mathbf{h}_{m,k_m^i}}{\left\|\mathbf{h}_{m,k_m^i}\right\|}, \quad i = 1,\ldots,Q. \tag{6}$$

At a user in Group II, the SINR is given by

$$\mathrm{SINR}_{\mathrm{II},m,u_m} = \frac{\left|\mathbf{g}_{m,u_m}^H \mathbf{w}_{\mathrm{II},m}\right|^2 P_{\mathrm{II}}}{\sum_{i=1}^{Q}\left|\mathbf{g}_{m,u_m}^H \mathbf{w}_{\mathrm{I},m,i}\right|^2 P_{\mathrm{I}} + \sigma^2}. \tag{7}$$

Since $\mathbf{w}_{\mathrm{I},m,i}$ is decided, this SINR can be maximized by maximizing the numerator. That is,

$$\text{maximize} \quad \left|\mathbf{g}_{m,u_m}^H \mathbf{w}_{\mathrm{II},m}\right|, \tag{8}$$

$$\text{subject to} \quad \mathbf{w}_{\mathrm{II},m} \perp \mathbf{h}_{m,k_m^i}, \quad \forall i = 1,\ldots,Q. \tag{9}$$

The optimal beamforming vector is given by

$$\mathbf{w}_{\mathrm{II},m} = \frac{1}{\left\|\widetilde{\mathbf{w}}_{\mathrm{II},m}\right\|}\widetilde{\mathbf{w}}_{\mathrm{II},m}, \tag{10}$$

where

$$\widetilde{\mathbf{w}}_{\mathrm{II},m} = \left(\mathbf{I} - \sum_{i=1}^{Q}\frac{\mathbf{h}_{m,k_m^i}\mathbf{h}_{m,k_m^i}^H}{\left\|\mathbf{h}_{m,k_m^i}\right\|^2}\right)\mathbf{g}_{m,u_m}$$

$$= \left(\mathbf{I} - \sum_{i=1}^{Q} \mathbf{w}_{\mathrm{I},m,k_m^i} \mathbf{w}_{\mathrm{I},m,k_m^i}^H \right) \mathbf{g}_{m,u_m}$$
$$= \left(\mathbf{I} - \mathbf{W}_{\mathrm{I},m} \mathbf{W}_{\mathrm{I},m}^H \right) \mathbf{g}_{m,u_m}. \tag{11}$$

Here $\mathbf{W}_{\mathrm{I},m} = [\mathbf{w}_{\mathrm{I},m,1}, \mathbf{w}_{\mathrm{I},m,2}, \ldots, \mathbf{w}_{\mathrm{I},m,Q}]$.

3.2. Motivations for the New Feedback Scheme. With perfect CSI, the optimal beamforming vectors for users in Group II are designed subject to the orthogonal constraints (9). The resulting beamforming vector (11) lies in the null space formed by the column vectors of $\mathbf{W}_{\mathrm{I},m}$ and is best aligned to the channel vector $\mathbf{g}_{m,u_m}$. However, with limited feedback, the exact value of $\mathbf{g}_{m,u_m}$ is not available at the BS.

In the conventional feedback scheme in [8], each $\mathbf{g}_{m,u_m}$ is decomposed into two components, that is, the gain and the direction, as follows:

$$\mathbf{g}_{m,u_m} = \left\| \mathbf{g}_{m,u_m} \right\| \vec{\mathbf{g}}_{m,u_m}, \tag{12}$$

where $\vec{\mathbf{g}}_{m,u_m} = \mathbf{g}_{m,u_m} / \| \mathbf{g}_{m,u_m} \|$ is the channel direction. The user computes the SINR and the channel direction, quantizes them according to a codebook and sends them back to the BS.

In conventional feedback schemes, a codebook-based quantizer with a codebook comprised of multiple sets of orthonormal vectors in $\mathscr{C}^L$ is used to quantize the channel direction $\mathbf{g}_{m,u_m} / \| \mathbf{g}_{m,u_m} \|$. Denote the codebook by $\mathscr{F}$,

$$\mathscr{F} = \bigcup_{b=1}^{B} \mathscr{V}^{(b)}, \tag{13}$$

where $\mathscr{V}^{(b)}$ is the bth orthonormal set in the codebook $\mathscr{F}$. There are L mutually orthogonal vectors in each $\mathscr{V}^{(b)}$, $b = 1, \ldots, B$. Thus, the codebook size can be denoted by $|\mathscr{F}| = BL$. For each weight vector $\mathbf{v}$ in the codebook $\mathscr{F}$, a distortion function $d(\mathbf{v}, \vec{\mathbf{g}}_{m,u_m})$ is defined as

$$d\left(\mathbf{v}, \vec{\mathbf{g}}_{m,u_m}\right) = 1 - \left| \mathbf{v}^H \vec{\mathbf{g}}_{m,u_m} \right|^2. \tag{14}$$

For the quantization of the channel direction, the member of $\mathscr{F}$ that has the smallest distortion is chosen as the feedback channel direction; that is,

$$\hat{\mathbf{g}}_{m,u_m} = \arg\min_{\mathbf{v} \in \mathscr{F}} d\left(\mathbf{v}, \vec{\mathbf{g}}_{m,u_m}\right). \tag{15}$$

For the quantization of the SINR, G bits are used to quantize this scalar value. For analysis convenience, the SINR is always assumed to be known perfectly by the BS. This assumption will be justified in Section 5 where the simulation results show that the throughput loss due to limited SINR feedback is marginal compared with the case in which perfect SINR feedback is available.

In the orthogonal beamforming for users in Group II, the optimal beamforming vectors are designed subject to orthogonal constraints which is illustrated in Figure 1. If $\mathbf{w}_{\mathrm{I},m}$ and $\mathbf{g}_{m,u_m}$ are given, the beamforming vector for $\mathbf{g}_{m,u_m}$ should lie in the null space of $\mathbf{w}_{\mathrm{I},m}$, that is, the disk perpendicular to $\mathbf{w}_{\mathrm{I},m}$. The best beam for u_m is $\mathbf{w}_{\mathrm{II},m}$ in Figure 1.

Figure 1: An example of orthogonal beamforming for MTCD gateways. $\mathbf{w}_{\mathrm{II},m}$ should lie in the null space of $\mathbf{h}_{m,k_m}$ and be best aligned with $\mathbf{g}_{m,u_m}$.

With limited feedback, if we can quantize $\mathbf{g}_{m,u_m}$ to the vector $\mathbf{w}_{\mathrm{II},m}$ as depicted in Figure 1, the BS obtains sufficient information to determine the beamforming vector. It is equivalent to say that if the quantization codebook $\mathscr{F}$ lies in the null space formed by the column vectors of $\mathbf{W}_{\mathrm{I},m}$, we are more likely to get a quantized version of $\mathbf{g}_{m,u_m}$ as close to $\mathbf{w}_{\mathrm{II},m}$ as possible.

In conventional feedback schemes, the codebook is predesigned and known at both the BS and each user. The vectors in $\mathscr{F}$ that lie out of null space of $\mathbf{W}_{\mathrm{I},m}$ provide redundant information and result in excessive feedback bits. Meanwhile it is not possible to change $\mathscr{F}$ flexibly according to a given $\mathbf{W}_{\mathrm{I},m}$. A new feedback scheme is needed in orthogonal beamforming for MTCD gateways.

3.3. Orthogonal Beamforming with Limited Feedback. We propose a new MBS strategy where a group of beamforming vectors for the MTCD gateways is generated by the BS and broadcasted through the downlink channel.

Suppose that the Q active users in Group I have been chosen and the beamforming matrix is denoted by $\mathbf{W}_{\mathrm{I},m}$. The

size of $\mathbf{W}_{\mathrm{I},m}$ is $L \times Q$. For an extra user from Group II, we generate N new candidate beamforming vectors as follows:

$$\begin{aligned} \mathbf{C}_{\mathrm{II},m} &= \frac{\widetilde{\mathbf{C}}_{\mathrm{II},m}}{\left\|\widetilde{\mathbf{C}}_{\mathrm{II},m}\right\|}, \\ \widetilde{\mathbf{C}}_{\mathrm{II},m} &= \left(\mathbf{I} - \mathbf{W}_{\mathrm{I},m}\mathbf{W}_{\mathrm{I},m}^{H}\right)\mathbf{U}_m, \end{aligned} \tag{16}$$

where $\mathbf{U}_m$ is a random matrix of size $L \times N$. Each column vector of $\mathbf{C}_{\mathrm{II},m}$ is a normalized weight vector. The set of the column vectors of $\mathbf{C}_{\mathrm{II},m}$ is denoted by $\{\mathbf{c}_{m,1}, \ldots, \mathbf{c}_{m,N}\}$.

For the scheduling of a MTCD gateway, the BS should transmit N pilot signals to let them choose the best beam among N beams, $\{\mathbf{c}_{m,1}, \ldots, \mathbf{c}_{m,N}\}$. The resulting beam selection problem can be given by

$$\max_{n} \mathrm{SINR}_{\mathrm{II},m,u_m,n}, \tag{17}$$

where the SINR is given by

$$\mathrm{SINR}_{\mathrm{II},m,u_m,n} = \frac{\left|\mathbf{g}_{m,u_m}^{H}\mathbf{c}_{m,n}\right|^2 P_{\mathrm{II}}}{\sum_{i=1}^{Q}\left|\mathbf{g}_{m,u_m}^{H}\mathbf{w}_{\mathrm{I},m,i}\right|^2 P_{\mathrm{I}} + \sigma^2}. \tag{18}$$

The SINR expressions in (17) and (18) are replaced by $\mathrm{SINR}_{\mathrm{II},m,u_m,n}$ in order to emphasize their dependence on n or $\mathbf{c}_{m,n}$, the beamforming candidate. The detailed process is described as follows.

(1) The BS broadcasts $\mathbf{c}_{m,1}, \ldots, \mathbf{c}_{m,N}$ when choosing a user in Group II by employing a series of the beamforming matrices, $\{[\mathbf{W}_{\mathrm{I},m}, \mathbf{c}_1], \ldots, [\mathbf{W}_{\mathrm{I},m}, \mathbf{c}_N]\}$ for N symbol durations. At the user u_m in Group II, the received signal when the nth pilot signal is transmitted is given by

$$y_{m,u_m} = \mathbf{g}_{m,u_m}^{H}\left(\sum_{q=1}^{Q}\mathbf{w}_{\mathrm{I},m,q}a_{m,q,n} + \mathbf{c}_{m,n}b_{m,n}\right) + v_{m,u_m}, \tag{19}$$

where $a_{m,q,n}$ and $b_{m,n}$ denote the nth pilot signals to the qth user in Group I and the user in Group II, respectively.

(2) After receiving all the N pilot signals, each user in Group II estimates the local SINR, namely, $\mathrm{SINR}_{\mathrm{II},m,u_m,n}$, with different beams, $n = 1, \ldots, N$. The $\mathrm{SINR}_{\mathrm{II},m,u_m,n}$s are assumed to be known perfectly at each user u_m.

(3) The users choose the best beam among all the pilot beams according to (17) and sends the index of the best beam $n_{u_m}^{*}$ to the BS.

$$n_{u_m}^{*} = \arg\max_{n\in\{1,\ldots,N\}} \mathrm{SINR}_{\mathrm{II},m,u_m,n}. \tag{20}$$

If there are more than one beam that are equally the best, the user randomly chooses one of them and feeds the index back.

(4) At each user, the SINRs of the best beam through each subcarrier are quantized using G bits as follows:

$$\mathrm{SINR}_{\mathrm{II},m,u_m,n_{u_m}^{*}} \longrightarrow \widehat{\mathrm{SINR}}_{\mathrm{II},m,u_m,n_{u_m}^{*}}. \tag{21}$$

Through subcarrier m, only the quantized SINR's of the best beam at each user, namely, $\widehat{\mathrm{SINR}}_{\mathrm{II},m,u_m,n_{u_m}^{*}}$, are sent back through the feedback channel.

In order to minimize the impact of the interference from the signals to users in Group I or maximize the throughput for users in Group II, the subcarrier allocation for Group II users can be carried out as follows:

$$\begin{aligned} u_m^{*} &= \arg\max_{u\in\{1,\ldots,U\}} \log_2\left(1 + \widehat{\mathrm{SINR}}_{\mathrm{II},m,u,n_u^{*}}\right) \\ &= \arg\max_{u\in\{1,\ldots,U\}} \widehat{\mathrm{SINR}}_{\mathrm{II},m,u,n_u^{*}}, \end{aligned} \tag{22}$$

and the corresponding beamforming vectors are selected as the final beam

$$n_{u_m^{*}}^{*}. \tag{23}$$

That is, for each subcarrier, the user in Group II who maximizes the achievable rate, namely, $\log_2(1 + \widehat{\mathrm{SINR}}_{\mathrm{II},m,u,n_u^{*}})$, or the $\widehat{\mathrm{SINR}}_{\mathrm{II},m,u,n_u^{*}}$ is to be selected. The set of subcarriers allocated to user u is given by

$$\mathscr{M}_u = \left\{m \mid u_m^{*} = u, m = 1, \ldots, M\right\}. \tag{24}$$

Thus, the estimated throughput for user u is given by

$$R_u = \sum_{m\in\mathscr{M}_u} \log_2\left(1 + \widehat{\mathrm{SINR}}_{\mathrm{II},m,u,n_u^{*}}\right). \tag{25}$$

4. Asymptotic Throughput Performance

In this section, we analyze the performance of the proposed MBS scheme asymptotically when $N \to \infty$. When $N \to \infty$, there is always a beam that is perfectly aligned to the channel direction of each user. From (11) and (16), if one column of $\mathbf{U}_m$ is aligned with $\mathbf{g}_{m,u_m}/\|\mathbf{g}_{m,u_m}\|$, the MBS scheme achieves the ideal performance with perfect CSI at the BS.

In this section the asymptotic throughput performance is first derived with perfect CSI feedback. In order to obtain a useful expression that characterizes the system performance with limited feedback, a lower bound on the asymptotic throughput performance is then derived.

Assumption 1. The elements of $\mathbf{h}_{m,k_m}$ and $\mathbf{g}_{m,u_m}$ are independent and identically distributed (i.i.d.) and CSCG random variables, $\mathscr{CN}(0, 1/L)$.

Hereafter, we consider performance analysis for users in Group II under the CSCG Assumption 1. For convenience, define the SNR at user u_m in Group II as

$$\gamma_{\mathrm{II},m,u_m} = \frac{\left\|\mathbf{g}_{m,u_m}\right\|^2 P_{\mathrm{II}}}{\sigma^2}. \tag{26}$$

Furthermore, define

$$\xi_m(\mathcal{Q}_m, u_m) = 1 - \frac{\left\|\mathbf{g}_{m,u_m}^H \mathbf{W}_{\mathrm{I},m}\right\|^2}{\left\|\mathbf{g}_{m,u_m}\right\|^2}. \quad (27)$$

Then, we have

$$\mathrm{SINR}_{\mathrm{II},m,u_m,\mathcal{Q}_m} = \frac{\xi_m(\mathcal{Q}_m, u_m)}{(1-\xi_m(\mathcal{Q}_m, u_m))(P_{\mathrm{I}}/P_{\mathrm{II}}) + (1/\gamma_{\mathrm{II},m,u_m})}. \quad (28)$$

In (28), $\mathrm{SINR}_{\mathrm{II},m,u_m}$, which is the SINR at user u_m in Group II, is replaced with $\mathrm{SINR}_{\mathrm{II},m,u_m,\mathcal{Q}_m}$ in order to emphasize its dependence on $\mathcal{Q}_m$ or $\mathbf{W}_{\mathrm{I},m}$. The proof of (28) is given in Appendix A.

Note that the SINR for a user in Group II is upper-bounded as

$$\mathrm{SINR}_{\mathrm{II},m,u_m,\mathcal{Q}_m} \le \min\left\{\frac{\xi_m(\mathcal{Q}_m, u_m)}{1-\xi_m(\mathcal{Q}_m, u_m)}\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}}, \gamma_{\mathrm{II},m,u_m}\right\}. \quad (29)$$

Assumption 2. The distribution of $1-\xi_m(\mathcal{Q}_m, u_m)$ can be approximated by a beta distribution with the parameters Q and $L-1$. Note that the approximation becomes more accurate when $L \to \infty$.

That is, we have

$$1-\xi_m(\mathcal{Q}_m, u_m) \sim \mathrm{B}(Q, L-1) \quad (30)$$

or

$$\xi_m(\mathcal{Q}_m, u_m) \sim \frac{1}{B(a,b)}(1-y)^{a-1}y^{b-1}, \quad 0 \le y \le 1, \quad (31)$$

where $B(a,b)$ is the B function and $a = Q$, $b = L-1$.

Now, we present numerical results to verify Assumption 2. In Figure 2, the number of users in Group I is set to $Q = 4$. Both the approximate and empirical distributions of $1-\xi_m(\mathcal{Q}_m, u_m)$ are illustrated for different number of transmitter antennas ($L = 10$, 20, and 40). When $L = 40$, approximation (30) fits well with the empirical distribution. Figure 3 shows the results when the number of users in Group I is set to $Q = 2$. The same transmitter antenna arrays are considered as those in Figure 2. It is shown that the approximate distribution is reasonable especially when there are less users in Group I.

The probability density function (PDF) of $z = \xi_m(\mathcal{Q}_m, u_m)/(1-\xi_m(\mathcal{Q}_m, u_m))$ is given by

$$\frac{\xi_m(\mathcal{Q}_m, u_m)}{1-\xi_m(\mathcal{Q}_m, u_m)} \sim f(z) = \frac{1}{B(a,b)}\frac{z^{b-1}}{(1+z)^{a+b}}, \quad 0 \le z < \infty \quad (32)$$

and the cumulative distribution function (CDF) by

$$F(z) = \Pr\left(\frac{\xi_m(\mathcal{Q}_m, u_m)}{1-\xi_m(\mathcal{Q}_m, u_m)} \le z\right) = 1 - I_{1/(1+z)}(a,b), \quad (33)$$

Figure 2: Distribution of $1-\xi_m(\mathcal{Q}_m, u_m)$, $Q = 4$ users in Group I.

Figure 3: Distribution of $1-\xi_m(\mathcal{Q}_m, u_m)$, $Q = 2$ users in Group I.

where $I_x(a,b)$ is the incomplete beta function. The CDF of the maximum SINR among U independent users in Group II is

$$\begin{aligned} F_{\max}(z) &= \Pr\left(\max_u \frac{\xi_m(\mathcal{Q}_m, u)}{1-\xi_m(\mathcal{Q}_m, u)} \le z\right) \\ &= F^U(z) = \left[1 - I_{1/(1+z)}(a,b)\right]^U. \end{aligned} \quad (34)$$

The outage probability of the maximum SINR is defined as

$$P_{\mathrm{out}}(\Gamma) = \Pr\left(\max_u \mathrm{SINR}_{\mathrm{II},m,u,\mathcal{Q}_m} \le \Gamma\right), \quad (35)$$

where Γ is the target threshold SINR.

Using the upper bound in (29), a lower bound on the outage probability of the SINR can be found. Since

$$\begin{aligned}\max_u \mathrm{SINR}_{\mathrm{II},m,u,\mathcal{Q}_m} &\le \max_u \min\left\{\frac{\xi_m(\mathcal{Q}_m,u)}{1-\xi_m(\mathcal{Q}_m,u)}\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}},\gamma_{\mathrm{II},m,u}\right\} \\ &\le \min\left\{\max_u \frac{\xi_m(\mathcal{Q}_m,u)}{1-\xi_m(\mathcal{Q}_m,u)}\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}},\max_u \gamma_{\mathrm{II},m,u}\right\},\end{aligned} \tag{36}$$

we have

$$P_{\mathrm{out}}(\Gamma) \ge 1-\left(1-F_{\max}\left(\frac{P_{\mathrm{I}}\Gamma}{P_{\mathrm{II}}}\right)\right)\left(1-F^{U}_{\chi^{(2L)}}\left(\frac{2L\sigma^2\Gamma}{P_{\mathrm{II}}}\right)\right), \tag{37}$$

where $F_{\chi^{(n)}}(z)$ denotes the chi-square CDF with n degrees of freedom. The proof of (37) in details is given in Appendix B.

The average achievable rate per subcarrier for Group II is

$$\bar{r} = \mathbb{E}\left[\log_2\left(1+\max_u \mathrm{SINR}_{\mathrm{II},m,u,\mathcal{Q}_m}\right)\right](\mathrm{bps/Hz}), \tag{38}$$

and the average throughput per user becomes

$$\mathbb{E}[R_u] = \frac{M}{U}\bar{r}. \tag{39}$$

We consider a lower-bound on the average achievable rate using the outage probability in (37). For a given target threshold SINR Γ, we have

$$\begin{aligned}\bar{r} &= \mathbb{E}\left[\log_2\left(1+\max_u \mathrm{SINR}_{\mathrm{II},m,u,\mathcal{Q}_m}\right)\right] \\ &\ge \Pr\left(\max_u \mathrm{SINR}_{\mathrm{II},m,u,\mathcal{Q}_m} \le \Gamma\right)\times 0 \\ &\quad + \Pr\left(\max_u \mathrm{SINR}_{\mathrm{II},m,u,\mathcal{Q}_m} > \Gamma\right)\log_2(1+\Gamma) \\ &= \left(1-P_{\mathrm{out}}(\Gamma)\right)\log_2(1+\Gamma).\end{aligned} \tag{40}$$

A tight lower-bound on $\bar{r}$ can be achieved if the lower-bound is maximized with respect to Γ, which results in

$$\mathbb{E}[R_u] \ge \frac{M}{U}\max_{\Gamma}\left(1-P_{\mathrm{out}}(\Gamma)\right)\log_2(1+\Gamma). \tag{41}$$

Note that the asymptotic throughput performance is first characterized by (39) using (28). This result is then lower-bounded in (41). We will use (41) as an indication of the system performance with limited feedback.

5. Simulation Results

In this section we present various numerical results to further investigate the performance of the proposed MBS scheme. The simulation settings are described as follows: the channel elements are independent and identically distributed (i.i.d.) CSCG random variables, that is, $\mathcal{CN}(0,1/L)$; the cellular radius is normalized as one; the base band channel is considered in the simulation without a specified radio frequency.

Figure 4: Throughput per MTCD gateway versus $U(Q = 1, L = 10, M = 128$, and $P_{\mathrm{II}}/\sigma^2 = 20$ dB).

In the following two subsections, we will first assume that the BS knows each users' SINR: in Section 5.1, the effect of increasing feedback bits, that is, increasing the number of candidate beams N, is investigated; in Section 5.2, the throughput performance of the proposed MBS is compared with that of the conventional feedback scheme. Then, in the last subsection, the effect of limited feedback of SINR is shown with different number of users in Group II. The simulation results are presented under Assumption 1 with $M = 128$, $P_{\mathrm{I}} = P_{\mathrm{II}}$, and $P_{\mathrm{II}}/\sigma^2 = 20$ dB.

5.1. Effect of Increasing Feedback Bits. Simulation results for different number of active users in Group I are presented to show the effect of increasing feedback bits. As conceived, the throughput per MTCD gateway increases as the increase of the number of beams generated by the BS, unlike the results in [8] where increasing channel shape feedback does not necessarily lead to better performance.

In Figure 4, the number of transmitter antennas is set to $L = 10$. Only single active user is considered in Group I. From Figure 4, when the number of beams increases from 1 to 10 beams with $U = 2$, we can see that the throughput increases about 80%. The lower bound of the asymptotic performance is also plotted in Figure 4. Theoretical result (41) gives a good prediction of the system performance with limited feedback.

In Figures 5 and 6, the number of active users in Group I are 2 and 4, respectively. As expected, theoretical result (41) can be used as a good indication of the system performance with limited feedback.

5.2. Comparison with Conventional Feedback Schemes. From the analysis in Section 3, the proposed MBS scheme achieves better performance than the conventional feedback schemes. In this subsection, numerical results are given to verify our analysis.

Figure 5: Throughput per MTCD gateway versus $U(Q = 2, L = 10, M = 128$, and $P_{\mathrm{II}}/\sigma^2 = 20$ dB).

Figure 6: Throughput per MTCD gateway versus $U(Q = 4, L = 10, M = 128$, and $P_{\mathrm{II}}/\sigma^2 = 20$ dB).

In Figure 7, the performances of the proposed feedback scheme and the conventional one are compared with an increasing number of users in Group II. The number of beams generated by the BS in our proposed scheme and the number of the vectors in the codebook of the conventional feedback scheme are the same and set to $N = 256$. The orthogonal beamforming matrices of the conventional feedback are generated using a method in [8]. As expected, the proposed feedback scheme outperforms the conventional one, especially when the number of active users in the Group

Figure 7: Throughput per MTCD gateway versus $U(N = 256, L = 6, M = 128$, and $P_{\mathrm{II}}/\sigma^2 = 20$ dB); the number of active users in Group I varies from up to bottom ($Q = 2$, 3, 4, and 5, for each pair of curves resp.).

I is small. In general, a throughput increase of about 10% over the conventional feedback scheme is achieved by the proposed MBS.

5.3. Effect of SINR Quantization. The analyses and simulations above are based on the assumption that the BS has perfect knowledge of each user's SINR. The effect of quantization of SINR on the performance is shown using numerical simulations in this subsection.

Figure 8 compares the cases of perfect and quantized SINR feedback with G-bit quantization, $G = 2, 3$, and 4. A simple scalar quantizer with equally spaced codebook is used to quantize each user's SINR on the best beam. The number of beams generated by the BS $N = 10$, and number of transmitter antennas $L = 10$. From Figure 8, we can see that the performance degradation is small. A magnified observation in Figure 8 shows that 4-bit feedback of SINR is sufficient for a performance as good as perfect feedback.

From Figure 8, the throughput loss due to limited feedback of SINR is shown to be considerably low, even by employing few number of feedback bits. The same observation is made in the conventional feedback of per user unitary and rate control (PU2RC). In [8], 3 bits for SINR feedback is found to be sufficient to make the capacity loss due to negligible SINR quantization.

6. Conclusions

In this paper, we propose a scheme to combine cellular networks and M2M networks. The devices in an M2M network access a cellular network through some gateways, which are assumed as secondary users in the cellular network. With limited feedback of CSI at the BS, the new MBS method is proposed for the MTCD gateways with an overlaid

Figure 8: Throughput per MTCD gateway versus $U(N = 10, L = 10, M = 128$, and $P_{\mathrm{II}}/\sigma^2 = 20$ dB).

transmission mode. The asymptotic performance of the proposed MBS method is characterized. Using a lower bound on the throughput with perfect CSI, we obtain a useful expression that shows the system performance with limited feedback. Simulation results show that, when the BS is equipped with 6 antenna elements, the throughput was improved by about 10% when the proposed MBS method is employed compared with the conventional ones.

Appendices

A. Proof of (28)

In order to make the proving process convenient, we omit the subscripts in the following derivations. Denote $\mathrm{SINR}_{\mathrm{II},m,u_m,\mathcal{Q}_m}$, $\mathbf{g}_{m,u_m}$, $\mathbf{w}_{\mathrm{II},m}$, and $\mathbf{W}_{\mathrm{I},m}$ by SINR, $\mathbf{g}$, $\mathbf{w}$, and $\mathbf{W}_{\mathrm{I}}$, respectively.

From (7), the SINR is

$$\mathrm{SINR} = \frac{\left|\mathbf{g}^H\mathbf{w}\right|^2 P_{\mathrm{II}}}{\left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2 P_{\mathrm{I}} + \sigma^2}. \tag{A.1}$$

From (10), the beamforming vector for the MTCD gateway is

$$\mathbf{w} = \frac{1}{\|\widetilde{\mathbf{w}}\|}\widetilde{\mathbf{w}}, \tag{A.2}$$

where

$$\begin{aligned}\widetilde{\mathbf{w}} &= \left(\mathbf{I} - \mathbf{W}_{\mathrm{I}}\mathbf{W}_{\mathrm{I}}^H\right)\mathbf{g},\\ \|\widetilde{\mathbf{w}}\| &= \sqrt{\widetilde{\mathbf{w}}^H\widetilde{\mathbf{w}}} = \sqrt{\mathbf{g}^H\left(\mathbf{I} - \mathbf{W}_{\mathrm{I}}\mathbf{W}_{\mathrm{I}}^H\right)^H\left(\mathbf{I} - \mathbf{W}_{\mathrm{I}}\mathbf{W}_{\mathrm{I}}^H\right)\mathbf{g}}\\ &\overset{(a)}{=} \sqrt{\mathbf{g}^H\left(\mathbf{I} - \mathbf{W}_{\mathrm{I}}\mathbf{W}_{\mathrm{I}}^H\right)\mathbf{g}} = \sqrt{\|\mathbf{g}\|^2 - \left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2}.\end{aligned} \tag{A.3}$$

Equation (a) is due to the following features of the matrix $A = (\mathbf{I} - \mathbf{W}_{\mathrm{I}}\mathbf{W}_{\mathrm{I}}^H)$:

(1) $A^H = A$

(2) $A^n = A, n = 1, 2, \ldots$.

The numerator of SINR is

$$\begin{aligned}\left|\mathbf{g}^H\mathbf{w}\right|^2 P_{\mathrm{II}} &= \frac{\left|\mathbf{g}^H\left(\mathbf{I} - \mathbf{W}_{\mathrm{I}}\mathbf{W}_{\mathrm{I}}^H\right)\mathbf{g}\right|^2}{\|\widetilde{\mathbf{w}}\|^2}P_{\mathrm{II}}\\ &= \frac{\left|\mathbf{g}^H\left(\mathbf{I} - \mathbf{W}_{\mathrm{I}}\mathbf{W}_{\mathrm{I}}^H\right)\mathbf{g}\right|^2}{\|\mathbf{g}\|^2 - \left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2}P_{\mathrm{II}}\\ &= \frac{\left|\|\mathbf{g}\|^2 - \left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2\right|^2}{\|\mathbf{g}\|^2 - \left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2}P_{\mathrm{II}}\\ &= \left(\|\mathbf{g}\|^2 - \left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2\right)P_{\mathrm{II}}.\end{aligned} \tag{A.4}$$

The denominator of SINR is

$$\left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2 P_{\mathrm{I}} + \sigma^2. \tag{A.5}$$

Thus, the SINR is

$$\mathrm{SINR} = \frac{\left(1 - \left(\left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2/\|\mathbf{g}\|^2\right)\right)}{\left(\left\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\right\|^2/\|\mathbf{g}\|^2\right)(P_{\mathrm{I}}/P_{\mathrm{II}}) + \left(\sigma^2/\|\mathbf{g}\|^2 P_{\mathrm{II}}\right)}. \tag{A.6}$$

Define $\xi = 1 - (\|\mathbf{g}^H\mathbf{W}_{\mathrm{I}}\|^2/\|\mathbf{g}\|^2)$, and $\gamma_{\mathrm{II}} = (\|\mathbf{g}\|^2 P_{\mathrm{II}})/\sigma^2$. The SINR expression becomes

$$\mathrm{SINR} = \frac{\xi}{(1-\xi)(P_{\mathrm{I}}/P_{\mathrm{II}}) + (1/\gamma_{\mathrm{II}})}. \tag{A.7}$$

Thus, we have

$$\mathrm{SINR}_{\mathrm{II},m,u_m,\mathcal{Q}_m} = \frac{\xi_m(\mathcal{Q}_m, u_m)}{(1 - \xi_m(\mathcal{Q}_m, u_m))(P_{\mathrm{I}}/P_{\mathrm{II}}) + \left(1/\gamma_{\mathrm{II},m,u_m}\right)}. \tag{A.8}$$

B. Proof of (37)

Consider

$$\begin{aligned}&P_{\mathrm{out}}(\Gamma)\\ &= \Pr\left(\max_u \mathrm{SINR}_{\mathrm{II},m,u,\mathcal{Q}_m} \le \Gamma\right)\\ &\ge \Pr\left(\min\left\{\max_u \frac{\xi_m(\mathcal{Q}_m, u)}{1 - \xi_m(\mathcal{Q}_m, u)}\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}}, \max_u \gamma_{\mathrm{II},m,u}\right\} \le \Gamma\right)\\ &= 1 - \Pr\left(\min\left\{\max_u \frac{\xi_m(\mathcal{Q}_m, u)}{1 - \xi_m(\mathcal{Q}_m, u)}\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}}, \max_u \gamma_{\mathrm{II},m,u}\right\} > \Gamma\right)\\ &= 1 - \Pr\left(\max_u \frac{\xi_m(\mathcal{Q}_m, u)}{1 - \xi_m(\mathcal{Q}_m, u)}\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}} > \Gamma\right)\\ &\quad\times \Pr\left(\max_u \gamma_{\mathrm{II},m,u} > \Gamma\right).\end{aligned} \tag{B.1}$$

Using the CDF of the maximum SINR among U independent users (34), the first multiplier in (B.1) is

$$\Pr\left(\max_u \frac{\xi_m(\mathcal{Q}_m, u)}{1-\xi_m(\mathcal{Q}_m, u)}\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}} > \Gamma\right) = 1 - F_{\max}\left(\frac{P_{\mathrm{I}}}{P_{\mathrm{II}}}\Gamma\right). \tag{B.2}$$

Now we calculate the second multiplier in (B.1). From (26) and Assumption 1, $\gamma_{\mathrm{II},m,u}$ is χ^2 distributed with $2L$ degrees of freedom and $1/2L$ variance. Denote $F_{\chi^{(n)}}(z)$ as the chi-square CDF with n degrees of freedom and unit variance. From [25],

$$F_{\chi^{(n)}}(z) = \int_0^z \frac{1}{2^{n/2}\Gamma((1/2)n)} t^{(n-2)/2} e^{-t/2} dt, \tag{B.3}$$

where $\Gamma(x)$ is the gamma function. The CDF of $\gamma_{\mathrm{II},m,u}$ is $F_{\chi^{(2L)}}(2Lz)$. Thus,

$$\Pr\left(\max_u \gamma_{\mathrm{II},m,u} > \Gamma\right) = 1 - F^U_{\chi^{(2L)}}\left(\frac{2L\sigma^2\Gamma}{P_{\mathrm{II}}}\right). \tag{B.4}$$

From (B.1), (B.2), and (B.4), we have

$$P_{\mathrm{out}}(\Gamma) \geq 1 - \left(1 - F_{\max}\left(\frac{P_{\mathrm{I}}\Gamma}{P_{\mathrm{II}}}\right)\right)\left(1 - F^U_{\chi^{(2L)}}\left(\frac{2L\sigma^2\Gamma}{P_{\mathrm{II}}}\right)\right). \tag{B.5}$$

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments that have helped in improving the overall quality of the paper. This work has been supported by the Specialized Research Fund for the Doctoral Program of the Ministry of Education of China (Grand no. 20120001120125) and the National Natural Science Foundation of China (Grants no. 61250001, no. 61231011, and no. 61231013).

References

[1] J. Zhang, L. Shan, H. Hu, and Y. Yang, "Mobile cellular networks and wireless sensor networks: toward convergence," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 164–169, 2012.

[2] G. Wang, X. Zhong, S. Mei, and J. Wang, "An adaptive medium access control mechanism for cellular based machine to machine (M2M) communication," in *Proceedings of the IEEE International Conference on Wireless Information Technology and Systems (ICWITS '10)*, pp. 164–169, September 2010.

[3] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184–192, 2012.

[4] 3GPP TR 23.888 v0.5.1, "System Improvements for Machine-Type Communications (Release 10)," July 2010.

[5] N. Tyler, B. Allen, and H. Aghvami, "Adaptive antennas: the calibration problem," *IEEE Communications Magazine*, vol. 42, no. 12, pp. 114–122, 2004.

[6] S. Nanda, R. Walton, J. Ketchum, M. Wallace, and S. Howard, "A high-performance MIMO OFDM wireless LAN," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 101–109, 2005.

[7] D. J. Love, R. W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, 2004.

[8] K. Huang, J. G. Andrews, and R. W. Heath, "Performance of orthogonal beamforming for SDMA with limited feedback," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 152–164, 2009.

[9] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478–1491, 2007.

[10] Y. Ko and C. Tepedelenlioglu, "Orthogonal space-time block coded rate-adaptive modulation with outdated feedback," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 290–295, 2006.

[11] Y. Ma, D. Zhang, A. Leith, and Z. Wang, "Error performance of transmit beamforming with delayed and limited feedback," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1164–1170, 2009.

[12] R. Bhagavatula and R. W. Heath, "Adaptive bit partitioning for multicell intercell interference nulling with delayed limited feedback," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3824–3836, 2011.

[13] C. R. Murthy, J. Zheng, and B. D. Rao, "Performance of quantized equal gain transmission with noisy feedback channels," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2451–2460, 2008.

[14] T. Wu and V. K. N. Lau, "Robust rate, power and precoder adaptation for slow fading MIMO channels with noisy limited feedback," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2360–2367, 2008.

[15] S. Ekbatani, F. Etemadi, and H. Jafarkhani, "Outage behavior of quasi-static fading channels with partial power control and noisy feedback," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 1556–1560, November 2007.

[16] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.

[17] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 458–462, 2007.

[18] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 406–417, 2003.

[19] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1543–1553, 2004.

[20] N. Sharma and L. H. Ozarow, "A study of opportunism for multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1804–1814, 2005.

[21] W. Choi, A. Forenza, J. G. Andrews, and R. W. Heath, "Opportunistic space-division multiple access with beam selection," *IEEE Transactions on Communications*, vol. 55, no. 12, pp. 2371–2380, 2007.

[22] J. Choi and J. Ha, "Orthogonal beamforming for overlay mode of OFDMA-based rural broadband wireless access," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '12)*, Paris, France, April 2012.

[23] J. Choi, *Optimal Combining and Detection: Statistical Signal Processing For Communications*, Cambridge University Press, New York, NY, USA, 2010.

[24] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.

[25] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 1995.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 234728, 9 pages
http://dx.doi.org/10.1155/2013/234728

Research Article

Multicast Capacity Analysis for High Mobility Social Proximity Machine-to-Machine Networks

Xin Guan

School of Information Science and Technology, Heilongjiang University, Harbin 150080, China

Correspondence should be addressed to Xin Guan; guanxin.hlju@gmail.com

Received 16 August 2013; Revised 25 October 2013; Accepted 28 October 2013

Academic Editor: Yan Zhang

Copyright © 2013 Xin Guan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wireless machine-to-machine (M2M) networks enable ubiquitous sensing and controlling via sensors, vehicles, and other types of wireless nodes. Capacity scaling law is one of the fundamental properties for high mobility M2M networks. As for high mobility M2M networks, vehicular ad hoc networks (VANETs) are a typical case. Since vehicles have social property, their moving trajectory is according to the fixed community. With the purpose of transmitting packets to different communities of VANETs and further improving the network capacity, we study the multicast capacity of bus-assisted VANETs in two scenarios: forwarding scenario and routing scenario. All the ordinary vehicles obey the restricted mobility model. Thus, the spatial stationary distribution decays as power law with the distance from the center spot of a restrict region of each vehicle. In forwarding scenario, all the buses deployed in all roads as intermediate nodes are used to forward packets for ordinary vehicles. In routing scenario, buses and ordinary cars construct a highway path supported by percolation theory to transmit urgent packets. Each ordinary vehicle randomly chooses $k-1$ vehicles from the other ordinary vehicles as receivers. For the two kinds of scenarios, we derived the upper bound and lower bound, respectively.

1. Introduction

In recent years, M2M networks have been a subject of intense interests due to the rapid evolution of wireless communication systems. M2M networks can be used to enable a wide variety of automated complex operations, which include advanced sensing, remote control, and monitoring technologies. M2M networks have many differences compared with conventional wireless networks because M2M is a machine-oriented networking technology. First of all, M2M networks commonly need to support a large number of devices. Second, power consumption of M2M devices is very important due to the limited battery capacity. Third, various data transfer delay requirements may exist depending on the M2M application being executed [1, 2]. Examples of M2M communications networks include vehicular ad hoc networks (VANETs) [3], underwater sensor networks (UWSNs) [4], wireless body area networks (WBANs) [5], and wireless mobile social networks [6–8].

Capacity scaling laws of wireless networks have attracted a lot of attention. Since Gupta and Kumar initiated the study of capacity scaling laws of ad hoc networks [9], the capacity of different types of wireless networks has been widely studied, such as sensor networks, ad hoc networks, cellular networks, vehicular networks, and so forth. Grossglauser and Tse showed how the mobility increases the capacity of ad hoc wireless network in 2002 [10], Li proposed a new method to calculate the multicast capacity of wireless ad hoc networks in 2009 [11], Garetto and Leonardi proved restricted mobility improves delay-throughput tradeoffs in mobile ad hoc networks [12], and Mao et al. calculated the multicast capacity for hybrid wireless networks [13] in 2012. There are also some people who work on the capacity of energy-constrained networks [14]; some people study the capacity of arbitrary networks [15] and inhomogeneous networks [16], some people let nodes cooperate with a hierarchical MIMO network [17], and some people use network coding [18] to improve the network capacity.

The capacity of vehicular ad hoc networks was first studied in 2007 by Pishro-Nik et al. [19]. They proposed a grid-like framework as shown in Figure 1 to demonstrate the streets of urban region. We have abstracted one real construction from

Figure 1: The real map of urban area.

the real urban map. However, we realized we also have to use statistic methods to calculate the capacity of high mobility social M2M networks. If the statistic methods were applied in the calculation, the detail of the real construction will be neglected in the calculation. Thus, there is no difference between real construction and grid-like construction. Furthermore, grid-like construction is easy to understand and convenient to be used in calculation. As a result, we choose the grid-like construction to be the geometrical construction in our calculation. Lu et al. extended their work and obtained the per-vehicle throughput $\Omega(1/\log(n))$ and delay $O(\log^2(n))$ [20]. However, Lu et al. only consider the scenario that source vehicles can only transmit packets to the vehicle which belongs to the same community as the source vehicle and they only consider the unicast transmission. There are rare studies that work on the calculation of capacity for different VANETs scenarios. The capacity of different typical ad hoc scenarios cannot be adopted for theoretical analysis for VANETs and cannot provide guidelines in designing VANETs. Therefore, we calculate the multicast capacity for bus-assisted VANETs in this paper.

Assume that n ordinary vehicles and n_b buses are deployed in a grid-like road framework. This construction is composed of m parallel lines with other intersected m parallel lines. Since the larger urban region means more number of vehicles, m increases linearly with n. Vehicles that have social proximity mobility always move in a localized area centered at the driver's home or work space and seldom move out. Thus, we let all the ordinary vehicles obey the restricted mobility model. As a result, the spatial stationary distribution of vehicles decays as the form of power law with the distance between the home point and each vehicle. Each ordinary vehicle randomly chooses $k1$ vehicles from the other ordinary vehicles as receivers. The network can be considered as unicast when k is 2 and as broadcast when k is $n1$.

In the forwarding scenario, all the buses deployed in all roads as intermediate nodes are used to forward packets for ordinary vehicles. The buses of each road were considered as one bus trajectory. The bus forward scheme was proposed in [21]. Each transmission flow transmits packets via two-hop relay scheme proposed in [22]. All the forward processes of buses are considered as one-hop. The packets could be transmitted directly from source to destination or be transmitted to an intermediate vehicle or bus, then be forwarded to the destination. In the routing scenario, we use buses and ordinary cars to construct the highway system. We use percolation theory to prove that there are highway path clusters that cross through the network vertically and horizontally. Thus, based on a simple routing protocol, the highway system can ensure the packets can be transmitted to destination located in any place of VANET. Simultaneously, the urgent TTL requirement can be satisfied.

To schedule the interference in MAC layer of wireless transmission, we use the protocol model [6] as our interference model in this paper. We assume the bandwidth of wireless transmission is W_a bits per time slot and the bandwidth of bus system is W_b bits per time slot. For simplicity, we also assume that there is only one wireless channel in the bus-assisted forwarding and routing and all the vehicles and buses have enough memory to buffer all the packets.

Our Main Contributions. With the purpose of transmitting packets to different communities of VANETs and further improving the network capacity, in this paper we derive the upper and lower bound of the multicast capacity for bus-assisted VANET in two kinds of scenarios.

In the forwarding scenario, the per vehicle capacity of bus-assisted VANET is as follows:

$$O\left\{\max\left[\frac{W_a}{k}, \min\left(\frac{W_b\sqrt{c}}{n\sqrt{k}}, \frac{W_a c}{kn}\right)\right]\right\} \quad \text{if } k = O(c),$$
$$O\left\{\max\left[\frac{W_a}{k}, \min\left(\frac{W_b}{n}, \frac{W_a c}{kn}\right)\right]\right\} \quad \text{if } k = \Omega(c) \tag{1}$$

and $w.h.p.$ scales of at least $\Omega((k-1)/\log(n))$.

In the routing scenario, we construct the highway system to transmit the urgent packets. We also derive the capacity of per vehicle via bus–assisted mode, it can achieve at most $O(1/kd)$ and cannot be lower than $\Omega(1/\sqrt{k}d)$.

The rest of the paper is organized as follows. In Section 2 we represent the network model in detail. The upper and lower bounds of multicast capacity by ad hoc forwarding are derived in Section 3. In Section 4 we calculate the upper and lower bounds of multicast capacity by bus forwarding. We combine the results of Sections 3 and 4 and obtain the multicast capacity for bus-assisted forwarding in Section 5. Section 6 constructs the highway path and derives the capacity of bus-assisted VANET in routing scenario. Section 7 discusses the results and the remaining challenges. Section 8 concludes this paper and reviews the results on multicast capacity for bus-assisted VANETs.

2. Network Model

2.1. Network Geometry. We use the grid-like construction (as shown in Figure 1) to represent the urban region. The construction comprises m parallel streets intersected with other m parallel streets. Each street considers a bus line. Let s denote the total number of road segments (street section between any two street lines) and let c denote the number of squares in the grid-like construction. Thus, $s = 2c = 2(m-1)^2$. The density of the network is $d = n/s = n/2(m-1)^2$. One has the experience that the center spots (such as workplaces, markets, and malls) always have high vehicular density than other locations. If we use a mathematical method to analyse this phenomenon, we can find out that the density of one location is steady. Furthermore, the density of the location would decrease as the distance from the center spot increases. Therefore, we can derive the density of one location according to its distance from the center spot. From Figure 1, we assume that Tier (1) is one of the center spots, so it has the highest vehicular density. According to the formulation calculated in [17], we can get the density of Tier (2) and Tier (3) based on their distance from Tier (1). Obviously, Tier (3) has lower vehicular density than Tier (2) due to its longer distance from Tier (1) than Tier (2). Larger urban region means the more number of vehicles, so we assume d is a constant to make the number of street segments s increase linearly with the number of vehicles n.

2.2. Mobility Model. Vehicles always move in a localized area centered at the driver's home or work space and seldom move out. Thus, we let all the ordinary vehicles obey the restricted mobility model. Each vehicle randomly selects one square as its home-point (as Tier (1) which is showed in Figure 1). Tier (2) was formed by the adjacent squares surrounding Tier (1). Tier (3) was formed so forth. The restricted mobility region is Tier (α), $\alpha \in \{1, 2, \ldots, \omega\}$, where Tier (ω) is the outermost tier of the mobility region of a vehicle. The stationary distribution of vehicles on the road segments of Tier (α) is denoted by π_α. Thus, the steady-state location of vehicles is modeled by power law function $\pi_\alpha = \alpha^{-\gamma}\pi_1$, where exponent $\gamma > 0$. Moreover $\pi_1 = 1/\sum_{\alpha=1}^{\omega}(16\alpha - 12)\alpha^{-\gamma}$ is calculated in [17]. All the buses are deployed in all roads as intermediate nodes. They were used to forward packets for ordinary vehicles in the forwarding scenario and construct highway system with ordinary vehicles in the routing scenario. The buses of each road construct one bus line. Thus, there are $2m$ bus lines in the network.

2.3. Communication and Interference Model. For simplicity, we assume that there is only one wireless channel in the bus-assisted forwarding and during a time slot t the wireless channel can only transmit W_a bits. The bandwidth of the bus system is W_b bits per time slot. Due to the interference of wireless transmission, one vehicle cannot transmit with two other vehicles at the same time slot. Let r denote the transmission range. The length of one segment is equal to r to ensure the transmission range covers the entire road segment. Thus, $r = (1/m - 1)$. To schedule the transmission flows, we adopt the protocol interference model. Protocol interference model schedule is defined as follows.

At each time slot, a transmission from vehicle i to vehicle j is successful only if the following inequality stands:

$$\left\| X_i(t) - X_j(t) \right\| \le r \tag{2}$$

and for any other vehicle l that transmits at t:

$$\left\| X_l(t) - X_j(t) \right\| \le (1 + \Delta)\, r, \tag{3}$$

where Δ is a guard factor defining a protection zone around the receivers and $x_i(t)$ denotes the location of vehicle i at the time slot t; we will add the missing definition in the paper.

2.4. Traffic Model and Relay Scheme. We consider n multicast flows existing in the network concurrently. Each packet is transmitted to $k-1$ destinations. Each vehicle is the source of one multicast flow and one of the destinations of another multicast flow. The network can be considered as unicast when k is 2 and as broadcast when k is $n1$. There are three different forwarding strategies in the bus-assisted VANETs: ad hoc forwarding, bus forwarding, and bus-assisted forwarding. With the first forwarding method, multicast flows transmit packets via two-hop relay scheme. With the application of two-hop relay scheme, a vehicle can transmit packets to a destination directly or relayed through one intermediate vehicle that has more contact opportunities with the destination vehicle and all the intermediate vehicles are ordinary vehicles. Since each vehicle has the restricted mobility region, the source and destination have the same home-point and k is less than or equal to the number of vehicles that have the same home-point. The second forwarding method uses bus as intermediate vehicle. The forwarding scheme of buses is discussed in [18]. In this forwarding method, the bus is used in intercommunity communication. Thus, the source and destination have different home-points. The third forwarding method combines the above two forwarding methods.

2.5. Definitions of Capacity. In this paper, capacity denotes the feasible throughput. The capacity of VANETs is defined as follows.

Definition 1 (feasible throughput). A throughput of $\lambda(n)$ bits per second for each vehicle is feasible if there is a spatial and temporal scheme for scheduling transmissions, and every vehicle can send $\lambda(n)$ bits per second on average to its chosen destination.

Definition 2 (capacity of vehicle network). The average capacity of vehicular network is of order $\theta(g(n))$ (we use Knuth's notation: given two functions $g(n) \ge 0$ where $f(n) = O(g(n))$ means $\limsup_{n\to\infty} f(n)/g(n) = c < \infty$; $f(n) = \Omega(g(n))$ is equivalent to $f(n) = O(g(n))$; $f(n) = \Theta(g(n))$ means $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$) bits per second if there are deterministic constants $c > 0$ and $c < c' < +\infty$ such that

$$\begin{gathered} \lim_{n\to\infty} \Pr\left(\lambda(n) = c\left(g(n)\right) \text{ is feasible}\right) = 1, \\ \lim_{n\to\infty} \inf \Pr\left(\lambda(n) = c\left(g(n)\right) \text{ is feasible}\right) < 1. \end{gathered} \tag{4}$$

Definition 3 (throughput capacity). Let $G(T)$ denote the amount of data received by all the vehicles during time T. A capacity throughput $\lambda(n)$ is feasible if there is a scheduling scheme for which the following properties hold:

$$\lim_{T\to\infty} \Pr\left(\frac{G(T)}{T} \geq \lambda\right) = 1. \quad (5)$$

3. Bounds in Multicast Capacity by Ad Hoc Forwarding

Forwarding scenario has two parts: ad hoc forwarding parts and bus forwarding parts. In this section, we discuss the bounds of capacity for pure ad hoc forwarding for the forwarding scenario. All the packets are transmitted to destination directly or via an ordinary vehicle. Since source vehicle and destination vehicle have the same home-point, k is less than or equal to the number of vehicles that have the same home-point. Broadcast cannot be achieved only with this forwarding method.

3.1. Upper Bound in Multicast Throughput Capacity by Ad Hoc Forwarding. We first calculate the upper bound of multicast throughput capacity by ad hoc forwarding. The following lemma is used in the calculation of upper bound of throughput capacity of ad hoc forwarding. This lemma was proved in [17].

Lemma 4. *Let p_a denote the probability of a road segment being active as follows:*

$$p_a = \frac{1}{(2\lceil 1+\Delta\rceil\lceil 2+\Delta\rceil)}, \quad (6)$$

where $\lceil x\rceil$ represents the smallest integer number greater than or equal to x.

According to Lemma 4, we first derive the upper bound in multicast throughput capacity by ad hoc forwarding.

Theorem 5. *For the social proximity VANETs, with two-hop scheme, the average per-multicast throughput capacity by ad hoc forwarding cannot be better than $1/(2kd\lceil 1+\Delta\rceil\lceil 2+\Delta\rceil)$.*

Proof. Let $G_d(T)$ denote the amount of data transmitted through direct transmission from source to destination during the time interval $[0, T]$ and let $G_r(T)$ denote the amount of data transmitted through relay transmission during the time interval $[0, T]$. According to Definition 3, throughput $\lambda(n)$ satisfies the following inequality:

$$\frac{G_d(T) + G_r(T)}{T} \geq kn\lambda(n) - \varepsilon, \quad (7)$$

where $\varepsilon > 0$ is an arbitrary and fixed number and $\varepsilon \to 0$ as $T \to \infty$. Let $K(T)$ denote the total transmit opportunities during $[0, T]$. The total number of transmitted packets must be less than the total number of transmit opportunities during a long time interval. Since the relay transmission needs the transmitting opportunities twice to transmit one packet, we have the following:

$$\frac{1}{T}K(T)W_a \geq \frac{1}{T}G_d(T) + \frac{2}{T}G_r(T). \quad (8)$$

Substituting (7) into (8), we have the following:

$$\frac{1}{T}K(T)W_a \geq \frac{1}{T}G_d(T) + 2\left(kn\lambda(n) - \varepsilon - \frac{1}{T}G_d(T)\right). \quad (9)$$

Sorting (9), we have the following:

$$\lambda(n) \leq \frac{(1/T)K(T)W_a + (1/T)G_d(T) + 2\varepsilon}{2kn}. \quad (10)$$

When $\varepsilon \to 0$ as $T \to \infty$

$$\lambda(n) \leq \frac{(1/T)K(T)W_a + (1/T)G_d(T)}{2kn}. \quad (11)$$

Due to the interference of wireless transmission, the total transmission must be less than the concurrent transmissions during time $[0, T]$. According to the law of large numbers, we have the following:

$$\lim_{x\to T}\frac{1}{T}K(T)W_a \leq sp_aW_a. \quad (12)$$

Similarly, we have the following:

$$\lim_{x\to T}\frac{1}{T}G_d(T) \leq sp_aW_a. \quad (13)$$

The two equalities hold when there is always a transmission flow exits on each unit of a concurrent transmissions group during each time slot. According to Lemma 4, by substituting (12) and (13) into (11), we have the following:

$$\lambda(n_v) \leq \frac{sp_a}{kn} = \frac{p_a}{kd} = \frac{1}{2kd\lceil 1+\Delta\rceil\lceil 2+\Delta\rceil}. \quad (14)$$

Thus, the theorem then follows. □

3.2. Lower Bound in Multicast Throughput Capacity by Ad Hoc Forwarding. To obtain the lower bound of average multicast throughput capacity, we introduce the following lemma which was proved in [17].

Lemma 6. *The number of vehicles whose mobility region contains road segment i is denoted by N. Thus, N scales as $O(\log(n))$, when $n \to \infty$, the probability approaches 1.*

With the two-hop relay scheme, a packet can be successfully transmitted only if there exists at least one source-destination pair or source-intermediate pair when the road segment is active. Since sources and destinations have the same home-point, there is one source-destination pair or source-intermediate pair with probability $(k-1)/2N$. According to Lemma 6, the probability of success transmission is at least $\Omega((k-1)/\log(n))$. Therefore, we derive the following theorem.

Theorem 7. *The throughput capacity of average per-multicast flow can be scaled at least $\Omega((k-1)/\log(n))$ w.h.p.*

4. Bounds in Multicast Capacity by Bus Forwarding

In this section, we calculate the bounds of multicast capacity by bus forwarding for the forwarding scenario. All the packets are relayed by bus, and all the ordinary vehicles can be the destinations of any source vehicle. With the bus forwarding, k can be equal to n to achieve broadcast in the network.

4.1. Upper Bound in Multicast Throughput Capacity by Bus Forwarding. We first calculate the upper bound in multicast throughput capacity by bus-to-bus transmissions. We introduce the Euclidean tree to demonstrate the bus transmission process of multicast flows and each segment of the bus transmission process used is one edge of a multicast tree. Let T_i denote the ith multicast tree and let $S(T_i)$ denote the number of segments the tree T_i will use. The total number of used segments of total transmissions in the network is denoted by $L = \sum_{i=1}^{n} S(T_i)$. To obtain the value of L, we introduce the following lemma which was proved in [20].

Lemma 8. *Given n nodes randomly and uniformly distributed in a 2-dimensional cube, divide the cube into c cells as Voronoi diagrams with the same side length. Each node transmits packets to k destination concurrent via base-station. The base-station forwarding of a transmission is considered as a Euclidean tree. When $k = O(m)$, with probability of at least $1 - 2e^{-n\theta^2/32}$, the total edge number of all Euclidean tree is $L \geq n\theta\sqrt{km}/16$. When $k = \Omega(m)$, with probability at least $1 - 2e^{-n/8}$, then $L \geq nm/4$.*

Different from the base-station connected by fiber, the ones connected by bus carry packets and move to the destination along roads represented by segments in the grid construction. However, if we consider the intersections of road segments as the vertex of Euclidean tree and each intersection belonging to a unique square, we can obtain the same conditions with Lemma 8. Thus, the results are suitable for the Euclidean tree of bus-to-bus transmissions. According to the above analysis, we derive the following corollary.

Corollary 9. *In the grid-like construction, if one uses Euclidean tree which represents the bus-to-bus transmissions, one can have the following results. When $k = O(m)$, with probability at least $1 - 2e^{-n\theta^2/32}$, the total edge number of all Euclidean tree is $L \geq n\theta\sqrt{km}/16$. When $k = \Omega(m)$, with probability at least $1 - 2e^{-n/8}$, then $L \geq nm/4$.*

Recall that there are s segments in the construction. Then according to Pigeonhole principle, when $k = O(c)$, there is at least one cell that will be used by at least $(n\theta\sqrt{kc}/16)/c$ flows with probability at least $1-2e^{-n\theta^2/32}$ and when $k = \Omega(c)$, there is at least one cell that will be used by at least $n_s/4$ flows with probability at least $1-2e^{-n/8}$. Let W_b denote the packets number of transmission by bus during one time slot; then we have the following theorem.

Theorem 10. *When $k = O(c)$, the per-multicast flow throughput capacity of bus-to-bus transmission is at most $W_b\sqrt{c}/n\theta\sqrt{k}$ w.h.p. When $k = \Omega(c)$, the per-multicast flow throughput capacity of bus-to-bus transmissions is at most $4W_b/n$ w.h.p.*

Theorem 10 is the capacity of bus-to-bus transmissions; buses have to transmit packets to ordinary vehicles at last. Then we calculate the up bound of the transmission between buses and ordinary vehicles for bus forwarding method. Recall that the active probability of one segment is p_a. The total transmission opportunities are at most sp_a. Each packet firstly is transmitted to bus then is transmitted to ordinary vehicles by bus. One multicast flow totally has k links. Thus, according to Pigeonhole principle, at least one link of a multicast has at most p_a/kn transmission opportunities. Then we derived the upper bound of the transmission between buses and ordinary vehicles for bus forwarding method.

Theorem 11. *The upper bound of the throughput capacity between bus and ordinary vehicles for bus forwarding is $W_a sp_a/kn = W_a p_a/kd$.*

Obviously, the minimum throughput of bus forwarding process and bus-to-vehicle process determines the throughput of whole bus forwarding. By summarizing Theorems 10 and 11, we derive the upper bound of multicast throughput capacity for bus forwarding.

Theorem 12. *The upper bound of the multicast capacity for bus forwarding is as follows:*

$$\begin{aligned} &O\left[\min\left(\frac{W_b\sqrt{c}}{n\sqrt{k}}, \frac{W_a c}{kn}\right)\right] \quad \text{if } k = O(c), \\ &O\left[\min\left(\frac{W_b}{n}, \frac{W_a c}{kn}\right)\right] \quad \text{if } k = \Omega(c). \end{aligned} \tag{15}$$

4.2. Lower Bound in Multicast Throughput Capacity by Bus Forwarding. By using the bus forwarding method, a packet can be successfully transmitted by bus only if there exists at least one bus-destination vehicle when the road segment is active. Reference to the proof of Lemma 6, we can easily prove the following lemma.

Lemma 13. *The number of vehicles whose mobility region contains bus line i is denoted by N_b. Thus, N_b scales as $O(\log(n))$ w.h.p.*

According to Lemma 13, we know that in any segment there is one bus-destination vehicle pair with probability $(k-1)/N_b$ and the probability of success transmission is at least $\Omega(k/\log(n))$. Therefore, we derive the following theorem.

Theorem 14. *The throughput capacity of average per-multicast flow can be scales at least $\Omega((k-1)/\log(n))$ w.h.p.*

5. Capacity Bounds for Bus-Assisted Forwarding

In this section, we will analyze the throughput capacity for bus-assisted forwarding. The analysis is based on the results

of ad hoc forwarding and bus forwarding derived above. With the purpose of transmitting packets to other communities that have different home-point and further improving the network capacity, both ad hoc forwarding and bus forwarding are used in bus-assisted forwarding. In particular, bus forwarding is the only way to transmit packets to other communities. According to the destination of packets, bus-assisted forwarding adaptively selects a better forwarding method. Therefore, the throughput capacity of bus-assisted forwarding cannot surpass the maximum capacity of ad hoc forwarding and bus forwarding. The maximum capacity of ad hoc forwarding and bus forwarding is optimum for bus-assisted forwarding. Similarly, bus-assisted forwarding has the same lower bound with ad hoc forwarding and bus forwarding. According to the analysis above, we can obtain the bounds of multicast capacity for bus-assisted forwarding as the following theorem.

Theorem 15. *The upper bound of the multicast capacity for bus-assisted VANET is as follows:*

$$O\left\{\max\left[\frac{W_a}{k}, \min\left(\frac{W_b\sqrt{c}}{n\sqrt{k}}, \frac{W_a c}{kn}\right)\right]\right\} \quad \text{if } k = O(c),$$
$$O\left\{\max\left[\frac{W_a}{k}, \min\left(\frac{W_b}{n}, \frac{W_a c}{kn}\right)\right]\right\} \quad \text{if } k = \Omega(c) \tag{16}$$

and w.h.p. scales of at least $\Omega((k-1)/\log(n))$.

6. Capacity of Highway Routing

The forwarding protocol can transmit packets with tiny delivery cost. However, the delay cannot be diminished by the protocol because it mainly depends on the velocity of the bus. For the urgent packets, the forward scenario cannot match the requirement. Thus, the capacity of forward scenario cannot satisfy the transmission of urgent packets. For the purpose of calculating the capacity of urgent packets transmissions in the bus-assisted VANET, we designed a basic routing protocol for the bus-assisted VANET. We call it the highway routing. Then we calculate the capacity with the highway routing.

6.1. Introduction of Percolation Theory. The highway system is constructed based on the percolation model on square lattice. Before we construct the highway, we first introduce the bond percolation of percolation theory.

Assume that some packets are generated on top of the network region. Will the packets be able to make their way from edge to edge and reach the bottom? This question is modeled mathematically as a two-dimensional network of $n\times n$ vertices, in which the edge between each two neighbors may be open (allowing the liquid through) with probability p or closed with probability $1-p$. For the purpose of transmitting packets all over the network, we have to derive the probability that an open path exists from the top to the bottom. For the regular square lattice, if there is an open path from the top to the bottom the open path from left to right of the network also exists.

Figure 2: Construction of highway path.

The open probability p of the edge is independent. The probability of an open path existing is determined by p. The single open path is not enough to ensure the all over network transmissions. As the number of vehicles increases to infinity, the network construction increases to infinity. Thus, there must be infinite open path clusters to ensure all over network transmissions.

By Kolmogorov's zero-one law, in the regular square lattice, for any given p the probability that an infinite open path cluster exists is either zero or one. The probability of p is an increasing function that was proved in [19]. It increases sharply from approach zero to one in a short span of P. Therefore, there must be a critical p (denoted by p_c) determining the probability P that an infinite open path cluster exists in the regular square lattice. When p is below p_c, the probability P is zero. When p is above p_c, the probability P is one and an infinite open path cluster exists in the regular square lattice.

6.2. The Construction of Highway. As in Figure 2, we assume that the intersection is the center of a virtual square. The road between two intersections is one edge. The virtual square lattice is denoted by red dotted lines. The black lines denote the open paths. To calculate the p_c in the square lattice of Figure 2, we introduce the definition of the coordination number which is denoted by z in this paper. It means the total number of neighbors of center intersection. Obviously in the regular square lattice in Figure 1, the coordination number is four. According to the result of Harry Kesten [20], the probability threshold p_c is $1/(z-1)$. Therefore, the open edge cluster exits only if the probability that the edge between two squares is open is larger than 1/3.

Packets are forwarded by intermediate vehicles. In the network, each edge between two squares is open if the road between the two squares has at least one vehicle. Thus, the probability of each road having at least one vehicle must be larger than 1/3 to make sure the open edge cluster exits in the network. For the regular road segment without

Figure 3: Construction of highway path.

buses, in any time slot the probability of not finding any vehicles is $\prod_{\alpha=1}^{\omega}(1-\pi_\alpha)N_i^{\alpha}$, where N_i^{α} denotes the road segment number of layer α. Thus, we have the probability that road segment i has at least one vehicle. It is also the open probability p_i of edge i as follows:

$$p_i = 1 - \prod_{\alpha=1}^{\omega}(1-\pi_\alpha)N_i^{\alpha}. \tag{17}$$

For any road segment, $p_i \geq \pi_1^2$. Obviously π_1^2 is less than 1/3. Thus, only ordinary vehicle cannot ensure the opening edge cluster in the network. The highway cannot be constructed by ordinary vehicles. However, the regular buses can significantly increase the probability that at any time slot the road segment has at least one vehicle. According to the real mobility trace collected by [21], we can know the probability p_b that in any time slot the road segment has a bus larger than 1/5. Adding p_a and p_b, we can derive the probability that the edge between two squares has at least one vehicle is bigger than 1/3. Therefore, with the bus-assisted roads, we can construct the highway path to transmit the urgent packets in the VANET.

6.3. Highway Routing Protocol. Based on the open edge cluster derivation above, we use basic routing protocol to transmit packets in VANETs. Following the protocol, all source vehicles upload packets to the highway path and then packets are transmitted through the highway path until they approach the destinations. Destination vehicles will download packets from highway path. Time slots will be well arranged to ensure the highway path has priority to occupy transmit opportunity.

Upload. Source vehicles upload the packets to highway path when there are some transmission opportunities. Otherwise, the source vehicle can add the packets to highway transmission flow when it is chosen as the intermediate vehicle of the highway path.

Routing. Packets are forwarded along the shortest highway path to the destination or the intermediate vehicle besides the destination. If the shortest bus path has a closed edge, then highway path will detour to avoid the closed path.

Download. Packets are forwarded to the destination vehicle through the highway path. The destination vehicle downloads the packet from the highway path when it has opportunity to access the link. Otherwise, the destination vehicle can get the packets from highway transmission flow when it is chosen as the intermediate vehicle of the highway path.

Figure 3 is used to show a simple routing process. Source vehicle S can be assumed as the highway path and transmit packets to destination vehicle D. When road a has no vehicle, the highway path detours to avoid road a. Then, packets will achieve the destination vehicle or the neighbor of destination D. Destination vehicle will get the packets or download them from its neighbor which is in the highway path.

6.4. Calculation of Capacity. When vehicle upload or download packet from the highway path, one packet occupies one transmission opportunity. That process is equivalent to transmit one packet through one edge. Thus, the upload and download process can be assumed as the first edge and the last edge of the highway path. The transmission opportunities are arranged by protocol, and thus an interference group can fully use the transmission opportunities. Each packet will cost one transmit opportunity to pass through one edge. The network has total sp_a transmission opportunities. When packets can be transmitted to each destination, they only pass through one edge. The capacity of each vehicle can achieve at most $(sp_a/kn)W_a$. The upper capacity bound of VANET with highway is identical with the forward scenario. However, the lower capacity bound of VANET with highway is different from forward scenario. To calculate the upper capacity bound, we introduce a lemma proved by Neely and Modiano in [22].

Lemma 16. *Given a square b in the lattice, the probability that a random highway path will be routed via the square s is at most $c\sqrt{k}\cdot(r/a)$, where c is a constant number.*

Therefore, a square b can be used by at most $nc\sqrt{k}\cdot(r/a)$ packets. We consider the protocol can give the opportunity to the busier square. Thus, the busiest square will have more transmission opportunities than other squares in the same interference group. Thus, the opportunities of the busiest square b must be more than the average. Thus, we can easily derive the lower capacity bound in each vehicle as follows:

$$\lambda(n) \geq \frac{p_a W_a}{n^2 c\sqrt{k}\cdot(r/a)} = \frac{W_a}{2cn^2\sqrt{k}\lceil 1+\Delta\rceil\lceil 2+\Delta\rceil}. \tag{18}$$

The above analysis is summarized by the following theorem.

Theorem 17. *With the application of highway protocol, the per-vehicle capacity of VANET can be achieved at most $O(1/kd)$ and cannot be lower than $\Omega(1/\sqrt{k}d)$.*

7. Discussion

Notice that $c = \Theta(m^2)$ and we did not use $\Theta(m^2)$ in the capacity results of bus routing method. The reason is that we use m to denote both road and bus lines in the grid-like construction in this paper, and the roads number is equal to the number of bus lines. However, the cell used in bus routing is constructed by bus lines not roads. If we use a different number of bus lines, the cell number will be different from the number of squares in the grid-like construction. Thus, for the capacity of scenarios with different bus lines, we only need to replace the value of c.

To calculate the bounds of multicast for the bus-assisted VANET, we assume that the TTL (time to live) of packets is infinite. However, in the realistic VANET, TTL is one of the most important characteristics of the packets in any kind of ad hoc network. Therefore, if we can tolerate the delay of transmission, the study of achievable capacity is also essential for the bus-assisted VANET. We will focus on the tradeoff between capacity and delay in the future work. Similarly, applying more real interference model is also essential such as physical interference model and Gaussian interference model. All of our results are derived under grid-like construction. A more realistic framework may close the gap of capacity between theoretic results and real value. We will consider all the remaining challenges in the future work.

We calculate the achievement per-vehicle capacity for the forwarding scenario and routing scenario of bus-assisted VANETs. The forwarding scenario can save lots of energy and transmission coasts to diminish the overhead of bus-assisted VANETs. The routing scenario can transmit packets within very little time to satisfy the urgent packets by sacrificing the overhead of the network. We just derive the performance. The selection of transmission scenario needs an additional protocol.

8. Conclusion

The capacity scaling law of high mobility M2M networks has been considered as one of the most fundamental issues. In this paper, we derive the upper and lower bounds of multicast capacity for high mobility social proximity M2M networks via bus-assisted forwarding method. In the routing scenario, we use buses and ordinary cars to construct the highway system for VANETs, which is a typical case of high mobility M2M networks. We use percolation theory to prove that there is a highway path cluster cross through the network vertically and horizontally. Therefore, the highway system can ensure the packets can be transmitted to destination located anywhere in VANETs. The per-vehicle capacity of routing scenario is also derived. At last we discussed how the different forwarding processes influence the results of capacity scaling law for high mobility M2M networks. Our work provides new insights for the design of bus-assisted VANETs as intermediate vehicle to relay packets.

Acknowledgment

This work is supported by Heilongjiang Province Education Department Foundation 12531Z007.

References

[1] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: architectures, standards, and QoS improvement," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, 2011.

[2] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive machine-to-machine communications: visions and potentials for the smart grid," *IEEE Network Magazine*, vol. 26, no. 3, pp. 6–13, 2012.

[3] H. Hartenstein and K. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, 2008.

[4] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, 2005.

[5] M. A. Hanson, H. C. Powell Jr., A. T. Barth et al., "Body area sensor networks: challenges and opportunities," *Computer*, vol. 42, no. 1, zpp. 58–65, 2009.

[6] G. Nan, Z. Mao, M. Li et al., "Distributed resource allocation in cloud-based wireless multimedia social networks," *IEEE Network Magazine*. In press.

[7] G. Nan, Z. Mao, M. Yu, M. Li, H. Wang, and Y. Zhang, "Stackelberg game for bandwidth allocation in cloud-based wireless live-streaming social networks," *IEEE Systems Journal*, no. 99, 2013.

[8] E. Palomar, A. Alcaide, E. Molina, and Y. Zhang, "Coalitional games for the management of anonymous access in online social networks," in *Proceedings of the 11th International Conference on Privacy, Security and Trust (PST '13)*, pp. 1–10, Tarragona, Spain, July 2013.

[9] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[10] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, 2002.

[11] X.-Y. Li, "Multicast capacity of wireless ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 950–961, 2009.

[12] M. Garetto and E. Leonardi, "Restricted mobility improves delay-throughput tradeoffs in mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5016–5029, 2010.

[13] X. Mao, X.-Y. Li, and S. Tang, "Multicast capacity for hybrid wireless networks," in *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '08)*, pp. 189–198, Hong Kong, May 2008.

[14] R. Zheng, "Asymptotic bounds of information dissemination in power-constrained wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 1, pp. 251–259, 2008.

[15] O. Goussevskaia, R. Wattenhofer, M. M. HalldórssoN, and E. Welzl, "Capacity of arbitrary wireless networks," in *Proceedings of the 28th Annual IEEE International Conference on Computer Communications (INFOCOM '09)*, pp. 1872–1880, Rio de Janeiro, Brazil, April 2009.

[16] G. Alfano, M. Garetto, and E. Leonardi, "Capacity scaling of wireless networks with inhomogeneous node density: lower bounds," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 1890–1898, Rio de Janeiro, Brazil, April 2009.

[17] C. Jiang, Y. Shi, Y. T. Hou, and S. Kompella, "On the asymptotic capacity of multi-hop MIMO ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1032–1037, 2011.

[18] Z. Li and L. C. Lau, "A constant bound on throughput improvement of multicast network coding in undirected networks," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1016–1026, 2009.

[19] H. Pishro-Nik, A. Ganz, and D. Ni, "The capacity of vehicular ad hoc networks," in *Proceedings of the Annual Allerton Conference*, september 2007.

[20] N. Lu, T. H. Luan, M. Wang, X. Shen, and F. Bai, "Capacity and delay analysis for social-proximity urban vehicular networks," in *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (INFOCOM '12)*, pp. 1476–1484, Orlando, Fla, USA, 25-30 March 2012.

[21] W. K. Lai, K.-T. Yang, and M.-C. Li, "Bus assisted connectionless routing protocol for metropolitan VANET," in *Proceedings of the 5th International Conference on Genetic and Evolutionary Computing (ICGEC '11)*, pp. 57–60, Xiamen, Cihna, September 2011.

[22] M. J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad hoc mobile networks," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1917–1937, 2005.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 870430, 9 pages
http://dx.doi.org/10.1155/2013/870430

Research Article

Survey of the DASH7 Alliance Protocol for 433MHz Wireless Sensor Communication

Maarten Weyn, Glenn Ergeerts, Luc Wante, Charles Vercauteren, and Peter Hellinckx

CoSys-Lab, Faculty of Applied Engineering, University of Antwerp, Paardenmarkt 92, 2000 Antwerp, Belgium

Correspondence should be addressed to Maarten Weyn; maarten.weyn@uantwerpen.be

Received 18 July 2013; Accepted 22 October 2013

Academic Editor: Jianhua He

Copyright © 2013 Maarten Weyn et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

433 MHz is getting more attention for Machine-to-Machine communication. This paper presents the DASH7 Alliance Protocol, an active RFID alliance standard for 433 MHz wireless sensor communication based on the ISO/IEC 18000-7. First, the major differences of 433 MHz communication compared to more frequently used frequencies, such as 2.4 GHz and 868/920 MHz are explained. Subsequently, the general concepts of DASH7 Alliance Protocol are described, such as the BLAST networking topology and the different OSI layer implementations, in a top-down method. Basic DASH7 features such as the advertising protocol, ad-hoc synchronization and query based addressing are used to explain the different layers. Finally, the paper introduces a software stack implementation named OSS-7, which is an open source implementation of the DASH7 alliance protocol used for testing, rapid prototyping, and demonstrations.

1. Introduction

Machine-to-Machine (M2M) communication can be organized using a wide area network (WAN), such as a mobile network, or a personal or local area network (PAN/LAN). Sensor communication is typically done using a PAN with a backbone connection using a wired or wireless WAN [1]. According to Beale and Morioka [1], the most frequently used technologies for wireless PAN are Wi-Fi (IEEE 802.11), Zigbee (which is based on IEEE 802.15.4), Z-Wave, and KNX-RF. The frequencies used by these technologies are 2.4 GHz, 868, and 915 MHz. Fadlullah et al. [2] also identify Bluetooth, Ultra Wide Band (UWB) and 6LoWPAN as prominent technologies suitable for M2M communication. Besides the aforementioned frequencies, 433 MHz is also gaining relevance in the area of M2M communication, as stated by Tuset-Peiro et al. [3].

Niyato et al. [4] also identify an neighborhood area network (NAN) which typically connects different homes using a so called concentrator to a base station of the WAN. They focus on home energy management. Such application consolidate the search for other low-power RF technologies which have a longer range than typical PAN technologies. This paper will introduce such a technology: the Dash7 Alliance protocol.

The DASH7 Alliance protocol (D7A) [5] is an active RFID alliance standard for 433 MHz wireless sensor communication based on the ISO/IEC 18000-7 standard maintained by the DASH7 Alliance. ISO/IEC 18000-7 defines parameters of the active air interface communication at 433 Mhz. D7A is built on top of an asynchronous Wireless Sensor Network (WSN) Media Access Control (MAC). In contrast to typical WSN standards such as ZigBee (built on top of IEEE 802.15.4), the DASH7 specification defines a full functional RFID tag. This means it does include high level functionality optimized for RFID applications. However, it can also be extended for non-RFID applications. In contradiction to legacy RFID systems [6], D7A supports tag-to-tag communication.

DASH7 uses a BLAST network technology. BLAST is an acronym for *Bursty, Light, Asynchronous, Stealth, Transitive*. *Bursty* means abrupt data transfer unlike streaming content such as video or audio. *Light* refers to a limited packet size (256 bytes). Multiple consecutive packets are however supported. *Asynchronous* is one of the key concepts of DASH7. Communication is command-response based

without any periodic synchronization. *Stealth* refers to the fact that DASH7 does not need any periodic address broadcasting, which means nodes can choose to communicate only in trusted environments. *Transitive* refers to the mobile transitional behavior of devices and/or tags.

This paper will first explain the major differences of 433 MHz communication compared to more frequently used frequencies in Section 2. Subsequently, it will describe the general concepts of the D7A specification in Section 3. Finally, it will introduce a software stack implementation OSS-7 in Section 4.

2. 433 MHz Based Sensor Communication

As stated before, 433 MHz is not the most widely used frequency for M2M communication, although it has some serious advantages. First of all, it is an unlicensed band which is almost worldwide applicable; however, it is not harmonized. The band spans from 433.05 to 434.79 MHz. Second, due to its frequency, it has better propagation characteristics opposed to higher frequencies. This is discussed later in this section. However, there are also some disadvantages. The wavelength at the center frequency 433.92 MHz is 69.14 cm. This means an $\lambda/4$ antenna has a length of 17.28 cm. Moreover, the bandwidth of the band is only 1.84 MHz. This limits the data rate of the communication at this frequency. Making 433 MHz ideal for low-power low data rate communication.

Some research has already been done in comparing the propagation aspects of 433 MHz towards other frequencies.

Tuset-Peiro et al. [3] compare the propagation performance of 433 MHz and 2.4 GHz and show that 433 MHz has a better communication range despite the effects of having a larger Fresnel zone. They also show that channel hopping will not solve multipath propagation effects since the channel coherence bandwidth is larger than the whole 433 MHz band.

Isnin [7] compares path propagation in multifloored building for 433, 868 and 1249 MHz. The 868 MHz band has a higher bandwidth then the 433 MHz band which enables a higher data rate. They show that in a waveguided corridor environment with line-of-sight, higher frequencies have an advantage. However, 433 MHz has a better penetration capability. This leads to a lower path loss-level for multi-floor propagation with the number of floor obstruction greater then two. Zhang et al. [8] state that it is very complex to model small scale indoor propagation.

As already stated, the regulations of 433 MHz transmissions are not harmonized.

In Europe, ETSI states for short-range devices (SRD) that the band from 433.05 till 434.790 MHz can be used for nonspecific SRD with 10 mW Effective Radiated Power (ERP) when the duty cycle is less than 10% or when then channel spacing is smaller than 25 kHz or with 1 mW ERP without duty cycle limitations [9].

In USA, FCC states that the field strength can only be 10995 μV/m at 3 meter for periodic applications and 4398 μV/m at 3 meter otherwise [10, 11]. This corresponds with an ERP of −14.4 dBm for periodic control applications and −22.36 dBm otherwise.

Table 1: D7A devices classes.

Device class	Transmits	Receives	Complete feature set	Wake-on scan cycle	Always on receiver
Blinker	*				
Endpoint	*	*		*	
Subcontroller	*	*	*	*	
Gateway	*	*	*		*

3. DASH7 Alliance Protocol

This section will describe the basic concepts of the DASH7 Alliance protocol (D7A).

D7A defines four different device classes as shown in Table 1. A device can switch between classes. A *blinker* device only transmits data and does not use a receiver. For this, a blinker, for example, cannot perform carrier sensing. A second class is an *endpoint*. This is a typical low power device which can transmit and receive data. An endpoint also supports wake-up events. This enables the device to receive requests and typically transmit a response. A *gateway* device is in most cases the device which connects the D7A network to another network. A gateway is obliged to support all D7A features and is never offline. It always listens unless it is transmitting. A *subcontroller* is a full featured device as well, but it is not always active. It uses wake on scan cycles, just like the endpoint devices. This will activate the device for short channel scanning.

D7A describes a full functional RFID tag. Every device supports one or more of the aforementioned+ device classes. For this a correctly configured tag is fully functional without the need for specific application code. Although, in most cases specific application code will be added.

DASH7 supports two communication models: pull and push. As in most RFID systems, dialogs between tags and interrogators are query response based (the pull model), as shown in Figure 1(a). This request response mechanism is described by the *D7A Query Protocol* (D7AQP). Data transfer initiated from the tags to the gateway on the other hand is based on the push model. This is shown in Figure 1(b). This approach can, for instance, be implemented as an automated message or beacon which is sent on specific time intervals. In D7A, this system is called *Beacon Transmit Series*.

DASH7 defines two types of frames: a *foreground frame* and a *background frame*. The foreground frames are regular messages which contain data or data requests. Background frames on the other hand are very short broadcast messages. Background frames are for instance used by the *D7A Advertising Protocol* (D7AAdvP) for rapid ad-hoc group synchronization. This D7AAdvP enables a low power wake-up of those tags which are interrogated in the pull model.

The D7A specification involves all 7 OSI layers [12] and describes the protocol in each of those layers. However, some of the features are still under revision or reserved for future extensions. Figure 2 gives an overview. The remainder of this section describes the D7A implementation of the application

(a) Pull model (e.g., query protocol)

(b) Push model (e.g., beaconing)

Figure 1: Communication models.

layer and how its features are supported by the underlying layers.

The D7A Query Protocol, the Advertising Protocol and the beaconing concept will be used as an example to clarify the different layers.

3.1. D7A Application Layer. The application layer typically describes the interaction between the application software and the communication stack. The application layer enables the software application, which is used by the user, to be able to access communication resources, identify communication partners, and activate certain communication concepts.

In D7A, this layer describes the so called *Application Layer Protocol* (ALP), the *file access protocol*, the *cryptographic authentication protocol*, and access to the *sensor subprotocol*. The ALP makes it possible to send specific data, structured by the application. This is in contrast to D7A specific data protocols which are described in other layers. The cryptographic authentication description is at the time of writing still under revision, but the underlying layers already support the necessary data fields. The sensor subprotocol is based on ISO 24151-7 [13] and specifies the representation of specific sensor data. The file system protocol describes the access methods (read, write, and execute) to the file system which is described in the presentation layer.

Table 2: File data access template.

File ID	Start byte offset	Byte accessing	Data
1 byte	2 bytes	2 bytes	X bytes

Table 3: Beacon transmit period list.

Channel ID	Command parameters	D7AQP call template	Next beacon
1 byte	1 byte	4 bytes	2 bytes

Accessing a file is based on the *File Data Access Template*, shown in Table 2. The File ID refers to a specific file (see Section 3.2: presentation layer). The other two values define the location in the file to be read from or written to. The data field is the actual data to be written.

A beacon transmit series uses an automated process in the Data Link Layer. This process is enabled, disabled and configured by a configuration file described in the presentation layer.

3.2. D7A Presentation Layer. Every D7A device needs to support a number of data elements. D7A has three types of data elements: *Indexed Short File Series Blocks* (ISFSB), *Indexed Short File Blocks* (ISFB), and *Generic file Blocks* (GFB). All of these blocks have their own permission code. This code defines the read, write an execute rights of the root user, current user and guests. An ISFSB is a collection of ISFB files. ISFB Files are stored as structured, mixed-data strings. ISFB IDs specify different files. For example, 0x00 specifies the network configuration, 0x01 specifies device capabilities and supported features, and 0x06 specifies a time controlled sequence for beaconing,. . .. The beaconing will be further explained in Section 3.6.

The data defined by the presentation layer for beaconing is shown in Table 3. The *Channel ID* is the physical layer Channel ID, defined in Section 3.7. The *Command Parameters* are bitfield parameters that map to data link layer and D7AQP parameters. The *D7AQP Call Template* defines the file or file series template which will be used as data. The *Next Beacon* is the number of ticks between the beacons. A tick is 2^{-10} seconds (0.9765625 ms) and is the basic time unit used by the D7A Protocol.

The network configuration settings (ISFB file with ID 0x00), determines the active device class (gateway, subcontroller, endpoint or blinker). It will also define if the device is capable of using Forward Error Correction (FEC), hi-rate channels, and other features which are used by the data link and physical layer.

3.3. D7A Session Layer. The session layer specifies which events may trigger session initiation or scheduling. It describes management and prioritization of multiple scheduled sessions and idle state of the data link layer. Moreover it supports a power level auto-scaling framework to optimize battery usage. This technique will adapt the transmission

Figure 2: OSI layers with DASH7 specification.

Figure 3: Nonarbitrated Two-Party Dialog Example.

power to find an optimum between power consumption and a stable wireless connection.

In a standard situation the subcontrollers and endpoints operate after a wake-on event. The wake-on events defined by D7A are automated channel scanning, automated beacon transmit series (both described in the data link layer), and any application layer relevant events (e.g., sensor event or a passive scanning).

The session layer creates a random session number for any new initiated session. This random number is used in the D7AQP as a dialog ID. The session layer also keeps track of a device's network state. The possible states are associated, scheduled, promiscuous, and unassociated.

A host supporting the D7A Advertising protocol in the network layer also has to implement a session stack to enable session scheduling. This session stack keeps track of the different initiated sessions. Ad-hoc sessions have a higher priority than scheduled sessions (such as beaconing) and are always added on the top of the stack.

3.4. D7A Transport Layer. Generally, the transport layer provides end-to-end communication services. In D7A, this is covered by the *D7A Query Protocol Transport Layer* (D7AQP). It is responsible for communication structuring, flow- and congestion control, and addressing beyond the subnet filtering. In this section, first the communication dialogs of D7AQP will be described. Afterwards the command structuring and finally collision avoidance.

3.4.1. D7AQP Dialogs. The transport layer supports two types of dialogs: *Nonarbitrated Two-Party Dialog* (NA2P) and *Arbitrated Two-Party Dialog* (A2P). The NA2P is a one-time dialog between a requester and a responder or responders. The requester is a single host which sends a single request. The responders can be any host addressed in the request. Besides the query itself, the requester sends an addressing method, a list of channel IDs which can be used for the responses and two time parameters (T_c,T_g). The *Response Completion Timeout* T_c, defines within which time frame the responders must complete their responses. The *Response Guard Time* T_g is a parameter used by the Carrier Sensing Multiple Access (CSMA) process of the data link layer, as explained in Section 3.6. An example of such a dialog is shown in Figure 3.

In contrast to a NA2P dialog, the A2P dialog is a persistent dialog between requester and responders. A2P enables the global and local addressing, which is an important feature of D7A. An example is shown in Figure 4. The first request includes global and local addressing. The global addressing notifies a large group of potential responders to stay alert for a query. The local addressing queries a subgroup of these potential responders. All responders have to respond within

Figure 4: Arbitrated Two-Party Dialog Example.

T_c, after T_c the next query with local addressing will be sent. When no new request is sent after T_c the dialog terminates. This method enables a D7A interrogator to query a large group of tags in a well-structured way. As discussed, in the upper layers an interrogator can send a query with a file read or write command. This method is also used to send an acknowledgment group query. For this, the response of all responders can be acknowledged using one method instead of sending individual acknowledgments.

3.4.2. Command Structuring. The transport layer also defines the structure of the command-response protocol (the *command structure*). The command structure is based on templates. It contains the addressing method, the command code and command extension.

In a request, the addressing method can be broadcast, unicast, multicast or anycast. In case of broadcast and unicast, the command contains a command extension and command data. In case of anycast, the command also includes a global query. A global query is a query using global addressing, as described in the previous subsection. In case of multicast, the command contains a global query and local query or a local query with an acknowledgment template. A response is always unicast and can contain command data or an error.

The command code contains a command type. It defines if the command is a response, an error response, a *Nonarbitrated Two-Party Dialog,* or an *Arbitrated Two-Party Dialog.* In the latter case, it defines whether it is an initial request, a intermediate request or a final request. These dialogs are explained in the previous subsection.

The command code also contains an opcode. It specifies if the command wants to announce a file or do an inventory or collection. An inventory means it will request a list of files. A collection means it will request the content of some file. The command can also define a so called application shell which means it encapsulates an application's specific data.

And finally, the command extension defines if CSMA is used for responses and which type of collision avoidance is used. This is explained in the next subsection.

3.4.3. Collision Avoidance. A third important component of the transport layer is collision avoidance (CA). In case a message is sent from the transport or higher layer, the transport layer is responsible for collision avoidance and flow control. The data link layer is responsible for the CSMA process as is explained in Section 3.6. The CSMA-CA process uses the T_C value from a response or from upper layers and calculates the *Collision Avoidance Timeout Period* T_{CA}:

$$T_{CA} = T_C - T_{\text{response}_{\text{duration}}}. \quad (1)$$

Figure 5: Example of the use of the D7A Advertising Protocol.

This is the time within a packet has to be transmitted to make sure it has been sent before T_C.

Three CSMA-CA and flow control models are defined by D7A: *Adaptive Increase No Division* (AIND); *Random Adaptive Increase No Division* (RAIND); *Random Increase Geometric Division* (RIGD).

AIND and RAIND use a fixed slot length which is approximate the duration of the transmission. In AIND, the CSMA process starts in the beginning of a slot. In RAIND, the CSMA process only starts after a random delay which is smaller than T_{CA}. In RGID the slot duration decreases following $T_{CA0}/2^{n+1}$ and CSMA only starts, as in RAIND, after a random delay.

3.5. D7A Network Layer. The network layer defines the *Background Network Protocol* which is used for the *D7A Advertising Protocol* (D7AAdvP) and the *Foreground Network Protocol* which is used for queries, responses and beacons.

3.5.1. Background Network Protocols. Background network protocols contain very short background frames (BF). Background frames are described in Section 3.6. Currently, only the D7A Advertising Protocol is defined as a background network protocol. D7AAdvP is used exclusively for rapid, ad-hoc group synchronization. It is a transmission only, broadcast protocol. D7AAdvP is used to notify hosts about a request which will be send in the future. The requester floods the channel with background frames which contains a time span until when the request will be sent. The responder hosts receive the background frame while listening for a background frame during background scanning (as described in Section 3.6), and can go to a sleep state until the request is planned to be received. The responder only has to receive one background frame to know this timing. This leads to a very low power consumption optimized method of ad-hoc synchronization. Figure 5 shows an example of such a synchronization train.

3.5.2. Foreground Network Protocols. Two foreground network protocols are defined by D7A: the *D7A Network Protocol* (D7ANP) and the *D7A Datastream Protocol* (D7ADP). The D7ANP is an addressable, routable protocol which is used for the D7A Query Protocol in the transport layer. The

Table 4: Background frame structure.

Subnet	Payload	CRC16
1 byte	3 bytes	2 byte

Table 5: Foreground frame structure.

Length	Headers	Payload	Footer	CRC16
1 byte	3–38 bytes	0–249 bytes	0–20 bytes	2 bytes

D7ADP is a generic data encapsulation protocol, which contains no routing or addressing information. It allows for maximum flexibility towards the upper layers. The D7ADP is, for example, used by the *Application Layer Protocol*.

The D7ANP supports network layer security and standard two hop routing. It implements unicast, broadcast, multicast and anycast addressing to support the D7AQP Dialogs.

3.6. D7A Data Link Layer. The data link layer (DLL) of D7A, specifies the data link filtering, addressing, the dialog models, MAC processing, frame construction and the field definitions.

3.6.1. Data Link Filtering. Incoming frames are filtered by three processes. The first one is a Cyclic Redundancy Check (CRC) validation with a 16 bit CRC16 field. The calculation uses the CCITT CRC16 polynomial. If the CRC16 validations passes, subnet matching or link quality assessment can be performed in an arbitrary order. The subnet is constructed with a 4 bit specifier and a 4 bit mask. If link quality assessments is enabled, the link budget should be higher than a predefined link quality threshold. This makes further frame processing possible.

3.6.2. Data Link Addressing. Related to addressing, the data link layer of the D7A protocol specifies an ISO 15963 [14] compliant Device ID manifesting in a fixed *Unique ID* (UID) and a dynamic network-unique *Virtual ID* (VID).

The UID is a EUI-64 (Extended Unique Identifier) compliant ID [15]. It contains a 24 or 36 bits Organizationally Unique Identifier (OUI) assigned by the IEEE Registration Authority based on a 40 or 28 bits serial number. The VID is a 16 bit ID which is supplied by the network administrator and should be unique within the network.

The data link layer only supports unicasting and broadcasting. If a target Device ID is present in the frame, it will be processed as a unicast message. The Device ID of the frame will be matched with the destination devices ID. The frame will only be processed by the upper layers when a match exists.

3.6.3. Frame Structure. As already mentioned in the network layer, the data link layer has background frames and foreground frames. A background frame is a fixed length 6 byte frame, proceeded by a sync word of class 0 (defined in Section 3.7). The structure of a background frame is shown in Table 4.

Figure 6: Overview of the foreground frame dialog model.

A foreground frame has a variable length, up to 255 bytes, proceeded by a sync word of class 1. The structure is shown in Table 5. The length byte is the total number of bytes, including the length byte and the 2 byte CRC.

The headers contain information related to optional data link layer security, address control (source and optional destination id), the subnet, and the estimated radiated power of the transmission (TX EIRP). This value can be used by the receiving node to estimate the link budget.

3.6.4. Data Link Dialog Models. Following the same concept of foreground and background frames, two related dialog models are defined.

A Background Dialog (not really a dialog) starts when a background frame starts transmitting and ends as soon as it finishes transmitting.

A Foreground Dialog is a dialog between devices and is used to support the D7A Foreground Network Protocols. An overview is shown in Figure 6. The *response contention period* T_C defines the time frame in which responses can transmit. The *listen period* T_L defines the time period after T_C in which the requester can send another request.

3.6.5. MAC Processing. The data link layer supports an automated scanning method. Using this scan scheduler the data link layer can scan for background or foreground frames.

In a background scan, the D7A device first checks the energy level. Only if the level is sufficient does it try to detect the correct sync word (class 0 for background scanning). In foreground scanning the device first searches for a sync word of class 1.

Using the automated *Channel Scan Series*, the data link layer can follow a predefined sequence of background or foreground scans. The series is defined by a list of channel IDs (explained in Section 3.7), scan types (background or foreground), a "Scan Detection Timeout" T_{SD}, and a "Time until next scan event" T_{NSE}. The T_{SD} is the duration the scan on a specified channel ID and type will be executed. The T_{NSE} is the time when the next scan event in the list will be executed.

The *Beacon Transmit Series* is another automated data link layer process. This process sends predefined beacons at specified time intervals and channel IDs.

The data link layer is also responsible for MAC channel guarding. A *Nonguarded Channel* (NGC) is a channel which does not require CSMA. A *Guarded Channel* (GC), however is a channel which does require CSMA. CSMA is also implemented by the data link layer. In this case CSMA-CA

Table 6: Spectrum ID allocation per channel class.

Channel class	Spectrum ID							
Base	—			00			—	
Normal	10	12	14	16	18	1A	1C	1E
Hi-rate	21		25		29		2D	
	—	23		27		2B		—
Blink	32			—			3C	
Spectrum	433.056 MHz					434.784 MHz		

ensures that a transmission is only executed when a channel is unguarded. After a transmission a guarding period T_G is specified in which no transmission can be performed from another host.

3.7. D7A Physical Layer. The physical layer defines the spectrum utilization and channels, the modulation, the symbol encoding, and the packet structure.

3.7.1. D7A Spectrum. The spectrum allocation and channels are shown in Table 6. The spectrum ID specifies which channels and channel classes are used. The *Channel ID* is the Spectrum ID which is logically OR'ed with the encoding option. The value is 0x80 in the case Forward Error Correction (FEC) is used, otherwise the value is 0x00. The base and normal class have a bit rate of 55.555 kbs. The hi-rate and blink class have a bit rate of 200 kbs. The normal class has a channel bandwidth of 216 kHz. The hi-rate and base class have a channel bandwidth of 432 kHz. The blink class has a channel bandwidth of 648 kHz. Gaussian frequency shift keying is used as modulation technique.

3.7.2. Symbol Encoding. As a last stage before transmission, a data whitening [16] technique based on a 9-bit psuedo-random generator (PN9 encoding [17]) is used to avoid a DC offset in the transmitted data. For this, a PN9 decoding stage will be the first decoding step at the receiver.

Before the data whitening step, an optional forward error correction (FEC) can be executed [18]. For FEC, D7 uses convolution encoding with constraint length 4, followed by a 32 bit interleaving executed on 2 bit symbols. The interleave/deinterleave process lowers the impact of bursty errors since it separates adjacent symbols.

3.7.3. Packet Structure. On the physical layer, the frame structure of the data link layer is preceded by a preamble and a sync word. A preamble of typically 32 bits on the base and normal channels and 48 bits on the hi-rate and blink channels of alternating 0 and 1's is used to enable the receiver to calibrate the data rate circuit. The preamble is followed by a sync word. The sync word can be of class 0 or class 1 and depends on the use of FEC. After the sync word, comes the packet payload, which is defined by the upper layers. In all cases D7A uses the big endian format (most significant byte first).

4. OSS-7

OSS-7 [19] is an ongoing software stack implementation of the DASH7 Alliance Protocol mainly developed by the CoSys-Lab of the University of Antwerp. OSS-7 is being developed with its main goal to provide a reference implementation of the D7A specification. This means that code clarity and structure is more important than performance.

OSS-7 is implemented in ANSI C to be as compatible as possible with Texas Instruments (TI) Code Composer Studio (CCS), gcc and mspgcc.

The OSI layers, as shown in Figure 2, are used as structure for the implementation. All the layers are implemented as pluggable layers, which makes it easier to benchmark other implementations of a specific layer (e.g., a new proposal for a MAC implementation).

Besides the OSI layers, there are two additional layers: Hardware Abstraction Layer (HAL) and the "Framework" layer. Both layers implement an API which can be called from any other layer. The HAL makes an abstraction of hardware aspects such as I/O, buttons, LEDs, CRC software or hardware implementations, UART, timers, and the Real Time Clock (RTC). Currently a HAL implementation is provided for TI MSP430, TI Stellaris (ARM Cortex-M4) and posix hardware. Support for ARM Cortex-M3 is planned for the near future. The framework API provides hardware independent functionality like logging and queueing which is used within the stack but can also be used by the application.

The PHY layer consists mostly of an abstract interface. This Radio Abstraction Layer (RAL) is required to be implemented by radio chip specific code. This plugin system effectively allows us to support multiple radio chips. One can switch between them at compile time. Currently TI's CC430 (which is system-on-chip composed of a TI MSP430 and an TI CC1101 sub 1 Ghz transceiver) and an external CC1101 (with communication through SPI) are supported. Support for STMicroelectronics SPIRIT1 RF chip will be implemented in the near future.

Figure 7 gives an overview of the file structure of OSS-7. The different OSI layers are visible. RAL and HAL can be extended to different hardware.

Besides the software stack itself, the OSS-7 project will also provide additional PC-based tooling for diagnostics, testing, and configuring tags running the stack.

4.1. Implementation. Figure 8 shows the devices which are used to implement and test OSS-7. Device (1) is a CoSys-Lab designed 3 × 3 cm CC430 based tag with PCB antenna. This device is typically used as an endpoint or blinker powered with a CR2032 battery for indoor sensor communication. Device (2) show a small extension board which is used to interface with the JTAG programmer and a serial connection. Device (3) is a WizziMote and WizziBoard from WizziLab powered by 2 AA batteries. Device (4) is the Texas Instrument EZ430-Chronos-433 which typically acts as endpoint or blinker. Device (5) consists of the nodes which are used in device (1) and device (3) but altered to be used as a subcontroller. Device (6) is a Raspberry Pi with an extension

Figure 7: File structure of OSS-7.

Figure 8: Devices used to implement OSS-7.

board and device (1) is altered to be used with an external antenna. This device is used as a gateway in our test settings.

As an example, devices (1) and (5) are used by Stevens et al. [20] to test robot localization using DASH7.

5. Summary

This paper gave an high-level overview of the DASH7 Alliance Specification and introduced OSS-7, a stack implementation for this specification. D7A is built for low-power sensor communication with integrated concepts such as global and local querying, automated beacon transmits series and automated channel scanning. D7A describes a full functional RFID device.

OSS-7 is embedded stack implementation for D7A which currently supports CC430, CC1011, and ARM Cortex-M4. It is build using a modular structure following the D7A's ISO layering and a separate hardware and radio abstraction layer. OSS-7 focuses on transparency and code clarity to be used as a tool to explain the specifications and test new protocol adaptations.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] M. Beale and Y. Morioka, "Wireless machine-to-machine communication," in *Proceedings of the 41st European Microwave Conference (EuMC '11)*, pp. 115–118, October 2011.

[2] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[3] P. Tuset-Peiro, A. Anglès-Vazquez, J. Lopez-Vicario, and X. Vilajosana-Guillen, "On the suitability of the 433 MHz band for M2M low-power wireless communications: propagation aspects," *Transactions on Emerging Telecommunications Technologies*, 2013.

[4] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, 2011.

[5] DASH7 Alliance Mode Specification, DASH7 Alliance Std., July 2013.

[6] H. Lehpamer, *RFID Design Principles*, Artech House, 2012.

[7] I. F. Isnin, *A study on wireless communication error performance and path loss prediction [Ph.D. dissertation]*, University of Plymouth, 2011.

[8] X. Zhang, T. Weston Burress, K. B. Albers, and W. B. Kuhn, "Propagation comparisons at VHF and UHF frequencies," in *Proceedings of the IEEE Radio and Wireless Symposium (RWS '09)*, pp. 244–247, IEEE, January 2008.

[9] European Telecommunications Standards Institute, "Electro-Magnetic Compatibility and Radio Spectrum Matters (ERM); Short Range Devices (SRD); Radio equipment to be used in the 25 MHz to 1000 MHz frequency range with power levels ranging up to 500 mW; Part 1: Technical characteristics and test methods," ETSI Standard EN 300 220-1, 2012.

[10] M. Loy, R. Karingattil, and L. Williams, "ISM-band and short range device regulatory compliance overview," Tech. Rep., Texas Instruments, 2005.

[11] "Part 15: radio frequency devices; 15.231: periodic operation in the band 40.66–40.70 MHz and above 70 MHz," Federal Communications Commission Std., 2011.

[12] H. Zimmermann, "Osi reference model—the iso model of architecture for open systems interconnection," *IEEE Transactions on Communications Systems*, vol. 28, no. 4, pp. 425–432, 1980.

[13] "Information technology—smart transducer interface for sensors and actuators—part 7: transducers to radio frequency identification (RFID) systems communication protocols and transducer electronic data sheet (TEDS) formats," ISO/IEC/IEEE Standard ISO/IEC/IEEE 21451-7, 2011.

[14] "Information technology—radio frequency identification for item management—unique identification for RF tags," ISO Standard ISO/IEC 15963, 2009.

[15] "Guidelines for 64-bit global identifier (eui-64)," Tech. Rep., IEEE Standards Association, 2013.

[16] E. M. B. Sobral, R. D. Lins, and Z. D. Lins, "Analyzing the propagation features of the Texas eZ430-RF2500 device," in *Proceedings of the 7th International Telecommunications Symposium*, 2010.

[17] G. Christiansen, "Design note dn509—data whitening and random TX mode," Texas Instruments, 2010.

[18] E. Shih, S.-H. Cho, N. Ickes et al., "Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, pp. 272–286, ACM, July 2001.

[19] "Dash7 open source stack: OOS-7," 2013, https://github.com/CoSys-Lab/dash7-ap-open-source-stack.

[20] J. Stevens, R. Berkvens, W. Loockx, and M. Weyn, "Robot localization with DASH7 technology," in *Proceedings of the 3rd International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT '13)*, Porto, Portugal, September 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 545362, 7 pages
http://dx.doi.org/10.1155/2013/545362

Research Article

The Performance of Multimessage Algebraic Gossip in a Random Geometric Graph

Gang Wang,[1] Zun Lin,[1] Wenyang Guan,[2] and Feng Wang[3]

[1] *School of Electronics and Information Engineering, Beihang University, Beijing, China*
[2] *School of Electronics Engineering and Computer Science, Peking University, Beijing, China*
[3] *Beijing Aeronautical Science & Technology Research Institute, Commercial Aircraft Corporation of China, Beijing, China*

Correspondence should be addressed to Wenyang Guan; guanwenyang@pku.edu.cn

Received 17 August 2013; Revised 21 October 2013; Accepted 22 October 2013

Academic Editor: Jianhua He

Copyright © 2013 Gang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gossip algorithm has been widely regarded as a simple and efficient method to improve quality of service (QoS) in large-scale network which requires rapid information dissemination. In this paper, information dissemination based on algebraic gossip in a random geometric graph (RGG) is considered. The n nodes only have knowledge about their own contents. In every time slot, each node communicates with a neighbor partner chosen randomly. The goal is to disseminate all of the messages rapidly among the nodes. We show that the gain of the convergence time is $O(n^{1/2} \log \varepsilon^{-1}/\log^{1/2} n)$ with network coding. Simulation results show that these bounds are valid for the random geometric graph and demonstrate that network coding significantly improves the bounds with the number of users increasing.

1. Introduction

With the increased demand for network resources and the expansion of the variety of information resources, information dissemination based on gossip becomes an effective way to improve quality of service (QoS) in large-scale network environment. Gossip algorithm is a simple and efficient algorithm which has good extensibility and robustness to adapt to the large-scale network environment well; it also reduces the redundancy of the retransmission to disseminate information in a wide range of distributed applications. Recently, gossip algorithms have been proposed for many distributed applications, including peer-to-peer (P2P) network, broadcast and multicast [1, 2], adhoc network routing [3], and distributed computation and parameter estimation [4, 5].

Network coding [6] is first proposed by Ahlswede et al. in 2000 to change the traditional store-and-forward routing. Network coding brought many potential advantages like improving network throughput, balancing network load and high bandwidth utilization, and increasing the network robustness and adaptability for multicast network. Chen and Tseng [7] proposed a distributed algorithm with random linear network coding [8], which is developed to effectively detect, locate, and isolate the Byzantine attackers in a wireless ad-hoc network.

In 2006, Boyd et al. boldly proposed a randomized gossip algorithm [9]. They showed that the convergence time in the random geometric graph (RGG) was $O(r^{-2} \log n)$, where r is the transmission radius threshold and n is the number of nodes. Based on the advantage of random linear network coding, Deb et al. [10] suggested an algebraic gossip algorithm that innovatively integrated network coding with a gossip algorithm and proved that the convergence time of the information dissemination improved more quickly than in a randomized gossip algorithm, needing only $O(n \log n)$ time to disseminate the messages in a completed graph. Subsequently, Mosk-Aoyama and Shah in [11] generalized the results of [10] and obtained the convergence time bound $O(n \log n)$ in expectation using an algebraic gossip algorithm in an arbitrary graph. Vasudevan and Kudekar [12] developed a uniform strong bound of $O(n \log^2 n)$ with algebraic gossip for an arbitrary graph with high probability. Borokhovich et al. used queuing theory as a novel approach for analyzing algebraic gossip [13]. They obtained an upper bound of $O(n\Delta)$

rounds for a graph with a maximum degree of Δ for all-to-all communication. An average connectivity degree cluster (ACDC) scheme gossip algorithm to improve the convergence speed and the accuracy of the consensus is proposed in [14]. The proposed algorithm can yield results that are 50% closer to the real average value than the referenced standard gossip and grid cluster gossip algorithms.

However, realistic networks are so complex that traditional topology graphs cannot match them except with the RGG. We analyze the performance of multimessage algebraic gossip in the RGG. We utilize the concept of conductance [15] to analyze the convergence time bounds. The conductance of a graph is a measure of the connectivity of the graph, which determines how fast a random walk on a graph converges to a uniform distribution. Our results show that the gain of the convergence time is $O((n^{1/2}\log\varepsilon^{-1})/\log^{1/2}n)$ with network coding.

The rest of this paper is structured as follows. In Section 2, we describe the model, the algorithm, and the protocols with a few preliminaries. We state our main theoretical results and also provide essential proofs of them in Section 3. In Section 4, simulation results and detailed analysis are given. Section 5 concludes this paper.

2. Preliminaries and Models

In this network model, there are n nodes, and each node has only one message. We assume that the set of messages is $\mathbf{M} = \{m_1, m_2, \ldots, m_n\}$, where node i stores the message in the initial time slot $t = 0$.

2.1. RGG Model. The RGG model $\mathbf{G}^d(n, r)$ is formulated as follows [16].

(i) Place n vertices on the region which are in random uniformly and independently distribution.

(ii) Connect two vertices (u, v) if and only if the distance between them is less than or equal to a threshold; that is, $d(u, v) \le r$, where r is a transmission radius threshold. Obviously, the connection probability between the nodes depends on their Euclidean distance.

We use adjacency matrix $\mathbf{A}_{ij}(t)$ to represent graph structure, where if nodes i and j are in their neighborhood, $\mathbf{A}_{ij}(t) = 1$, for $i \ne j$, else $\mathbf{A}_{ij}(t) = 0$. Moreover, $\mathbf{N}_i = \{j \in \{1, 2, \ldots, n\} : \mathbf{A}_{ij}(t) \ne 0\}$ represents the set of neighbor nodes of node i, and $d_i = |\mathbf{N}_i|$ represents the degree of node i, which means the number of neighbor nodes.

2.2. Synchronous Time Model. A time model determines the time of multimessage dissemination. We use the synchronous time model as in [11]. In the gossip algorithms, round is the time unit during which nodes exchange their messages with each other. Assume one round is divided into n consecutive time slots. In each time slot, each node selects one of its neighbor nodes randomly and independently and shares information with it.

2.3. Gossip Algorithm. In time slot t, any node i can select neighbor nodes uniformly, randomly and independently with the probability $p = 1/d$, where d represents a node's degree [13]. In the RGG, the transition probability matrix of the node [17] is

$$P_{ij} = \begin{cases} \dfrac{1}{2}, & \text{if } j = i, \\ \dfrac{1}{2d}, & \text{if } j \in \mathbf{N}_i, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Here $\mathbf{N}_i$ is the set of node i's neighbors. Equation (1) presents the transition probability in the RGG without network coding. When node j is one neighbor of node i, (1) that ensures this neighbor is selected randomly and uniformly. The transition probability expectation of node j is equal to that of node i when chosen as a transmitted node. Then we can use it to derive the transition probability with network coding in the RGG.

There are three kinds of gossip algorithms—Push, Pull, and Exchange, which are defined as follows, respectively.

(i) Push: in this model, a message is transmitted from a transmitting node (*the caller*) to a receiving node (*the called*). Thus, the communication process is initiated by the transmitting node.

(ii) Pull: in this model, a message is transmitted from a transmitting node (*the called*) to a receiving node (*the caller*). Thus, the communication process is initiated by the receiving node.

(iii) Exchange: here, a message is transmitted from a transmitting node (*the caller*) to a receiving node (*the called*), and at the same time, the other message is transmitted from a transmitting node (*the called*) to a receiving node (*the caller*).

2.4. Gossip Protocol. Here we use the random linear network coding technology as the gossip protocol. Random linear network coding encodes each message before sending it. We take each message as a vector from the finite field $\mathbf{F}_q$ of size q. So if the message size is m bits, the dimensions of message vector can be represented as $r = \lceil m/\log_2 q \rceil$. All the initial messages are linearly independent vectors in the finite field $\mathbf{F}_q$, and we take $\mathbf{A}_i(t)$ to represent the matrix of a coding coefficient vector of nodes i in time slot t.

When coding coefficient vector of any node i increases to n, it indicates that node i can recover all the messages and also means that wireless communication process has achieved convergence.

In the rest of this paper, we integrate the exchange algorithm with random linear network coding to analyze the convergence time of multimessage dissemination.

2.5. Communication Criteria. First, consider the local convergence time $\mathbf{T}_{\text{local}}^{\text{EX,RLC}}(n) = \mathbf{min}\{t : \dim(\mathbf{A}(t)) = n, \exists i \in \mathbf{V}\}$, representing the shortest time that there is a node i containing all of the messages in the set $\mathbf{M}$.

Second, consider the global convergence time [18] $\mathbf{T}_{\text{global}}^{\text{EX,RLC}}(n, \varepsilon) = \mathbf{inf}\{t : \mathbf{Pr}(\bigcup_{i=1}^{n}\{\dim(\mathbf{A}(t)) \ne n\}) < \varepsilon, \exists i \in \mathbf{V}\}$,

representing the time point at which all of the nodes receive all of the messages with high probability $1-\varepsilon$, where $\varepsilon > 0$.

3. Main Results

Based on the above model, we analyze the benefit of network coding in multimessage dissemination in the RGG.

Before transmission, each sender and receiver encodes the necessary messages stored in the buffer. At each time slot, each sender i randomly selects the encoded message to send to one of its neighbors j and receives the encoded message from node j at the same time.

Firstly, we use the following definitions similar to [17].

Definition 1. The "**Basic Class**" $\mathbf{B}(t)$ is the set of matrices satisfying the following properties.

(i) If $a_{ij}^{i}(t) \neq 0$, then $a_{ij}^{j}(t) \neq 0$, for all $a_{ij}^{i}(t) \in \mathbf{A}_i(t)$, $a_{ij}^{i}(t) \in \mathbf{A}_i(t)$;

(ii) nonzero submatrices of $\mathbf{A}_i(t)$ and $\mathbf{A}_j(t)$ satisfy column full rank.

If the matrices satisfy condition (i) but do not satisfy condition (ii), then the set of them can be defined as the "**Virtual Class**" denoted by $\widetilde{\mathbf{V}}(t)$.

If neither of the conditions (i) or (ii) are satisfied, then the matrices belong to the set defined as the "**Residue Class**" denoted by $\mathbf{R}(t)$.

Let a node $j \in \mathbf{N}_i$. Suppose that P'_{ij} represents the transition probability of the nodes i, j that satisfy $\mathbf{A}_i(t), \mathbf{A}_j(t) \in \mathbf{R}(t)$ and $\overline{P_{ij}}$ represents the transition probability of the nodes i, j that satisfy $\mathbf{A}_i(t), \mathbf{A}_j(t) \in \widetilde{\mathbf{V}}(t)$. Therefore, the relation between P'_{ij} and $\overline{P_{ij}}$ satisfies the following.

Lemma 2. *Consider*

$$\overline{P_{ij}} \approx \frac{2}{(d_i - 2)\, P'_{ij}}. \tag{2}$$

Proof of Lemma 2. According to the definition of the RGG and (1), we can easily derive $\overline{P_{ij}} \approx (\Delta_i(k)/(d_i - \Delta_i(k)))P'_{ij}$, where $\Delta_i(k)$ is the number of "**Virtual Class**" in the set $\mathbf{N}_i$ and k is the maximum degree of node i.

Suppose the probability of $\Delta_i(k) = 1$ is p_k; therefore, the probability of $\Delta_i(k) = 2$ is p_k^2. The rest may be deduced by analogy; the probability of $\Delta_i(k) = k$ is p_k^k. p_k satisfies the equality $p_k + \cdots + p_k^k = 1$; if k is large enough, then $p_k \approx 1/2$ and

$$\mathbf{E}\left(\Delta_i(k)\right) = \frac{1}{2} + \frac{2}{2^2} + \cdots + \frac{k}{2^k} \approx 2. \tag{3}$$

Therefore, with the help of expectation of $\Delta_i(k)$, the transition probability $\overline{P_{ij}} \approx (2/(d_i - 2))P'_{ij}$ is obtained. □

Theorem 3. *For the RGG, the conductance of the static algebraic gossip algorithm with multimessage dissemination is* $\widehat{\mathbf{\Phi}}(p) = O(n^{3/2}/(\log^{1/2} n))$.

Proof. According to the definition of conductance, we can have k-conductance for any $k \in [1, n]$ as follows:

$$\begin{aligned}\mathbf{\Phi}_k(p) &= \min_{\mathbf{S}\subset\mathbf{V},\mathbf{0}<\pi(\mathbf{S})<k/n} \frac{\mathbf{Q}\left(\mathbf{S},\overline{\mathbf{S}}\right)}{\boldsymbol{\pi}(\mathbf{S})} \\ &= \min_{\mathbf{S}\subset\mathbf{V},\mathbf{0}<\pi(\mathbf{S})<k/n} \frac{\sum_{x\in\mathbf{S},y\in\overline{\mathbf{S}}} \boldsymbol{\pi}(x)\, P(x,y)}{\boldsymbol{\pi}(\mathbf{S})}.\end{aligned} \tag{4}$$

Here $\boldsymbol{\pi}(\mathbf{S})$ is the probability density of $\mathbf{S}$ under the stationary distribution π and $\mathbf{Q}(\mathbf{S},\overline{\mathbf{S}})$ is the sum of $\mathbf{Q}(v,u)$ over all $(v,u) \in \mathbf{S}\times\overline{\mathbf{S}}$.

According to the cut size between $\mathbf{S}$ and $\overline{\mathbf{S}}$ $\mathbf{Cut}_{\mathbf{\Phi}}(\mathbf{S},\overline{\mathbf{S}}) = 3n^2r^3$ in [19], (4) can be converted as follows:

$$\begin{aligned}\mathbf{\Phi}_k(p) &= \min_{\mathbf{S}\subset\mathbf{V},\mathbf{0}<\pi(\mathbf{S})<k/n} \mathrm{O}\left(\frac{n}{k}\right) \sum_{x\in\mathbf{S},y\in\overline{\mathbf{S}}} \mathrm{O}\left(\frac{1}{n}\right)\mathrm{O}\left(\frac{1}{nr^2}\right) \\ &= \min_{\mathbf{S}\subset\mathbf{V},\mathbf{0}<\pi(\mathbf{S})<k/n} \mathrm{O}\left(\frac{1}{knr^2}\right)\mathbf{Cut}_{\mathbf{\Phi}}\left(\mathbf{S},\overline{\mathbf{S}}\right) \\ &= \mathrm{O}\left(\frac{nr}{k}\right).\end{aligned} \tag{5}$$

From [20], we can obtain the wireless transmission radius in the RGG of two dimensions as $r(n) = \mathrm{O}(\sqrt{\log n/n})$. Then, from (5) we can obtain k-conductance as follows:

$$\mathbf{\Phi}_k(p) = \mathrm{O}\left(\sqrt{\frac{n\log n}{k}}\right). \tag{6}$$

Therefore, the conductance is

$$\widehat{\mathbf{\Phi}}(p) = \sum_{k=1}^{n-1} \frac{k}{\mathbf{\Phi}_k(p)} = \sum_{k=1}^{n-1} \frac{k^2}{\sqrt{n\log n}} = \mathrm{O}\left(\frac{n^{3/2}}{\log^{1/2} n}\right). \tag{7}$$

□

Theorem 4. *For any $\varepsilon > 0$, the convergence time of the static algebraic gossip algorithm with multimessage dissemination is*

$$\mathbf{T}_{global}^{EX,RLC}(n,\varepsilon) = O\left(\frac{n^{3/2}\log\varepsilon^{-1} - n^{1/2}\log\varepsilon^{-1}}{\log^{1/2} n}\right). \tag{8}$$

Proof. Assume that in the initial time slot $t = 0$, $\mathrm{rank}(\mathbf{A}_i(t)) = 1$, for all $i \in \mathbf{V}$, where $\mathbf{V}$ is the set of all of the nodes, $t_k = \mathbf{inf}\{t : \mathbf{C}(t) \geq k\}$ and $\mathbf{C}(t)$ is the largest cardinality in $\mathbf{B}(t)$. Our goal is to make $\mathrm{rank}(\mathbf{A}_i(t_n)) = n$, for all $i \in \mathbf{V}$.

According to the definition of t_k, we divide the total convergence time into k time periods as follows:

$$[t_1, t_2), [t_2, t_3), \ldots, [t_k, t_{k+1}). \tag{9}$$

Let I_{ij} be a characteristic function which takes the value of one if node i transmits message to node j; otherwise it is zero. When node i exchanges messages with node j, we can obtain the inequality as follows:

$$\mathbf{E}\left(I_{ij}\right) + \mathbf{E}\left(I_{ji}\right) \geq P'_{ij} + \overline{P_{ij}}. \tag{10}$$

Now consider the general time slot $t \in [t_k, t_{k+1})$; the accumulative rank $\Delta(t) = \sum_{i=1}^{n}(\text{rank}(\mathbf{A}_i(t)) - 1)$ should satisfy the following inequality:

$$\Delta(t+1) - \Delta(t) \geq \sum_{i \in V} \sum_{j \in V, j \neq i} \left(I_{ij} + I_{ji}\right). \tag{11}$$

By taking expectations on both sides of the above inequality and using (11), we have

$$\begin{aligned}
\mathbf{E}(\Delta(t+1) - \Delta(t)) &\geq \sum_{\substack{i,j \in \mathbf{V}, j \neq i \\ \mathbf{A}_i(t), \mathbf{A}_j(t) \in \mathbf{R}(t)}} P'_{ij} + \sum_{\substack{i,j \in \mathbf{V}, j \neq i \\ \mathbf{A}_i(t), \mathbf{A}_j(t) \in \overline{\mathbf{V}}(t)}} \overline{P_{ij}} \\
&= \sum_{\substack{i,j \in \mathbf{V}, j \neq i \\ \mathbf{A}_i(t), \mathbf{A}_j(t) \in \mathbf{R}(t)}} P'_{ij} + \sum_{\substack{i,j \in \mathbf{V}, j \neq i \\ \mathbf{A}_i(t), \mathbf{A}_j(t) \in \mathbf{R}(t)}} \frac{2P'_{ij}}{d_i - 2} \\
&\geq \sum_{\substack{i,j \in \mathbf{V}, j \neq i \\ \mathbf{A}_i(t), \mathbf{A}_j(t) \in \mathbf{R}(t)}} P'_{ij}\left(1 + \frac{2}{k-2}\right) \\
&\geq \left(1 + \frac{2}{k-2}\right) n\mathbf{\Phi}_k(p).
\end{aligned} \tag{12}$$

For the convenience of our analysis, we use $g(t) = \Delta(t+1) - \Delta(t) - (1 + 2/(k-2))n\mathbf{\Phi}_k(p)$. Therefore, we quote the definition of $Z_k(t)$ in [20] $Z_k(t) = \sum_{i=t_k}^{t-1} g(i)\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}}$; then we can easily get $Z_k(t_k) = 0$.

At the same time, we can conclude that

$$\begin{aligned}
&\mathbf{E}\left(Z_k(t+1) \mid Z_k(t)\right) \\
&= \mathbf{E}\left(\sum_{i=t_k}^{t} g(i)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}} \mid \sum_{i=t_k}^{t-1} g(i)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}}\right) \\
&= \mathbf{E}\left(\sum_{i=t_k}^{t-1} g(i)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}} \mid \sum_{i=t_k}^{t-1} g(i)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}}\right) \\
&\quad + \mathbf{E}\left(g(t)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}} \mid \sum_{i=t_k}^{t-1} g(i)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}}\right) \\
&= Z_k(t) + \mathbf{E}\left(g(t)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}} \mid \sum_{i=t_k}^{t-1} g(i)\,\mathbf{1}_{\{\mathbf{i} < t_{k+1}\}}\right) \\
&\geq Z_k(t).
\end{aligned} \tag{13}$$

Equation (13) showed that $Z_k(t)$ is submartingale. Moreover, according to the property of submartingale, $\mathbf{E}(Z_k(t+1)) \geq \mathbf{E}(Z_k(t+1)) = 0$ is obtained.

We unfold the above inequality to get the result as

$$\begin{aligned}
&\mathbf{E}\left(\sum_{i=t_k}^{t_{k+1}-1} (\Delta(i+1) - \Delta(i))\right) \\
&\quad \geq \mathbf{E}\left(\sum_{i=t_k}^{t_{k+1}-1} \left(1 + \frac{2}{k-2}\right) n\mathbf{\Phi}_k(p)\right), \\
&\mathbf{E}(\Delta(t_{k+1}) - \Delta(t_k)) \\
&\quad \geq \left(\frac{1+2}{k-2}\right) n\mathbf{\Phi}_k(p)\,\mathbf{E}(t_{k+1} - t_k).
\end{aligned} \tag{14}$$

According to (14) and Theorem 5 in [11], (12) can be derived as follows

$$\begin{aligned}
\mathbf{E}(t_n) &\leq \frac{k-2}{k}\sum_{k=2}^{n-1} \frac{\mathbf{E}(\Delta(t+1) - \Delta(t))}{n\mathbf{\Phi}_k(\mathbf{p})} \\
&\quad + \frac{\mathbf{E}(\Delta(t+1) - \Delta(t))}{n\mathbf{\Phi}_k(\mathbf{p})}\bigg|_{k=1} \\
&\leq \sum_{k=1}^{n-1} \frac{2k}{n\mathbf{\Phi}_k(\mathbf{p})} - \sum_{k=2}^{n-1} \frac{2}{n\mathbf{\Phi}_k(\mathbf{p})} \\
&= \frac{2}{n}\widehat{\mathbf{\Phi}}(\mathbf{p}) - \frac{4}{n}\sum_{k=2}^{n-1} \frac{1}{n\mathbf{\Phi}_k(\mathbf{p})}.
\end{aligned} \tag{15}$$

By Markov's inequality, (15) implies that

$$\mathbf{P}\left(t_n > \frac{\log \varepsilon^{-1}}{n}\left(\widehat{\mathbf{\Phi}}(\mathbf{p}) - \sum_{k=2}^{n-1} \frac{1}{\mathbf{\Phi}_k(\mathbf{p})}\right)\right) < \varepsilon. \tag{16}$$

Therefore we can conclude that the convergence time of the static algebraic gossip network model is expressed as

$$\mathbf{T}_{\text{global}}^{\text{EX,RLC}}(n, \varepsilon) = \mathrm{O}\left(\frac{\widehat{\mathbf{\Phi}}(\mathbf{p}) - \sum_{k=2}^{n-1} 1/(\mathbf{\Phi}_k(\mathbf{p}))}{n} \log \varepsilon^{-1}\right). \tag{17}$$

According to (6) and (7), we can conclude that the convergence time of the mobile algebraic gossip model is

$$\mathbf{T}_{\text{global}}^{\text{EX,RLC}}(n, \varepsilon) = \mathrm{O}\left(\frac{n^{3/2}\log \varepsilon^{-1} - n^{1/2}\log \varepsilon^{-1}}{\log^{1/2} n}\right). \tag{18}$$

□

Because the conductance is the same as in (6), we conclude that the convergence time of the static gossip algorithm is as follows.

Corollary 5. *For any $\varepsilon > 0$, the convergence time of the static gossip network model is*

$$\mathbf{T}_{global}^{EX,RLC}(n, \varepsilon) = O\left(\frac{n^{3/2}\log \varepsilon^{-1}}{\log^{1/2} n}\right). \tag{19}$$

The results are summarized in Tables 1 and 2 for clarity. From these tables, we observe that network coding directly affects the convergence time. Next, we will check and verify these results by simulation.

Table 1: Comparison of algorithms in the RGG.

	k-conductance	Conductance	Convergence time
Static gossip	$O\left(\sqrt{\frac{n\log n}{k}}\right)$	$O\left(\frac{n^{5/2}}{\log^{1/2} n}\right)$	$O\left(\frac{n^{3/2}}{\log^{1/2} n}\right)$
Static algebraic gossip	$O\left(\sqrt{\frac{n\log n}{k}}\right)$	$O\left(\frac{n^{5/2}}{\log^{1/2} n}\right)$	$O\left(\frac{n^{3/2}\log\varepsilon^{-1} - n^{1/2}\log\varepsilon^{-1}}{\log^{1/2} n}\right)$

Table 2: Gossip in different graph models.

Graph	Complete	Star	Ring	Bounded degree	RGG
Convergence time	$O(n\log n)$	$O(n\log n)$	$O(n)$	$O(n)$	$O\left(\frac{n^{3/2}\log\varepsilon^{-1} - n^{1/2}\log\varepsilon^{-1}}{\log^{1/2} n}\right)$

4. Simulation

In this section, simulation results are presented to support our theoretical analysis. In our RGG simulation, it is assumed that the maximal communication range for each node is set to $O(\sqrt{\log n/n})$. All the nodes do not perform as the message relay any more, because they are outside of the communication radius. The simulation network is implemented in a fully distributed way, which guarantees that nodes have no knowledge of the other's state (e.g., location, caller/called state, etc.). Therefore, both successful and unsuccessful communication attempt are added to the convergence time.

The simulation parameters are set as follows:

(i) simulation software: MATLAB R2010b,

(ii) simulation area: unit square,

(iii) number of nodes: 100–200 with step 10,

(iv) maximum number of rounds for each case: 2000,

(v) runs for each case: 100,

(vi) time unit: round number.

Here, the meanings of the abbreviations are set as follows:

(i) SAG: static algebraic gossip,

(ii) SG: static gossip,

(iii) GCT: global convergence time, and its time unit is a round number,

(iv) LCT: local convergence time, and its time unit is a round number,

(v) W.H.P: with high probability.

To analyze the performance of the static algebraic gossip algorithm, we built, simulator to investigate the relation between convergence time and other typical parameters. The simulation results are as follows.

Firstly, we analyze the relation between the global convergence time and W.H.P $1-\varepsilon$. From Figure 1, it can be observed that the global convergence time increases as the parameter ε decreases. Moreover, GCT for each algorithm rapidly grows as W.H.P $1-\varepsilon$ increases from 0.1 to 1. Obviously, when W.H.P $1-\varepsilon$ is less than or equal to 0.7, the gap of the global convergence time between two algorithms in Figure 1 becomes smaller as W.H.P $1-\varepsilon$ increases, which proves that network coding has great advantages for the convergence time in this condition. However, when W.H.P $1-\varepsilon$ is greater than or equal to 0.7, the gap maintains relatively stable, meaning that network coding has little influence on the convergence time in this condition.

Thus, we can choose an appropriate value of $\varepsilon = 0.1$ for the convenience and the precision of the next simulations.

Secondly, we analyze the relation between the global convergence time, the local convergence time, and the number of nodes in Figures 2 and 3.

As shown in Figures 2 and 3, both the global convergence time and local convergence time go up with the increasing number of nodes. As the number of nodes increases, the global convergence time and local convergence time of static algebraic gossip algorithm are less than those of static gossip algorithm. This phenomenon results from the network coding, which effectively and obviously reduces the GCT and LCT when the number of nodes becomes larger. It can be observed from Figure 2 and Figure 3 that network coding has a relatively larger benefit on the local convergence time than the global convergence time.

Finally, we can conclude some relations from the figures. In the beginning periods of communication, network coding significantly lowers the retransmission probability. It can explain why the gap is relatively larger in the initial phase in Figure 1 and why network coding has a relatively larger impact on the local convergence time than the global convergence time in Figures 2 and 3. However, in the subsequent periods of communication, new messages are increasingly difficult to transmit, so network coding has less help to improve the performance. The constant gap in the later phase shown in Figure 1 and the differences between the global convergence time and local convergence time in Figures 2 and 3, respectively, can illustrate network coding has less help to improve the performance.

5. Conclusion and Future Work

In this paper, we studied the performance of multimessage dissemination based on the algebraic gossip algorithm in the RGG. We considered the network with n nodes which only

Figure 1: The relation between GCT and W.H.P $1-\varepsilon$.

Figure 2: The relation between GCT and the number of nodes.

Figure 3: The relation between LCT and the number of nodes.

have knowledge about their own contents. In every time slot, each node communicates with a randomly chosen neighbor partner by the exchange algorithm. Our goal is to disseminate all of the messages rapidly among the nodes. We utilized the concept of conductance to analyze the convergence time bounds. The results showed that the bounds are significantly improved from $O(n^{3/2}\log\varepsilon^{-1}/\log^{1/2}n)$ to $O(n^{3/2}\log\varepsilon^{-1} - n^{1/2}\log\varepsilon^{-1}/\log^{1/2}n)$. Next, through the simulations, we verified that these bounds are valid for the random geometric graph and network coding significantly improves the convergence time bounds of multimessage dissemination in the RGG.

However, there are still a number of open questions that require further research and improvement. One is to find a quantitative description of the local convergence time in a specific network environment and compare it to the global convergence time. There are still no related literatures about local convergence time quantitatively expressed so far. Compared to global convergence time, local convergence time is more susceptible to external factors. The bound of the local convergence time is not easy as $1/n$ of the bound of the global convergence time. Once the external factors change, the impact on local convergence time is relatively large. Comparing and analyzing the simulation of local convergence time and global convergence time in different topologies to conclude some related rules may be a feasible step. The other important direction is to research the influence of nodes' mobility. Due to the static network environment, the connectivity between nodes is fixed. That may lead to the problem where some neighbors of a node have very low transition probabilities, resulting in a sharp increase in the propagation times. Nodes' mobility could bring a lot of potential opportunities, like improving the network throughput, increasing network coverage and so on. There are more and more researches about nodes' mobility to solve the problem of information dissemination. Therefore, we will analyze the case with nodes' mobility in future work. The start point is if nodes' mobility can help change the connectivity between nodes over time to impact transition probability.

Acknowledgments

This work was supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant no. 61221061) and by the National Natural Science Foundation of China (Grants nos. 61271194, 60972007, 61250001, and 61231013).

References

[1] C. Gkantsidis and P. R. Rodriguez, "Network coding for large scale content distribution," in *Proceedings of the 24th Annual*

Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05), pp. 2235–2245, Miami, FLa, USA, March 2005.

[2] P. T. Eugster, R. Guerraoui, S. B. Handurukande, P. Kouznetsov, and A.-M. Kermarrec, "Lightweight probabilistic broadcast," *ACM Transactions on Computer Systems*, vol. 21, no. 4, pp. 341–374, 2003.

[3] Z. J. Haas, J. Y. Halpern, and L. Li, "Gossip-based ad hoc routing," *IEEE/ACM Transactions on Networking*, vol. 14, no. 3, pp. 479–491, 2006.

[4] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[5] K. Jung, D. Shah, and J. Shin, "Distributed averaging via lifted Markov chains," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 634–647, 2010.

[6] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[7] J.-Y. Chen and Y.-Y. Tseng, "Distributed intrusion detection of byzantine attacks in wireless networks with random linear network coding," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 758340, 10 pages, 2012.

[8] T. Ho, M. Médard, R. Koetter et al., "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.

[9] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[10] S. Deb, M. Médard, and C. Choute, "Algebraic gossip: a network coding approach to optimal multiple rumor mongering," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2486–2507, 2006.

[11] D. Mosk-Aoyama and D. Shah, "Information dissemination via network coding," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 1748–1752, Seattle, Wash, USA, July 2006.

[12] D. Vasudevan and S. Kudekar, "Algebraicgossip on arbitrary networks," http://arxiv.org/abs/0901.1444.

[13] M. Borokhovich, C. Avin, and Z. Lotker, "Tight bounds for algebraic gossip on graphs," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '10)*, pp. 1758–1762, Austin, Tex, USA, June 2010.

[14] Y. Li, Z. Zhou, and T. Sato, "A cluster-based consensus algorithm in a wireless sensor network," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 547124, 14 pages, 2013.

[15] B. Bollobás, *Modern Graph Theory*, vol. 184, Springer, New York, NY, USA, 1998.

[16] M. D. Penrose, *Random Geometric Graphs*, vol. 5, Oxford University Press, Oxford, UK, 2003.

[17] D. Shah, "Gossip algorithms," *Foundations and Trends in Networking*, vol. 3, no. 1, pp. 1–125, 2008.

[18] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: design, analysis and applications," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM '05)*, pp. 1653–1664, Miami, Fla, USA, March 2005.

[19] C. Avin and G. Ercal, "On the cover time and mixing time of random geometric graphs," *Theoretical Computer Science*, vol. 380, no. 1-2, pp. 2–22, 2007.

[20] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Proceedings of the Stochastic Analysis, Control, Optimization and Applications*, pp. 547–566, Birkhäuser, Boston, Mass, USA, 1998.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 895078, 11 pages
http://dx.doi.org/10.1155/2013/895078

Research Article

Trajectory Optimization of Packet Ferries in Opportunistic Social-Based Machine-to-Machine Networks

Xin Guan

School of Information Science and Technology, Heilongjiang University, Harbin 150080, China

Correspondence should be addressed to Xin Guan; guanxin.hlju@gmail.com

Received 16 August 2013; Revised 25 October 2013; Accepted 28 October 2013

Academic Editor: Yan Zhang

Copyright © 2013 Xin Guan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine-to-machine (M2M) communication is an emerging paradigm to connect a large number of devices through wireless technologies. In recent application scenarios, devices are normally carried by people in social contexts. Like the behavior of people, the moving activity of devices always happens in a specific area, which corresponds to some specific communities. How to guarantee the higher packet delivery ratio while reducing forwarding delay in such M2M networks is a challenging issue and has not been well addressed yet. In this paper, we put some special mobile devices (called postmen) in the opportunistic social-based M2M networks and their main responsibility is to carry packets for normal devices which belong to specific communities. Our work focuses on the optimization of the moving trajectory by considering the minimum transmission delay. We formulate the optimal issue into semi-Markov Decision Process model. The decision process includes two parts: packets-choosing strategy and trajectory of packet-ferrying determination strategy. By maximizing the individual reward of a postman, its optimal trajectory will be found. Furthermore, the proposed solution guarantees the packet delivery ratio and delay for isolated communities. Simulation results show that the proposed packet-ferrying solution outperforms the two existing ferrying solutions in terms of packet delivery ratio.

1. Introduction

With the development of wireless technology [1–3], the progress in short range networking, growth of wireless mobile networks, and advances in device networking have allowed the development of a new technology, machine-to-machine (M2M) communications [4], which has recently received considerable attention [5, 6]. Features such as self-organization, ease of deployment, low cost, and infrastructurelessness bring to wireless M2M networks the advantages of robustness, easy maintenance, economy, and flexibility. Therefore, M2M networks are envisioned to be the key technology for next generation wireless networks [7]. An M2M network consists of a large number of self-organized and self-configured nodes. In some real application scenarios, nodes are carried by people; the behavior of human being reflects on the node. Then, these nodes which are carried by people have the social property.

As we know, mobile social network belongs to a typical application branch of M2M network, so, we call it social-based M2M network. Due to the sparse distribution of nodes, it is difficult to guarantee permanent forwarding path between the source and destination nodes [8]. In such case, nodes need to cooperate with each other to achieve the task of multihop forwarding. In mobile social networks, normally, mobile nodes are carried by people [9]. Therefore, the behavior and moving activity of nodes are similar to human activities, which are usually in specific areas [10]. For instance, university students always move around the university campus. From a sociological point of view [11], we call the activity area community.

In a network which is composed by some communities, the communities are typically isolated from each other. For example, the source and destination nodes belong to the same community; it is easy to achieve the goal of packet delivery, so this issue is not discussed in our work. This paper only considers the situation where the source and destination nodes belong to the different communities. In most existing solutions, source node prefers to choose the neighbor node which has higher probability to reach the destination [12] or

the nodes which have higher cluster coefficient [13]. However, there are some instinct limitations in these solutions, and the delivery ratio is difficult to be guaranteed.

Some researchers proposed to set particular mobile nodes (i.e., supernodes) to finish the forwarding task among the isolated communities [14, 15]. These supernodes do not belong to any community and they only need to take packets from different communities [16]. However, these solutions only consider the scenario of stationary nodes [17] or assume the supernodes move according to fixed routes. Packet ferry issue was first proposed in [17]; the authors only considered one ferry node works in sparse mobile wireless networks. The main idea of [17] is to introduce nonrandomness in the movement of nodes and exploit such nonrandomness to help forward packet between isolated communities for social M2M networks. In [18], ferry calculates the fixed trajectory which follows TSP algorithm according to the parameters of community such as coordinate. Then ferry moves along the concluded trajectory and forward packets. Since in different time slot community has different packet generate rate, the request of forward service is dynamic in each community. However, ferry only moves along the concluded trajectory neglecting the dynamic request. Thus, this algorithm wastes a lot forwarding ability of ferry and has very low efficiency. The authors of [19] start with a Markov Decision Problem (MDP) formulation which produces the optimal ferry route that minimizes the average data delivery delay. While this formulation, in principle, enables optimal ferry route design, it is numerically intractable for moderate to large size problems.

In this paper, we call a supernode as postman. Then, how to optimize the moving trajectory for a postman to forward packet efficiently is a key issue. Our basic idea is that, given a packet delivery ratio objective, we minimize the delivery delay as the basis of optimizing the moving trajectory of the postman. In our work, we assume the postman is rational. Towards the forwarding task, the only goal of a postman is to obtain the maximum reward. That is, when the postman forwards packets to a destination between communities, reward can be obtained. Meanwhile, we assume all packets have expired time (TTL) and the forwarding reward and the residual time of packet exist with linear and proportional relationship. It means, within packet TTL, if the postman delivers packet earlier, more reward can be obtained. Otherwise, the postman obtains nothing for a expired packet. To maximize the reward, postman would try to forward packets as soon as possible and reduce the delivery delay. When postman has taken several packets and some destinations are different, how to order the delivery sequence and further optimize the moving trajectory are the main goals in this paper. As we know, in sparse social-based M2M networks, the permanent link between isolated devices is hard to achieve. In such case, some researchers proposed the ferry based solution for packet delivery. And this approach has been proved that it is useful in the real application scenario. As for our proposed solution, the postman device makes moving decision based on the status of the isolated communities; it is fully distributed algorithm. There is no need to introduce any additional centralized algorithm or equipment. From these points, we can see that our proposed solution is useful and practical.

We formulate the optimal trajectory issue as a semi-Markov Decision Process to maximize the reward. The decision process is divided into two decision strategies. Firstly, when a postman arrives at a community, there are several packets that need to be forwarded in this community. Due to the buffer limitation and the requirement of successful delivery, the postman needs to make decision on which packets it would carry according to the attributes of packets. Secondly, after the above decision was made, the postman needs to decide how to move among the different destination communities, and optimize the packets delivery sequence and the moving trajectory. By optimizing the moving trajectory, our solution ensures the packet delivery ratio while reducing the delivery delay.

The remainder of this paper is organized as follows. Section 2 describes the problem formulation of our proposed packet ferry protocol and introduces our network model. In Section 3, we present the optimal model of postman mobility trajectory by using semi-Markov Decision Process. Section 4 models the multiple postman working scenario and concludes the optimal number of postmen which are working in the same network area. Section 5 provides our simulation results. We conclude our work in Section 6.

2. Problem Formulation

In this paper, we adopt the university campus as network model. According to the activity area of the people who belong to the university campus, we divide the entire campus into several subareas or communities, for example, library, dormitory, and so forth. In [14], the authors assume that in each community, there exists at least one mobile node which is appointed as gateway node. Furthermore, gateway nodes never move to the outside of the community. In practical scenario, it is difficult to find such special node to be gateway for the community. Furthermore, the buffer requirement and bandwidth overhead become too heavy for gateway node, since all the nodes within the same community send packets to it. Normally, in mobile social networks, mobile devices are carried by people, it is difficult to find such people who like to do such kind of voluntary job unselfishly. In comparison, this paper introduces one gateway with fixed position for each community. Its only mission is to store the packets which come from other nodes and forward the packets to the packet ferry. We assume that the gateway has the same communication ability as normal mobile node. In addition, each gateway is equipped with finite buffer. Since the number of communities is limited, the implementation of gateway node would not lead to too much additional cost for the entire network. I set one fixed position device for each community. All devices which are in the same community send the packets to this fixed position device. Ferry device only needs to know the location information of fixed position device; it also means the location of community is known for all ferry devices.

Regarding the normal mobile nodes in community, the communication radius is r and the buffer size is b. Mobile

node generates packet to the random destination within a fixed time interval (e.g., ΔT). In order to achieve the goal of packet transportation between different communities, we exploit some ferry nodes in the network. Since ferry node has only one mission of packet delivery, we call it as postman which moves among the communities. In the following section, we optimize the moving trajectory for postman by using semi-Markov Decision Process. By considering the real application scenario, we assume the buffer of postman is finite and the communication radius is r.

In order to describe our work more clearly, we make the following constraints for our network model.

First, all the packets are sent to the gateway of each community. In our work, we ignore the community-detection issues, which means each node knows its belonging community. Secondly, a postman knows the community that the packet should be ferried to. Note that this assumption needs a unique ID for each node. Within time interval ΔT, mobile node generates a packet and assigns destination randomly. In order to save buffer and bandwidth, we initialize active time for each packet, which is denoted by time to live (TTL). If the packet reaches destination before deadline, it is deemed as a positive transmission. In contrast, the transmission is negative and the packet would be erased without retransmission. When a postman reaches a gateway node, it communicates with the gateway node to exchange information. The moving velocity of postman is constant, and its buffer size is fixed and limited.

3. Optimized Trajectory for Postman

3.1. Design Framework. In this section, we discuss how to formulate the moving trajectory of individual postman into semi-Markov Decision Processes (SMDPS). By using the expected discount criterion of SMDPS, we conclude the optimal expression of decision function. Furthermore, based on the optimal expression, we conclude the optimal moving trajectory of postman. Our main goal is to minimize transmission delay of packet and, meanwhile, guarantee the required packet delivery ratio. Here, the optimal trajectory for postman is defined as the route by which the corrected packet is distributed to the destination gateway as quick as possible. We assume that postman is pursuing reward by ferrying packets. By combining packet TTL and reward of postman into consideration, the linear relationship exists between TTL and reward. That is, the less time a postman spends on delivering packet to the destination gateway before TTL, the more reward it would obtain. If the packet is expired, the postman gains nothing. In such case, we conclude the optimal moving trajectory of postman by maximizing the reward, which also means the packets are transmitted with the minimum delay.

3.2. Overview of Proposed Protocol Modeled by SMDP. Since the decision-making process of postman only relies on the current state of the networks, independent from the previous state, postman with the same current states would make the identical decisions to determine what it should do in the next state. Therefore, we formulate this issue as a semi-Markov Decision Process (SMDP), which is composed by the the following six tuples: $\{S, A(i), P_{ij}(a), T(\cdot \mid i, a, j), r(u, i, a, j, t), V, i, j \in S, a \in A(i)\}$. Here, S means the state set. Towards the state $i \in S$, at time point i, the available decision set is $A(i)$. When the system model is in the decision-making time point i and adopts the decision a, its transition probability from state i to j is $P_{ij}(a)$ and the transition time from i to j is the nonnegative random variables of distributed function $T(\cdot \mid i, a, j)$. Under the condition of locating in transition state j and transition time is t, the system would gain $r(u, i, a, j, t)$ as reward. The reward function is shown by

$$r(u, i, a, j, t) = r_1(i, a, j) + \delta_t(u) r_2(i, a, j) + r_3(i, a, j) u. \tag{1}$$

Towards the given function $\delta_t(u)$, only in case $u = t$, the value of given function is 1. Otherwise, in case $u \neq t$, the value of given function is 0. Equation (1) denotes that, when the current state is i and the decision maker adopts the decision a; in the next decision time, the state would transfer to j and transition time is t. Decision maker gains the instantaneous reward $r_1(i, a, j)$ at the beginning of transition. In the process of transition, decision maker gains reward $r_3(i, a, j)$ at the transition finishing time point ($u = t$); it gains the instantaneous reward $r_2(i, a, j)$. V means criterion function of SMDP model.

We adopt the total reward discount criterion in our SMDP model, where time continuously changes and the discount factor is a nonnegative real number (denoted by α). It means that, at time t, the unit reward is only equivalent to $e^{-\alpha t}$ times of time 0. Towards $(i, a) \in \Gamma$, $j \in S$, $t \geq 0$, $\alpha > 0$, we have

$$r_\alpha(i, a, j, t) = \int_0^t e^{-\alpha u} d_u r(u, i, a, j, t). \tag{2}$$

Equation (2) means if the current state is (i, a), the next state would be in j. In such case, under the condition which the sojourn time is t, decision maker would gain the reward which is shown by the following equation:

$$r_\alpha(i, a, j, t) = r_1(i, a, j) + r_2(i, a, j) e^{-\alpha t} + r_3(i, a, j) \frac{1 - e^{-\alpha t}}{\alpha}. \tag{3}$$

When the decision maker is in the nth decision cycle, the reward should be counted from the beginning of nth cycle; it is denoted by $r_\alpha(X_n, \Delta_n, X_{n+1}, t_n)$. Furthermore, when time goes from the beginning of nth to the end (0 point), the reward would be denoted by $e^{-\alpha T_n} r_\alpha(X_n, \Delta_n, X_{n+1}, t_n)$. As for any strategy π and state $i \in S$, in case the decision maker adopts strategy π and in the state i, we can conclude the expected reward from the beginning to the end time point of nth cycle. It is denoted by $E_{\pi,i}\{e^{-\alpha T_n} r_\alpha(X_n, \Delta_n, X_{n+1}, t_n)\}$.

Thus, we conclude that, under the strategy π, the expected discount total reward of N stages starting from the initial state i is shown by

$$V_{N,\alpha}(\pi, i) = \sum_{n=0}^{N-1} E_{\pi,i}\left\{e^{-\alpha T_n} r_\alpha\left(X_n, \Delta_n, X_{n+1}, t_n\right)\right\}, \quad i \in S. \tag{4}$$

The optimal function of expected discount total reward of N stages is shown by

$$V_{N,\alpha}(i) = \sup\left\{V_{N,\alpha}(\pi, i)\, \pi \in \Pi\right\}. \tag{5}$$

3.3. Optimal Model of Postman Mobility Trajectory. Since users in mobile social networks are always located in the specific communities which depend on their living styles or the personal hobbies, the guaranteed delivery means that the packet is distributed within TTL, and we introduce postman to forward packets between different communities. In our work, to avoid the additional cost caused by the postmen, we need to control the number of postmen and their buffer sizes in a limited level. In such case, it is difficult to guarantee the immediate delivery service by a postman when a packet is just generated. We assume that the postmen are working for pursuing maximized rewards. First of all, before a postman takes packet, it needs to ensure the packet could reach the destination community within TTL. Otherwise, they cannot obtain the reward. Also, there exists linear relationship between reward and TTL, which means the more residual time is spent, the higher reward will be obtained. In this way, by maximizing the reward of postman, we conclude the minimum packet transmission delay and also optimize the packet ferry trajectory of the postman. The semi-Markov Decision Process includes two steps. Firstly, when a postman is meeting the gateway of a specific community, the number of packets which are waiting to be delivered is larger than the residual buffer size of the postman; the postman needs to select which packets it should take. The decision is based on the destination and residual alive time of the packet. Secondly, since packets have their individual destinations, the postman needs to optimize its moving trajectory by the use of semi-Markov Decision Process.

To make the optimizing process more clearly, we choose one postman to express our semi-Markov Decision Process. We assume this postman has taken m units packets and the initial buffer size is n ($n > m$). It is located at x_0. The attributes of m packets are denoted by $X_1, X_2, \ldots, X_m$, respectively. Here, X_i means the attribute vector (x_i, t_i) of ith packet, x_i is the destination community of packet i and t_i is the residual alive time of packet i. Regarding the gateway of x_0, there are s candidate packets in its buffer and their attributes are denoted by $Y_1, Y_2, \ldots, Y_s$. Here, Y_i means the attribute vector (y_i, t_i') of ith packet and y_i is the destination community of packet i, t_i' is the residual alive time of packet i. By combining attributes of postman and gateway, we define the state vector as $(n, m, s, x_0, X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_s)$ for our semi-Markov Decision Process. Each state corresponds to a vector; all the state vector are included in set S. We denote the parameter a of state i as $i \to a$.

In our model, each decision step leads the postman to transfer from the current state; we define $a := (i, j)$ as one decision. Also, let $A(i)$ denote all of the decisions coming from i. Towards the transfer probability $P_{ij}(a)$, in case the subscript of i and j is one to one correspondence with the two components of a, $P_{ij}(a) = 1$; otherwise $P_{ij}(a) = 0$. Towards time distribution function $T(\cdot \mid i, a, j)$, a corresponds to i and j. Therefore, $T(\cdot \mid i, a, j)$ represents the time interval for the postman to move from i to j. We assume the velocity of postman is constant; then, $T(\cdot \mid i, a, j)$ is only related to x_0. As mentioned in previous section, reward function $r(u, i, a, j, t)$ is composed by three terms. In our decision model, the reward is only payed in the time point when packet reached destination. It is shown by

$$r(u, i, a, j, t) = \delta_t(u)\, r_2(i, a, j). \tag{6}$$

Furthermore, we have

$$r_\alpha(i, a, j, t) = r_2(i, a, j)\, e^{-\alpha t}. \tag{7}$$

Substitute (7) into (4) and (5), respectively; we have the following two substitution equations:

$$V_{N,\alpha}(\pi, i) = \sum_{n=0}^{N-1} E_{\pi,i}\left\{e^{-\alpha(T_n+t_n)} r_2\left(X_n, \Delta_n, X_{n+1}\right)\right\}, \quad i \in S,$$
$$V_{N,\alpha}(i) = \sup\left\{\sum_{n=0}^{N-1} E_{\pi,i} \frac{\left\{e^{-\alpha(T_n+t_n)} r_2\left(X_n, \Delta_n, X_{n+1}\right)\right\}}{\pi \in \Pi}\right\}. \tag{8}$$

Here, t_n is the sojourn time of nth decision; $T_n = \sum_{i=1}^{n-1} t_i$. We use $V_{N,\alpha}(i)$ as the decision criterion function for our model. Since the transfer probability is 1, we have

$$V_{N,\alpha}(i) = \sup\left\{\sum_{n=0}^{N-1} \frac{\left\{e^{-\alpha(T_n+t_n)} r_2\left(X_n, \Delta_n, X_{n+1}\right)\right\}}{\pi \in \Pi}\right\}. \tag{9}$$

In our model, sojourn time is the moving time between two gateways of postman. Due to the moving velocity is a constant, we assume the velocity of postman is V. In such case, the moving time between $i \to x_0$ and $j \to x_0$ is the sojourn time when the postman is at i:

$$t = \frac{\left|(i \longrightarrow x_0) - (j \longrightarrow x_0)\right|}{V}. \tag{10}$$

When the postman is moving among $a_0, a_1, a_2, \ldots, a_n$ continuously, we conclude the total sojourn time; it is shown by

$$T = \sum_{i=1}^{n-1} t_i. \tag{11}$$

As we mentioned previously, if $x = y$, $\delta(x, y) = 1$. Otherwise, if $x \neq y$, $\delta(x, y) = 0$. We have

$$r_2\left(X_n, \Delta_n, X_{n+1}\right) = \sum_{i=1}^{m} \delta\left(X_n \longrightarrow x_i, X_{n+1} \longrightarrow x_0\right)$$
$$\left(X_n \longrightarrow t_i - \frac{\left((X_{n+1} \longrightarrow x_0) - (X_n \longrightarrow x_0)\right)}{V}\right). \tag{12}$$

Here, $r_2(X_n, \Delta_n, X_{n+1}) > 0$. Then, we conclude $[X_n \rightarrow t_i - (((X_{n+1} \rightarrow x_0) - (X_n \rightarrow x_0))/V)] > 0$. In this section, we introduce the packet choosing strategy. We assume a postman has taken p units packets. In a specific gateway, there are q candidate packets that are waiting for forwarding. Currently, the postman is in state H. In case, $p < n$, where n is the initial buffer size, the postman would try to take each candidate packet sequentially. When the postman is trying to take ith packet, we calculate the maximum expected discounted total reward as

$$V_{p+1,\alpha}(H) = \sup\left\{\sum_{j=0}^{p}\left\{e^{-\alpha(T_j+t_j)}r_2\left(X_j, \Delta_j, X_{j+1}\right)\right\}\right\}. \quad (13)$$

In this time, we assume the postman has only taken Y_i and the current state is X_0. According to specific order, the packets will be forwarded to the different destinations and the state sequence is $X_1, X_2, \ldots, X_{p+1}$. Towards the total reward function, if each term is greater than 0, we define this expected reward as effective. Meanwhile, we conclude one expected discounted total reward $V_{i,1}$. Then, we change the packet forwarding order to calculate the expected reward for the second time. If this order is also effective, we have the expected discounted total reward $V_{i,2}$. According to this ergodic rule, we calculate the correspondence expected reward of u effective order, $V_{i,1}, V_{i,2}, \ldots, V_{i,u}$. Then we have

$$V_{p+1,\alpha}(H)_i = \max\left\{V_{i,1}, V_{i,2}, \ldots, V_{i,u}\right\}. \quad (14)$$

We let all the candidate packets fitting to be taken and calculate every maximum expected discounted total reward, they are shown by $V_{p+1,\alpha}(H)_1, V_{p+1,\alpha}(H)_2, \ldots, V_{p+1,\alpha}(H)_q$. If every value of the given $V_{p+1,\alpha}(H)_i$ is 0, the postman cannot choose any candidate packet. Otherwise, the postman chooses the packet Y_c which corresponds to the maximum $V_{p+1,\alpha}(H)_c$ to be taken. In this time, the postman has taken $p + 1$ units packets. We change the order of Y_1 and Y_c in the state vector; the state vector is shown by $(n, m, s, x_0, X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_s)$ and also m plus 1. If $m = n$, the packet choosing process is done. Otherwise, the postman repeats the choosing process in the rest of the candidate packets.

When the postman finishes the packet choosing process, it has taken m packets and the state is shown by $X_1, X_2, \ldots, X_m$. We substitute them into $\sum_{j=0}^{m}\{e^{-\alpha(T_j+t_j)}r_2(X_j, \Delta_j, X_{j+1})\}$ according to the specific order when they are forwarded to the destination. If this specific order can obtain the ordered expected discounted total reward V_1, the postman ergodic all the orders to calculate the effective ordered expected discounted total reward $V_1, V_2, \ldots, V_u$ and further obtain the maximum value V_c as the optimal ordered expected discounted total reward to adjust state vector $(n, m, s, x_0, X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_s)$ to match the specific order to obtain V_c. In this way, we determine order of the packets which are taken by the postman. From the packet point of view, this order is the forwarding priority. From the postman point of view, this order is the optimal moving trajectory for the packet ferrying.

4. Optimal Number for Multiple Postmen Working Together

As we know, the capability (e.g., buffer size and bandwidth) of each postman is limited, and in the given time, it only can forward the fixed number of packets by using its maximum capability. Therefore, when large number of nodes distributed in the sparse wireless networks, or the packet generation ratio is too high of each node, single postman is hard to guarantee the effective packet forwarding. In such case, multiple postmen should be implemented into the networks. Based on this aspect, under the scenario of given network scale and parameters, how to conclude the optimal number of postmen is an important issue for multiple postmen working in the same area. By using the optimal number of postmen, every postman can work under the condition of optimal packet forwarding load.

As for the given sparse network scale, we define the following necessary parameters for calculating the optimal number of postmen. We set the network running time as t and the TTL packet as T. The moving velocity of postman is V and the initial buffer size is n. There are N communities in our network model and M nodes are distributed in the different communities. We denoted by X_i the ith node and by X the set which includes all nodes. Y_i is the ith community and also denoted the set of nodes which belonged to community Y_i. Y is the set which including all communities. We denoted $\rho(Y_i, Y_j)$ the shortest path from Y_i to Y_j, especially when $i = j$ and $\rho(Y_i, Y_j) = 0$. Towards for all $X_m \in Y_i$, $X_l \in Y_j$, we have the following:

$$\rho\left(X_m, X_l\right) = \rho\left(Y_i, Y_j\right). \quad (15)$$

In order to describe the different scenarios of shortest path existing, we have the following:

$$\delta\left(Y_i, Y_j\right) = \begin{cases} 0, \\ 1. \end{cases} \quad (16)$$

In equation (16), in case $i = j$ or there are other communities between the shortest path of Y_i to Y_j, we have $\delta(Y_i, Y_j) = 0$. Otherwise, in case $i \neq j$ and there is no other community between the shortest path of Y_i to Y_j, we have $\delta(Y_i, Y_j) = 1$. As for all $X_m \in Y_i$, $X_l \in Y_j$, we have $\delta(X_m, X_l) = \delta(Y_i, Y_j)$.

Here, we assume that S_X is the σ algebra of X. As for all $D \in S_X$, $|D|$ denotes the number of nodes of D. In D, we assume one arbitrary node X_m, it generates one packet. Then the probability of this packet is $P(X_m, X_k)$, which the destination is X_k. $r(D)$ is the sparseness measure of field D. We have

$$r(D) = \frac{\sum_{i \neq j}^{Y_i \in D, Y_j \in D} \delta\left(Y_i, Y_j\right)\rho\left(Y_i, Y_j\right)}{\sum_{i \neq j}^{Y_i \in D, Y_j \in D} \delta\left(Y_i, Y_j\right)}. \quad (17)$$

In our work, we consider the packet generation time interval of each node according to poisson distribution. In the network, node X_m generates packet according to poisson distribution with the intensity $\lambda(X_m)$. We consider λ_m as the packet generation intensity for node X_m. In network

field D, we conclude that within system running time t, the mathematical expectation of total number of packet generation is shown by the following equation:

$$E(D,t) = \sum_{X_m \in D} \lambda(X_m) t. \tag{18}$$

Especially, as for all m, l, $m, l \in \{1, \ldots, M\}$, $\lambda_m = \lambda l = \lambda$. Since $E(D,t) = |D|\lambda t$, we assume in field D the probability of arbitrary node X_m generates the packet whose destination is X_k is $P(X_m, X_k)$. Within time interval t, the total number of packet generation is shown by the following equation:

$$W(D) = \sum_{X_m \in D, X_k \in X} P(X_m, X_k) \rho(X_m, X_k) \lambda(X_m) t. \tag{19}$$

Especially, when D is the complete set of X. We have

$$W = \sum_{X_m \in X, X_k \in X} P(X_m, X_k) \rho(X_m, X_k) \lambda(X_m) t. \tag{20}$$

Equation (20) denoted the total working load within time interval t. We also conclude the average update number of hops for each packet and it is shown by the following equation:

$$u = \frac{1}{r(D)} \sum_{X_m \in D, X_k \in D} P(X_m, X_k) \rho(X_m, X_k). \tag{21}$$

As for each packet, the average transmission delay is shown by

$$\bar{t} = \frac{1}{v} \sum_{X_m \in D, X_k \in D} \rho(X_m, X_k) P(X_m, X_k). \tag{22}$$

Further, we have the packet carrying time and it is denoted by $t_s = (T - t_w - \bar{t})/2$; the packet carrying redundancy is $r_t = \bar{t}/t_s$. In the running time interval t, the maximum ferrying capability of postman is $P_w = r_t r_s v_t$. In such case, in network field D, the optimal number of postmen is shown by the following equation:

$$N_r(D) = \frac{W(D)}{P_w}. \tag{23}$$

Especially, during the above process, when D is always the complete set of X, $N_r = W/P_w$ is the perfect state of optimal number of postmen. We substitute (23), we conclude the unary function whose unknown number is N_r. In such case, by solving the N_r, we can conclude the optimal number of postmen in our proposed network. The expression is shown by

$$t_w = \frac{E(D,t) r(D)}{N_r r_s u v_t}. \tag{24}$$

In (24), N_r is the number of postmen and r_s is the expected buffer occupation of postman. In the ideal state, we assume that and under the situation of optimal number of postmen working in the given network field, each postman would work with full load. We have $r_s = n$. However, most of the time, postman cannot work under the condition of full load; in such case, we set a tolerable space occupancy rate and it is denoted by $0 < \varepsilon < 1$ and $r_s = \varepsilon_n$. Based on these descriptions, we have the following equation to show the result of optimal number of postmen:

$$\begin{aligned} N_r = \Bigg(& \sum_{X_m X_k} \left(P(X_m, X_k) P(X_m, X_k) \lambda(X_m) \right) \\ & \times \left(\frac{T}{2} - \frac{E(X,t) r(D)}{N_r r_s u V_t} \right) \Bigg) \\ & \times \left(r_s \sum_{X_m X_k} \rho(X_m, X_k) P(X_m, X_k) \right)^{-1}. \end{aligned} \tag{25}$$

5. Performance Evaluation

In this section, we compare our proposed solution with other three existing packet ferry forwarding protocols. In Section 2, we have overviewed the existing packet ferry protocols which we compared with our proposed solution. In order to obtain the simulation results which are close to the real application scenario, we use the Reality Mining which comes from MIT as our simulation dataset [20]. Furthermore, we use the ONE [21] as our platform for simulation. The ONE is a simulation environment that is capable of generating node movement using different movement models, routing messages between nodes with various DTN routing algorithms and sender and receiver types, and visualizing both mobility and message passing in real time in its graphical user interface. In our simulations, we focus on delivery ratio and the average delivery delay as the key metrics to evaluate the proposed packet-ferrying solution. We define the delivery ratio as the ratio of the number of delivered packets to the total number of packets in our simulation process.

5.1. Reality Mining Data Analysis. Reality Mining comes from MIT Media Laboratory which collected data from 94 volunteers. The dataset contains the following items. Call logs, bluetooth devices (the distance between any two close devices is 5 meters), cell tower IDs, application usage, and phone status. The Reality Mining was collected using mobile phones preinstalled with several pieces of software that recorded and sent the researcher data during the months between September 2004 and June 2005.

Since our main focus is on the scenario in which the nodes have different belonging communities and how to make the ferry forwards packets more effectively, we filter the encountering time and cellular tower IDs which are recorded by subjects between all cellular towers and subjects. Through the location of cellular towers, we can obtain the location information of subjects. Unfortunately, in Reality Mining only the cellular towers that are associated with MIT have location information. Therefore, we choose the areas which are covered by the cellular towers and these areas location information is known for our work.

5.2. Simulation Parameter Setting. In MIT Reality Mining, our simulation runs for 24 hours to collect data. We choose 50 subjects which have the complete information and typical community character as the normal mobile nodes among all 94 subjects. According to the locations of cellular towers in campus map, we divided the whole map into 7 communities. By using the information of contacts with cellular towers of Reality Mining, we obtain the belonging relationship for different communities in the given time period. In our simulations, we set one gateway for each community and the buffer size of gateway has no limitation. We put 5 ferries working in the network to forward packets between 7 isolated communities. Since the moving activity is unknown in community for us, we assume that the normal node is adopted by the Random Way point model as its moving rule. When normal nodes encounter the gateway, they send their packets to the gateway. If some nodes move to another community, they execute the same packet sending action after determining the community belonging. The packets are generated randomly, and the destinations of packets are also randomly assigned. The total simulation time is 24 hours (86900 seconds). We set the initial 400 seconds as the warming-up time period and the last 500 seconds as the terminal time period. In these two special time periods, we do not collect data for final simulation results. The detailed simulation parameters are shown as follows: the velocity of normal node is 5 meter/second; the velocity of ferry node is 15 meter/second; the communication range of normal node, ferry, and gateway is 150 meter. Firstly, we simulate all the protocols under the condition in which the buffer size is 10 units and the packet expired time is 60 minutes. Secondly, based on the different buffer size and packet expired time, different packets distribution model, we compared the performance of all packet forwarding protocols.

5.3. Impacts of Different Packet Generation Rate. In order to observe the different performance under the condition of different packet generation rate, we simulate all packet forwarding protocols in which the number of packet is from 100 to 2000 packets and the packet generation interval is uniform. The simulation results are shown by Figures 1 and 2. In these two figures, since our proposed protocol and HRC protocol are all based on the social characteristic and the distribution of nodes also corresponds to the social networks, from Figure 1, we can see that, under the condition of different packet generation rate, our proposed protocol and HRC show better performance of packet delivery ratio than that of DRR. When the number of packets is larger than 400, our proposed protocol and HRC are all better than FIMF in terms of packet delivery ratio. Furthermore, when the number of packets is larger than 600, the performance of our proposed protocol is also better than that of HRC, and according to the increasing of packet generation ratio, our protocol shows better performance evidently. The main reason is that our proposed protocol optimizes the trajectory of postman for packet ferrying. On the contrary, the trajectory of ferry node of HRC is moving according to the fixed routine. This is also the reason why our protocol shows better performance than

Figure 1: Comparison of delivery ratio under different message number.

that of others in terms of packet delivery delay in Figure 2. In Figure 2, we can see that the packet delivery delay is the highest compared with other three protocols. This is because DRR prefers to provide service for the nodes which have the larger packet generation ratio instead of the nodes whose buffers have more packets. Since the ferry node of HRC protocol moves according to the fixed trajectory, the movement cycle is fixed. In such case, the influence of packet generation ratio is lower in terms of packet delivery delay for HRC protocol. It leads to the result that when the number of packet is less than 500, packet delivery delay of HRC is higher than that of FIMF. On the contrary, when the number of packet is larger than 500, HRC is lower than FIMF. In the traditional scenario, packet delivery delay should increase according to the packet generation ratio increasingly. For example, in Figure 2, the curves of DRR and FIMF are obtained under the condition that, when the number of packet is less than 600. However, in Figure 2, besides our proposed protocol, the packet delivery delay of other three compared protocols is going down according to the packet generation ratio increasingly when the number of packets is larger than 700. This is because when we count up the average delivery delay, we only consider the packet delivery delay which is successful forward to the destination. However, when the number of packet is larger than 700, delivery ratio of other three compared protocols have decreased rapidly according to the increase in packets. It leads to the results that as for the packets which have the predominance of space and time, they can be forwarded by lower delivery delay. On the contrary, the packets which have no predominance would expire before the TTL and cannot be counted. Towards our proposed protocol and HRC, postman provides forwarding

Figure 2: Comparison of average delay under different message number.

Figure 3: Comparison of delivery ratio under different buffer size.

Figure 4: Comparison of average delay under different buffer size.

services for the whole community. Therefore, the increase in packet generation would not impact the delivery delay evidently.

5.4. Impacts of Buffer Size. In Figure 1, we can see that, when the number of packets is less than 500, the delivery ratio of all the four simulated protocols is stable. In order to obtain the comprehensive result of buffer size influence, we choose 600 as the number of packets to simulate the impacts under different buffer size scenarios. We set the buffer size as changing from 5 to 14 units and also, 1 unit means one packet buffer occupation. Figures 3 and 4 show the performance of delivery ratio and delay, respectively, under the changing of buffer size. From these two figures, we can see that, when the number of packets is 600, as for our proposed protocol, the delivery ratio is the highest compared with other ferrying protocols. When buffer size is less than 10 units, the delivery ratio of other three protocols becomes dropping down evidently. The delivery ratio of FIMF is lower than those of HRC and DRR. Since as for FIMF, the ferry node provides forwarding service until normal nodes collect enough packets. In this way, if the buffer size of ferry node is small, the delivery ratio of FIMF is becoming lower. Towards HRC and DRR, they provide forwarding service based on the periodic mode; therefore, their dropping down magnitude is almost the same. Specially, since HRC provides forwarding service based on community structure, its delivery ratio is higher than that of DRR protocol.

From Figure 3, we can see when buffer size of postman is larger than 10 units, besides our proposed protocol, delivery ratio of other three protocols is becoming stable. Meanwhile in Figure 4, the average delivery delay of these three protocols is also becoming stable. It means that besides our proposed protocol, when buffer size is 10, these three compared protocols are in the buffer saturation status. However, as for our proposed protocol, when the buffer size is 4 units, it

has fall in the saturation status and enough buffer for packet forwarding.

5.5. Impacts of Packet Expired Time (TTL). From Figure 3, we can see that when buffer size is larger than 9 units, all protocols are becoming stable in terms of delivery ratio. In order to show the influence of packet expired time (TTL) more clearly, we set the buffer size as 8 units and the number of packets as 600. we simulate all protocols under the condition of TTL to be 30 minutes to 120 minutes to show the performance in terms of packet delivery ratio and delivery delay; the results are shown in Figures 5 and 6.

In Figure 5, we can see that our proposed protocol and HRC show better performance since they are designed based on social community. When TTL is less than 50 minutes, the delivery ratio is in the worst case for these two social-based protocols. In our proposed protocol, we optimize the trajectory of postman; it makes better performance compared with HRC in terms of delivery ratio under the different TTL. As for FIMF and DRR protocol, when TTL is lower than 100 minutes and 70 minutes, respectively, the delivery ratio of them is becoming lower. Meanwhile, when TTL is larger than 70 minutes, DRR shows better performance than FIMF. This is because FIMF needs to wait for normal nodes to collect enough packets to start the forwarding process.

From Figure 6, we observe that as for DRR protocol, under the different TTL, it always in the higher delivery delay. And our proposed protocol is still lower than the other three compared protocols. Furthermore, the TTL changing from 30 to 120 minutes almost cannot impact the performance of our proposed protocol.

5.6. Impacts of the Distribution for Packet Destination. In order to make our simulation results much closer to the real application scenario, we make the destination of packet obey the normal distribution and according to the changing of parameter δ to simulate the proposed protocols. In the simulation process, we set the parameter δ changing from 1 to 10. According to the probability density function of normal distribution, we know that if δ is smaller, the destination of packets is more concentrated. The simulation results are shown by Figures 7 and 8.

In Figure 7, we can observe that parameter δ almost cannot impact the performance of our proposed protocol in terms of packet delivery ratio. When δ is larger than 6, delivery ratio of FIMF is becoming lower according to the increase in δ. This is because FIMF provides forwarding service for normal node; when there are several nodes which are located in the network and have much more packets than other normal nodes, FIMF prefers to provide all forwarding services to these special nodes. This also is the reason why DRR has the lower delivery ratio when δ is larger than 7. As for HRC protocol, when δ is less than 4, its delivery ratio increases according to the increase in δ. When δ is larger than 4, the delivery ratio of HRC is tending to be stable. This is because the postman of HRC is moving according to the periodic mode. Towards our proposed protocol, we optimize the trajectory of postman; it leads to the results that before the moving activity, postman determines the optimal traveling trajectory. Therefore, the distribution parameter δ has hard impact on the delivery ratio performance of our protocol.

From Figure 8, we observe that as for all four compared protocols, the packet delivery delay is hard to be impacted

Figure 5: Comparison of delivery ratio under different TTL.

Figure 6: Comparison of average delay under different TTL.

Figure 7: Comparison of delivery ratio under different parameter δ.

Figure 8: Comparison of average delay under different parameter δ.

by distribution parameter δ. DRR protocol still has the highest delivery delay. Our proposed protocol shows the best performance in terms of delivery delay. However, when δ is less than 4, since the destination of packet is too concentrated, it makes the delivery delay higher according to the decrease in δ.

6. Conclusion

Packet routing and forwarding is one of the most important issues for opportunistic social-based M2M networks. In this paper, we focus on the scenario that nodes are distributed in the sparse area. This scenario is a real application case for social-based M2M networks. By analyzing the moving behavior of supernode (called postman in this paper), we optimize the moving trajectory of postman for opportunistic social-based M2M networks. We formulate the optimization issue as a semi-Markov Decision Process model to produce the delivery sequence and conclude the optimal moving trajectory of postman. We combine the pursuing reward attribute of postman and the TTL of packet into one reward function. We achieve the goal of minimizing delivery delay while improving the delivery ratio of packets. Simulation results show our ferrying solution performs better than other existing schemes in opportunistic social-based M2M networks.

Acknowledgment

This work is supported by the Heilongjiang Province Education Department Foundation 12531Z007.

References

[1] G. Nan, Z. Mao, M. Li et al., "Distributed resourceallocation in cloud-based wireless multimedia social networks," *IEEE Network Magazine*. accepted.

[2] G. Nan, Z. Mao, M. Yu, M. Li, H. Wang, and Y. Zhang, "Stackelberg game for bandwidth allocation in cloud-based wireless live-streaming social networks," *IEEE Systems Journal*, 2013.

[3] E. Palomar, A. Alcaide, E. Molina, and Y. Zhang, "Coalitional games for the management of anonymous access in online social networks," in *Proceedings of the 11th International Conference on Privacy, Security and Trust (PST '13)*, pp. 1–10, July 2013.

[4] G. Lawton, "Machine-to-machine technology gears up for growth," *Computer*, vol. 37, no. 9, pp. 12–15, 2004.

[5] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: architectures, standards, and QoS improvement," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, 2011.

[6] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive machine-to-machine communications: visions and potentials for the smart grid," *IEEE Network Magazine*, vol. 26, no. 3, pp. 6–13, 2012.

[7] N. Tekbiyik and E. Uysal-Biyikoglu, "Energy efficient wireless unicast routing alternatives for machine-to-machine networks," *Journal of Network and Computer Applications*, vol. 34, no. 5, pp. 1587–1614, 2011.

[8] C. Liu and J. Wu, "An optimal probabilistically forwarding protocol in delay tolerant networks," in *Proceedings of the 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '09)*, pp. 105–114, 2009.

[9] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," in *Proceedings of the 29th IEEE Conference on Information Communications (INFOCOM '10)*, pp. 1–9, San Diego, Calif, USA, March 2010.

[10] X. Lin, R. Lu, X. Liang, and X. Shen, "STAP: a social-tier-assisted packet forwarding protocol for achieving receiver-location privacy preservation in VANETs," in *Proceedings of the IEEE Conference on Information Communications (INFOCOM '11)*, pp. 2147–2155, Shanghai, China, April 2011.

[11] F. Tinnies, "Community and Society: (GEMEINSCHAFT AND GESELLSCHAFT)," 1887.

[12] Q. Yuan, I. Cardei, and J. Wu, "Predict and relay: an efficient routing in disruption-tolerant networks," in *Proceedings of the 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '09)*, pp. 95–104, May 2009.

[13] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANETs," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 32–40, September 2007.

[14] W. Zhao, M. Ammar, and E. Zegura, "A message ferrying approach for data delivery in sparse mobile Ad Hoc networks," in *Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MoBiHoc '04)*, pp. 187–198, May 2004.

[15] M. M. B. Tariq, M. Ammar, and E. Zegura, "Message ferry route design for sparse ad hoc networks with mobile nodes," in *Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '06)*, pp. 37–48, May 2006.

[16] J. Fan, Y. Du, W. Gao, J. Chen, and Y. Sun, "Geography-aware active data dissemination in mobile social networks," in *Proceedings of the 7th IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS '10)*, pp. 109–118, November 2010.

[17] T. He, K. W. Lee, and A. Swami, "Flying in the dark: controlling autonomous data ferries with partial observations," in *Proceedings of the 11th ACM International Symposium on Mobile ad Hoc Networking and Computing (MobiHoc '10)*, pp. 141–150, 2010.

[18] Y. Chen, W. Zhao, M. Ammar, and E. Zegura, "Hybrid routing in clustered DTNs with message ferrying," in *Proceedings of the 1st International MobiSys Workshop on Mobile Opportunistic Networking (MobiOpp '07)*, pp. 75–82, San Juan, Puerto Rico, USA, June 2007.

[19] A. Mansy, M. Ammar, and E. Zegura, "Deficit round-robin based message ferry routing," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference (GLOBECOM '11)*, pp. 1–5, December 2011.

[20] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, CRAWDAD data set cambridge/haggle (v. 2009-05-29), May 2009, http://crawdad.cs.dartmouth.edu/meta.php?name=cambridge/haggle.

[21] http://www.netlab.tkk.fi/tutkimus/dtn/theone/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 928528, 8 pages
http://dx.doi.org/10.1155/2013/928528

Research Article

A Vector Algebraic Algorithm for Coverage Compensation in Hybrid Wireless Sensor Networks

Ning-ning Qin,[1,2] Li-xia Guo,[1] Zhi-guo Ding,[2,3] and Bao-guo Xu[1,2]

[1] *Key Laboratory of Advanced Process Control for Light Industry, Jiangnan University, Wuxi 214122, China*
[2] *The Engineering Research Center of IOT Technical Application, School of IoT Engineering, Jiangnan University, Wuxi 214122, China*
[3] *School of Electrical, Electronic and Computer Engineering, Newcastle University, Newcastle, UK*

Correspondence should be addressed to Ning-ning Qin; ningning801108@163.com

Received 8 July 2013; Revised 7 November 2013; Accepted 7 November 2013

Academic Editor: Jianhua He

Copyright © 2013 Ning-ning Qin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In wireless sensor networks, coverage holes are caused by energy depletion at some nodes, and the aim of this paper is to study how to utilize the redundant nodes with remaining energy. Particularly, this paper proposes a vector algebra based algorithm by exploring redundant nodes as an extra dimension for coverage compensation. This algorithm consists of two parts. One is to find the locations of potential redundant nodes for coverage compensation; and the other is to opportunistically select the best redundant nodes by jointly considering the hole boundaries and the remaining energy of nodes. Simulation results are provided to demonstrate that the proposed algorithm minimizes the energy consumption when repairing the holes for full coverage. Furthermore, compared with other algorithms, the proposed one exhibits better performance in terms of moving distance and energy consumption.

1. Introduction

Wireless sensor networks (WSNs) are composed of numerous small-size sensors that have various features, such as low cost, light weight, high mobility, and capabilities for sensing, computing, and communications. The sensor has the same capabilities as the machine does. In fact, WSNs are one of typical application of machine to machine communications. The "smart home" is just an example. It is not only a good application based on WSN but also an instance of machine to machine communications. With these features, WSNs have been widely used in many fields, including intelligent transportation, military applications, infrastructure monitoring, antiterrorism disaster relief, environmental monitoring, wireless healthcare, and target tracking [1–3].

Coverage is an important metric of the quality of service (QoS) in WSNs [4, 5] and is a fundamental issue since each sensor is energy constrained. The reason why some areas are not covered is due to random deployment of sensors, node energy depletion, or environmental damages. Consequently, coverage holes appear in various parts of WSNs [6, 7], significantly reducing the QoS and the lifetime of the network. An important observation is that a number of inactive redundant nodes also exist in the network, which results in the simultaneous coexistence of two types of sensors, one with extra energy and the other with drained batteries. In hybrid WSNs, such a difficulty can be overcome by moving mobile redundant sensors to the locations at which coverage holes occur.

Many recent studies on patching holes by exploring mobile nodes are based on geometry. Reference [8] proposed the Coverage Hole Patching Algorithm (CHPA), which selects the patching point along the perpendicular bisector between any two adjacent hole boundary nodes. The distances between the selected point and the corresponding two boundary nodes are equal to the communication radius. This algorithm can be easily implemented due to its low complexity but suffers low resource efficiency. To overcome this shortcoming, the two-point vertical determination method was modified to a three-point approach in the Patching Algorithm Triangle by Triangle based on mobile node (PATT) [9]. Compared with CHPA, PATT can achieve better stability

and higher coverage but does not consider the use of the existing redundant sensors. Reference [10] proposed the use of a vector algebra distributed method to achieve effective deployment of redundant nodes. This algorithm effectively uses the existing redundant resources while maintaining a sufficient level of coverage. However, [10] did not consider the energy consumption caused by long-distance moving.

In order to use such redundant node resources, this paper proposes a novel mobile node scheduling algorithm, namely, a Dispatching Algorithm based on vector algebra at mobile node (DAVM), for patching holes in hybrid WSNs. Localization algorithms are independently carried out at both static and mobile nodes. Particularly, the first step of the proposed algorithm is to find where patching position is. Then the mobile sensors adjacent to boundaries calculate their moving distance and direction towards the patching position. Static nodes will select their best patching redundant nodes individually. The criterion for node selection is based on the largest remaining energy which is determined by moving distance and initial energy of redundant node. At last, the selected nodes will be activated and asked to move towards the patching positions. Those nodes are not utilized before they are activated. DAVM provides an effective method to utilize redundant nodes. It is worthy pointing out that redundant nodes calculate their corresponding patching positions by using the local and one-hop positioning information. And the proposed node scheduling method can not only avoid patching conflict caused by more than one redundant node nearby the boundaries but also minimize the energy consumption caused by node moving by inviting the nodes close to the boundaries. Simulation results are provided to demonstrate the superior performance of the proposed scheduling algorithm by comparing it with some existing schemes.

2. Preliminaries

2.1. Network Model. Consider a hybrid wireless sensor network which consists of randomly deployed n sensor nodes. These sensors comprise a set C, including both static nodes and mobile redundant nodes. Static nodes do not have energy harvesting ability, whereas mobile redundant nodes cannot move and sense until they are activated. Mobile redundant nodes have the same capacities as static nodes in terms of communication, sensing, and data processing. Moreover, mobile redundant nodes can have thea priori information about their own initial energy. The binary perception model is adopted at all of the nodes [11]. According to [12–14], the communication radius (R_c) is assumed to be twice the coverage radius (R_s) in order to obtain a connected network. Each node $u(x_u, y_u)$ can obtain the relative positions of adjacent nodes by using its directional antenna and algorithms in [15, 16]. The probability that the sensing discs of any three sensors would intersect at a single point is close to zero because of the random deployment of the nodes. No more than one sensor is located at the same point. And every sensor node should have at least two neighbors to avoid the isolated coverage or teardrop shaped coverage. In addition, the location information of the sink node is known to all other nodes.

Figure 1: Coverage holes.

We assume that the coverage holes are closed and inside the monitoring region. We can obtain all the information about the boundary nodes of the coverage hole by using hole detection algorithm [17]. In [17], the maximum simple complex is built according to network topology. When a node is adjacent to the edges of the maximum simple complex network, this node is a hole boundary node but not vice versa.

2.2. Relative Definitions

Definition 1 (coverage hole). If a continuous area is not covered by any sensor node in WSNs, this area is defined as a coverage hole [19]. As shown in Figure 1, the shadow area is a coverage hole. Given a set of boundary nodes $S = \{S_i \mid i = 1, 2, \ldots, \text{Num}\}$, where $S \subset C$, $S_i \in S$ is a hole boundary node.

Definition 2 (hole boundary neighbor). The two nodes $S_i, S_j \in S$ are mutually hole boundary neighbors if $d(S_i, S_j) \leq R_c$ where $d(\cdot)$ is the 2D Euclidean distance between the two nodes.

Definition 3 (hole boundary intersection). If two hole boundary nodes intersect at one point within a hole that is beyond the scope of the third node's sensing range, then this intersection point is called a hole boundary intersection.

3. Vector Algebra Principles

The fundamental function of DAVM is to calculate the potential patching locations. In consideration of the existing coverage hole, vector algebra can be used to move the

(a) $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) < \pi$ (b) $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) < \pi$

(c) $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) > \pi$ (d) $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) > \pi$

Figure 2: Mobility azimuth of R.

redundant nodes. Without any constraints, more than one redundant node can patch the hole. Therefore, choosing the best node is another challenging problem considered in DAVM. To be considered effective, DAVM needs to guarantee that the patching node covers the hole without splitting the hole; that is, the patching node does not generate new holes [20, 21].

Thus, DAVM consists of two parts: (1) calculating the displacement of the redundant node, including the direction and the distance; and (2) identifying the hole boundary node and scheduling the best redundant node.

3.1. Displacement of Redundant Node

(i) Moving Direction. Consider a part of hole boundary that includes S_1 and its neighbors S_2 and S_3, as illustrated in the following example. Given a redundant node $R(x_R, y_R)$, their sensing circles of $R(x_R, y_R)$ intersect with S_1 and S_3 at $a(x_a, y_a)$ and $b(x_b, y_b)$ as shown in Figure 2. The following can be calculated:

$$\begin{aligned} \overrightarrow{V_1} &= (x_a - x_R)\vec{i} + (y_a - y_R)\vec{j}, \\ \overrightarrow{V_2} &= (x_b - x_R)\vec{i} + (y_b - y_R)\vec{j}, \\ \overrightarrow{V_R} &= \overrightarrow{V_1} + \overrightarrow{V_2}, \end{aligned} \tag{1}$$

where $\theta(\vec{i}, \vec{j})$ is defined as the mobility azimuth that is counterclockwise from $\vec{i}$ to $\vec{j}$. If V_2 is regarded as the reference vector, the mobility azimuth of R is defined as $-\theta(\overrightarrow{V_2}, R)$ or simply as $\theta(R)$. Let the auxiliary vector $\overrightarrow{V}$ satisfy $\theta(R) = \theta(\overrightarrow{V})$ and $\| \overrightarrow{V} \| = \| \overrightarrow{V_R} \|$. The following four cases are considered for the calculation of $\overrightarrow{V}$, as shown in Figures 2(a)–2(d). Then, $\theta(R)$ is obtained.

Case 1. When $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) < \pi$, $\forall P_k$, if $d(R, P_k) \le R_s$, there is $\overrightarrow{V} = \overrightarrow{V_R}$.

Case 2. When $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) < \pi$, $\exists P_k$, if $d(R, P_k) > R_s$, there is $\overrightarrow{V} = -\overrightarrow{V_R}$.

Case 3. When $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) > \pi$, $\forall P_k$, if $d(R, P_k) \le R_s$, there is $\overrightarrow{V} = -\overrightarrow{V_R}$.

Case 4. When $\theta(\overrightarrow{V_1}, \overrightarrow{V_2}) > \pi$, $\exists P_k$, if $d(R, P_k) > R_s$, there is $\overrightarrow{V} = \overrightarrow{V_R}$.

(ii) Moving Distance. The moving distance of R, L_0 is dependent on both R and hole boundary intersections. As shown in Figure 3, the hole boundary node S_2 exists; its sensing range intersects the ranges of S_1 and S_3 at $P_1(x_{P_1}, y_{P_1})$ and $P_2(x_{P_2}, y_{P_2})$, respectively. The calculation flow of L_0 is shown as follows:

$$\begin{aligned} \overrightarrow{V_{RP_k}} &= \left(x_{P_k} - x_R\right)\vec{i} + \left(y_{P_k} - y_R\right)\vec{j}, \\ \cos\theta\left(\overrightarrow{V_{RP_k}}, \overrightarrow{V}\right) &= \frac{\overrightarrow{V_{RP_k}} \bullet \overrightarrow{V}}{\left|\overrightarrow{V_{RP_k}}\right| \times \left|\overrightarrow{V}\right|}, \\ L_0 &= L_{RP_k} \times \cos\theta\left(\overrightarrow{V_{RP_k}}, \overrightarrow{V}\right) \\ &\quad + \sqrt{{R_s}^2 - {L_{RP_k}}^2 + {L_{RP_k}}^2 \cos^2\theta\left(\overrightarrow{V_{RP_k}}, \overrightarrow{V}\right)}, \quad \text{where} \\ L_{RP_k} &= \sqrt{\left(x_R - x_{P_k}\right)^2 + \left(y_R - y_{P_k}\right)^2}. \end{aligned} \tag{2}$$

The new position R' can be determined by using $\theta(R)$ and L_0. Two new positions are calculated by P_1 and P_2. An R' that satisfies $d(R', P_k) \le R_s$ $(k \in [1, 2])$ is chosen as the new position. Therefore, the moving distance should be calculated by the chosen R'.

3.2. Best Redundant Node. More than one node may try to participate into patching because of the high mobility of redundant nodes. To avoid patching collision, the best redundant node should be selected in terms of energy and moving distance. Thus, the evaluated remaining energy $E_r(i)$ of the redundant node R_i can be calculated as follows:

$$E_r(i) = \alpha E_0(i) - \beta L_0(i), \tag{3}$$

Figure 3: Moving distance based on the new position R'.

where $E_0(i)$ is initial energy of node i, $\alpha(\alpha \geq 1)$ is the energy endurance parameter; β is the consumption of energy per unit distance, that is, the energy consumption parameter, which is governed by the given work environment and the physical properties of the nodes.

Let E_{th} be the threshold energy required for working. The best moving node R_{best} can be selected if $E_r(R_{\text{best}})$ is larger than both E_{th} and that of the other nodes.

4. DAVM

In DAVM, the hole boundary node S_j broadcasts the information regarding the hole. Any redundant node R_i that receives this message runs the displacement algorithm and replies to S_j. Then, S_j activates the best redundant node R_{best} by using the feedback messages from the redundant nodes. And node R_{best} moves to a new location. Let the queue Q_{all} record the displacements and energy information of the redundant nodes. The procedures of the displacement algorithm and the best redundant node schedule are presented in detail in the following sections.

4.1. Displacement Algorithm. When a redundant node R_i receives the information on the hole, the following steps are carried out.

Step 1. Calculate $L_0(i)$, $\theta(i)$, and $E_r(i)$.

Step 2. Reply $L_0(i)$, $\theta(i)$, $E_0(i)$, and $E_r(i)$ to S_j.

Step 3. Start the timer T.

Step 4. If T is out of time, proceed to Step 6.

Step 5. If R_i receives the activated command from S_j, it will be activated and move to the new position based on $(\theta(i), L_0(i))$; otherwise, proceed to Step 4.

Step 6. Enter a dormancy state.

4.2. Best Redundant Node Schedule. The hole boundary node S_j runs the following.

Step 1. Calculate the value of $\theta(\overrightarrow{V_{S_jS_k}}, \overrightarrow{V_{S_jS_i}})$, where S_i is the right hole boundary neighbor and S_k is the left one. If $\theta(\overrightarrow{V_{S_jS_k}}, \overrightarrow{V_{S_jS_i}}) < \pi$, proceed to Step 2; otherwise, select the next hole boundary node counterclockwise until the selected node S_j satisfies $\theta(\overrightarrow{V_{S_jS_k}}, \overrightarrow{V_{S_jS_i}}) < \pi$.

Step 2. S_j broadcasts the hole information.

Step 3. Wait for the feedback messages from the redundant nodes and update Q_{all}.

Step 4. Based on Q_{all}, S_j selects the best redundant node R_{best}.

Step 5. Send activated command to R_{best}.

Step 6. Update the set S.

4.3. DAVM for Multiple Holes. In applications, it is a fact that there is more than one hole in WSN because of random deployment. If these holes are separated and satisfy the network model in Section 2.1, DAVM can still work effectively through patching them one by one. It is not overlooked that the patching order of holes might affect the results of DAVM. That is, the set of nodes used for holes patching may not be the same if the holes are patched in different orders. But the local optimality can be obtained when patching multiple holes in sequence, because DAVM can select the best redundant node from the remaining redundant node.

Table 1: Simulation parameters and calculations.

Parameters	Calculation description
Coverage hole region	A N-polygon about 12 meters on a side, where $N \in [6, 29]$
Area of hole S	$36N/\tan(\pi/N)$ (m^2)
Hole-boundary node number	N
Redundant node number M	$144N/\pi R_s^2 \tan(\pi/N)$
Initial energy $E_0(i)$ of redundant node i	$E_0(i) \in [90, 110]$ (J)
Communication radius R_c	16 m
Coverage radius R_s	8 m
Energy endurance parameter α	1 (network without endurance ability)
Energy consumption parameter β [18]	1 J/m
Energy threshold E_{th}	60 J

5. Simulation

5.1. Simulation Environment. The simulations are implemented by using Matlab. DAVM is evaluated and compared with the coverage optimization algorithm (COA) mentioned in [10]. Note that COA is also based on vector algebra. In consideration of the randomness in WSNs, the simulation is performed 20 times with the same topology for each algorithm in each scenario according to Monte Carlo theory. The simulation focuses on a normal N-polygon hole, and the N vertices are the N hole boundary nodes. The length of the edge is constant; thus, more N corresponds to a larger hole. For the multiple holes, DAVM can be usedpatching these holes one by one.

Furthermore, the initial energies and positions of the redundant nodes are random in the coverage hole. The simulation parameters and calculations are given in Table 1.

5.2. Simulation Analysis

5.2.1. Coverage Evaluation. The coverage ratio is the ratio of the covered area to the area of hole. Figure 4 shows the relationship between the coverage ratio and the number of mobile nodes Sum in different N-polygon holes, where $N \in [17, 29]$. After running DAVM, the coverage ratio approximately increases linearly with an increasing number of mobile redundant nodes. Repeated experiments reveal that the increasing trend slows down with the enlarging hole using linear fitting method. Let the line slope K be the trend, in which the coverage varies with the number of mobile nodes; obviously, $K \geq 0$. In the experiment, the number of edges of hole N increases from 17 to 29, but K decreases from 0.0551 to 0.0178. Thus, for a larger area of the hole, the coverage gain of additional nodes is not significant, because more additional nodes are needed to compensate the larger coverage losses.

Figure 4: Number of mobile nodes versus coverage ratio.

Figure 5: Evaluation on number of mobile nodes.

5.2.2. Evaluation of the Total Number of Mobile Nodes. As shown in Figure 5, the number of activated redundant mobile nodes is simulated for various hole sizes. The number of redundant mobile nodes activated Sum increases with an enlarging hole in both algorithms. Compared with DAVM, COA requires smaller mobile nodes for a larger hole ($N >$

Figure 6: Evaluation on average moving distance L.

Figure 7: Evaluation of average remaining energy E, where coordinates (X, Y, Z) refer to β, N, and E.

12), which is due to the fact that COA can only guarantee 90% coverage. On the other hand, DAVM is to achieve approximately 100% coverage but requires the activation of more mobile nodes. Or in other words, DAVM can provide a better performance than COA in terms of hole compensation with a price of more system complexity.

5.2.3. Evaluation of Average Moving Distance. The moving distance of the mobile node directly affects the energy consumption. To analyze the average moving distance, several algorithms are simulated to evaluate the energy consumption in different holes. The average moving distance L is defined as follows:

$$L = \frac{1}{\text{Sum}} \sum_{i=1}^{\text{Sum}} L_0 \left(R_{\text{best},i}\right). \tag{4}$$

As shown in Figure 6, the average moving distance L shows an obvious pulsation when the area of the hole increases (i.e., $N \uparrow$) because COA does not consider the remaining energy of nodes. In DAVM, L stably ranges from 5 to 6, which is smaller than the range from 6 to 8 in COA. This finding verifies that the best redundant node schedule balances the energy.

5.2.4. Evaluation of Average Remaining Energy. Figure 7 shows the average remaining energy with the hole size N and β. The average remaining energy E is defined as follows:

$$E = \frac{1}{\text{Sum}} \sum_{i=1}^{\text{Sum}} E_r \left(R_{\text{best},i}\right). \tag{5}$$

DAVM takes the remaining energy of nodes into the consideration when designing the scheduler, which is important to effectively prevent larger moving distances of nodes and hence avoid consuming excess energy. As shown in Figure 7, the average remaining energies of both algorithms fluctuate minimally with various N and β. Compared with COA, DAVM has smoother and higher energy surface, and the fluctuation range is maintained within 1%. As result, the decrease of the network lifetime brought by using DAVM would be less than the use of COA. With an increase in β, E decreases, especially in COA. Therefore, DAVM is less sensitive to β.

5.2.5. Evaluation of Average Computation Time. The comparison of average computation time of nodes under several conditions is shown in Table 2, and Figure 8 presents the corresponding box plots. As shown in Figure 8, the computation time ranges from 0.304 s to 0.399 s for different N-polygon holes. Given a certain β, DAVM has consistently shorter computation time than COA. The performance of DAVM is more effective than COA.

6. Conclusions

This paper has proposed a novel algorithm to combat coverage holes by using redundant nodes in a wireless sensor network. The proposed algorithm achieves a balanced tradeoff between the initial energy and the moving energy consumption and chooses the best redundant node to patch the hole. Simulation results have been provided to show that DAVM is superior to COA for several aspects. However, DVAM may not be effective to the scenarios with sparse redundant nodes. A promising future direction is to focus on coverage holes with sparse redundant nodes.

Figure 8: Box plots of simulation results for average computation time of each node.

Table 2: Average computation time of each node.

Scenario		Average computation time of each node (sec)					
		DAVM			COA		
#	N	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 1$	$\beta = 2$	$\beta = 3$
1	6	0.393	0.376	0.367	0.382	0.367	0.329
2	8	0.376	0.344	0.379	0.312	0.378	0.364
3	10	0.388	0.380	0.304	0.369	0.347	0.360
4	12	0.372	0.395	0.372	0.356	0.395	0.392
5	14	0.338	0.367	0.366	0.361	0.346	0.360
6	16	0.368	0.348	0.388	0.327	0.320	0.388
7	18	0.363	0.393	0.336	0.329	0.340	0.330
8	20	0.390	0.399	0.347	0.362	0.391	0.312

Acknowledgments

All authors thank the effort from Dr. Yang Le. This research was financially supported by the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (no. JUSRP11129), the Doctoral Fund of Ministry of Education of China (no. 20100093120007), and the National Natural Science Fund of China (no. 61304214).

References

[1] Y. Younghwan and D. P. Agrawal, "Mobile sensor relocation to prolong the lifetime of wireless sensor networks," in *Proceedings of the IEEE 67th Vehicular Technology Conference (VTC '08)*, pp. 193–197, Singapore, May 2008.

[2] J. Kanno, J. G. Buchart, R. R. Selmic, and V. Phoha, "Detecting coverage holes in wireless sensor networks," in *Proceedings of the 17th Mediterranean Conference on Control and Automation (MED '09)*, pp. 452–457, Thessaloniki, Greece, June 2009.

[3] J. Teng, T. Bolbrock, G. Cao, and T. la Porta, "Sensor relocation with mobile sensors: design, implementation, and evaluation," in *Proceedings of the IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems (MASS '07)*, pp. 1–9, Pisa, Italy, October 2007.

[4] Z.-J. Han, Z.-B. Wu, R.-C. Wang, L.-J. Sun, and F. Xiao, "Novel coverage control algorithm for wireless sensor network," *Journal on Communications*, vol. 32, no. 10, pp. 174–184, 2011.

[5] X. Gu, L. Sun, J. Guo, and F. Xiao, "An equilibrium optimization strategy on coverage energy consumption in wireless sensor networks," *Chinese Journal of Sensors and Actuators*, vol. 23, no. 11, pp. 1627–1632, 2010.

[6] D. R. Raymond, R. C. Marchany, M. I. Brownfield, and S. F. Midkiff, "Effects of denial-of-sleep attacks on wireless sensor network MAC protocols," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 367–380, 2009.

[7] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[8] J. Yao, G. Zhang, J. Kanno, and R. Selmic, "Decentralized detection and patching of coverage holes in wireless sensor networks," in *Intelligent Sensing, Situation Management, Impact Assessment, and Cyber-Sensing*, Proceedings of SPIE, Orlando, Fla, USA, April 2009.

[9] L.-M. Wang, F. Li, and Y. Qin, "Resilient method for recovering coverage holes of wireless sensor networks by using mobile nodes," *Journal on Communications*, vol. 32, no. 4, pp. 1–8, 2011.

[10] Y. P. Deng and C. P. Wu, "A coverage optimization algorithm based on mobile nodes in wireless sensor networks," *Application Research of Computers*, vol. 29, no. 8, pp. 3134–3137, 2012.

[11] N. Heo and P. K. Varshney, "Energy-efficient deployment of intelligent mobile sensor networks," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 35, no. 1, pp. 78–92, 2005.

[12] X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration in wireless sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys '03)*, pp. 28–39, New York, NY, USA, November 2003.

[13] H. Zhang and J. C. Hou, "Maintaining sensing coverage and connectivity in large sensor networks," *Wireless Ad Hoc and Sensor Networks*, vol. 1, no. 1-2, pp. 89–124, 2005.

[14] D. Tian and N. D. Georganas, "Connectivity maintenance and coverage preservation in wireless sensor networks," *Ad Hoc Networks*, vol. 3, no. 6, pp. 744–761, 2005.

[15] X. Wang, Z. Wang, and B. O'Dea, "A TOA-based location algorithm reducing the errors due to non-line-of-sight (NLOS) propagation," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 1, pp. 112–116, 2003.

[16] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 37, no. 6, pp. 1067–1080, 2007.

[17] F. Yan, P. Martins, and L. Decreusefond, "Connectivity-based distributed coverage hole detection in wireless sensor networks," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference (GLOBECOM '11)*, pp. 1–6, Houston, Tex, USA, December 2011.

[18] A. Sekhar, B. S. Manoj, and C. Siva Ram Murthy, "Dynamic coverage maintenance algorithms for sensor networks with limited mobility," in *Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications (PerCom '05)*, pp. 51–60, Kauai Island, Hawaii, USA, March 2005.

[19] H. Su and Y. Wang, "A self-healing algorithm without location information in sensor networks," *Chinese Journal of Computers*, vol. 32, no. 10, pp. 1957–1970, 2009.

[20] J. Liu, Y. Feng, H. Y. Xu, J. S. Xue, and M. R. Qian, "Greedy approximation algorithm of patching hole in wireless sensor networks," in *Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet '12)*, pp. 2604–2608, Yichang, China, April 2012.

[21] K. Yang, Q. Liu, S. K. Zhang, J. Li, and D. L. Weng, "Hole recovery algorithm based on mobile inner nodes in wireless sensor networks," *Journal on Communications*, vol. 9, no. 33, pp. 116–117, 2012.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 814267, 11 pages
http://dx.doi.org/10.1155/2013/814267

Research Article

Damping Based Traffic Allocation in Wireless Machine-to-Machine Communications Networks

Ruiguo Yu,[1,2] Zhenxia Yuan,[1,2] Mankun Zhao,[1,2] Mei Yu,[1,2] and Xu Lu[1,2]

[1] *School of Computer Science and Technology, Tianjin University, Tianjin 300072, China*
[2] *Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin 300072, China*

Correspondence should be addressed to Mei Yu; yumei@tju.edu.cn

Received 18 August 2013; Accepted 10 October 2013

Academic Editor: Yan Zhang

Copyright © 2013 Ruiguo Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine-to-machine (M2M) technology that allows machine devices to communicate directly with each other without human intervention has emerged as a cutting-edge technology for next-generation communication. As its applications become more and more popular, the number of users grows quickly and the quality of service needs to be improved. The traffic allocation problem may lead to many other problems such as low utilization ratio of the bandwidth and the jitter. When we tried to analyze the load on different paths using the simplest multipath routing algorithms, we found that the load changes just like simple harmonic motion, which inspires us to use physic and mathematic methods and tools to analyze the loading balance problems. And then, we designed a new traffic allocation algorithm. Compared with other algorithms, the algorithm proposed in this paper has much wider applicability. And the performance comparison experiment shows that DBTAs can really achieve improvement in decreasing the packet delay variance.

1. Introduction

Machine-to-machine (M2M) technology that allows machine devices to communicate directly with each other without human intervention has emerged as a cutting-edge technology for next-generation communication [1]. M2M technology enables the next generation of wireless sensor networks and global connectivity to billions of processes, devices, and machines through the Internet and it is considered the most promising solution for revolutionizing future applications [2]. However, the wireless M2M networks are composed of a large number of small and power-limited machine devices, which can only transmit low-power signals. Due to the propagation loss between the source devices and the destination devices, it is very challenging to guarantee the quality of service (QoS) requirements of the destination devices [3]. Recently, considerable effort has been put into this problem. Here, we discuss several issues that need to be considered in the design of a framework of wireless M2M communications network to meet the above requirements.

The first issue pertains to the MAC protocol, which plays a key role in determining the network throughput, delay, power consumption, and so forth. A large number of MAC protocols have been proposed for sensor networks and personal area networks, with energy efficiency as the primary design criterion. The most common of these protocols are the contention-based carrier sense multiple access (CSMA) scheme [4] and the schedule-based time division multiple access (TDMA) scheme [5].

The second issue concerns the routing path of data from end devices to an M2M gateway, which is the main content of this paper. Network design has traditionally followed the principle of layering. Previous work in this area has focused largely on the design of link-level protocols [6], and some previous papers address routing for such a network [7]. A large number of studies have shown that the routing protocols based on the criterion of minimum number of hops cannot meet the requirements of M2M network for throughput and guarantee the quality of service.

In recent years, many wireless M2M communications network routing protocols have been proposed: the multiradio link quality source routing protocol [7] is a multichannel routing protocol, which uses weighted cumulative expected time as the routing performance; predictable wireless routing

protocol [8] by comparing the data packet error rate and other network state parameters to select the optimal path under certain circumstances; single transceiver device and multichannel routing protocol [9] support multichannel resources in the same area at the same time without disturbing each other so that the network performance can be enhanced; routing protocols combined traditional DSR [10] with the ETX [11] criterion are limited to enhance the throughput, so high-throughput routing protocol makes various improvements to enhance the network throughput; based on the ETX criterion RF-aware routing protocol introduces RF-aware routing information as the judgment of the path.

According to how many paths a routing algorithm uses, we can divide all routing algorithms into two classes: single-path routing and multipath routing [12]. Single-path protocol is simple but prone to path overheating (such as multiple clients sending data to the Internet through the same router; all data will be passed along the same path when using a single-path protocol), so that network performance and stability are difficult to be guaranteed; multipath protocols establish multiple paths to transfer data and distribute traffic on multiple paths; thereby they can enhance network throughput, lower packet loss rate transmitted over the network, improve delay and jitter performance indicators, and are better able to achieve network quality of service requirements.

However, most multipath algorithms pay much attention to the multipath discovery and multipath maintenance, while ignoring the traffic allocation problem, which is just the problem we focus on. The traffic allocation problem may lead to many other problems such as low utilization ratio of the bandwidth and the jitter. When we tried to analyze the load on different paths using the simplest multipath routing algorithms we found that the load changes just like simple harmonic motion, which inspires us to use physic and mathematic methods and tools to analyze the loading balance problems [13]. And then we designed a new traffic allocation algorithm. Compared with other algorithms, the algorithms proposed in this paper have much wider applicability.

The structure of this paper is as follows. In Section 1, we first analyze the present algorithms of the wireless M2M communications networks. And an appropriate physical model for the network of multipath traffic allocation problem has been built in Section 2. In Section 3, we combine the conclusions deduced from the physical model with a more mature routing algorithm and design some new multipath routing protocol traffic distribution algorithms for wireless M2M communications networks. The performance comparison experiments are noted in Section 4. And in Section 5 we give the conclusion of this paper.

2. Theoretical Model

2.1. Network Model. We use the graph model to represent wireless M2M communications networks in this paper: wireless network routers and end devices are abstracted as vertices in the graph; the wireless or wired links used by wireless network nodes to communicate with each other are abstracted as edges in the graph. In this model we ignore the phenomenon that the wireless network transmission capacity may be unequal in different directions, because its effects on flow allocation and load balancing are very small; we can simplify the theoretical model and use an undirected graph $G = \{V, E\}$ to represent the abstraction of wireless M2M communications networks.

There are two original intentions to design new algorithms in this paper. Firstly, reduce the network transmission delay jitter and secondly load balancing. The rules of the action and reaction between the path load condition and its delay are the starting point to design algorithms in this paper. Path load condition is related to the number of data packets per unit time passing, and the number of packets per unit time passing in the path affects the path delay through the impact on the queue length of each node in the path. By the above, we can infer that the queue lengths of the nodes in the path are the bridge linking the load on the path and the path delay.

Therefore, in the theoretical model proposed in this paper, each vertex in the graph uses an integer to describe the queue length of the network node. In addition, the reason for the wait queue of the network node is that a node needs a period of time that cannot be ignored to forward a data packet. Therefore, in the theoretical model, we will also add a constant as the time nodes need to forward a data packet. It needs to be noted that the time for forwarding data packets cannot be ignored, but since the source node can send new packets to the path at any rate, we assume that the time for the source node sending data packets can be ignored. The purpose of this assumption is to design the algorithm in this model that can increase network throughput as much as possible. Its practical significance is that it is usually the wireless network clients that initiated to send data in wireless M2M communications networks; in this case, the first hop node of the path is typically the wireless access point becuase multiple clients may send data packet to access point simultaneously; we can abstract the clients through the same access point connecting to the wireless M2M communications networks as a source node, and the time of the source node sending data packets can be ignored.

In summary, this theoretical model can be described as

$$\begin{aligned} M &= \{G, c\}, \\ G &= \{V, E\}, \\ V &= \{v \mid v = (id, l_{id})\}, \\ E &= \{(u, v) \mid u, v \in V\}. \end{aligned} \tag{1}$$

The meaning of the model: the model M is composed of the network topology G and the constant value c that represents the time required to forward data packet; network topology graph G consists of the vertex set V and the edge set E; the elements in the vertex set V are a binary group (id, l_{id}), where id is the unique number of vertex and l_{id} is the current queue length of the network node whose unique number is id; the elements of the set of edges E are a tuple (u, v); u,v are different elements in the vertex set V. In the following subsequent sections of this section, we will use this model to calculate the network transmission delay and its changes.

2.2. Theorems

Theorem 1. *One assumes that none of vertices on the path is in other data packets transmission paths and the queue length is* 0 *when passing a data packet on the path* $v_1, v_2, \ldots, v_n$*; then the forward time of the data packets from* v_1 *to* v_n *is* $(n-2)\,c$.

Proof. Based on the assumption that the time for the source node sending data packet to initiate the transmission can be ignored, the residence time of the data packet on the path is the product of the number of intermediate nodes in the path and the time for each intermediate node forwarding the data packet. In order to facilitate the description, c is identified as the unit of time; that is, $c = 1$. When a path is designed for a source node, the lower bound of delay is decided by the number of hops, which is a constant only related to the path itself. □

Theorem 2. *Considering the data packet transmission path* $v_1, v_2, \ldots, v_n$*, if the path is used exclusively for* v_1 *to* v_n*, the path delay changes only related to the queue length of* v_2.

Proof. Node does not receive data packets from the nodes outside the path, so in unit time, one node can only receive a data packet except v_2, and at the same time they can forward a packet, so the queue length will remain unchanged and have nothing to do with the path delay variation. □

Theorem 3. *Considering the data packet transmission path* $v_1, v_2, \ldots, v_n$*, when the source node* v_1 *sends data packets to the target node* v_n *with velocity* v*, the increasing rate of path delay is* $v - 1$.

Proof. Given c for the unit time, the queue length of node v_2 changes $v - 1$ per unit time; the queuing time variation of the next data packet in the node v_2 is $v - 1$. According to Theorem 2, the path delay also changes $v - 1$ per unit time. □

2.3. Theoretical Model. According to these three theorems, we can associate network transmission model with vibration model in physics [14, 15]. First of all, we express the path delay as the location x; the rate of change of the path delay must correspond to the speed $v = dx/dt$; at the same time according to Theorem 3, we can find that the rate of source node sending data packet is proportional to the rate of change of the path delay by making a simple linear transformation. Therefore, the rate of the source node sending data packets to the path can be represented as speed in physics. This paper will describe the following network transmission model by using the physics of vibration model to achieve the design goal below:

(i) the difference of delay time of the two paths $D_1 - D_2$ can be described as the oscillator equilibrium position offset x in the vibration model;

(ii) the rate of change of the difference between the two path delays $d(D_1 - D_2)/dt$ can be described as the rate of change of the oscillator equilibrium position offset in vibration model $v = dx/dt$, and according to Theorem 3, we can use the rate of the path receiving new data packet to represent the rate of change.

Therefore, we can make the difference of the two path delays change in accordance with harmonic vibration model in physics by changing the rate of data packets sent to paths in accordance with the following formula:

$$\frac{d^2x}{dt^2} = -kx. \tag{2}$$

Here, k is a constant and it will determine the cycle of simple harmonic motion. And d^2x/dt^2 is the rate of change of the source node sending data packets to two different paths in dual-path network transmission model. We can intuitively analyze the meaning of this formula as follows: when the delay on one path is greater than on another path, we will reduce the sending data packet rate of the greater one and increase the rate of the smaller one. This is also the main idea of the most heuristic path selection and flow allocation algorithm.

Obviously, using the formula (2) and other similar traffic shaping policies in the qualitative analysis would lead to endless vibration load on the different paths until the end of the network transmission task. We try to design an algorithm that can make the load vibration gradually flatten and eventually remain stable at the load balancing point. As a motivation, we consider introducing the concept of resistance into the system, so that the load vibration model with its initial state transition similar to the harmonic vibration can transform into damping vibration state in finite time. The physics model used to explain this is shown in Figure 1.

In physics, the significance of the friction is to impede the movement of objects, that is, to hinder the change of objects location. In the network data transfer model, change in the location of objects is related to the change of path delay. Intuitively, the significance of the introduction of friction into the network data model is to maintain the delay on the path unchanged. From the qualitative point of view, the idea to design the algorithm is on the one hand to transfer traffic to the path with lower delay and on the other hand to inhibit this shift to some extent. Through the interaction of these two aspects, we can ultimately achieve the purpose of vibration tending to stationary.

In the following, we will introduce the flow allocation algorithm damping based traffic allocation algorithm (DBTA) in detail. One thing that should be pointed out is that the flow allocation algorithm according to the vibration model designed in this paper is also an algorithm framework that can be used in other network protocols. Therefore, the algorithms proposed in this paper have wider applicability.

3. Damping Based Traffic Allocation Algorithm

3.1. Damping Based Traffic Allocation Algorithm (DBTA). We design a new algorithm to make the network transmission model run in accordance with damping vibration law. Consider the differential equations describing the damping of vibration:

$$\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + \omega_0^2 x = 0. \tag{3}$$

Figure 1: Damping vibration.

In this differential equation, x is the displacement of the oscillator from the equilibrium position, that is, the difference of the two path delays in the network transmission model; dx/dt is the velocity of oscillator, that is, the difference of the source node sending data packet rate to the two paths, while d^2x/dt^2 represents the acceleration of the oscillator, that is, the rate of change of the difference of the source node sending data packet rate to the two paths. The core of the proposed algorithm is computing the value, of d^2x/dt^2, adjusting the rate sending data packets to two different paths in accordance with the value, and achieving the goal of the adjustment path delay. And γ and ω_0^2 are system parameters; that is, damping vibration attenuation rate cycle can be controlled by setting these two parameters, where γ is a negative number.

Based on the assumption of the network environment in this paper, there is only one data sender and only one recipient in the network and there are two paths from the sender to the recipient, and then you can get the above three theorems. Because the issue of this paper is the network load oscillation and the purpose of the algorithm is to eliminate this phenomenon, we omitted other complex situations to simplify the problem.

According to Theorem 2, we learn that the path delay is only related to the queue length of the next hop node of the source node in this hypothetical network environment. In dynamic source routing protocol [10], each node only interacts with its next node to receive confirmation, so the design algorithm based on dynamic source routing protocol is easy to access the queue length information of the next hop node of the source node and then to estimate path delay. We have illustrated how to obtain all things needed to be calculated including the path delay and the data packet rates of the source node sending to different paths. Taking into account in practical applications, the source node sending rate usually has an upper limit; it may be decided by the amount of data that needed to be sent or by the hardware capacity. After joining this assumption, the process of the allocation algorithm based on the flow of damping vibration process is as Algorithm 1.

Algorithm 1 is a damping based traffic allocation algorithm. This algorithm makes the load on the path vibrate in the form of damping vibration by adjusting the change rate of the data packets sending to different paths, and ultimately the delay on the two paths is converging. In physics, acceleration is proportional to the size of force and the work in this paper is precisely to control acceleration.

What should be noted is the set of parameters ω_0 and γ; they can be set according to the actual requirements of the application, given the damped vibration function:

$$x = e^{-\gamma t/2}\left(Ae^{i\omega_\gamma t} + \overline{A}e^{-i\omega_\gamma t}\right). \tag{4}$$

Here, $\omega_\gamma = \sqrt{\omega_0^2 - \gamma^2/4}$. It can be seen that ω_0 determines the period of vibration. The increases of ω_0 will lead to the shortening of the period of vibration, resulting in severe vibration of the path load in the initial stages of the data transmission. So according to actual needs, if the period of the vibration you can tolerate is t seconds, it will be appropriate to set ω_0 to $2\pi/t$. And the decay rate of amplitude is determined by γ; the greater the value of γ is, the faster the decay is. However, large γ will still lead to larger delay difference of the two paths in the vibration stationary position. If the delay difference of the two paths need to be reduced to x after t seconds, you should make $\gamma = -(2\ln x)/t$.

3.2. Improved Damping Based Traffic Allocation Algorithm (Improved-DBTA). As the actual situations which DBTA applied to are relatively simple, we proposed more complete improved-DBTA on the basis of DBTA to make the algorithm able to achieve better results in a complex network environment. Firstly, improved-DBTA no longer applies only to two paths but to any multiple paths; secondly, improved-DBTA will no longer be simple to use the queue length of the second node on the path as the evaluation of the path but take the path delay which the target node feedback as a basis for adjustment of the rate of sending data to the path.

The delay in the network transmission mainly consists of four parts: the processing delay, queuing delay, sending delay, and propagation delay. To simplify the problem, assuming that the size of packet in the network is similar and the path complexity is analogous to each other, then the processing delay, sending delay, and propagation delay can be regarded as constant. Therefore, the delay of packets in the network is $x = C + x_q$, where C is a constant and x_q is the queuing delay. Therefore, the rate of change of the total propagation delay is decided by the rate of change of queuing delay:

$$\frac{dx}{dt} = \frac{dx_q}{dt}. \tag{5}$$

Considering a simple path connecting the source node and destination node, the rate of the source node sending data to the path is r_{input} packages per second, and the maximum output rate of this path is R data packages per second; then there will be $r_{\text{input}} - R$ packets blocked in the queue in a second, so that the path delay of the packets increases $k(r_{\text{input}} - R)$, where k is a constant. Therefore, we can obtain the following relationship:

$$\frac{d^2x}{dt^2} = \frac{d^2x_q}{dt^2} = k\frac{dr_{\text{input}}}{dt}. \tag{6}$$

```
while There are still data packets need to send do
    Check the path cache of the source node, and find two paths L_1, L_2 leading to the target node.
    if L_1, L_2 are found then
        In the next unit of time, send data packets to L_1 at the rate of v_1 and send data packets to L_2 at the rate of v_2; and
        get the queue length information of the next hop node from the received confirmation; update d_1, d_2;
    else
        Enable path discovery mechanism to find the two path L_1, L_2 leading to the target node, and get the queue length of
        the next hop node of the source node on the path d_1, d_2. At the same time set the maximum sending rate for the path
        with lower delay: v_1 = max_v; and set the minimum sending rate for the path with higher delay: v_2 = 0
    end if
    Calculate Δ = −ω_0^2 (d_1 − d_2) − γ (v_1 − v_2);
    Set v_1 = v_1 + Δ/2, v_2 = v_2 − Δ/2;
end while
```

Algorithm 1: Damping based traffic allocation algorithm (DBTA).

According to Newton's second law, the vector force added to the objects is the reason for the change of momentum of objects:

$$F = \frac{dp}{dt} = m\frac{dv}{dt} = m\frac{d^2x}{dt^2}. \tag{7}$$

If only we considered the situation in one dimension, we can use the sign to indicate the direction of the vector and use the absolute value to indicate the size of the vector; then we can get

$$F = m\frac{d^2x}{dt^2}. \tag{8}$$

Inspired by this, consider the rate of change of the path input rate dr_{input}/dt as the force F in physics. Whether it is in theoretical physics or engineering, we can use such a mathematical model to represent damping: the size of the force and speed changes synchronously but in the opposite direction. In mechanics, friction is one such force. In a variety of circumstances, the friction is seen as proportional to the speed of the movement of objects; that is,

$$F = -cv. \tag{9}$$

There, c is the coefficient of friction. And then consider the situation in the network; assuming that there is a simple path between the source node and destination node, the rate of the source node sending data to this path is r_{input} and the maximum output rate of the path is R. If we can find the delay increases on this path in the process of transferring data, then it can be concluded that $r_{\text{input}} > R$; in this case, the source node should be appropriate to reduce the rate sent to the path; otherwise the delay will be more and more large. Conversely, if the delay on the path decreases, we can know that $r_{\text{input}} < R$; this time, the source node can appropriately increase the rate sent to this path to improve transmission efficiency. Further, when the path delay variation is greater, the change of sending rate will be greater, but the change should be in the opposite direction. Therefore, the change rate of the velocity of the source node sending data to the path and the rate of change of the path delay have similar links to the movement of objects suffering friction and the object velocity. This also confirms from another point that the network transmission model is linked to the variables in physics.

The process of the improved damping based traffic allocation algorithm in this section is as follows: first, set the equilibrium position of the path delay d_{balance} according to the maximum output rate R, then calculate the average path propagation delay d of the previous period and d' of the previous two periods, and finally, calculate the increment of the output rate of the source node the next unit of time relative to the last period.

Consider

$$r_{\text{input}} - r'_{\text{input}} = -\gamma\left(d - d'\right) - \omega_0^2\left(d - d_{\text{balance}}\right). \tag{10}$$

There, $\gamma \in (0, 1)$, $\omega_0 \in (\gamma/2, 1)$ are the parameters used to control the damping vibration frequency and decay rate and r'_{input} is the output rate of the source node in previous period. So far, the network transport protocol, improved-DBTA, used in single-path has finished presentation. In the design of the algorithm, for the convenience of description, we assume that the node outside the path will not send packets to the path and the case is too simple and not realistic. However, the node outside the path sends packets to the path nodes, which works as an external force to the damping vibration of the oscillator, and that may change the vibration amplitude, frequency, and the equilibrium position but still maintains the process of damped vibration. Then, consider extending the algorithm so that it can use multiple paths to transmit packets. Assuming in the path discovery phase, we found k paths from the source node to destination node: $p_1, p_2, \ldots, p_k$ and then we represented the delay on the k paths with the symbol $d_1, d_2, \ldots, d_k$ and the input rate of the source node to the k paths with the symbol $r_1, r_2, \ldots, r_k$, and the average input rate of the k paths is r_{avg}, so

$$r_{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} r_i. \tag{11}$$

Meanwhile we use d_{avg} to denote the average delay on these k paths; that is

$$d_{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} d_i. \tag{12}$$

In addition, the symbol $d_{\max}$ stands for the maximum value of difference between the path delays; that is,

$$d_{\max} = \max_{1\le i<j\le k} \left|d_i - d_j\right|. \tag{13}$$

The goal of the proposed algorithm is to minimize Δd_{avg} and $d_{\max}$. To achieve these two objectives, the algorithm is divided into two parts in this paper. In this paper, these two parts were referred to as flow control section and flow distribution section.

In the flow control section, we take all the k paths as one path and then use the single-path algorithm previously described, where Δd_{avg} is the displacement of the oscillator, and we can control the sending rate of the source node to minimize Δd_{avg}.

In the flow distribution section, divide the path into two parts, recursively, and then take the difference of the average delay time between these two parts as a displacement of the oscillator to minimize $d_{\max}$. Detailed description of the algorithm is shown in Algorithm 2.

Algorithm 2 is the process of the improved-DBTA. In order to simplify the algorithm, we take the first path as the first part and the rest of the path as the second part during dividing the paths into two parts recursively. If the number of the rest paths is greater than 1, then in the second division, we take the second path as one part and the other paths are composed of the remainder and so on. In the case of using this classification method, we take the difference of the average delay time between the two parts as the displacement relative to the equilibrium position of the damped oscillator, so we can deduce all formulas in the algorithm.

3.3. Adaptive Damping Based Traffic Allocation Algorithm (Adaptive-DBTA). In improved-DBTA, we need to estimate the maximum output rate of the path in advance. It is relatively easy to determine the maximum output rate of the path when the type of nodes is single in the network and the communication between nodes is similar. However, usually the network conditions would be more complex in the actual applications, and, therefore, the maximum output rate of the path is difficult to estimate in advance. Even if the effects of the incorrect estimated maximum output rate of the path on the algorithm are not significant (from the model perspective, the difference between the estimated value and actual value can be filled with a hypothetical data stream, which is equivalent to a constant force with unchanged size and direction applied to the oscillator in the vibration model), we still hope to be able to design a simple algorithm to find the balance point of the sending rate adaptively.

Supposing a path p from the source node to destination node, the source node sends data to the destination node at the rate of r_1 and the change of the rate of the path ps delay is Δd_1 in the unit time. While the source node sends data at the rate of r_2 and the change of the rate of the path ps delay is Δd_2 in the unit time, we can know that if r_1 is more away from the maximum output rate R of the path p than r_2, then $|\Delta d_1| > |\Delta d_2|$. As described in the previous section, we can assume the difference of the sending rate of source node and the actual maximum output rate of the path directly proportional to the rate of change of the path delay; that is,

$$(R - r_1) \times (\Delta d_2) = (R - r_2) \times (\Delta d_1). \tag{14}$$

It can be introduced by

$$R = \frac{r_1 \times \Delta d_2 - r_2 \times \Delta d_1}{\Delta d_2 - \Delta d_1}. \tag{15}$$

In adaptive-DBTA, we record the average sending rate of the source node and multipath average delay in the first two periods of time and calculate the equilibrium position in accordance with the above formula, and then, the next period of time, the source node sends data at the sending rate of the calculated equilibrium position. In order to facilitate description, use the symbol d_i^k to express the average delay d in the ith time on the kth path and use r_i^k to denote the average rate of the source node sending to the kth path. The specific process of algorithm is shown in Algorithm 3.

Algorithm 3 is the process of adaptive-DBTA. The method distributing the load to each path after calculating the total sending rate of previous period is the same as improved-DBTA. In addition, both in improved-DBTA and adaptive-DBTA, the delay information of the paths used in the calculation has some degree of lag, but due to calculation methods, adaptive-DBTA will have a greater lag than improved-DBTA. This is also a major drawback of the adaptive-DBTA. However, adaptive-DBTA does not need to set complex parameters and thus can be more convenient in practical network applications.

4. Experiment

In this paper, we use the NS-2 network simulation software to do the experiments. We do a slight modification on DSR protocol: record the multiple paths in the path finding stage and then use this design algorithm to choose the right path for data packages to be sent. The aim of the simulation experiments is to view whether the flow allocation algorithms designed in this paper are able to reduce the path delay and throughput fluctuations or not.

In the experiments, we established 100 nodes randomly at first and then established 40 streams with different data sending rates. The parameters of NS-2 simulation are as follows: experiment scene: 2000 m × 2000 m, number of nodes: 100, MAC type: 802.11, IFq type: drop − tail/pri − queue, IFq length: 20, and simulation time: 200 seconds.

In the DBTAs network simulation experiments, we selected some particular paths for comparison to observe the difference between DSR protocol and the flow allocation algorithm in delay when collecting the experimental results.

Figure 2 is the delay comparison chart over time of a path whose absolute value of delay is approximately equal to avoid

(1) Set the parameters d_{balance}, γ and ω_0 according to the actual application requirements as well as the average maximum output rate of every path;
(2) Set $r_{\text{avg}} = r_1 = r_2 = \cdots = r_k = 0$, $r'_{\text{avg}} = r'_1 = r'_2 = \cdots = r'_k = 0$;
(3) Set $d_1 = d_2 = \cdots = d_k = 0$, $d'_1 = d'_2 = \cdots = d'_k = 0$;
(4) **while** There are still data packets need to send **do**
(5) Calculate $d_{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} d_i$, $d'_{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} d'_i$, and $s_i = \frac{1}{k-i+1}\sum_{j=i}^{k} d_j$, $s'_i = \frac{1}{k-i+1}\sum_{j=i}^{k} d'_j$ in the tth period;
(6) Calculate the average sending rate of the source node in the tth period: $r_{\text{avg}} = r'_{\text{avg}} - \gamma\left(d_{\text{avg}} - d'_{\text{avg}}\right) - \omega_0^2\left(d_{\text{avg}} - d_{\text{balance}}\right)$, and calculate the total sending rate of the source node in the tth period: $r_{\text{total}} = k \times r_{\text{avg}}$;
(7) Calculate the input rate of the first path: $r_1 = r'_1 + \frac{1}{k}\left(r_{\text{total}} - r'_{\text{total}} - (k-1)\gamma\left((d_1 - s_2) - (d'_1 - s'_2)\right) - (k-1)\omega_0^2\left(d_1 - s_2\right)\right)$
(8) Calculate the input rate of the ith path:
$$r_i = r'_i + \frac{1}{k-i+1}\left(r_{\text{total}} - \sum_{j=1}^{i-1} r_i - \left(r'_{\text{total}} - \sum_{j=1}^{i-1} r'_j\right) - (k-i)\gamma\left((d_1 - s_2) - (d'_1 - s'_2)\right) - (k-i)\omega_0^2\left(d_1 - s_2\right)\right)$$
for $i = 1, 2, \ldots, k-1$ and $r_k = r_{\text{total}} - \sum_{j=1}^{k-1} r_j$
(9) In the tth period, the source node sends packet to the ith path at the rate of r_i. After this period set $r'_i = r_i$, $r'_{\text{avg}} = r_{\text{avg}}$, $d'_i = d_i$, and calculate the average delay d_i of the ith path in this period;
(10) Set $t = t + 1$;
(11) **end while**

Algorithm 2: Improved damping based traffic allocation algorithm (improved-DBTA).

(1) Set the parameters γ and ω_0 according to the actual application requirements;
(2) Gradual increase the rates to each path in the time period of 1 to 3 to send data, and record the sending rate and the average delay of three time periods;
(3) Set $d_1 = d_2 = \cdots = d_k = 0$, $d'_1 = d'_2 = \cdots = d'_k = 0$;
(4) **while** There are still data packets need to send **do**
(5) Calculate $d_{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} d_t^k$, $d_{t-2}^{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} d_{t-2}^i$, $d_{t-3}^{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} d_{t-3}^i$, and $s_i = \frac{1}{k-i+1}\sum_{j=i}^{k} d_{t-1}^j$, $s'_i = \frac{1}{k-i+1}\sum_{j=i}^{k} d_{t-1}^j{}'$ in the tth period;
(6) Calculate the average sending rate of the source node in the tth period: $r_t^{\text{avg}} = \frac{\left(r_{t-2}^{\text{avg}} \times \left(d_{t-2}^{\text{avg}} - d_{t-3}^{\text{avg}}\right) - r_{t-3}^{\text{avg}}\left(d_{t-1}^{\text{avg}} - d_{t-2}^{\text{avg}}\right)\right)}{\left(2d_{t-2}^{\text{avg}} - d_{t-1}^{\text{avg}} - d_{t-3}^{\text{avg}}\right)}$, and if the denominator in the equation is 0, then $r_t^{\text{avg}} = r_{t-1}^{\text{avg}}$ and calculate the total sending rate of the source node in the tth period $r_{\text{total}} = k \times r_{\text{avg}}$
(7) Calculate the input rate of the first path: $r_1 = r'_1 + \frac{1}{k}\left(r_{\text{total}} - r'_{\text{total}} - (k-1)\gamma\left((d_1 - s_2) - (d'_1 - s'_2)\right) - (k-1)\omega_0^2\left(d_1 - s_2\right)\right)$
(8) Calculate the input rate of the ith path:
$$r_i = r'_i + \frac{1}{k-i+1}\left(r_{\text{total}} - \sum_{j=1}^{i-1} r_i - \left(r'_{\text{total}} - \sum_{j=1}^{i-1} r'_j\right) - (k-i)\gamma\left((d_1 - s_2) - (d'_1 - s'_2)\right) - (k-i)\omega_0^2\left(d_1 - s_2\right)\right) \text{ for } i = 1, 2, \ldots, k-1$$
and $r_k = r_{\text{total}} - \sum_{j=1}^{k-1} r_j$
(9) In the tth period, the source node sends packet to the ith path at the rate of r_i. After this period set $r'_i = r_i$, $r'_{\text{avg}} = r_{\text{avg}}$, $d'_i = d_i$, and calculate the average delay d_i of the ith path in this period;
(10) Set $t = t + 1$;
(11) **end while**

Algorithm 3: Adaptive damping based traffic allocation algorithm (Adaptive-DBTA).

the effects of the absolute value of delay on the volatility. From the experimental results, we can see that the load vibration using the DBTA algorithm is much smoother than that when just using DSR protocol. The effect of DBTA algorithm on delay jitter is obvious.

Figure 3 is the throughput comparison chart over time of a path whose throughput is approximately equal to avoid the effects of the size of throughput on the volatility.

In contrast to the delay fluctuations, we still use similar methods and choose the path whose absolute value of delay is approximately equal to avoid the effects of the absolute value of delay on the volatility. The transfer delay result is shown as Figure 4.

In the third part of the experiment, we used our own program written in C language to compare the difference of improved-DBTA and adaptive-DBTA in the delay and

Figure 2: The transfer delay of DSR and DBTA.

Figure 3: The throughput of DSR and improved-DBTA.

throughput. Similar to the principle of NS-2 principle, we used priority queue to implement event-driven simulation. The aim of the simulation is to study the difference between adaptive-DBTA and improved-DBTA, and the throughput comparison results are shown in Figure 5.

From the above experimental results, we can see, compared with the improved-DBTA, that the throughput jitter of adaptive-DBTA is slightly serious; this result is because each time adaptive-DBTA directly adjusts the sending rate to the equilibrium position calculated, while improved-DBTA gradually adjusts the sending rate to simulate the process of damped vibration. The comparison of these two algorithms in terms of path delay results is shown in Figure 6.

From the above experimental results, we can see, compared with the improved-DBTA algorithm, that the delay jitter of the adaptive-DBTA is slightly serious. And it is related to the experimental results of adaptive-DBTA throughput jitter. The delay of the two algorithms trends to increase, which is also related to the growth trend of throughput.

In addition, we also compared improved-DBTA with DCJOTA in [16], and the purpose of the experiment is to check the difference of effect between the algorithm in this paper and the other algorithms. The results of these two algorithms contrast experiment are shown in Figure 7.

In addition, in the last part of the experiment, we did an experiment to compare the difference in the path delay jitter between improved-DBTA, the DCJOTA, the OPNET in [17], and the COTA in [18]. The path delay comparison results of all these algorithms are shown in Figure 8.

From the above experimental results, we can see that the control effect of the improved-DBTA on the throughput and delay jitter is similar to DCJOTA and COTA algorithms, all of which are indeed achieving the aim to control the delay jitter and are better than the effect of OPNET algorithm.

5. Conclusion

Wireless M2M communications network is a promising network access technology and has broad prospect of applications. The multipath transmission and load balancing problems that related to service quality and throughput of wireless M2M communications network have become a hot spot issue for many researchers. In many studies, people used complex mathematical models to analyze the problem. But we think that the physical model may be more applicative and easy to understand.

In this paper, we analyzed the traffic allocation problem of wireless M2M communications networks and built an appropriate model using physics method. We proposed new traffic allocation algorithms based on damping model, and the algorithms can reduce the network transmission delay jitter and load balancing. Here, we simply applied the damping model of classical mechanics; the more in-depth

Figure 4: The transfer delay of DSR and improved-DBTA.

Figure 5: The throughput of improved-DBTA and adaptive-DBTA.

Figure 6: The transfer delay of improved-DBTA and adaptive-DBTA.

Figure 7: The throughput of improved-DBTA and DCJOTA.

Figure 8: The throughput of DCJOTA, COTA, OPNET and improved-DBTA.

and effective application of the model to analyze and solve problems remains to be explored.

Acknowledgments

This research is supported by the National Science Funds of China Grant no. 61170305 and is part of Project of Education Information Strategy supported by Educational Commission of Tianjin, China, 2013, under Grant no. x2013-005.

References

[1] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, 2011.

[2] N. Tekbiyik and E. Uysal-Biyikoglu, "Energy efficient wireless unicast routing alternatives for machine-to-machine networks," *Journal of Network and Computer Applications*, vol. 34, no. 5, pp. 1587–1614, 2011.

[3] D. Wang, L. Bai, X. Zhang et al., "Collaborative relay beamforming strategies for multiple destinations with guaranteed QoS in wireless machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 525640, 10 pages, 2012.

[4] O. D. Incel, L. van Hoesel, P. Jansen, and P. Havinga, "MC-LMAC: a multi-channel MAC protocol for wireless sensor networks," *Ad Hoc Networks*, vol. 9, no. 1, pp. 73–94, 2011.

[5] T. Shon, H. H. Choi, and E. J. Kim, "Joint routing and scheduling in multi-channel capillary machine-to-machine networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–9, 2013.

[6] J. So and N. H. Vaidya, "Multi-channel MAC for ad hoc networks: handling multi-channel hidden terminals using a single transceiver," in *Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MoBiHoc '04)*, pp. 222–233, May 2004.

[7] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking (MobiCom '04)*, pp. 114–128, October 2004.

[8] R. Draves, J. Padhye, and B. Zill, "Comparison of routing metrics for static multi-hop wireless networks," in *Proceedings of the ACM SIGCOMM Computer Communication Review*, vol. 34, pp. 133–144, September 2004.

[9] J. So and N. H. Vaidya, *A Routing Protocol for Utilizing Multiple Channels in Multi-Hop Wireless Networks with a Single transceiver*, University of Illinois at Urbana-Champaign, 2004.

[10] D. B. Johnson, D. A. Maltz, and J. Broch, "DSR: the dynamic source routing protocol for multi-hop wireless ad hoc networks," *Ad Hoc Networking*, vol. 5, pp. 139–172, 2001.

[11] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," *Wireless Networks*, vol. 11, no. 4, pp. 419–434, 2005.

[12] M. Mosko and J. J. Garcia, "Multipath routing in wireless mesh networks," in *Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS '06)*, pp. 741–746, 2006.

[13] Z. Yuan, R. Yu, M. Zhao et al., "Vibration model based multipath traffic allocation algorithm for wireless Mesh networks," *Journal of Information and Computational Science*, vol. 10, no. 5, pp. 1253–1260, 2013.

[14] S. O. Reza Moheimani and A. J. Fleming, *Piezoelectric Transducers for Vibration Control and Damping*, Springer, 2010.

[15] A. D. Nashif, D. I. G. Jones, and J. P. Henderson, *Vibration Damping*, John Wiley and Sons, New York, NY, USA, 1985.

[16] Z.-G. Chen, F. Zeng, and Q.-H. Li, "Delay-constrained and jitter-optimized traffic allocation for multipath routing in wireless mesh networks," *Journal on Communications*, vol. 32, no. 1, pp. 1–8, 2011.

[17] X. Chang, "Network simulations with OPNET," in *Proceedings of the Winter Simulation Conference Proceedings (WSC '99)*, vol. 1, pp. 307–316, December 1999.

[18] L. Qian, N. Song, and X. Li, "Detection of wormhole attacks in multi-path routed wireless ad hoc networks: a statistical analysis approach," *Journal of Network and Computer Applications*, vol. 30, no. 1, pp. 308–330, 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 849572, 12 pages
http://dx.doi.org/10.1155/2013/849572

Research Article

Efficient and Secure Resource Management in Home M2M Networks

Yue Lai, Jiawen Kang, and Rong Yu

School of Automation, Guangdong University of Technology, Guangzhou 51006, China

Correspondence should be addressed to Rong Yu; yurong@gdut.edu.cn

Received 18 July 2013; Accepted 9 October 2013

Academic Editor: Yan Zhang

Copyright © 2013 Yue Lai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, M2M networking is an active research topic in both academic and industrial fields. Home M2M networks, as an important branch of M2M networks, have attracted wide concerns for home application. In this paper, we first present a brief overview of home M2M networks. Then we introduce a hierarchical architecture of home M2M networks and describe the approach of devices management and services discovery. After that, we focus on the security authentication with centralized node and an improved method of available bandwidth measurement. With the centralized node, our system is more efficient and secure, because of efficient resource management and users' anonymity protection. Compared to existing bandwidth measurement methods, our method takes up less additional bandwidth, saves more time, and has a higher accuracy for measurement. Experiments are carried out to demonstrate the effectiveness and accuracy of the proposed scheme.

1. Introduction

Nowadays, home networks are rapidly developing to contain more and more digital living devices or terminals, including laptops, TVs, PCs, iPads, printers, smart phones, speakers, smart switches, and cameras. Machine-to-machine (M2M) technology allows machines or devices to directly communicate with each other without human intervention. It holds huge potentials for services improvement in a wide range of industries and daily life, such as smart grid, smart home, consumer electronics, security, and surveillance [1–7]. With the embedded devices widely used in home environment, home networks will shift from current human-to-human communications or machine-to-human communications to machine-to-machine pattern. Home M2M networks are characterized by low power, low cost, and low human intervention, which has huge potentials to improve service quality and reduce cost.

Home M2M network is a typical class of heterogeneous networks, which consists of a backbone network and multiple subnetworks. In the backbone network, as shown in Figure 1, there is a central machine called home gateway (HGW). It is used to connect the home network to outside world (e.g., Internet) and manage the whole network. The subnetworks can be classified into three complementary M2M subnetworks: body areas, personal areas, and local areas. Each subnetwork may be designed for a specific application. They work in a self-organized manner. There is a subgateway (SGW) in each subnetwork, which works as an endpoint to connect the subnetwork to the HGW as well as the backbone network. Both the HGW and SGW are logical entities. Their functionalities can be physically implemented in a single device (i.e., a cognitive gateway) [1].

Home M2M networking enables people to remotely control and monitor their homes in a more convenient way. It facilitates people to utilize portable devices, which are connected to Internet, to manage their in-home devices. However, machines in home environment are normally small and inexpensive, which put several constraints on M2M communications, including energy, computation, storage, and bandwidth. These constraints pose a number of unique challenges in the design of home M2M networks. For the aforementioned problem, some solutions have been proposed in the literature [1].

In the body areas, personal areas, or local areas, many applications highly depend on multimedia sharing. Three major standards, Universal Plug and Play (UPnP), Digital Living Network Alliance (DLNA) and Intelligent Grouping and Resource Sharing (IGRS), have been defined to facilitate multimedia discovering, searching, and sharing in home

Figure 1: The framework of home M2M networks.

M2M networks [8–10]. These standards aim to establish a world of interconnected machines and a flexible environment of multimedia sharing. The specific contents of these standards will be explained in Section 3.

Although the previous solutions are valid in home M2M networks, they are not efficient and have some deficiencies, such as information repetition, information redundancy, security problem, and link overlap. In this paper, we present a centralized devices management approach, security authentication with centralized node and an improved bandwidth measurement method. These technologies can achieve the goal of efficient and secure resource management in home M2M networks.

The rest of the paper is organized as follows. We outline the system architecture in Section 2. In Section 3, we illustrate an approach of centralized node devices management and services discovering. Sections 4 and 5 introduce efficient security authentication with centralized node in detail and present an available bandwidth measurement method, respectively. Finally, the conclusion and the future work are presented in Section 6.

2. System Architecture

Home M2M networks can carry audio and video information and a variety of high-speed data transmission applications, including multimedia transmission and sharing, video monitoring, and home remote control. Home M2M networks can support data transmission of cable (such as Ethernet) and wireless (such as WLAN). It is based on IP network protocol and widely supports the interconnection between different devices in the home networks environment. It also supports the transparent transmission of multimedia and data services under varieties of network topologies.

In this paper, the communication protocols of home networks contain relevant protocols from link layer to application layer, whose functional structure is shown in Figure 2. The network connection supports IP protocol, Ethernet and wireless LAN transmission. There are some application modules on the top of the network connection, including the applications of multimedia transmission and sharing, video monitoring, remote control, and cross-network communication. The application of multimedia transmission and sharing contains the functional components such as devices management, media management, media transmission, and media formats. This application also contains QoS and corresponding link protection mechanisms. It can achieve, in home environment, the goals of autodiscovery, autoconfiguration and automatic management among a variety of online devices. It establishes a seamless connection and implements comprehensive control, data transmission and sharing.

In the physical layer and data link layer, home M2M networks can use wired or wireless way. When using wireless way, wireless LAN standards (such as IEEE802.11b, IEEE802.11g, IEEE802.11a, 3G, 2G, and LTE) should be followed. When using wired way, IEEE802.3 should be followed. Its physical layer reaches the requirement of data transmitted rate, which is 10 Mbps/100 Mbps. In the backbone network, the devices are combined to form the main family network by the gateway. It not only achieves the goal of networks configuration and management, but also joints the in-home networks and external networks.

(i) The communication protocols of transport layer should use TCP/IP protocol, and all the multimedia and data devices at home should support the TCP/IP and UDP protocols.

(ii) In the application layer, media and device management protocol provides an interoperability framework, which supports a variety of open standards. The framework is based on IEC 62481-1, and it defines the devices and software infrastructures to support interoperable building blocks, including the physical media, network transmission, media formats, streaming protocols, connection protection mechanism and QoS mechanisms. Through a seamless and interoperable network, this framework integrates Internet, mobile devices, and broadcast into the applications of home environment.

As shown in Figure 2, the applications are based on device management, media management, and media transmission protocol, including two modes: decentralized mode and centralized mode. In the decentralized mode, the devices, which need to share data or resources, can be automatically discovered and identified by each other. Then they establish a connection and transmit resources without the coordination and control of other nodes.

Based on the above framework, the multimedia transmission and sharing can be divided into three stages: devices management, media management and media transmission.

(i) Device management is embodied in device addressing, network declaration, device discovery, service discovery, and call operation. Among them, the Dynamic Host Configuration Protocol (DHCP) addressing mechanism should be preferentially used in the device addressing. Devices can also manually set the IP addressing to ensure that there is no conflict between its address and other addresses.

Figure 2: The home networks architecture.

(ii) Media management is reflected in multimedia content identification, management, and distribution, which should support the IEC 62481-1, UPnP AV 1.0, and UPnP protocols.

(iii) Media transmission, namely, the transmission of multimedia content, should support the IEC 62481-1, HTTP 1.0/1.1, IEC 62481-2 (media format protocol), optional support for DTCP-IP (connection protection agreement), and RTP/RTSP (media transfer protocol).

However, when there is a centralized node in the home M2M networks, the data or resources can be centrally managed. After modifying parts of the technical contents of decentralized mode, home networks with centralized node mode can be compatible with the decentralized mode.

(i) Devices management contains the functions of devices addressing centralized node declaration, media devices registration and cancellation, devices discovery, service discovery, and call operation. Devices can also manually set the IP addressing to ensure that there is no conflict between its address and other addresses.

(ii) Media management is reflected in multimedia contents identification, management, and distribution. The centralized mode mainly provides centralized devices directory, media resources, and information service.

(iii) The media transmission of centralized node mode is the same as the decentralized mode.

As shown in Figure 2, there are still two parts in the framework: bandwidth measurement and security authentication. It is apparent that bandwidth measurement plays an important role in improving performance of system. In the heterogeneous home M2M networks, we need to find an efficient and flexible audio and video codec technology. Then it can utilize the limited bandwidth resource and achieve high-speed media information interconnection between different terminals in a more reasonable and efficient way. Furthermore, in order to coincide with the implementation of codec technology in the home networks, it also needs to improve appropriate technology of bandwidth monitoring and develop relevant description files of terminal devices. Measurement of network bandwidth allows us to know the current situation of remaining bandwidth. Then the data can be adaptively transmitted according to the remaining bandwidth. So we can take advantage of the limited bandwidth resources in more reasonable and more efficient way. In this paper, we propose an improved algorithm for measuring the network bandwidth, which is efficient and low complex. More details are described in Section 5.

Wireless technologies will considerably facilitate the home M2M networking. But security issues are the main challenges. Security authentication is the basis for the entire security architecture. In this paper, we propose an efficient wireless IP authentication method with user anonymity. If the machines want to obtain services from other legal machines, they first need to be authenticated by the centralized node. In the authentication process, the machines send real private information to the centralized node for authentication. After the security authentication, they can obtain a legal status. Then they can directly connect to the target machines by temporary keys. Compared to other existing methods, our method has the characteristics of higher safety, practicality, and efficiency.

As mentioned above, there are five functional modules in the home networks architecture. The internal connection between the different modules can be concluded as follows. The module of multimedia and data network application is based on the modules of media transmission protocol and communication protocols. The module of bandwidth measurement gets the available bandwidth of system in a high-speed and efficient way, and then the system can adaptively

adjust the data transmission among different modules. What's more, the module of security authentication provides the protection of data transmission and communication among modules.

3. Devices Management and Services Discovery

Devices management is composed of device registration, device cancellation, devices and service discovery, invocation operation, event trigger, and so forth. Devices management is divided into decentralized mode and centralized mode.

3.1. Devices Management and Resource Management with Decentralized Node. In the mechanism of decentralized devices management, it needs to automatically discover, identify, and control the sharing resource. Currently, there are mainly three technologies to achieve the management as follows [1].

UPnP is essentially a technology in the application layer. It is based on TCP/IP protocol stacks in the lower layers. UPnP has two logical devices [11]: controlled devices (CDs) and control points (CPs). A CD can be a physical or logical network node to provide services. A CP can be a controller that discovers and controls devices. Once a new device is added to a network, it will firstly obtain its IP address and then advertise its services to CPs via a multicast Simple Service Discovery Protocol (SSDP). Once a new CP is added to a network, it shall search for devices of interest in the network. When a CP discovers a device, it may obtain the device's description file via the device's URL message. The CP keeps monitoring the state changes of the CDs. For home multimedia applications, the UPnP audio/video (AV) architecture is the standard for interoperability of multimedia systems among multimedia appliances. This architecture is based on the UPnP processes of discovery, description, control, and eventing, which facilitates a CP coordinating the flows of AV content between the source and sink devices. As a consequence, a consumer can retrieve any shared multimedia by running the CP in her own device and browsing the target media server.

In fact, based on the original network framework, the DLNA standard is an intermediate layer. In particular, the UPnP AV architecture is adopted by the DLNA Home Networked Device Interoperability Guidelines for media management and control between networked devices. End users also employ the UPnP mechanism to discover interoperable server devices in the network. The M2M connection is performed automatically without human intervention. The UPnP mechanism is used to search media content. The key of the DLNA protocol is media management. The DLNA standard defines two main classes of devices: digital media servers (DMSs) and digital media players (DMPs). A DMS is responsible for acquiring, recording, storing, and sharing the media content. A DMP does online searching and plays the media in a home network.

IGRS intends to offer seamless resource sharing and service collaboration among devices for communication terminals, computers, and consumer electronics for users at homes, offices, and public areas. IGRS includes three components: core protocols, application profiles, and basic applications. The core protocols define device grouping and the interaction between clients and services. The application profiles specify the service description. One of the key features of IGRS is taking into account the security issue. IGRS defines two layers of security protocols: the tunnel between devices and the session between users and services. IGRS categorizes the security mechanisms into three levels: identity and messages confirmation mechanism based on the symmetric key cryptosystem; identity authentication, encrypted message transmission, and message authentication based on the public key cryptosystem; identity authentication, encrypted message transmission, and message signature based on a trusted third party.

Although the previous technologies are good ways to manage the resource, they are not efficient and have some deficiencies, including information repetition, information redundancy, and link overlap. Moreover, there are still some problems, especially the security problems. For UPnP, there are two security flaws: the first one, the buffer has not been checked and restricted. The illegal users may get the privilege of controlling the entire system. It is very dangerous that the attackers get control of the computer ports, which may cause serious problems. The second one is that illegal user sends instructions to control a computer to download important data from system. Besides, the illegal user attacks the system, which may result in system crash. For the immaturity of DLNA and IGRS, the wide application cannot be accomplished. There are still some bottlenecks in the development. So we present a centralized management approach as follows.

3.2. Devices Management and Resource Management with Centralized Node. The devices management with centralized node is a centralized management approach, including three logical devices: centralized node (CN), controlled devices (CDs), and control points (CPs).

The CN refers to a device with the functions of security authentication, devices registration, and devices and services search as well as media information. The CN in the home M2M networks can monitor the devices declaration and registration information of CPs. It collects the information of devices and services and generates services registry list. The CN provides the CPs with accessing interfaces of service invocation and updates services registry list in the service description files. In particular, in this paper, the CN also has the function of security authentication. All the devices in the home M2M networks must be confirmed as safe devices before they begin to be linked or served. More details about the process of security authentication will be introduced in Section 4.

When the CN exists in the networks, the CN can finish multicast by the standard address and the port of 239.255.255.250:1900. It broadcasts its status and services to CDs and CPs. The CDs and CPs can be informed of the existence and basic information of CN by monitoring the packets of multicast ports.

When the CDs are online and they monitor the online declaration of CN, at the first time, they need to unicast registration information, which provides device information to the CN for security authentication and registration. After obtaining a legal identity, the devices periodically send the

Table 1: Scheme comparison.

Content	Centralized node	Secure	Efficiency	Information repetition or redundancy	Bandwidth aware
UPNP	No	No	Low	Yes	No
DLNA	No	No	Low	Yes	No
IGRS	No	Security protocols	Modest	Yes	No
Proposed scheme	Yes	Centralized authentication	Hight	No	Yes

current status information to the CN and update the service status. When a device is going to exit the networks, it unicates a cancellation to the CN. Then the CN spreads the cancellation to all the devices and informs other CDs that this device has exited the networks.

What's more, the registration and cancellation process of CPs are similar to the process of CDs. Compared with the decentralized node mode, the CN mode has the advantages of highly efficiency, no repeated information, and more safety.

3.3. Device and Service Discovery. If a CP is on the line, it needs to monitor the multicast ports and find the CN. Then it sends the request information to CN for searching devices and discovering services. The processes of device discovery and service sharing are shown in Figures 3 and 4.

After monitoring the declaration of CN, the CPs get the basic information of CN, such as devices' Universally Unique Identifier (UUID), devices type, service type, and device description file URL. The legal CPs or CDs register in the CN. CPs send the searching instructions to CN for finding out target devices. When the CN receives a legal request information, it will read the registration list and find out the target device, which matches with the request information. Then the CN sends response information to CPs. The CPs obtain the response information, and then they invoke related devices registry list to obtain all the basic information of registered devices in the CN. Thereby, they find the corresponding devices and services in the home M2M networks.

3.4. Session Establishment. After the requester finds the service provider, through the security authentication of CN, the CN provides the information of requester to provider, including the location, corresponding user identity, and authentication information. Then the service provider confirms the identity of requester by the previous information and determines the access rights of requester.

A session is the basis of service invocation. On one hand, the session indicates the security mechanism of service invocation. On the other hand, during the process of service invocation, if one part has message to inform others, it can use the notification mechanism. The session security mechanism is validated by the CN.

3.5. Service Invocation. If the service requester is successfully authenticated by the CN, then it is able to establish a session with the service provider and obtain the permissions of service. Through the interfaces of service description file provided by service provider, the service requester can use the service with the Simple Object Access Protocol (SOAP).

3.6. Session Ending. When the requester wants to exit the service, it can send a removed notification for ending the session. Moreover, the provider can also take the initiative to end the session and give the right of management back to the CN for the next turn of service.

In this part, a table is used to clarify the main features of proposed scheme, which is compared with UPNP, DLNA, and IGRS. The key parameters of comparison are listed in Table 1. From the table, we can see that the performance of proposed devices management approach is much better than the existing devices management schemes, such as UPNP, DLNA, and IGRS.

4. Efficient Security Authentication with Centralized Node

The security authentication is the basis for the entire security system in home M2M networks. When we design wireless authentication protocols, we need to consider the following three main factors: (1) the client cannot use the algorithm with a large amount of calculation because of limited computing ability of the devices. (2) As the limitation of bandwidth in the wireless communication networks, it should reduce the length and number of exchanged information. (3) For the identity confidentiality, it should minimize the possibility of user's exposure [12].

For the previous security requirements, in this paper, we propose an efficient and secure authentication mechanism with anonymity. This mechanism achieves authentication and generates one-time keys between service requester and service provider. It is composed of three entities, that is, service requester (SR), centralized node (CN), service provider (SP).

Usually, there are two kinds of authentication polices without CN in home networks. One is that all the devices in the home networks share the same key. All the devices monitor and search the services by the same key. This policy is simple but unsafe and vulnerable. The other policy is that all the devices in the home networks have their own keys. Any device in the home networks must send its own public key to every other device to ensure the information encryption and transmission. Although it is safe, it is inefficient and complicated. For example, if a device has changed its public key, it has to multicast its new public key to every other device. It will be vulnerable and take up a lot of bandwidth resource. However, the devices in our policy have their own

Figure 3: The process of device and service discovery.

keys and a unique ID number provided by the CN. The devices only need to be authenticated by the CN and establish the communication with CN. Then they can be served in an efficient way without repeatedly mutual authentication among all the devices.

4.1. Symbols. The symbols, which may be used in the mechanism, are shown in Table 2.

4.2. Protocol Description. In this paper, there is a CN which discovers and manages devices in the home M2M networks. The node has a certification signed by the Certificate Authority (CA). The certification includes the public key and private key of CN, that is, KU_{CN} and KR_{CN} and other information as well as CA's signature.

In the home M2M networks, a legal user must be registered in the CN. There is a big secret random number N with the length more than 100 bits in the CN. When a device is artificially added in the CN, the CN assigns a unique

Table 2: Symbols.

Symbols	Meaning
$(X)_K$	The message X is symmetrically encrypted by the key K
$E_{K(X)}$	The message X is asymmetrically encrypted by the key K
$h(X)$	One-way hash function
ID_A	The identification of A
KU_A	The public key of A
KR_A	The private key of A
T_A	The timestamp of entity A
$g \in Z_p^*$	An element with q order of Z
P	A big prime

Figure 4: The process of device sharing.

ID number to the device for calculating the key $PW_{SR\backslash SP} = h(ID_{SR\backslash SP}||N)$. The CN calculates an equation as follows [12]:

$$r = h(ID_{CN}||N) \oplus h(ID_{CN}||N) \oplus ID_{CN} \oplus ID_{SR\backslash SP}. \quad (1)$$

Then the CN sends a smart card, which stores the information of PW_{CN}, ID_{CN}, r, and h, to SP or SR by the safe channel.

As there are many kinds of devices and applications in the home M2M networks, in order to simplify the process, we take printing service as an example. In this scenario, the service requester (SR) is a laptop and the service provider (SP) is a printer. The steps of authentication are briefly described as follows.

4.3. The Steps of Registration

Step 1. As shown in Figure 5, we should manually enter the password of SP, that is, PW_{PR}. Then the printer symmetrically encrypts its own ID (i.e., ID_{PR}) by the PW_{PR}. Finally, it sends $(ID_{PR})_{PW_{PR}}$ to the CN.

Step 2. When the CN receives the information from printer, it checks the password whether it is right. If the checking is passed, the CN decrypts the information to get the ID_{PR}. Then the CN generates a time stamp T_{CN} and symmetrically encrypts its own ID and public key by the PW_{PR}, that is, ID_{CN} and KU_{CN}. Finally, it sends the $(ID_{CN}, KU_{CN})_{PW_{PR}}$ and T_{CN} to the printer.

Step 3. After the printer receives information from CN, the printer checks whether the T_{CN} is within valid date for countering repeated attacks. If the checking is not passed, the printer will ignore the information. Otherwise, the printer symmetrically decrypts the information to get the ID_{CN} and KU_{CN} by the PW_{PR}. The mutual authentication between SP and CN is finished. Then the printer generates a new time stamp T_{PR} and asymmetrically encrypts its public key KU_{PR} and basic information (defined as Infor) by the KU_{CN}. The Infor includes the information of service type, service status, and specific functions. Finally, the printer sends the $E_{KU_{CN}}(\text{Infor}, KU_{PR})$ and T_{PR} to CN.

Step 4. The CN gets the information from SP, and it checks the T_{PR} too. After checking, the CN decrypts the information by its private key KR_{CN} and adds printer's ID_{PR} and other basic information to the registry list.

For SP, the process of authentication and registration is finished by the previous four steps. Actually, the authentication and registration of SR are very similar to SP's process. But the ID, time stamp, public key and password of laptop are ID_{LA}, T_{LA}, KU_{LA}, and PW_{LA}, respectively. Finally, the CN decrypts the information by its private key KR_{CN} and adds laptop's ID_{LA} and other basic information to the registry list.

4.4. The Steps of Invocation Services. When a legal SR wants to be serviced, for example, the laptop needs to print. The steps of service are as follows. The process of services invocation is shown in Figure 6.

Step 1. The laptop generates a time stamp T_{LA} and asymmetrically encrypts its ID_{LA} and request (i.e., Req) by the KU_{CN}. Then it sends the T_{LA} and $E_{KU_{CN}}(\text{Req}, ID_{LA} \text{ and } KU_{LA})$ to the CN.

Step 2. The CN receives the information and checks that whether the T_{LA} is effective. If T_{LA} is passed, the CN will decrypt the information by its private key KR_{CN} and read the request from SR. At the same time, the CN searches corresponding SPs from registry list and gets the ID_{PR} and KU_{PR} from an available SP. Then it generates a time stamp T_{CN} and asymmetrically encrypts the ID_{PR} and KU_{PR} of printer. Finally it sends the $E_{KU_{LA}}(ID_{PR} \text{ and } KU_{PR})$ and T_{CN} to the laptop.

Step 3. When the laptop gets the information from CN, it will check the T_{CN} too. If the T_{CN} is effective, the laptop will decrypt the information by its private key KR_{LA} and get KU_{PR} of printer. Then it generates a time stamp T_{LA} and asymmetrically encrypts its ID_{LA} and KU_{LA} by the KU_{PR}. It directly sends the T_{LA} and $E_{KU_{PR}}(ID_{LA} \text{ and } KU_{LA})$ to the printer.

Step 4. The printer receives the information and checks the T_{LA}. If the checking is passed, the printer will choose a secret random r_s and compute the $b = g^{rs} \in Z_p^*$. Meanwhile, it generates a time stamp T_{PR} and calculates the temporary key $K_{Temp} = h(T_{PR} \oplus b)$. Then it sends the $E_{KU_{LA}}(ID_{PR} \text{ and } K_{Temp})$ and T_{PR} to the laptop.

Step 5. The laptop gets the information from the printer and then checks the T_{PR}. If T_{PR} is effective, the laptop will decrypt the information by its KR_{LA} and get K_{Temp}. It generates a

FIGURE 5: The steps of registration.

FIGURE 6: The steps of invocation services.

time stamp T_{LA} and symmetrically encrypts the DATA by the K_{Temp}. It sends the T_{LA} and $(DATA)_{K_{Temp}}$ to the printer.

Step 6. The printer receives the information and checks the T_{LA}. If the T_{LA} is not effective, then the printer will ignore the information; otherwise, it will decrypt the information by K_{Temp} and get the DATA to execute the task of printing.

4.5. Performance Evaluation and Analysis. To analyze the security performance of the previous policies without CN, we assume that an illegal user steals the public key and frequently sends a lot of illegal information to the target devices. It will bring much interference to the system and waste a lot of bandwidth. But our policy, during the service, adopts one-time keys, which are changing in every single time to minimize the possibility of key exposure. It can protect the true identity of device and achieve the users' anonymity. Meanwhile, the devices must pass the legal authentication before they are able to request the services. This policy not only improves the efficiency but also enhances the security.

Additionally, the time stamp used in above steps can improve the efficiency and safety. If the time stamp is checked to be illegal, the entities will not carry out the following steps including encryption and decryption. Thereby it can improve the system efficiency and reduce the waste of resources. What is more important, it can also prevent repeating attacks when the time stamp is checked as illegal.

4.6. Simulation. In this section, we provide some numerical examples to illustrate the performances of the previous proposed policies. In the simulation, we assume that there are some legal data that need to be transmitted in 1000 seconds.

In this paper, the total value of legal data is set as 750 MB, and the bandwidth of system is set as 1 Mbps. We use Poisson model to characterize the data packets arrival process. The length of packets ranges from 1 to 200 KB randomly. The legal and illegal users transmit data in the same time. But once the illegal users begin to attack, the illegal data is five times the amount of legal data during the transmission. During the process of legal data transmission, the number of illegal users is increasing gradually. The number of illegal users increases from 1 to 8. We assume that there is only an illegal user after 50th seconds, and then the number of illegal users multiplies by two in each 100 seconds. Finally, there are 8 illegal users in the 350th seconds. After that, the number of illegal users always keeps in maximum value, which is 8, till the 1000th seconds. If there is no measures to protect the system, the bandwidth of system will be seriously occupied.

As shown in Figure 7, the vertical axis shows the sum of valid transmitted data. And the horizontal axis shows the time. In this simulation, we evaluate the performance of different policies by the sum of valid transmitted data. The larger value of the sum is, the better performance of the policy is. It means that the system is less influenced by the attacks of illegal users.

From Figure 7, the line of centralized node with authentication almost linearly increases during 1000 seconds. It is because that the invalid data of illegal users is detected and filtered by the process of registration in CN and the authentication among devices. The target device receives all the legal data without interference, which is about 740 MB. As the total value of valid data is 750 MB, the ratio of received valid data is over 99%. The decentralized node with authentication almost linearly increases to 602 MB in 1000th second, in which the ratio of received valid data is about 80%. Without CN's management, the invalid data is only filtered by the authentication among devices. So some invalid data is still received by the target device, which takes up a part of available bandwidth. In addition, the centralized node without authentication is similar to the decentralized node with authentication, which is linearly increasing to 340 MB in 1000th second. Due to the use of registration in CN, a portion of invalid data is filtered by the centralized node. So the ratio of received valid data is about 45.3%. However, it is shown that the decentralized node without authentication transmits the minimum valid data, whose final value is only 100 MB. The ratio of received valid data is only 13%. When the illegal users increas, this curve has a quick decay. It is because that the target device receives both valid and invalid data in the same time. But the illegal data takes up most of the bandwidth, which has a deep impact on legal data transmission. The policy of centralized node with authentication is much better than the other three policies, which demonstrates the superiority of our security authentication policy.

5. Bandwidth Measurement

Bandwidth measurement is another enabling technology to enhance the networking efficiency. Usually, there are many data transmission applications in the home M2M networks.

Figure 7: Four kinds of security authentication policies performance.

Figure 8: The framework of media transmission system.

We take the multimedia transmission as an example to illustrate the bandwidth measurement algorithm.

5.1. System Model for Adaptive Multimedia Transmission. The basic framework, which has been adopted for the media transmission system, is shown in Figure 8. We can see that there are mainly two modules in the system: available bandwidth measurement module (ABWM module) and scalable video coding module (SVC module) [13]. Video data is encoded by the SVC module and then sent to the network according to the current measurement result of the ABWM module. SVC is also called layered coding, whose main concept is to divide the original video stream into two or more streams, called layers.

Before the transmitter sends the video stream to the receiver, available bandwidth measurement algorithm (e.g., Pathload [14], PathChirp [15], etc.) will be used to measure

the available bandwidth between the transmitter and receiver. Then the base layer data will be sent to the client to ensure that the receiver can get basic video image information. After that, the enhancement layers will be sent according to the current available bandwidth. If the available bandwidth between the transmitter and receiver is very low, the transmitter will not send enhancement layers or just send a few enhancement layers to the receiver. Otherwise, if the available bandwidth between the transmitter and receiver is high, then the transmitter will send all the enhancement layers to the receiver to ensure that the receiver can receive the video with the best quality.

5.2. Available Bandwidth Measurement Algorithms

(1) The Definition of Available Bandwidth. Available bandwidth [16, 17] is the maximum data transmission rate that can be provided for a new single connection without affecting the transmission rate of the existing connections.

Assuming that the cross-traffic stays constant in a short time T, u_i denotes link Li's utilization ratio in time period T, and then the available bandwidth on the path P can be defined as follows:

$$A = \min_{i=1,\ldots,N} C_i \left[1 - u_i\right]. \tag{2}$$

(2) the Category of The Available Bandwidth Measurement Algorithms. Usually, the available bandwidth measurement algorithms are divided into two classes: direct probing and iterative probing.

Probe Gap Model (PGM) is one of the direct probing models. It estimates the available bandwidth through the changes of the gap in the probing packet pair. The source host sends the probing packet train at the rate of R_i, and the rate of the probing packet train that the destination host receives will be R_0, and then the available bandwidth can be calculated by

$$A = C_{\mathrm{TL}} - R_i \left(\frac{C_{\mathrm{TL}}}{R_0} - 1\right), \tag{3}$$

PGM model has two advantages: simple implementation and low computational complexity, but it only works when $R_i > A$. There is also a premise when applying PGM model: the tight link bandwidth C_{TL} must be known beforehand.

Iterative probing measures the available bandwidth via high speed probing packet trains which will cause congestion in the path. Bursty packet trains will be sent to the network. If the rate between two packets is higher than the available bandwidth, congestion will occur in the path instantaneously. The timing space between probing packets will be changed. We can estimate the available bandwidth by analyzing these delay characteristics of probing packets. The measurement process is actually an iterative process of changing the rate of the packet trains continually. Therefore, iterative probing is also called (PRM) Probe Rate Model. Unlike PGM, iterative probing does not require the tight link bandwidth C_{TL} to be known. The rate of the packet trains can be changed linearly or following some kind of rules. In the PRM measurement process, when $R_0 < R_i$, we will consider that $R_i > A$; else if $R_0 = R_i$, then $R_i \leq A$. Additionally, $R_0 > R_i$ is impossible. Change R_i for iterative probing, and find the critical point of $R_0 = R_i$. The rate at this point is the available bandwidth A.

(3) The PathChirp Algorithm. PathChirp [15] is a kind of iterative probing. PathChirp estimates the available bandwidth along a path by sending a number of packet chirps from transmitter to receiver and then the receiver, will get the available bandwidth of the path by analyzing the delay characteristic of the packets. The transmitter sends the chirps as shown in Figure 9. One chirp consists of N exponentially spaced packets. Ideally, the queuing delay of the packets that the receiver received will monotonously increase from packet K. But due to bursty traffic, queuing delays will typically not increase monotonically within a chirp. Figure 10 shows the queuing delays of a typical chirp train [15]. We refer to such a plot as a queuing delay signature.

PathChirp uses the shape of the signature to make an estimation $E_k^{(m)}$ of the per-packet available bandwidth. It then takes a weighted average of the $E_k^{(m)}$'s corresponding to each chirp m to obtain estimates $D^{(m)}$ of the per-chirp available bandwidth:

$$D^{(m)} = \frac{\sum_{k=1}^{N-1} E_k^{(m)} \Delta_k}{\sum_{k=1}^{N-1} \Delta_k}. \tag{4}$$

Finally, it makes estimation of the available bandwidth by averaging the estimates $D^{(m)}$ obtained in the time interval $[t - \tau, t]$.

Compared with other existing measurement algorithms, PathChirp has less influence on network, consumes less time for measurement, and has a high accuracy. It is suitable for both wired or wireless measurement environments.

(4) Experiment Results. We use PathChirp algorithm to measure the available bandwidth in home network. In this measurement, the link bandwidth of the home work is about 85 Mbps. One host runs the transmitter program, and another host runs the receiver. We also add some cross-traffic (about 10 Mbps) during the measurement period. Table 3 and Figure 11 show the result.

In Figure 11, the vertical axis shows the values of available bandwidth. The horizontal axis shows the times. In this paper, we carry out five measurements by using the PathChirp algorithm. The circle points are on behalf of the values of actual available bandwidth, and the star points are on behalf of PathChirp algorithm's measurement values. We evaluate the performance of PathChirp algorithm by comparing the measurement values with the values of actual available bandwidth.

For comparison, we use the actual values minus the corresponding measurement values. From Table 3 and Figure 11, we can see that the maximum and minimum of errors are 6.23 MB and 0.145 MB, respectively. The average error value of measurements is 3.38 MB. In short, the measurement results are closed to the real value of the network. This means

Figure 9: Chirp probe train.

Figure 10: A typical queuing delay signature.

Figure 11: Experiment result.

Table 3: Experiment result.

Measurement	1	2	3	4	5
Path Chirp (Mbps)	71.77	68.93	80.21	73.31	78.855
Actual available bandwidth (Mbps)	78.0	72.0	83.0	78.0	79.0

that the accuracy of the measurement algorithm is high. The PathChirp algorithm can basically reflect the current situation of the available bandwidth. Therefore, we can use it to measure the available bandwidth and monitor the network status in home networks.

6. Conclusion and Future Work

This paper proposes an efficient and secure resource management framework in home M2M networks. In this paper, there is a centralized node in the system architecture. We present a security authentication approach, which is a policy of centralized authentication. It can protect the true identity of devices and achieve the requirement of users' anonymity. Our system is more efficient and secure, because of efficient management of keys and resource allocation. We also adopt an available bandwidth measurement algorithm. It can adaptively measure the bandwidth and dynamically adjust the resource, which has less influence on network, consumes less time for measurement, and has a higher accuracy. Due to the above technologies, we can achieve the goal of efficient and secure resource management in home M2M networks. For future work, we will add the digital signatures of devices to information, which can further improve the security of the system.

Acknowledgments

This research is partially supported by programs of NSFC (grant nos. 61370159, U1035001, U1201253, and 61203117), the Opening Project of Key Lab. of Cognitive Radio and Information Processing (GUET), Ministry of Education (grant no. 2011KF06), and key research projects supported by the Department of Science and Technology of Guangdong Province (grant nos. 2012B090600053, 2011A090100039, 2011B090300108, 2011A080803009, 2011A011305003, 2011B090400360, and 2012B031800215).

References

[1] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: architectures, standards, and QoS improvement," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, 2011.

[2] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive machine-to-machine communications: visions and potentials for the smart grid," *IEEE Network Magazine*, vol. 26, no. 3, pp. 6–13, 2012.

[3] Q. Liu, S. Leng, Y. Mao, and Y. Zhang, "Optimal gateway placement in the smart grid Machine-to-Machine networks," in *Proceedings of the IEEE GLOBECOM Workshops (GC Wkshps '11)*, pp. 1173–1177, December 2011.

[4] Y. Zhang, S. Gjessing, H. Liu, H. Ning, L. Yang, and M. Guizani, "Securingvehicle-to-grid communications in the smart grid," *IEEE Wireless Communications Magazine*. In press.

[5] D. He, C. Chen, S. Chan, Y. Zhang, J. Bu, and M. Guizani, "Secure service provision in smart grid communications," *IEEE Communications Magazine*, vol. 50, no. 8, pp. 53–61, 2012.

[6] R. Yu, Y. Zhang, S. Gjessing, C. Yuen, S. Xie, and M. Guizani, "Cognitive radio based hierarchical communications infrastructure for smart grid," *IEEE Network*, vol. 25, no. 5, pp. 6–14, 2011.

[7] K. Suzuki and M. Inoue, "Home network with cloud computing for home management," in *Proceedings of the 15th IEEE International Symposium on Consumer Electronics (ISCE '11)*, pp. 421–425, June 2011.

[8] Universal Plug and Play, Understanding Universal Plug and Play: A White Paper, June 2000, http://upnp.org/resources/whitepapers.asp.

[9] DLNA Overview and Vision Whitepaper 2007, http://www.homeelectronics.jp/whitepaper/3.pdf.

[10] IGRS, Information Device Intelligent Grouping and Resource Sharing QoS Specification for Wireless UWB networks (draft), 2008.

[11] Ministry of Industry and Information Technology of the People's Republic of China, "The Communication Protocols of Multimedia and Data Networks," People's Republic of China.

[12] J. Zhu and J. Ma, "An efficient authentication protocol with anonymity for wireless IP networks," *Journal-China Institute of Communications*, vol. 25, no. 6, pp. 12–18, 2004.

[13] P. Amon, H. Li, A. Hutter, D. Renzi, and S. Battista, "Scalable video coding and transcoding," in *Proceedings of the IEEE International Conference on Automation, Quality and Testing, Robotics*, vol. 1, pp. 336–341, May 2008.

[14] M. Jain and C. Dovrolis, "End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput," in *Proceedings of Passive and Active Measurements Workshop (PAM '02)*, pp. 295–308, August 2002.

[15] V. J. Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "Pathchirp: efficient available bandwidth estimation for network paths," in *Passive and Active Measurement Workshop*, vol. 4, 2003.

[16] N. Hu and P. Steenkiste, "Evaluation and characterization of available bandwidth probing techniques," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 6, pp. 879–894, 2003.

[17] M. Jain and C. Dovrolis, "End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput," in *ACM SIGCOMM Computer Communication Review*, vol. 32, no 4, pp. 295–308, August 2002.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 871607, 12 pages
http://dx.doi.org/10.1155/2013/871607

Research Article

Passive Synchronization Based Energy-Efficient MAC Protocol over M2M Wireless Networks

Pranesh Sthapit and Jae-Young Pyun

Department of Information and Communication Engineering, Chosun University, Gwangju 501-759, Republic of Korea

Correspondence should be addressed to Jae-Young Pyun; jypyun@chosun.ac.kr

Received 19 July 2013; Revised 4 October 2013; Accepted 4 October 2013

Academic Editor: Zhong Fan

Copyright © 2013 P. Sthapit and J.-Y. Pyun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Providing diverse quality-of-service (QoS) with ultra-low power consumption and mobility support is the most important and challenging issue in machine-to-machine (M2M) networks. In a typical beacon-enabled network, nodes need to wake up for receiving periodic beacon in order to maintain synchronization. Here, we present a new MAC protocol called passive-synchronization-based energy-efficient MAC (PSMAC) that synchronizes nodes in their sleep state by using interrupt generated from the proposed radio-triggered hardware. In order to activate the radio-triggered hardware, the synchronization frame is broadcast, but in a separate channel called synchronization channel, so that interrupts are not generated during data transmission. This synchronization frame is also used for providing the fast and energy-efficient association. Network information is embedded into the synchronization frame so that mobile nodes can learn about their neighbors just by scanning the synchronization channel. Furthermore, by positioning the beacon period after the contention access period, PSMAC provides fast and preemptive slot allocation that can handle diverse QoS requirements. We compare PSMAC with IEEE 802.15.4 and show that PSMAC has much better performance in node association time, energy efficiency, and faster data delivery at the cost of additional radio-triggered hardware and a dedicated channel.

1. Introduction

Machine-to-machine (M2M) communications are the information exchanges among machines without any human interaction. The possibility to establish networks of automated devices without human intervention facilitates the creation of new applications that can make the world smarter and more efficient, for example, smart buildings, smart cities, smart grids, e-health, or automotive applications, among many other examples. M2M communications are characterized by low power, low cost, and low human intervention [1]. A typical M2M network would have large number of nodes; thus, sharing radio resource efficiently while maintaining sufficient quality-of-service (QoS) for reliable communications becomes an essential and challenging requirement [2, 3]. In many M2M applications, saving energy for machines is more important than increasing the throughput because machines usually transmit a small amount of data and are battery powered. One of the critical issues of M2M is how to deal with large number of accesses from large amount of machines while maintaining low power consumption with tolerable latency. Furthermore, some applications such as the automotive applications, robotic networks, and e-health need support for mobility [4]. Hence, the design of a smart and efficient MAC protocol remains a key requirement for successful deployment of any M2M network.

MAC for M2M should support different types of data such as continuous, periodic, burst, and aperiodic along with different levels of QoS [2]. The main schemes of MAC protocols are grouped into either CSMA/CA or TDMA protocols [5]. The CSMA/CA protocols support large number of devices but encounter heavy collisions and extra energy consumption. On the other hand, TDMA protocols provide good solutions to the collision and idle listening problems. However, common TDMAs are required to receive beacons periodically in order to synchronize their clocks. Since the radio on a sensor node is usually the component that uses the most energy, the node should go into sleep state (turn-off transceiver) as long as possible [6]. Another possible alternative to save energy is to use an additional radio-triggered

Figure 1: The superframe structure of IEEE 802.15.4.

hardware, which can interrupt the sleeping node at the event of interest [7]. Gu and Stankovic proposed radio-triggered wakeups for energy saving [8]. The idea is to use only passive components in order to collect energy from ongoing radio transmissions, as it is done in RFID technology. When the RF signal power induced at the receiving antenna is large enough, a microcontroller in the sensor node is interrupted to wake up its radio component for data communication. Similarly, a wireless identification and sensing platform (WISP) [9] is a commercially available RFID tag, which supports sensing and computing. WISP allows an exciting exploration of a new battery-free form of wireless sensor networking.

In the literature, some MAC schemes have been proposed for M2M to attach end device with the base station (BS), such as the adaptive-traffic-load- (ATL-) slotted MACA in [10], the access class barring in [11], and the frame-based hybrid MAC scheme for M2M networks in [5]. However, these schemes are insufficient. To achieve successful M2M communications, quality-of-service- (QoS-) guaranteed provisioning is the most important requirement. In [12], a QoS-guaranteed M2M massive access management scheme is proposed to reduce energy consumption and end-to-end delay. The IEEE 802.15.4 standard that exhibits a desirable feature for low-power M2M space has been analyzed as a platform for M2M. However, there are several limitations in meeting the requirements of M2M in terms of power consumption, ability to handle mobility, and also flexibility and adaptability [13]. IEEE 802.15.4 lacks support for mobility because it spends a lot of time for association. Also, in IEEE 802.15.4, regardless of presence of data, nodes need to wake up in every beacon period for synchronization draining significant energy. Also, IEEE 802.15.4 allows only up to seven guaranteed time slots (GTSs), which is not enough for M2M applications. The high latency associated with the GTS allocation is also a significant problem. Furthermore, GTS allocation does not consider priorities among different applications. High-priority application can be blocked when all GTSs are already allocated to low-priority applications. In this paper, we propose a hybrid MAC protocol called passive-synchronization-based energy-efficient MAC (PSMAC) for M2M networks, which will combine the benefits of both contention-based and reservation-based protocols. PSMAC supports diverse QoS along with mobility, but with minimal power consumption. To avoid periodic wakeups and decrease the association time, we introduce a new period called synchronization period (SP) in the superframe structure. During SP, coordinators broadcast synchronization frames (SFs) on a separate channel called synchronization channel (SC). SF is used by the unassociated nodes for association and by the already associated nodes for synchronization. SF activates the mounted radio-triggered circuit (triggered only by the SF) interrupting the sleeping node. Thus, a node is synchronized without even turning on the main communication radio saving significant amount of energy. Similarly, nodes can learn about the neighbors by just scanning SC enabling prompt association. Furthermore, by positioning the beacon period after the contention access period, PSMAC provides fast and preemptive slot allocation that can handle diverse QoS requirements. The proposed PSMAC solution attempts to provide mobility, flexibility, scalability, and adaptability, combined with ultra-low power consumption. The PSMAC is sophisticated MAC protocol design that will be beneficial in the next-generation M2M applications.

The rest of this paper is organized as follows. Section 2 presents the brief description on association procedure of beacon-enabled IEEE 802.15.4 followed by network model in Section 3. Sections 4 and 5 show the design and main features of the proposed PSMAC protocol. Section 6 explains QoS features of PSMAC: prompt association, fast and preemptive slot allocation, and prioritized random backoff. Section 7 presents performance evaluation using NS-2. Finally, we conclude the paper in Section 8.

2. Association in Beacon-Enabled IEEE 802.15.4

In IEEE 802.15.4, there are mainly two types of devices: full function device (FFD) and reduced function device (RFD). FFD can support all the network functions and can operate as a coordinator or an end device. RFD can only be used as an end device. The IEEE 802.15.4 standard supports three kinds of topologies: star, peer-to-peer, and cluster tree topologies, which can operate on beacon and non-beacon-enabled modes.

In the beacon-enabled mode, communication is synchronized and controlled by a network coordinator, which transmits periodic beacons. The periodic beacons are used to synchronize the attached devices, to identify the personal area network (PAN), and to describe the structure of the superframe. The superframe shown in Figure 1 may consist of active and inactive periods. The active part of the superframe consists of two groups: the contention access period (CAP) and the optional contention free period (CFP). All

Figure 2: Channel scanning mechanism in beacon-enabled IEEE 802.15.4.

Figure 3: Association message exchange in IEEE 802.15.4.

communications must take place during the active part, and devices can sleep to conserve energy in the inactive part. The structure of superframe is determined by coordinators using two parameters: superframe order (SO) and beacon order (BO). SO is used to determine the duration of superframe, whereas BO defines the beacon interval. Increasing both BO and SO increases transmission latency and decreases system throughput due to intensive channel contention, whereas simultaneously decreasing BO and SO increases energy consumption and reduces available bandwidth [14]. Any FFD could be a PAN coordinator. FFD should perform the energy detection (ED) scan in each channel for the duration of t_{scan} symbols to detect the peak energy and choose an appropriate channel for data transmission. Then, it carries out an active scan to locate any coordinator transmitting beacon frames within its personal operating space (POS). During the active scanning, it first sends out a beacon request command and waits for the duration of t_{scan} symbols. If a beacon could not be detected in t_{scan}, the FFD believes that there is no coordinator in its POS and can construct its own PAN by broadcasting its periodic beacons. After a PAN has been initialized, any device in the POS can join it. In order to start association, an end device needs to discover coordinators in the surrounding. In beacon-enabled IEEE 802.15.4, two types of channel scanning operations are performed by end devices.During scan, nodes are deprived from data communication and must discard all data frames received. As shown in Figure 2(a), passive scan is performed by a node, which is just turned on or that failed to find its coordinator through the orphan scan. During passive scan, a device searches for beacon in each channel for the duration of t_{scan} symbols and records the beacon frames received in each channel. If no beacon is detected, the device starts another passive scan after a period of time. Orphan scan allows a device to attempt to relocate its coordinator following a loss of synchronization. The device shall first send the orphan notification command frame and waits for coordinator realignment command frame for at most *MAC response wait time* symbols as shown in Figure 2(b). This procedure is repeated until it receives the coordinator realignment frame or all of the available channels are scanned. If the device is unable to find its parent through the orphan scan, it looks for a new parent by performing passive scan. Once the node finds the suitable coordinator, it starts association procedure by sending a request for association as shown in Figure 3. In beacon-enabled IEEE 802.15.4, the association requires a time-consuming channel scanning before the association message exchange as discussed above. For a mobile node, association procedure is even worse because the mobile node has to go through the orphan scan first followed by the passive scan and then only can start the association message exchange. We use the term reassociation to explicitly denote the association procedure used by the mobile node. There are some efforts done to minimize the association duration in IEEE 802.15.4. In IEEE 802.15.4 e [15], optional fast association (FastA) is defined, which allows a device to associate in a reduced duration. However, most of the efforts, such as [16–18], are limited to decreasing the duration of association message exchange or coordinator discovery, whereas the channel scanning part has been left untouched so far. The association duration can be drastically shortened if nodes are somehow prevented from scanning every single available channel. Similarly, there are several works on multichannel solutions, but they are limited to throughput improvement or beacon collision avoidance [19–21].

3. Network Model

We consider M2M network consisting of one base station (BS), k coordinators, and n number of end devices. Moreover, coordinator or FFD is assumed to have fixed static position, while end device or RFD can be mobile. The basic network design is shown in Figure 4, which models a typical home area network (HAN) [2, 3]. All end devices are attached with a coordinator. The coordinator collects data from the surrounding nodes and delivers them to the BS. The BS is the sink node or control center. The distance between a coordinator and an end device is assumed to be of several meters, whereas the distance between a coordinator and the BS can be of several kilometers. In this paper, we only consider the communication between end devices and a coordinator and proposed a MAC protocol for communication between them. Also, the beacon-enabled star topology is considered. However, the communication between the coordinator and the BS is assumed to be done by using some long-distance wireless technology such as 3G connection. The end device can be any machine such as smart meter, smart TV, sensors, and body sensors attached to human body. In our network model, we assume a heterogeneous scenario where all of the devices have a different amount of data with different priorities and QoS requirements.

Figure 4: Network model.

4. System Design

Figure 5 shows the physical architecture of sensor node adapting the proposed PSMAC. As shown in the figure, apart from a normal RF module, a new hardware, that is, a passive radio sensor, is added to the sensor node. Radio sensor is a passive device and does not have decoding capacity [7]. Its sole purpose is to generate interrupt whenever the signal within its operational range is detected. A periodic signal transmitted at the beginning of every superframe activates the radio sensor generating an interrupt. The periodic interrupt synchronizes (keeps track of superframe timing) node in the sleep state. This feature results in continuous operation of sensor node in the sleep mode without losing synchronization and instantly waking-up upon the need for data communication [22].

Unlike the other conventional MACs, which use separate frequency band for the radio sensor [8], PSMAC uses the same frequency band. However, a dedicated channel is assigned for synchronization.

Figure 5: Sensor node physical architecture.

4.1. Motivation behind Radio Sensor. In low-power wireless networks, waking up the receiver in every superframe to listen to a beacon is a significant drain on energy. One solution to avoid node waking in every beacon period is to maintain internal clock and sleep through a number of beacon periods without losing synchronization [13]. However, there are two problems associated with the skipping of beacons. The first problem is the clock drift which may lead to synchronization error if the beacon is skipped for long time. However, this problem can be solved by listening to beacon at certain intervals of superframes. The second but major problem is how to communicate with sleeping nodes. If a coordinator wants to request data from a sleeping node, the only option is to wait until the node wakes up and receives a beacon. These two problems motivated the use of radio sensor. The radio sensor has following two functions.

(i) *Synchronization*. The radio sensor allows nodes to skip beacon. The SF transmitted by a coordinator in the synchronization channel causes the radio sensor to generate interrupt. The interrupt synchronizes node in the sleep state.

(ii) *Wake-Up Node*. When a coordinator wants to communicate with a sleeping node, the SF is broadcasted in the synchronization channel twice. Two consecutive interrupts mean that a node should wake up getting ready for the communication.

In the first sight, it may seem that the radio sensor is extra hardware, which means extra cost. However, the close examination of Figure 6 reveals that the cost of radio sensor is less than few dollars as it is made of few registers and capacitors, a diode, and a transformer. The investment of few dollars has potential to prolong node's life by few years.

Figure 6: Radio sensor schematic.

4.2. Hardware Architecture of Radio Sensor. The schematic of the radio sensor is shown in Figure 6. The radio sensor designed here does not need any power supply for its operation [7]. The designed hardware first collects and stores energy from the radio signals and generates the sufficient output voltage to interrupt the node. The resistor R1 and the capacitors C1 and C2 form a bandpass filter filtering out the radio frequencies different from the wake-up signal frequency and preventing the false alarms. The resistance of R1 and the capacitance of C1 and C2 depend on the property of the antenna and the wake-up signal. A step-up voltage transformer is added to boost the input voltage received from the antenna. The transformer increases the peak voltage of the received signal to a certain level making it easier to reach the required threshold voltage for triggering an interrupt. Resistor R2 is for impedance matching preventing the circuit from reflecting back the received signal. Furthermore, in case the instantaneous signal is not sufficient to drive the interrupt, the capacitor C3 is added to the circuit to store energy over a period of time and drive the interrupt when sufficient energy is accumulated on the capacitor. The circuit requires some time; that is, few milliseconds are needed to accumulate sufficient voltage to trigger interrupt.

5. Overview of Proposed Protocol

The proposed PSMAC has been designed to deliver continuous, periodic, aperiodic, and emergency data supporting diverse QoS requirements of M2M energy efficiently along with the support of mobility. In this section, we discuss how IEEE 802.15.4 is analyzed and how a new MAC protocol is developed fulfilling all of the requirements of M2M network.

5.1. Superframe Structure. In order to support all data types of Table 1, PSMAC combines both contention-based and contention-free schemes keeping priority intact. Figure 7 shows the format of the superframe defined by the coordinator. A superframe consists of mainly four periods: SP, CAP, beacon period (BP), and CFP. To avoid the periodic wake-up for synchronization and to support mobility, we introduced SP. CFP is divided into TDMA slots of equal length called dedicated time slot (DTS) for handling continuous data. Any device requesting for DTS or wishing to communicate during the CAP competes with other devices using the CSMA-CA mechanism. All transactions are completed by the time of the next network BP. Nodes are assigned DTS during CFP and notified through the beacon. However, the unused DTSs form the inactive period during which both the coordinator and nodes can sleep. All contention-based transactions are completed before the BP begins. Also, each device transmitting in a DTS ensures that its transaction is completed before the end of the CFP. The structure of this superframe is described by the values of BO and SO. However, BO and SO in PSMAC are totally different than those of IEEE 802.15.4. Here, BO determines the interval of the beacon period from the beginning of the superframe (i.e., the length of CAP), and SO describes the interval at which the coordinator shall transmit its synchronization frames (i.e., the length of superframe) as shown in Figure 7. The structure of superframe is determined by coordinators using ($0 \leq BO \leq SO \leq 14$), which determines length of superframe and CAP. However, when $BO = SO$, there is no CFP period; thus, there is no BP also. $SO = 0$ and $SO = 14$ give superframe durations of 15.36 ms and 251.69 s (assuming same parameters as IEEE 802.15.4 at data rate of 250 Kbps). Since the length of superframe can be varied from 15.36 ms to 251.69 s, the protocol can support large number of nodes. However, according to the requirement, the optimal value of BO and SO should be chosen to balance the tradeoff between energy consumption and latency. Thus, the superframe admits flexible configuration according to application demand. Furthermore, the protocol is pure CSMA/CA without superframes when $BO = SO = 15$.

Table 1: Data nature and priority in M2M application.

Emergency	Priority	Data types	Examples
Yes	1	Emergency	Emergency vital signs
No	2	Continuous	EEG, ECG, and EMG
	3	Periodic	Sensors smart meter
	4	Entertainment	Audio and video
	5	Others	File transfer

5.2. Synchronization Period. The superframe begins with the synchronization period, and generally nodes are in sleep state during SP. SP is used for the synchronization of currently associated nodes and for the association of a newly joining node. In our proposed PSMAC scheme, we exploit the channel switching capability of CC2420 radio hardware [23]. A dedicated separate channel called SC is assigned for synchronization. For example, IEEE 802.15.4 has 16 channels (11 through 26) for 2.4 GHz frequency, channel 26 (interference free from 802.11) can be used for synchronization, and the rest are used for communication [24]. At the beginning of SP, coordinator switches to SC and broadcasts the synchronization frame and switches back to its original DC as shown in Figure 8. The synchronization frame consists of basic information of coordinator and superframe such as the network identifier, the communication channel of the M2M, and the lengths of CAP and superframe. Assigning the dedicated channel for synchronization has the following advantages.

Figure 7: The superframe of the proposed PSMAC.

Figure 8: Channel switching mechanism of coordinator in the proposed PSMAC.

(i) *Prevention from False Alarm.* Only the synchronization frame transmitted in the synchronization channel is capable of activating the radio sensor and triggering the interrupt.

(ii) *Quick Association.* Unassociated nodes can learn about neighbors just by listening to the synchronization channel only.

The duration of the SP should be long enough to accumulate two consecutive synchronization frames (explained later in Section 5.2.3). The synchronization frame has mainly the following three functions.

5.2.1. Network Initialization. Our intelligent MAC design prevents nodes from scanning each and every available channel for association. When a coordinator is initialized, it performs the passive scan (similar to IEEE 802.15.4) in the SC for the durations of t_{scan}. The scan duration should be long enough to receive every SF in the surrounding. During the passive scan, the coordinator records the time instance of every SF it receives. If no SF is received during the passive scan, coordinator assumes that there is no coordinator in the surrounding, and it can start its superframe instantly. However, if any SF is received, the coordinator schedules its own superframe after the SP duration from the time instance of the last SF it received. The whole process is shown in the flowchart of Figure 9. However, in the rare situation, two or more coordinators can start the passive scan at the same time resulting in the collision of SFs. In this situation, the neighboring nodes (those performing the passive scan) acknowledge the collision of SFs, and coordinators start new passive scan after waiting for the random amount of time. The scheduling algorithm ensures that the new SF does not collide with SF of existing coordinators and maintains the safe distance of SP. Coordinator is flexible to select its own nonoverlapped data channel, but it must broadcast its SF in the SC.

Figure 9: Flowchart for superframe scheduling in PSMAC.

5.2.2. Node Association. An unassociated node switches its radio channel to SC and performs the passive scan for the duration of t_{scan} symbols. Once it finds the suitable coordinator, it switches its radio channel the same as the DC of the coordinator (and remains there) and sends request for association similar to IEEE 802.15.4. Once the node is associated, it turns off its receiver and leaves synchronization task to the passive radio sensor. Note that only the coordinator changes its radio channel to SC in every SP and switches to DC at the end of SP.

5.2.3. Synchronization. SF causes radio sensor to generate an interrupt. Multiple HANs can coexist, and each coordinator transmits their its SF. But nodes are synchronized by the interrupt received during their SP only and are ignored if the interrupt is received elsewhere. However, it may be possible that SF is lost and node is unable to receive interrupt during

its SP. The missing of interrupt has to be handled separately for the following two cases.

(i) When BO < SO, there exists BP when BO < SO. Thus, whenever a node misses an interrupt, it wakes up during the BP and synchronization is restored by receiving the next beacon. Similarly, a mobile node performs association procedure if it loses the beacon for a certain amount of time.

(ii) When BO = SO, since there is no BP, the missing of interrupt has to be handled differently. In this case, node switches its channel to SC and listens to the SF in the next superframe. If the SF is received from the coordinator, then the synchronization is restored. However, if the SF cannot be received for a certain amount of time, then the node starts association procedure to find a new parent.

Although nodes are synchronized only by the interrupts received during their SP, some interference from the unwanted signal may trigger the interrupt during SP. Thus, to avoid false synchronization, nodes receive beacon (for BO < SO) or SF (for BO = SO) at the regular interval of certain superframes.

Thus, by introducing the SP, now the nodes can continue to sleep until they need to transmit data. But this implementation may prevent the coordinator from communication with the nodes because the nodes are skipping the beacons. The situation has been cleverly handled with the following technique. If a coordinator wants to communicate with a sleeping node, it transmits two consecutive SFs interrupting nodes two times. Whenever a node is interrupted two times during its SP, it must wake up immediately. Thus, the particular node can be reached.

5.3. Communication. The data communication is done in both CAP and CFP periods. CAP is used for periodic data, emergency alarm, and command frames. CFP is used for continuous data, that is, when nodes have to transmit in every superframe. The nodes intending to transmit during CAP use prioritized CSMA/CA algorithm for channel access. During CAP, the coordinator collects DTS requests and allocates DTS based on the priority. The coordinator assigns first available slot to the highest-priority data. The coordinator broadcasts a beacon in the BP. The beacon is used for synchronizing all of the network nodes and notifying DTS allocation status. The length of a DTS slot should be long enough to transmit a data frame and receive an ACK. The total number of DTSs supported is more than 7 and depends on value of BO and SO. A single node can be given one or more DTSs depending on its requirement. The assigned DTS remains with the node until and unless coordinator or node explicitly cancels it. However, if all DTSs are exhausted, higher-priority data can preempt lower-priority data. Thus, every node with DTS must verify its DTS and its position through beacon.

Only the devices waiting for communication have to wake up to receive the beacon from the coordinator. When the coordinator wants to send data to a sensor, it allocates DTS and notifies through the beacon. The designated sensor listens in the assigned time slot to receive the data and replies with the ACK ending the process. A node sends periodic or emergency data during the CAP and continuous data in CFP using DTS. After receiving the data frame, the coordinator replies with the ACK frame. The transceiver is active only during the communication and, for the remainder of the time, is in sleep mode for power saving.

6. QoS Features and Analysis

6.1. Association. Let aBaseSuperframeDuration be the number of symbols forming a superframe when SO = 0, and let t_{scan} symbols be the time spent to scan the SC. The total time spent for association is the sum of time spent in channel scan and the time spent in the association message exchange (Asso_{msg}). The total time spent for channel scan (t_{scan}) and node association (t_{asso}) are given by

$$t_{\text{scan}} = \text{aBaseSuperframeDuration} \times \left(2^{\text{SO}} + 1\right), \quad (1)$$

$$t_{\text{asso}} = t_{\text{scan}} + \text{Asso}_{\text{msg}}. \quad (2)$$

As shown in (1), the association time in PSMAC does not depend on the number of available channels; thus, it saves signification amount of time and energy. A device that wants to associate with a coordinator can be new or orphan. IEEE 802.15.4 has separate procedures for the association of new and orphan devices. However, in PSMAC, there is no orphan scan, and all unassociated devices perform passive scan in SC only.

6.2. Data Priority and Transmission Delay in CAP. Priority of data is determined by the degree of their importance and is divided into different priority classes as summarized in Table 1. However, in practice, the priority classification should accommodate application requirements.

From Figure 11, the packet transmission delay, T_l, can be calculated as

$$T_l = T_{\text{bo}} + T_{\text{data}} + T_{\text{wtx}} + T_{\text{ack}} + T_{\text{ifs}}, \quad (3)$$

where T_{bo} is the total backoff time, T_{data} and T_{ack} are transmission times for data and ack packets, respectively, T_{wtx} is transmitter warm-up time, and T_{ifs} is time of interframe pace (IFS). As shown in (3), the only variable parameter in the equation is T_{bo}. Thus, prioritized random backoff is applied, and the number of backoff slots, N_{bos}, is randomly chosen for emergency and nonemergency data as follows:

$$N_{\text{bos}} = \begin{cases} \text{random}\left[0, \dfrac{2^{\text{BE}}}{2} - 1\right] & \text{if emergency,} \\ \text{random}\left[0, 2^{\text{BE}} - 1\right] & \text{else,} \end{cases} \quad (4)$$

where BE denotes the backoff exponent. Hence, emergency data enjoy higher probability to win and dominate the transmission opportunity.

6.3. Prompt DTS Allocation and Scheduling. For real-time applications, assuring data delivery on time and initiating

(a) The case that data are generated during CAP

(b) The case that data are generated after CAP

Figure 10: Delay associated with DTS allocation.

Figure 11: Transmission delay for CAP.

data transmission immediately after sensor deployment are the most important issues. As shown in Figure 7, the positions of SP and BP enable a device to quickly get associated with a coordinator and immediately request DTS allocation during CAP. Then, the coordinator allocates DTS, and the device transmits data in its assigned TDMA slot. Unlike IEEE 802.15.4, all of these processes are done in sequence and within one superframe interval. The comparison of allocation delay between proposed and IEEE 802.15.4 is shown in Figures 10(a) and 10(b) and derived as

$$T_{dl} = \begin{cases} T_{\text{cap}} - t_{dr}, & t_{dr} < \text{CAP}, \\ T_{sf} + T_{\text{cap}} - t_{dr}, & t_{dr} \geq \text{CAP}, \end{cases} \tag{5}$$

where t_{dr} is the time instance in CAP during which data are generated and T_{sf} and T_{cap} are the lengths of the superframe and CAP, respectively. The delay becomes the worst when the data are generated right after CAP because a node has to wait for the next CAP for allocation request.

The coordinator collects the DTS requests and serves the highest-priority data first. The coordinator sorts all the requests and allocates higher priority first from the first available slot as shown in Figure 12(a). If remaining DTSs are insufficient to admit all requests, higher-priority data preempt lower-priority data as shown in Figure 12(b).

Figure 12: Priority-based DTS allocation.

7. Performance Evaluation

The simulation analysis is performed using network simulator NS-2. Nodes are deployed in a 500 m × 500 m field as shown in Figure 13. Eight end devices form the star topology with a coordinator in the center. In all the simulation, SO is the same as BO (for IEEE 802.15.4), whereas the CAP, CFP, and superframe of PSMAC were set the same as with IEEE 802.15.4. Unless mentioned, the superframe of PSMAC

Table 2: Network parameters and values.

Parameter	Value
Assumed power supply	3 V
Reception power	56.5 mW
Transmission power	48 mW
Idle power	2.79 mW
Sleep power	30 μW
Radio range	40 m
Routing	AODV
Frequency band	2.4 GHz
Radio data rate	250 Kbps
Number of channels	16
Buffer size	10 packets
Packet size	100 bytes
Traffic	CBR and emergency
Node	**Packet interval**
Nodes 1, 7, and 8	0.2 secs
Node 2	1 sec
Node 3	2 secs
Node 4	3 secs
Node 5	4 secs
Node 6	No data

Figure 13: Network topology used in the simulation.

is made the same as with IEEE 802.15.4 with BO = SO = 4 (245.76 ms). In the simulation time of 100 secs, all nodes start data transmission. The simulation is run for 600 secs. Network parameters and the data traffic attached with end devices are listed in Table 2. The parameters of RF module are taken from CC2420 datasheet [23].

7.1. Association Time. Figure 14 shows the network initialization time and the node re/association times for various values of BO obtained from NS-2 simulations. In IEEE 802.15.4, the PAN coordinator performs the passive scan in all available channels, but in PSMAC only SC is scanned. Thus, the network initialization time is shorter in PSMAC. As shown in the figure, the time required by PSMAC for re/association is much lower because only the SC is scanned for the association procedure. However, in the case of IEEE

Figure 14: Total time spent for association.

802.15.4, it scans all available 16 channels spending significant amount of time and energy. Also, from the figure, we can see that the time required for both re/association is the same for PSMAC because there is no orphan scan, and the same passive scan procedure is used for both re/association. At BO = 3, PSMAC is able to decrease the network initialization time by 15.92 times, the node association time by 6.19 times, and the reassociation time by 16.59 times, which is a great achievement in itself.

7.2. Energy Consumption. The energy saving achieved by PSMAC was studied by performing various experiments. The simulation results showed that, in the heavy traffic, IEEE 802.15.4 and PSMAC have comparable energy consumption. However, in low traffic, the energy saving by PSMAC becomes significant. Energy consumption obtained by varying the upload traffic is shown in Figure 15. The message inter-arrival period of 0 implies no data upload at all and interarrival period of 1 implies that data are uploaded once in every second. As seen from Figure 15, PSMAC consumes the least energy when nodes have no data to upload because nodes in PSMAC do not need to wake up in every superframe if there is are data. In the case of absence of data traffic, PSMAC saved the energy consumption by 13.85 times as compared to IEEE 802.15.4. The skipping of beacons accounts for the energy saving. We observed that the lesser the data traffic is, lesser the energy consumption in the case of PSMAC is.

7.3. Latency. Figure 16 shows DTS allocation delay, which is the time duration from the data generation to the confirmation of DTS via beacon. For this analysis, superframes of both protocols were divided into 16 slots, while slot 0 and slot 8 were allocated for beacon positions of IEEE 802.15.4

Figure 15: Average energy consumption.

Figure 16: Latency experienced in slot allocation.

and PSMAC, respectively. The graph shows allocation delays, while data generation time is moving from slot 1 toward slot 16 of superframe. As observed in Figures 10(a) and 10(b), slot allocation in IEEE 802.15.4 is not confirmed until the next BP. However, PSMAC enables fast allocation just after CAP as well as prompt transmission within current superframe. In both cases, delay becomes the worst when the data are generated right after CAP because a node has to wait for the next CAP for allocation request. In this analysis, we observe that PSMAC reserves DTS almost twice faster than IEEE 802.15.4.

Figure 17 shows the average transmission delay and average maximum delay when nodes with the emergency data contend during CAP. In this simulation, all of the end devices (Figure 13) were given the same constant bit rate (CBR) traffic, except that node 6 generated 10 emergency packets during

Figure 17: Latency experienced in emergency traffic.

Figure 18: Network topology used for the mobile environment.

the simulation of 600 secs. The CBR traffic is varied from the message interval of 0.2 to 1 sec. In PSMAC, an emergency node has a high probability to choose short backoff duration, resulting in low-access delay. The latency of emergency data was decreased by 50% in PSMAC as compared to IEEE 802.15.4.

7.4. Association Success Rate and Throughput. To observe how PSMAC handles mobility, a simulation was performed. In this simulation, a BS, 8 coordinators, and one mobile node are used as shown in Figure 18. Nodes are placed 40 m away from each other, and the arrow heads indicate the direction of movement of the mobile node. In the simulation time of 100 secs, node 9 starts data transmission at the rate of 4 Kbps. In the simulation time of 110 secs, node 9 starts to move while transmitting data and stops when it reaches its original position. The simulation environment simulates a person walking in a building. Success rate of association is calculated as the ratio of successful associations to the total number of possible associations. In our simulation model, there were

Figure 19: Association success rate at 0.5, 1, and 1.5 m/s.

Figure 20: Throughput observed at mobility of 1.5 m/s, data rate of 4 Kbps, and BO = 4.

total 8 associations possible before the mobile node comes to rest. PSMAC increases the successful association rate of a mobile node by providing quick passive discovery of a coordinator. Percentage of successful associations at different BO(s) and node speeds are shown in Figure 19. We observed that, in case of IEEE 802.15.4, even a slight mobility has a significant negative impact on association. At the human walking speed of 1.5 m/s, IEEE 802.15.4 has poor success rate of association even at the lower values of BO, and at BO = 5, nodes were completely unable to associate. However, in the case of PSMAC, association rate was 100% until SO = 4 (equivalent to BO = 4 in IEEE 802.15.4), and even at SO = 5, PSMAC could successfully perform 7 associations out of 8. However, at node speed of 1.5 m/s, PSMAC was completely unable to associate at SO ≥ 9.

The effect of cell change on throughput observed at the BS when BO = 4 and at the node speed of 1.5 m/s is shown in Figure 20. Throughput of PSMAC drops while performing cell change because packets cannot be transmitted in the periods of passive discovery and association. However, nodes quickly get associated with a new coordinator and restart data transmission. Nodes can buffer packets. Therefore, buffered data are also transmitted after new association is completed, resulting in increased throughput immediately after association. The average throughput of PSMAC was of 3490 bps, and the packet delivery ratio was 87%. However, in case of IEEE 802.15.4, average throughput was just 1180 bps, and packet delivery ratio was 29% as nodes were unable to associate in every cell change resulting in poor performance.

8. Conclusion

In this paper, we proposed PSMAC that supports diverse QoS along with mobility, but with minimal power consumption. PSMAC uses two separate channels: synchronization channel and data channel. Synchronization period is added in the superframe during which coordinators change their radio channel to the synchronization channel and broadcast the synchronization frame. A radio-triggered hardware attached to node, tuned to be activated by the synchronization frame only, synchronizes the nodes in their sleep state and conserves energy. Similarly, nodes can learn about neighbors by just scanning the synchronization channel which enables fast association. Furthermore, by positioning the beacon period after the contention access period, PSMAC provides fast and preemptive slot allocation that can handle diverse QoS requirements. Our simulation results demonstrated that our protocol is highly energy efficient, takes less time for association, supports diverse QoS, and also has low latency, which can be adapted for various M2M applications.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Republic of Korea Government (MEST) (no. 2011-0013062).

References

[1] G. Botter, J. Alonso-Zarate, L. Alonso, F. Granelli, and C. Verikoukis, "Extending the lifetime of M2M wireless networks through cooperation," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, IEEE, 2012.

[2] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: architectures, standards, and QoS improvement," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 44–52, 2011.

[3] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[4] B. H. Lee and S. L. Kim, "Mobility control for machine-to-machine LTE systems," in *Proceedings of the 11th European Wireless Conference 2011-Sustainable Wireless Technologies (European Wire-less)*, pp. 1–5, VDE, 2011.

[5] Y. Liu, C. Yuen, J. Chen, and X. Cao, "A scalable hybrid MAC protocol for massive M2M networks," http://arxiv.org/abs/1301.4315.

[6] P. Sthapit and J. Pyun, "Medium reservation based sensor MAC protocol for low latency and high energy efficiency," *Telecommunication Systems*, vol. 52, no. 4, pp. 2387–2395, 2013.

[7] P. Sthapit and J. Pyun, "Effects of radio Triggered Sensor MAC protocol over wireless sensor network," in *Proceedings of the 11th IEEE International Conference on Computer and Information Technology (CIT '11)*, pp. 546–551, September 2011.

[8] L. Gu and J. A. Stankovic, "Radio-triggered wake-up for wireless sensor networks," *Real-Time Systems*, vol. 29, no. 2-3, pp. 157–182, 2005.

[9] A. P. Sample, D. J. Yeager, P. S. Powledge, A. V. Mamishev, and J. R. Smith, "Design of an RFID-based battery-free programmable sensing platform," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 11, pp. 2608–2615, 2008.

[10] G. Wang, X. Zhong, S. Mei, and J. Wang, "An adaptive medium access control mechanism for cellular based machine to machine (M2M) communication," in *Proceedings of the IEEE International Conference on Wireless Information Technology and Systems (ICWITS '10)*, September 2010.

[11] CATT, "R2-100182: access control of MTC devices," in *Proceedings of the 3GPP-TSG RAN WG2 Meeting 68bis*, January 2010.

[12] S. Y. Lien and K. C. Chen, "Massive access management for QoS guarantees in 3GPP machine-to-machine communications," *IEEE Communications Letters*, vol. 15, no. 3, pp. 311–313, 2011.

[13] N. F. Timmons and W. G. Scanlon, "An adaptive energy efficient MAC protocol for the medical body area network," in *Proceedings of the 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace and Electronic Systems Technology, Wireless (VITAE '09)*, pp. 587–593, May 2009.

[14] S. A. Gopalan and J. Park, "Energy-efficient MAC protocols for wireless body area networks: survey," in *Proceedings of the International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT '10)*, pp. 739–744, October 2010.

[15] IEEE P802.15.4e-2012, "Part 15.4: Low-Rate Wireless Personal Area Networks (WPANs), Amendment 1: MAC sub-layer," February 2012.

[16] F. Bashir, W. S. Baek, P. Sthapit, D. Pandey, and J. Y. Pyun, "Coordinator assisted passive discovery for mobile end devices in IEEE 802.15.4," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC '13)*, January 2013.

[17] F. Zhang, F. Wang, B. Dai, and Y. Li, "Performance evaluation of IEEE 802.15.4 Beacon-enabled association process," in *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications Workshops (AINAW '08)*, pp. 541–546, March 2008.

[18] C. Chaabane, A. Pegatoquet, M. Auguin, and M. B. Jemaa, "Energy optimization for mobile nodes in a cluster tree IEEE 802.15.4/ZigBee network," in *Proceedings of the Computing, Communications and Applications Conference (ComComAp '12)*, pp. 328–333, January 2012.

[19] F. Osterlind and A. Dunkels, "Approaching the maximum 802.15.4 multi-hop throughput," in *Proceedings of the 5th ACM Workshop on Embedded Networked Sensors (HotEmNets '08)*, pp. 1–6, Charlottesville, Va, USA, June 2008.

[20] E. Toscano and L. L. Bello, "Multichannel superframe scheduling for IEEE 802.15.4 industrial wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 2, pp. 337–350, 2012.

[21] A. W. Chen, B. Shih, C. Chang, and C. Chen, "Enhanced mac channel selection to improve performance of ieee 802.15.4," *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 12, pp. 5511–5526, 2010.

[22] J. Ansari, D. Pankin, and P. Mähönen, "Radio-triggered wake-ups with addressing capabilities for extremely low power sensor network applications," in *Proceedings of the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, September 2008.

[23] Texas Instrument, "CC2420 Data Sheet," http://www.ti.com/lit/ds/symlink/cc2420.pdf.

[24] N. C. Tas, C. Sastry, and Z. Song, "IEEE 802.15.4 throughput analysis under IEEE 802.11 interference," in *Proceedings of the International Symposium on Innovations and Real Time Applications of Distributed Sensor Networks*, 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 584707, 8 pages
http://dx.doi.org/10.1155/2013/584707

Research Article

A Game-Theoretic Analysis of Multiple Protocol Data Flows in Hierarchical M2M Communication Networks

Gang Wang, Jie Leng, and Lin Bai

School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Lin Bai; l.bai@buaa.edu.cn

Received 13 August 2013; Accepted 3 October 2013

Academic Editor: Jianhua He

Copyright © 2013 Gang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A hierarchical machine-to-machine (M2M) communication network, where multiple heterogeneous devices compete for transmission on congested links, is considered. In such a network, the cluster headers gather the network coding and routing messages from multiple M2M devices and forward them to the cellular network. We propose a novel discriminatory user-based utility mechanism for network coding users and routing users. The price of anarchy (PoA), which is used to characterize the worst-case efficiency bounds, is analyzed. The simulation results show that network coding could promote the efficiency to a certain extent.

1. Introduction

With the increasing development of the mobile communications and the expansion of the internet applications, the machine-to-machine (M2M) technologies are driving demands. The M2M communication networks are carried out as an integration of different techniques including transducer technologies, computer software, and data communications. In particular, the M2M communications provide diverse application scenarios compared to those of traditional communications among humans, where numerous sensor nodes and massive data are in need. The features, that are infrequent data transmissions and functions of group-based operations, are inherent from human-to-human communications [1].

Since a great number of self-organized nodes offer the information sources or the command execution without human intervention, these nodes can also be considered to carry the selfish or self-benefited features to take actions [2]. Thus, the M2M network is more likely to get congested especially when massive selfish nodes act to maximize each individual benefit. In order to overcome such a problem, the idea of game theory over the strategic behaviors of selfish users can be further studied [3–5].

With the advent of network coding [6], nodes are allowed to process the incoming independent information flows before forwarding them to communication networks. Network coding has been acknowledged to offer benefits along various dimensions of communication networks such as wireless resources, throughput, and security [7]. Particularly, in [8], it has been confirmed that as an effective means of efficient reliable data dissemination, network coding plays a key role in developing the M2M communication networks. In addition, combined with traditional routing schemes, alternative strategies can also be carried out in M2M communications [9].

Using the well-known game theory together with the network coding, in this paper, we propose a game model with the existence of Nash equilibrium (NE) [10] in M2M communication networks. In particular, we analyze the price of anarchy (PoA) in a cognitive M2M communication system with a hierarchical structure and cluster headers, where the tremendous information which contains the data packets and signaling messages is issued by a significant number of mobile devices. Assuming that the number of source devices in machine type communication (MTC) is the same as the number of nodes in the network, the nodes can be regarded as the network users who can also be classified as network coding users or routing users depending on different transmission patterns. The main idea is that different behaviors of different sorts of users, which are considered to maximize each individual surplus, can influence the efficiency of the whole system.

Figure 1: Hierarchical system in cognitive M2M communication.

Consider a cloud-based M2M network in Figure 1, where a number of machine nodes transmit message to the same amount of receivers through one simple link. Each M2M device selects a nearby cluster header as its desired access point. Cluster header gathers the total amount of members that belong to it and then broadcasts the information to each member and the enhanced Node Bs (eNBs). After that, the eNBs constructs the connections between the two clusters, and each device selects its own strategies. With the inferences of the PoA which is adopted to measure the efficiency of the system, the network coding and routing can be regenerated to improve the PoA, whether the link is congested or not. The formulations of PoA under the conditions of the link being congested or unrestricted can be obtained. It can be observed that the clues are concluded to induct the resources allocation to achieve a better efficiency according to the consistence of different sorts of users.

The rest of the paper is organized as follows. In Section 2, the related works with respect to the resource allocation methods in M2M communication, network coding, and pricing mechanisms are summarized. Section 3 proposes the network model including the game model, calculating functions, and some definitions. Section 4 illustrates the transmission game and the corresponding conclusions. Simulations in Section 5 show the envisioned curves of the PoA under different conditions. The paper is finally concluded in Section 6 with future works.

2. Related Work

A cognitive M2M communication system has been proposed in [11], in which the cluster headers gather the traffic information from multiple M2M devices using cognitive radio technologies in cellular bands and forward it to the cellular networks. In this case, the number of M2M devices that directly access the cellular network is decreased and the cellular network only needs to manage the transmission among the header nodes and the reduced M2M devices. It has been confirmed in [11] that the congestion or overload of the control signal can be avoided or resolved using the corresponding structure, while an effective way has been considered in [12] to promote energy efficiency. In [13], the 3GPP concept and the network operation have been presented where different solutions have been provided to release the congestion in the radio as well as the core network, by allocating different priorities for different users to access the network. Based on the 3GPP standard, the overview of the M2M and LTE network architectures has been presented in [14]. However, the detailed operations and the corresponding results or influences have not been well demonstrated. In order to develop efficient mechanisms to handle congestion settlements due to signaling messages which are issued by massive numbers of mobile devices, [15] has proposed the mechanisms for signaling congestion settlements to minimize the frequency of attempts or disallow the connection from the devices to the network. In addition, it has been demonstrated in [16] that different protocols can be carried out to improve controlling congestion and overload problems over MTC devices.

As to the network coding, it has been confirmed in [17] that resource allocation games with routing can be utilized to improve the throughput in a single line bottleneck link of wireless networks. This work has been extended in [18], where joint intersession network coding has been first considered and the discriminatory pricing functions of routing and network coding which is based on the affine marginal cost (AMC) pricing mechanism in [17] have been developed. Note that since the link capacity has not been considered in the above literature, the efficiency and optimization of resource allocations may not be achieved for the network coding in M2M networks.

Similar to the market regulation of price in economics, the pricing mechanism is an effective means to allocate resources. The widely known pricing mechanism is the AMC mentioned above. However, as the AMC does not consider the number of users sharing the resources, the average cost-sharing (ACS) pricing mechanism has been developed in [19]. Note that ACS takes not only the total rates through the sharing link but also the number of sharing users into consideration. Thus, the ACS pricing mechanism is more appropriate for the issues as we will discuss.

3. Network Model

Let us simplify the M2M communication architecture in Figure 1 to a congestion control model which is shown in Figure 2.

Figure 2: A link shared by K users, containing one network coding flow and R routing flows.

We assume that there are R $(R \geq 1)$ routing users together with N $(N \geq 2)$ network coding users, while the total number of users is $K = N + R$. The set of coding users is denoted by $N = \{1, 2, \ldots, N\}$. For each user, $i \in N$, r_{NC} is the coding rate of each network coding user. For all routing users, $q = \sum_{i=1}^{R} r_i$ represents the sum of routing rates, while r_i is the transmission rate of routing user $i \in R = \{1, 2, \ldots, R\}$.

Nodes W and Z represent two cluster headers. Each user is offered with the system information that contains the utility functions and the price or cost formulations. A congestion communication link between two intermediate nodes, W and Z, is given by $(W; Z)$. The intermediate node W performs the linear network coding on packets of N NC users. The node W delivers network coding packets together with routing packets to node Z, and then node Z forwards these packets to the corresponding destinations. Thus, the sum rate over link $(W; Z)$ becomes $q + r_{NC}$. Notice that the link will be overloaded if it exceeds the fixed link capacity C, that is, $q + r_{NC} = r_{NC} + \sum_{i=1}^{R} r_i \leq C$.

In the following, we will introduce the game model and its related computing methods or functions.

3.1. Game Model

(i) Players: users in set K.

(ii) Strategies: transmission rates $\sum_i r_i$ and data transmission options.

(iii) Payoffs: surplus of one user $S_i(r_i)$ and the whole system $\sum_{i=1}^{K} S_i(r_i)$.

Although each user prefers to make a decision to choose a transmission mode between routing and network coding to acquire its own payoffs, maximizing the payoffs of the whole system is the main target.

3.2. Utility, Cost, and Price Functions. The utility function $U(r_i)$ quantifies the user i's payoff in terms of its desired transmission rate r_i; here $i \in K$. In this paper, we propose a novel discriminatory user-based utility mechanism, which is defined as follows.

Definition 1. For each network coding user, its transmission rate is r_{NC}, and then the utility function is $U(r_{NC}) = \gamma_{NC} \cdot r_{NC}$. For each routing user $i \in R$, the utility function is $U(r_i) = \gamma_R \cdot r_i$. Here, both γ_{NC} and γ_R are utility parameters, which are in the interval (0,1].

Obviously, the utility parameters depend on the transmission mode. Without the loss of generality, the utility function is assumed to be nonnegative, concave, increasing, and differentiable, which is the same as the one defined in [17].

The function $C(r_i)$ is a monetary cost for utilizing the link resource, which is defined as follows.

Definition 2. For each user $i \in K$, if its transmission rate is r_i, the cost function $C(r_i) = r_i \cdot p(\cdot)$. Here, $p(\cdot)$ denotes the price function depending on the transmission mode.

In the following, we formulate the resource allocation problem in the hierarchical M2M communication networks under the proposed discriminatory user-based utility mechanism. An efficient rate allocation can be characterized as an optimal solution of the following aggregate surplus maximization with network coding and routing (SMNCR) problem [18]:

SMNCR Problem

$$\begin{aligned} S\left(\sum_i r_i\right) &= \sum_i U(r_i) - \sum_i C(r_i) \\ &= \sum_{i=1}^{N} \left(\gamma_{NC} \cdot r_{NC} - C(r_i)\right) \\ &\quad + \sum_{i=1}^{R} \left(\gamma_R \cdot r_i - C(r_i)\right) \\ &\text{s.t.} \quad r_i \geq 0, \quad i = 1, \ldots, K. \end{aligned} \tag{1}$$

The ACS pricing mechanism takes both the numbers of sharing users and the total rates through the sharing link into consideration. Reference [19] defines it as follows.

Definition 3. Assume there are n users. To each user i, $i = 1, 2, \ldots, n$, its transmission rate is given by r_i. Then, the ACS pricing function is defined as $p(\sum_i r_i / n)$.

Notice that the linear price functions are the only price functions that can satisfy the four axioms of additivity, rescaling, consistency, and positivity for cost-sharing systems at the same time [20, 21]. Suppose that $p(x) = a \cdot x$, where the parameter a is a constant real number and $a > 0$. To all network coding users, the transmission rate is r_{NC}, while they share the same link resource. Hence, the cost function becomes

$$C(r_{NC}) = r_{NC} \cdot p\left(\frac{r_{NC}}{N}\right) = \frac{a \cdot r_{NC}^2}{N}. \tag{2}$$

Then, to routing user $i \in R$, the cost function is given by

$$C(r_i) = r_i \cdot p\left(\frac{\sum_i r_i}{R}\right) = \frac{a \cdot r_i \cdot q}{R}. \tag{3}$$

In the next section, we will analyze the existence of NE [8] and formulate the corresponding efficiency in terms of the PoA.

3.3. NE and PoA. The selfish nature of players leads to undesirable and inefficient network performances. The selfish behavior is mainly presented by choosing the strategies to maximize its own payoffs without the consideration of the utility of the whole system. An NE is an action profile with the property that no single player can obtain a higher payoff by deviating unilaterally from the profile.

The NE of a game can be defined as a rate vector $\mathbf{r}^*$ such that for any user $i \in K$, we have $S(r_i^*, \mathbf{r}_{-i}^*) \geq S(r_i, \mathbf{r}_{-i}^*)$, for all $r_i \geq 0$. Here, $\mathbf{r}_{-i}^*$ denotes the vector of all rates chosen by users other than ith. The existence of NE in the proposed game model is analyzed as follows.

Suppose that for user $i \in K$, its transmission rate is r_i. At first, we analyze the existence of r_i^*. According to (1) and Definitions 1 and 2, the utility function is $U(r_i) = \gamma_i \cdot r_i$ and the cost function is $C(r_i) = r_i \cdot p(r_i)$; then we can obtain

$$\begin{aligned} F(r_i) &= \frac{\partial S_i(r_i, \mathbf{r}_{-i})}{\partial r_i} = \frac{\partial U(r_i)}{\partial r_i} - \frac{\partial C(r_i)}{\partial r_i} \\ &= \gamma_i - p(r_i) - r_i \frac{\partial p(r_i)}{\partial r_i}. \end{aligned} \tag{4}$$

The condition to maximize the individual surplus $S(r_i)$ is given by $F(r_i) = \partial S(r_i, \mathbf{r}_{-i})/\partial r_i = 0$.

Furthermore, since $\partial F(r_i)/\partial r_i = -2\partial p(r)/\partial r_i < 0$, $F(r_i)$ is a monotone decreasing function. Then, we can infer that the transmission rate r_i^* exists and becomes the only solution when the price function $p(r_i)$ is given.

Based on the above analysis, we can conclude that an NE solution exists. Here, we only consider the two blocks of resources allocated to two sets of NC and routing users, respectively, instead of performing an individual resource allocation to each user. As a result, given the block of resource allocated to the routing users, each individual routing user's transmission rate depends on other routing users. Thus, NE is nonunique for this game model under the ACS pricing mechanism. After that, we can assess the efficiency of the network resource allocation by the PoA. The PoA is the worst-case efficiency of the NE among all possible choices of parameters, that is, the number of users, the utility, cost, and price functions.

Let $S^O(\mathbf{r})$ be an optimal solution of *SMNCR Problem* and let $S^*(\mathbf{r})$ be the NE of the game under the same system parameters. We can define the PoA as follows:

$$\text{PoA} = \frac{\min_r S^*(\mathbf{r})}{\max_r S^O(\mathbf{r})}. \tag{5}$$

The main purpose in this paper is to find the appropriate transmission strategies by analyzing the PoA under the discriminatory user-based utility mechanism. A detailed discussion is offered in the next section.

4. The PoA Analysis

In this section, we aim to find the effectiveness of the discriminatory user-based utility mechanism in the proposed game model. The PoA analysis is divided into three parts, i.e., the NC users' NE solutions, the routing users' NE solutions, and the optimal solutions of *SMNCR Problem*. Assume that the NC users have the priority to determine their coding rate. After that, each routing user select their own rate either under the conditions of the congestion or not. The optimal solutions may also adapt to the condition. Finally, the formulation of PoA is acquired.

Proposition 4. *For each network coding user $i \in N$, its transmission rate is denoted by $r_i = r_{NC}$, the NE solution is given by $r_{NC}^* = N \cdot r_{NC}/2a$, and the corresponding surplus of network coding group becomes $S_{NC}^* = N^2 \cdot \gamma_{NC}^2/4a$.*

Proof. According to Definitions 1 and 2, (1), and (2), we have

$$S_i(r_{\text{NC}}) = U(r_{\text{NC}}) - C(r_{\text{NC}}) = \gamma_{\text{NC}} \cdot r_{\text{NC}} - a \cdot r_{\text{NC}} \cdot \frac{r_{\text{NC}}}{N}. \tag{6}$$

Taking the derivative of (6) with respect to r_{NC} and we can show that

$$S_i'(r_{\text{NC}}) = \gamma_{\text{NC}} - 2 \cdot a \cdot \frac{r_{\text{NC}}}{N}. \tag{7}$$

Let the above equation be zero and we can obtain $r_{\text{NC}}^* = N \cdot \gamma_{\text{NC}}/2a$. Then, the surplus of the user becomes $S_i^* = N \cdot \gamma_{\text{NC}}^2/4a$, and the accumulation surplus of the network coding group is given by

$$S_{\text{NC}}^* = N \cdot S_i^* = \frac{N^2 \cdot \gamma_{\text{NC}}^2}{4a}. \tag{8}$$

□

Proposition 5. *Without congestion, one has*

$$PoA \geq \frac{N^2 \cdot \gamma_{NC}^2 \cdot (R+1)^2 + 4R^2 \cdot \gamma_R^2}{N^2 \cdot \gamma_{NC}^2 \cdot (R+1)^2 + R \cdot \gamma_R^2 \cdot (R+1)^2}. \tag{9}$$

Proof. According to Definitions 1 and 2, (1), and (3), for routing user $i \in R$, we can show that

$$\begin{aligned} S_i(r_i) &= U(r_i) - C(r_i) \\ &= \gamma_R \cdot r_i - a \cdot r_i \cdot \frac{\sum_i^R r_i}{R} \\ &= \gamma_R \cdot r_i - a \cdot r_i \cdot \frac{q}{R}. \end{aligned} \tag{10}$$

Taking the derivative of (10) with respect to r_i, we have

$$\frac{\partial S_i(r_i)}{\partial r_i} = \gamma_R - a \cdot \frac{q}{R} - a \cdot r_i \cdot \frac{1}{R}. \tag{11}$$

Since without any congestion, each routing user just selects its own rate to maximize (10). Due to the symmetry in [22], it is given that $q^* = R \cdot r_i^*$. Let (11) be zero and $r_i^* = R \cdot \gamma_R/a(R+1)$. Then, the surplus $S_i^* = R \cdot \gamma_R^2/a(R+1)^2$ and the accumulation surplus of the routing group is given by

$$S_R^* = R \cdot S_i^* = \frac{R^2 \cdot \gamma_R^2}{a(R+1)^2}. \tag{12}$$

According to (1), the total surplus of the system becomes

$$\begin{aligned} S\left(\sum_i r_i\right) &= \sum_{i=1}^{N}\left(\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}} - C\left(r_i\right)\right) \\ &\quad + \sum_{i=1}^{R}\left(\gamma_R\cdot r_i - C\left(r_i\right)\right) \\ &= N\cdot\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}} - N\cdot a\cdot r_{\mathrm{NC}}\cdot\frac{r_{\mathrm{NC}}}{N} \\ &\quad + \gamma_R\cdot\sum_i^R r_i - a\cdot\sum_i^R r_i\cdot\frac{\sum_i^R r_i}{R} \\ &= N\cdot\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}} - a\cdot r_{\mathrm{NC}}^2 \\ &\quad + \gamma_R\cdot q - a\cdot\frac{q^2}{R}. \end{aligned} \tag{13}$$

While the network is not yet congested, the q and r_{NC} are independent of each other. Thus, taking the derivative of q or r_{NC}, respectively, the optimal solution can be obtained as follows:

$$\begin{gathered} \frac{\partial\left(N\cdot\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}} - a\cdot r_{\mathrm{NC}}^2\right)}{\partial r_{\mathrm{NC}}} = 0, \\ \frac{\partial\left(\gamma_R\cdot q - a\cdot q^2/R\right)}{\partial q} = 0, \\ \Downarrow \\ r_{\mathrm{NC}}^{\mathrm{O}} = \frac{N\cdot r_{\mathrm{NC}}}{2a}, \\ q^{\mathrm{O}} = \frac{R\cdot r_R}{2a}. \end{gathered} \tag{14}$$

Then, the surplus maximums of two groups are, respectively, calculated as

$$\begin{aligned} S_{\mathrm{NC}}^{\mathrm{O}} &= N\cdot\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}}^{\mathrm{O}} - a\cdot\left(r_{\mathrm{NC}}^{\mathrm{O}}\right)^2 = \frac{N^2\cdot\gamma_{\mathrm{NC}}^2}{4a}, \\ S_R^{\mathrm{O}} &= \sum_{i=1}^{R} U\left(r_i\right) - \sum_{i=1}^{R} C\left(r_i\right) \\ &= \gamma_R\cdot q^{\mathrm{O}} - a\cdot\frac{\left(q^{\mathrm{O}}\right)^2}{R} = \frac{R\cdot r_R^2}{4a}. \end{aligned} \tag{15}$$

According to (8), (12), (15), and the definition of PoA, we have

$$\begin{aligned} \mathrm{PoA} &= \frac{\min_r S^*\left(\mathbf{r}\right)}{\max_r S^{\mathrm{O}}\left(\mathbf{r}\right)} \\ &\geq \frac{\sum_{i=1}^{K} U\left(r_i^*\right) - \sum_{i=1}^{K} C\left(r_i^*\right)}{\sum_{i=1}^{K} U\left(r_i^{\mathrm{O}}\right) - \sum_{i=1}^{K} C\left(r_i^{\mathrm{O}}\right)} \\ &\geq \frac{N^2\cdot\gamma_{\mathrm{NC}}^2/4a + R^2\cdot\gamma_R^2/a(R+1)^2}{N^2\cdot\gamma_{\mathrm{NC}}^2/4a + R\cdot r_R^2/4a}. \end{aligned} \tag{16}$$

□

Proposition 6. *In congestion, one has*

$$PoA \geq \frac{\left(N^2\cdot\gamma_{NC}^2/4a\right) + \left(\left(2a - N\cdot\gamma_{NC}\right)\left(2\gamma_R - 2aC + N\cdot\gamma_{NC}\right)/4a\right)}{\left(a\left(R{+}1\right)/R\right)\cdot\left[\left(NR\cdot\gamma_{NC} - R\cdot\gamma_R + 2aC\right)/2a\left(R+1\right)\right]^2 + C\cdot\gamma_R - \left(aC^2/R\right)}. \tag{17}$$

Proof. Once the congestion occurs, the worst NE solution is considered as the total resource for all the routing users is allocated to only one user [14], that is, $r_i^* = q^* = C - r_{\mathrm{NC}}^* = C - N\cdot\gamma_{\mathrm{NC}}/2a$, $r_1^* = \cdots = r_{i-1}^* = r_{i+1}^* = r_R^* = 0$. Therefore, in this case, the pricing function becomes $p(r_i^*)$. According to (10), the surplus of the routing group is given by

$$\begin{aligned} S_R^* &= \gamma_R\cdot r_i^* - a\cdot r_i^*\cdot\frac{q^*}{1} \\ &= \frac{\left(2a - N\cdot\gamma_{\mathrm{NC}}\right)\left(2\gamma_R - 2aC + N\cdot\gamma_{\mathrm{NC}}\right)}{4a}. \end{aligned} \tag{18}$$

Then, the global optimal surplus is obtained as follows:

$$\begin{aligned} S\left(\sum_{i\in K} r_i\right) &= N\cdot\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}} - a\cdot r_{\mathrm{NC}}^2 \\ &\quad + \gamma_R\cdot q - a\cdot\frac{q^2}{R} \\ &= N\cdot\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}} - a\cdot r_{\mathrm{NC}}^2 \\ &\quad + \gamma_R\cdot\left(C - r_{\mathrm{NC}}\right) - a\cdot\frac{\left(C - r_{\mathrm{NC}}\right)^2}{R}. \end{aligned} \tag{19}$$

Taking the derivative of (19) with respect to r_{NC}, we can have

$$r_{\mathrm{NC}}^{\mathrm{O}} = \frac{NR\cdot\gamma_{\mathrm{NC}} - R\cdot\gamma_R + 2aC}{2a\left(R+1\right)}, \tag{20}$$

while the optimal surplus is given by

$$\begin{aligned} S^{\mathrm{O}}\left(\sum_i r_i\right) &= N\cdot\gamma_{\mathrm{NC}}\cdot r_{\mathrm{NC}}^{\mathrm{O}} - a\cdot\left(r_{\mathrm{NC}}^{\mathrm{O}}\right)^2 \\ &\quad + \gamma_R\cdot\left(C - r_{\mathrm{NC}}^{\mathrm{O}}\right) - a\cdot\frac{\left(C - r_{\mathrm{NC}}^{\mathrm{O}}\right)^2}{R} \end{aligned}$$

$$= \frac{a(R+1)}{R} \cdot \left[\frac{NR \cdot \gamma_{NC} - R \cdot \gamma_R + 2aC}{2a(R+1)}\right]^2 + C \cdot \gamma_R - \frac{aC^2}{R}. \quad (21)$$

In addition, according to the definition of PoA, (8), (18), and (21), we have

$$\begin{aligned} \text{PoA} &= \frac{\min_r S^*(\mathbf{r})}{\max_r S^O(\mathbf{r})} \\ &\geq \frac{\sum_{i=1}^K U(r_i^*) - \sum_{i=1}^K C(r_i^*)}{\sum_{i=1}^K U(r_i^O) - \sum_{i=1}^K C(r_i^O)} \\ &\geq \left(\frac{N^2 \cdot \gamma_{NC}^2}{4a} + \frac{(2a - N \cdot \gamma_{NC})(2\gamma_R - 2aC + N \cdot \gamma_{NC})}{4a}\right) \\ &\quad \times \left(\frac{a(R+1)}{R} \cdot \left[\frac{NR \cdot \gamma_{NC} - R \cdot \gamma_R + 2aC}{2a(R+1)}\right]^2 + C \cdot \gamma_R - \frac{aC^2}{R}\right)^{-1}. \end{aligned} \quad (22)$$

□

5. Simulation and Numerical Results

In this section, we evaluate the performances of ACS pricing mechanism with choices of different parameters including N, R, γ_{NC}, and γ_R. It is assumed that the link capacity $C = 1$ and the pricing parameter $a = 1$. The utility ratio $\eta = \gamma_{NC}/\gamma_R$ is defined to formulate the PoA.

Firstly, let us suppose that the number of the NC users equals that of routing users, that is, $N = R$. Here, we consider that the PoA when $N = R$ varies from 2 to 100 with $\eta = 4, 2, 1, 0.5$, respectively. The corresponding results are shown in Figure 3.

As illustrated in Figure 3, when the utility parameter becomes $\eta = 1$ and $\eta = 2$, respectively, the corresponding PoA approaches to 1, which means that the efficiency of the system is maximized. When $\eta = 2$, with the increased number of users, the PoA is approximately 77%. On the other hand, when $\eta = 4$ (i.e., $\gamma_{NC} = 4\gamma_R$), the PoA is decreased to approximately 51%, which is surprisingly lower than those of the other three cases. The result shows that the utility gap between network coding and routing cannot be too large.

Let $N = 2$, where η is the counterpart of the β in [15]. In Figure 4, it is shown that when $N = 2$, the PoA is reduced to almost 5%, which indicates that as the proportion of NC users is lower than that of routing users, the ACS pricing mechanism could not provide better performance compared to the AMC pricing mechanism [13].

Figure 3: When $N = R$, the PoA varies with the N or R and η.

Figure 4: Efficiency (PoA) at an NE of 100 randomly generated resource allocation game scenarios when $N = 2$.

Under the same conditions, it is worthy to compare our proposed schemes to those in [14], where only two coding users are considered. With the number of the routing users varying from 0 to 100, as R is not very large, for example, $R < 10$, the uniform utility parameter mechanism functions well as shown in Figure 5. However, the increasing of routing users leads to a sharp decline of PoA if $\eta = 1$ or 1/2 as shown in Figure 5. Specifically, it also shows that as $\eta = 2$ or 4, the PoA becomes much more stable. In addition, when $R > 10$, which is always the case, the discriminatory utility mechanism is in advantage obviously.

Moreover, we make a research into the question whether the proportions of coding users or routing users in the whole users have an influence on PoA. The results are illustrated, respectively, in Figure 6, where the higher proportion of the

Figure 5: The PoA varies when $N = 2$ and R is from 0 to 100.

Figure 6: The PoA varies with the proportion of coding users or routing user.

network coding user leads to the larger PoA, that is, enlarging the performance of networks. It demonstrates that the proportion of users choosing network coding transmission mode has a direct effect on the efficiency of utilizing network resource.

6. Conclusion and Future Work

In this paper, we developed a novel discriminatory utility game model based on ACS pricing mechanism by assigning network coding users and routing users with the different utility parameters. The detailed analysis was provided to study the influences on the PoA due to different transmission strategies of the users. The simulation results demonstrated that the newly proposed utility mechanism contributes to the promotion of PoA. Furthermore, we found that the proportion of users choosing network coding acts as a key parameter to the PoA results, while the network coding could promote the efficiency of utilizing network resource to a certain extent. In reference to these conclusions, some useful methods or clues can be taken in the management of M2M network resources. Although only one transmission is considered among M2M devices in this paper, the results can be extended to repeated games in future works.

Acknowledgments

This work was supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation

of China (Grant no. 61221061) and the National Natural Science Foundation of China (Grant no. 61271194, no. 60972007, no. 61250001, and no. 61231013).

References

[1] 3GPP Organizational Partners, Study on RAN Improvements for Machine-Type Communications, 3GPP Support Office, Valbonne, France, 2011, http://www.3gpp.org/ftp/specs/html-INFO/37868.htm.

[2] Y. Wu, S. Deng, and H. Huang, "Optimal routing control in disconnected machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 963758, 11 pages, 2012.

[3] X. Zheng, P. Martin, W. Powley, and K. Brohman, "Applying bargaining game theory to web services negotiation," in *Proceedings of the IEEE 7th International Conference on Services Computing (SCC '10)*, pp. 218–225, Miami, Fla, USA, July 2010.

[4] Z. F. Mao, G. F. Nan, and M. Q. Li, "A dynamic pricing scheme for congestion game in wireless machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 840391, 9 pages, 2012.

[5] D. Xie, Q. Sun, Q. Zhou, Y. Qiu, and X. Yuan, "An efficient clustering protocol for wireless sensor networks based on localized game theoretical approach," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 476313, 11 pages, 2013.

[6] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[7] M. Muriel and A. Sprintson, Eds., *Network Coding: Fundamentals and Applications*, Academic Press, 1st edition, 2011.

[8] M. Zorzi, A. Gluhak, S. Lange, and A. Bassi, "From today's INTRAnet of things to a future INTERnet of things: a wireless- and mobility-related view," *IEEE Wireless Communications*, vol. 17, no. 6, pp. 44–51, 2010.

[9] L. Popova, A. Schmidt, W. Gerstacker, and W. Koch, "Network coding assisted mobile-to-mobile file transfer," in *Proceedings of IEEE Australasian Telecommunication Networks and Applications Conference (ATNAC '07)*, pp. 110–116, Christchurch, New Zealand, December 2007.

[10] K. Nishimura and J. Friedman, "Existence of Nash equilibrium in *n* person games without quasi-concavity," *International Economic Review*, vol. 22, no. 3, pp. 637–648, 1981.

[11] H. K. Lee, D. Kim, Y. Hwang, S. Yu, and S. L. Kim, "Feasibility of cognitive machine-to-machine communication using cellular bands," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 97–103, 2013.

[12] W. Liu, S. Zhang, and J. Fan, "A diagnosis-based clustering and multipath routing protocol for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 504205, 11 pages, 2012.

[13] R. Vaidya, C. Yadav, J. Kunkumath, and P. Yadati, "Network congestion control: mechanisms for congestion avoidance and recovery," in *Proceedings of the 1st International Conference on Wireless Technologies for Humanitarian Relief*, pp. 199–207, New York, NY, USA, 2011.

[14] J. L. Chen, H. C. Hsieh, and Y. T. Larosa, "Congestion control optimization of M2M in LTE networks," in *Proceedings of the 15th International Conference on Advanced Communication Technology*, pp. 823–8827, PyeongChang, Korea, January 2013.

[15] 3GPP Organizational Partners, System Improvements for Machine-Type Communications, 3GPP Support Office, Valbonne, France, 2012, http://www.3gpp.org/ftp/specs/html-INFO/23888.htm.

[16] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: potential, challenges, and solutions," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 178–184, 2012.

[17] R. Johari and J. N. Tsitsiklis, "A scalable network resource allocation mechanism with bounded efficiency loss," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 992–998, 2006.

[18] A. H. Mohsenian-Rad, J. Huang, V. W. S. Wong, S. Jaggi, and R. Schober, "A game-theoretic analysis of inter-session network coding," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, Dresden, Germany, June 2009.

[19] G. Wang and X. Dai, "The price of anarchy of network coding and routing based on an ACS pricing mechanism," *Chinese Journal of Electronics*, vol. 22, no. 3, pp. 567–571, 2013.

[20] Y. Chen and J. Zhang, *Design of Price Mechanisms for Network Resource Allocation via Price of Anarchy*, Mimeo, New York University, 2005.

[21] D. Samet and Y. Tauman, "The determination of marginal cost prices under a set of axioms," *Econometrica*, vol. 50, no. 4, pp. 895–909, 1982.

[22] A. H. Mohsenian-Rad, J. Huang, V. Wong, S. Jaggi, and R. Schober, "Inter-session network coding with strategic users: a game-theoretic analysis of network coding," 2009, http://arxiv.org/pdf/0904.2921.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 239170, 10 pages
http://dx.doi.org/10.1155/2013/239170

Research Article

A Dual-Cycle Architecture in Cognitive M2M Wireless Networks

Ying Li,[1] Xu Dong,[1,2] Lifeng Wang,[1] Xiaobo Guo,[1] and Lin Bai[3]

[1] *Institute of China Electronic System Engineering Corporation, Beijing 100141, China*
[2] *Institute of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China*
[3] *School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Lin Bai; l.bai@buaa.edu.cn

Received 13 August 2013; Accepted 17 October 2013

Academic Editor: Jianhua He

Copyright © 2013 Ying Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dual-cycle cognitive model (DCCM) is an effective means to develop the future large-scale cognitive wireless networks. In this paper, the DCCM is considered to construct cognitive wireless netwroks, in which the structure and implementation of service-oriented radio architecture (SORA) are expounded. Using SORA, a DCCM-based wireless resource management is considered in design and implementation of cognitive wireless networks. It shows that our proposed DCCM-based SORA is able to provide loosely coupled, reusable, and scalable implementations in machine-to-machine (M2M) heteogeneous wireless networking environment, while the intelligence of network can be improved.

1. Introduction

Cognitive radio technology has been developed during the last decade and played a crucial role in wireless communications. However, large-scale and heterogeneous wireless network environments have posed us big challenges with respect to the design and implementation of cognitive wireless networks. Since the cognitive wireless networks with various air interfaces require a great deal of sensors to monitor the surrounding environment, the interactions between these sensors and heterogeneous wireless networks may cause troubles in the process of system integration. Moreover, multiple hierarchies drive the design and management of cognitive wireless networks to be even more complex. Accordingly, a series of schemes have been considered in design and implementation of cognitive wireless networks.

Mitola proposed the observe-orient-plan-decide-act-learn (OOPDAL) loop to support the concept and architecture of cognitive radio in 1999 [1]. The OOPDAL loop is an abstract model to describe the operation process of a cognitive radio system, which is also known as the extension of Boyd's observe-orient-decide-act (OODA) [2]. In addition, Thomas designed his cognitive model based on the OODA loop [3]; Balamuralidhar divided the communication process into five phases to build the sense-analyze-decide-reconfigure-communicate (SADRC) loop [4]; Fortuna modified the OOPDAL loop and proposed the sense-plan-decide-act-learn-policy (SPDALP) loop [5]. All these loops above provide the basic models in order to construct cognitive radio systems. Nevertheless, they may not be able to describe the whole operation process of cognitive wireless networks. Therefore, an alternative model has to be considered before we make a further discussion on the implementation of cognitive wireless networks.

As a key component of cognitive radio system, the software-defined radio (SDR) platform has been well studied. The SDRs are usually implemented with tightly coupled software communication architectures (SCA) [6], which lack flexibility and compatibility in the integration of heterogeneous components. On the other hand, since a variety of cognitive radio systems were developed on different software and hardware development platforms, the integration of heterogeneous cognitive systems is drawing wide attentions [7]. Particularly, building a large-scale and multinode test bed over wide areas becomes an urgent need for the research on future cognitive wireless networks and has already been concerned by the Europeans [8]. Fortunately, the machine-to-machine (M2 M) technology provides us with an efficient

method to solve these problems. In [9], using the idea of service-oriented architecture (SOA) [10], a novel M2 M architecture named service-oriented radio architecture (SORA) has been developed to design and construct cognitive radio systems. In this respect, Virginia Tech has also built a cognitive radio open-source system (CROSS) to integrate cognitive engines based on different intelligent algorithms [11].

Although the SORA has provided us with a new way to implement the cognitive wireless networks, it may not be practical since only simple scenarios are considered on the application of SORA [10]. Thus, it is desired to consider the design and implementation of SORA in more realistic and complicated network scenarios. In this paper, using dual-cycle cognitive model (DCCM), a general structure of complex cognitive wireless networks is constructed with SORA. It shows that our proposed DCCM-based SORA is able to provide loose-coupling, reusability, and scalability implementations, especially in M2M heterogeneous wireless networking environments.

The rest of this paper is organized as follows. In Section 2, we will propose a generic model for multiple-hierarchies cognitive wireless networks named DCCM. In Section 3, the SORA will be introduced in detail. In Section 4, an intelligent wireless resource management system will be designed by the DCCM and implemented by the SORA. Finally, in Section 5, we conclude the paper.

2. Dual-Cycle Cognitive Model

2.1. Hierarchy of Cognitive Wireless Networks. In the long-term development on networking techniques, different types of networks come forth. They may adopt different access technologies, different hardware, and different network operators and lead to heterogeneous architectures. Fortunately, cognitive radio technologies provide compelling solutions, with which the networks could connect to each other. Therefore, the cognitive wireless networks, including the technology and architecture, have become popular research topics.

Among the researches of cognitive wireless network architecture, E2R is famous for its end-to-end reconfigurability [12]. Its aim is to design, develop, and test the configurable terminal and their suspending system structures, furthermore, to provide a systemic network solution for users, service providers, and the network providers. In order to implement these goals, the E2R proposes reconfiguration management plane (RMP). The RMP includes three fundamental function entities, that is, the reconfiguration manager (RCM), radio reconfiguration support function (R-RSF), and user equipment reconfiguration support function (U-RSF). These entities are located distributedly in different hierarchies of the network management field.

In order to combine the heterogeneous networks, IEEE 1900.4 has been developed [13], where the distributed decision-making architecture is used to optimize the resource utilization in a "network-terminal" format. With the spectrum management system, two dominating entities are defined within the standard: network reconfiguration manager (NRM) and terminal reconfiguration manager (TRM), both of which are components of the composite wireless network (CWN) that cooperate with each other to manage the network.

In addition, there exist other different cognitive wireless network architectures, including the dynamic intelligent management of spectrum for ubiquitous mobile network (DIMSUMNet) [14], and the OverDRiVE designed by the European Union [15], and the cognitive wireless cloud (CWC) proposed by NICT [16].

Through the investigation of various cognitive wireless network architectures mentioned above, several common characteristics can be concluded as follows.

(1) The cognitive wireless networks all work in a hierarchical structure, which are abstracted as the following layers: core networks, access networks, and terminals. Function entities on different layers are considered with different operation characteristics: the management and control between the core networks and the access networks are coarse grained, of large time scale, and intelligent, while those between the access networks and the terminals are fine grained, of small time scale, and adaptive.

(2) Although the heterogeneity exists among different cognitive wireless networks, their cognition actions are almost similar and can be generally concluded to three phases, that is, sensing, decision, and action, which are regarded as the basic cycle behind all the cognitive wireless networks.

2.2. Dual-Cycle Cognitive Model. Obviously, according to the above analysis, OOPDAL loop or OODA loop cannot completely describe the behaviors of cognitive wireless networks. However, considering the structure of both loops, it is possible to combine them into a dual-cycle cognitive model (DCCM) that corresponds to the multiple-hierarchies structure of cognitive wireless networks.

Dual-cycle consists of the decision loop and the execution loop, where the former and latter loops are located on the up-level and the down-level of the cognitive wireless network, respectively. As shown in Figure 1, the dual-cycle cognition runs at the hierarchical network, while the core network is managed by the core level functional entity (C-FE), the access networks are managed by the radio access level functional entities (R-FE), and the terminals perform with the terminal level functional entities (T-FE).

The decision loop is responsible for global network management and control, aiming at the global target. Figure 2 illustrates the detailed phases of the decision loop, including observe, orient, plan, decide, and act components. Note that the learn component is different from that in Mitola's OOPDAL, due to its operation in an offline manner in a loose-coupling way. Thus, we name it OOPDAL loop.

Using the database, the decision loop is able to gain information of global environment under the constraint of the global target to make orientations and plans. The plans are accordant to the whole variation and global change to form an optimal strategy. The strategy provides the up-layered

C-FE: Core level functional entity
R-FE: Radio access level functional entity
T-FE: Terminal level functional entity

Figure 1: Illustration of the double-cycle mode.

Figure 2: Illustration of the decision loop.

adjustment and down-layered adjustment, while the upper-layered one is sent to the act component in this loop and the other is sent to the execution loop. As the background component, using the continuously updated database, the learn function renews the knowledge storage regularly.

The execution loop is responsible for local network management and control, aiming at the local target. Figure 3 illustrates the detailed phases of the execution loop, including observe, orient, decide, and act components. It receives the delivered policy as the local target and then makes the local plan with the observation of the local environment. Note that the execution loop is only performed within the local networks. After the loop operation is terminated, the local information is then sent to the upper decision loop.

The dual-cycle (decision and execution loops) configuration is tightly coupled, as shown in Figure 4. Since two loops manage different network parameters, take different network responsibility, and aim at different adjustment targets, from which we could consider them independent. In the other prospect, the execution loop sends local information to the decision loop, while the decision loop sends policy to the execution loop, and thus they are connected to shape a whole cognition process.

Figure 3: Illustration of the execution loop.

3. Service-Oriented Radio Architecture

Our proposed SORA is an open architecture, where the principles and methodologies of service orientation are

Figure 4: Illustration of the double-cycle relation.

considered to design and integrate radio communication systems. Using SORA, different radio function modules are encapsulated into the network-base components (namely, services). Since radio services are published and discovered over networks, users are able to invoke and orchestrate different required services to implement various radio communication applications.

Two key issues are considered to define the SORA. Firstly, SORA is a distributed and modular system architecture, where the radio function modules can be deployed as services over networks. Secondly, SORA provides an open architecture, where the radio services adopt open and standardized interface descriptions and interoperation contracts. Overall, SORA is able to provide radio systems a number of advantages, such as loose-coupling, reusability, scalability, reconfigurability, and the capability of cooperative development.

All these characteristics of SORA lead to the significant improvement of flexibility and compatibility during the processes of radio system design, development, and implementation, especially for cooperative research, test, and exploitation.

3.1. Framework. The framework of our proposed SORA has been expounded in [9], where three terms are considered, namely, radio infrastructures, services, and applications. Specifically, radio infrastructures provide the basic function of radios (e.g., encoder, modulator-demodulator (MODEM), RF module, antenna, etc.), services act the infrastructure functionalities (i.e., blocks of the radio system that can be assembled), and radio communication applications are implemented through design and integration of various services. The three terms of the framework are specified as follows.

(1) Radio infrastructures: it is known that different wireless networks (e.g., 2G/3G, Wi-Fi, and WiMAX) exist in the current wireless environment, simultaneously. In such an environment, the radio infrastructures can be characterized by the heterogeneity, as they could be based on different development platforms and may belong to different operators and networks. Three components are considered as the infrastructures, including antennas, RF modules, and multiple programmable processors.

(2) Services: using functionality description, different radio infrastructures are encapsulated and abstracted from the corresponding software/hardware platforms. The abstracted infrastructure is regarded as a service to be published, discovered, and invoked over networks. In our proposed architecture, the services are classified into four different categories, namely, waveform services, information security services, sensing services, and cognition services.

(3) Applications: through the service integration, a virtual radio is constructed to execute the different applications of radio communication.

According to the framework of our proposed SORA systems, we can show that the service layer plays a crucial role, since different radio infrastructures are encapsulated and abstracted as services to support the heterogeneity of underlying development platforms. Comparing to the traditional radio architectures, the services are deployed distributedly over different networks with our proposed SORA configuration, while a virtual radio system is constructed through integration of appropriate services to meet the requirements of different users. From this perspective, SORA is regarded as an agile virtualization method to assemble radio function modules in a network-based concept.

The implementation of SORA is based on a series of web service standards, including the web service description language (WSDL) for service description, the simple object access protocol (SOAP) for service interaction, and the universal description, discovery, and integration (UDDI) for service registration, publishing, discovery, and integration, respectively. The functions of these standards will not be introduced for simplicity, which have already been specified in [9].

3.2. Application Method. Three components are considered for SORA to construct the wireless communication systems, which are client, service provider, and registry. The client is regarded as a user who employs one or more services to implement a certain communication task according to its requirement. The service provider publishes various services into the registry in order to be discovered by the client. The registry acts as a public server for service publication and discovery. Through interactions among the above three components, an SORA-based wireless communication system can be constructed. The process of the construction is usually divided into four steps, including the service encapsulation, publication, discovery, and orchestration.

(1) Service encapsulation: all the service providers need to encapsulate their radio function modules as the invokable web services. Then, the WSDL documents need to be generated to define relevant service functionalities.

(2) Service publication: the WSDL documents are published in a UDDI service registry over the IP core network, where the registry is developed by a third party for public service publication and discovery.

(3) Service discovery: the client accesses to the service registry by utilizing the SOAP messages to find out the required services in the service list. Correspondingly, the locations, interfaces, and operations of different services are discovered using the WSDL documents.

(4) Service orchestration: according to the order of the signal processing in the communication system, different discovered services are orchestrated by the client. Then, the services are invoked through the SOAP messages, where a virtual SORA-based CR access network is finally constructed.

4. Intelligent Wireless Resource Management

Complicated heterogeneous wireless networks have posed big challenges to wireless resource management. Firstly, dynamic wireless environment requires the wireless communication systems possess intelligence, where the capabilities of sensing, reconfiguration, and learning are taken into account. Secondly, multiple-hierarchies-based structure leads the management process to be even more complex, where a suitable model is desired to guide the system design. Moreover, a flexible implementation architecture is needed to integrate different components of heterogeneous systems with insurable interactions. Considering the above problems, the proposed DCCM and SORA are carried out to design and implement the wireless resource management systems in heterogeneous environments.

4.1. DCCM-Based System Design. Note that the DCCM has been proposed in Section 2. In this section, it is used to design an intelligent wireless resource management system. According to the DCCM, the intelligent wireless resource management system could also be divided into three layers corresponding to the three hierarchies of cognitive wireless networks, that is, core layer, access layer, and terminal layer. The core layer includes two entities: advanced radio resource management (ARRM) and cognitive engine (CE). The access layer includes the local radio resource management (LRRM) entity, while the terminal layer includes the terminal decision management (TDM) entity. In Figure 5, the dash line represents data information flow, while the full line denotes the management and control flow.

Using the knowledge obtained from the CE, the ARRM provides different access strategies for all the managed access networks. These strategies are then sent to the LRRM and TDM in the form of advice. LRRM mainly allocates the spatial, the temporal, and spectral resources based on the received strategies. In addition, TDM performs the decision and reconfiguration based on the LRRM and ARRM.

The cognitive loop in the scenario acts as a regular and steady manner. The corresponding working flow is shown in Figure 6. In the figure, steps 1 to 14 are explained as follows.

(1) LRRM1 reports the load info to the CE.

(2) LRRM2 reports the load info to the CE.

(3) LRRM1 reports the load info to the ARRM and submit handover request.

(4) ARRM requests for other RANs' load info from CE.

(5) CE replies for the request.

(6) ARRM requests for the user's preference info from CE.

(7) CE replies for the request.

(8) ARRM requests for policy info from CE.

(9) CE replies for the request.

(10) ARRM delivers the handover message to LRRM1.

(11) ARRM delivers the handover message to LRRM2.

(12) ARRM delivers the handover message to TDM.

--> Data info
⟶ m and c info
ARRM: Advanced radio resource management
ICE: Information collection and extraction
PAD: Policy assembly derivation
PEE: Policy efficiency evaluation
RANS: Radio access network selection
PB: Policy base
CE: Cognitive engine
DB: Data base
KB: Knowledge base
LRRM: Local radio resource management
RANM: Radio access network measurement
RRA: Radio resource allocation
TDM: Terminal decision management
TIC: Terminal information collection
TR: Terminal reconfiguration

Figure 5: Illustration of cognitive network architecture.

(13) TDM replies with the channel info and user info to LRRM2.

(14) TDM replies with QoE to CE.

And the seven operations are explained as follows.

(a) Establish the RAN2 and 3 as the target switch network.

(b) Some users in RAN1 are switched to RAN2.

(c) Permit the handover from the RAN1 to RAN2, and generate the handover strategy.

(d) Permit to access.

(e) Channel allocation.

(f) Waveform reconfiguration.

(g) Learn.

4.2. SORA-Based Implementation. Next, with an example provided, we illustrate how to use SORA to implement the intelligent wireless resource management system. The scenario of the example is shown in Figure 7. In this scenario, the whole network consists of five parts, including two operators, a research institute, a service provider, and a service registry, which are deployed over wide areas. It shows that each part is considered with an individual network connected to an IP core network. Two radio access networks (RANs), that is, RAN1 and RAN2, are carried out with Operator A, while another radio access network RAN3 is carried out with Operator B. Note that each of the RAN includes the relevant LRRM entities. It is also known that the service provider is used to provide ARRM service, while the CE services (including learning service, data base, and knowledge base) are provided by the research institute. As a result, the three RANs form a heterogeneous wireless network, where the various services are integrated to manage the wireless resources by SORA. Using the above architecture, the heterogeneity of the network is able to be overcome, where the management information can be interacted without any compilation. Thus, the flexibility and reusability of the system can be greatly improved.

The corresponding process of service integration can be divided into three steps, including service publication, service discovery, and service invocation, which are summarized as follows.

(1) Service publication: at first, the service provider and research institute encapsulate their entities, including the ARRM and CE, as web services. Then, the WSDL documents of each service are generated and published in the service registry for discovery.

(2) Service discovery: the LRRMs in different RANs as clients access to the service registry to find out the needed services in the service list. Correspondingly, the locations, interfaces, and operations of different services are discovered using the WSDL documents.

(3) Service invocation: according to the order of the working flow mentioned above, different discovered services are orchestrated by the clients. Then, the services are invoked through the SOAP messages, where a virtual SORA-based wireless resource management system is finally constructed.

4.3. Simulation. In this subsection, the quality of experience-(QoE-) [17] based simulation is carried out to demonstrate the advantages of the proposed DCCM and SORA in wireless resource management. It shows that the QoE can be significantly improved by using our proposed architecture. Note that three heterogeneous wireless access networks (i.e., RAN1,

FIGURE 6: DCCM working flow.

FIGURE 7: Illustration of SORA-based implementation method.

Figure 8: Simulation scenario.

RAN2, and RAN3) are considered in the proposed structure, where multiple terminals with cognitive ability can access to different wireless access networks due to networks' dynamic change and the needs of users, as shown in Figure 8.

As indicated in Figure 8, different wireless resources are assigned to different RANs, while the spectrum utilization efficiency of a certain RAN is considerably high (i.e., F3 for RAN3), while that of others is low (i.e., F1 and F2 for RAN1 and RAN2, resp.). Therefore, in this scenario, the main goal of wireless resources management is to balance the load of the wireless access network, where the spectrum utilization can be improved to ensure good QoE for users. According to the goal, a load balancing algorithm is considered to guarantee the QoE for different users.

QoE refers to the users' comprehensive feeling with respect to the equipments, networks, and systems. For heterogeneous wireless networks, users may present different preferences for different network. For example, some users target on security mobile networks, while other users prefer low cost of broadband access networks. If the system has been developed with the cognitive structure, it could be able to adopt the method of artificial intelligence to predict subjective preference of the users with different strategies.

According to the above scenario, we assume that the heterogeneous wireless network environment contains I wireless access networks with different air interface technologies. The wireless resources of downlink and uplink for each RAN are limited to the upper bounds ($S_{\mathrm{DL,max}}(i)$ and $S_{\mathrm{UL,max}}(i)$), respectively, which are related to the time slots ($t_{\mathrm{DL}}(i)$ and $t_{\mathrm{UL}}(i)$). The access services are provided to J cognitive terminals, where L wireless access modes are considered for each cognitive terminal. Note that different terminal requires different service QoS. To guarantee the QoS, the adaptive modulation and coding modes are adopted to satisfy the data rates R_{DL} and R_{UL} according to the downlink and uplink SNRs ($z_{\mathrm{DL}}(i,j)$ and $z_{\mathrm{UL}}(i,j)$).

The QoE guaranteeing load balancing algorithm includes two aspects: (1) the requirements of QoS, where the minimum average data rates ($R_{\mathrm{DL,min}}(j)$ and $R_{\mathrm{UL,min}}(j)$) are mainly considered; (2) the subjective preference of the users, which is forecasted according to the preference knowledge ($K(i,j)$) acquired from the CE. The preference knowledge, which consists of a series of QoE scores, indicates a user's preferences for different networks. When the user prefers a network more, the relevant QoE score is higher. When we calculate the average QoE score, the scores of the chosen networks for each user are accumulated.

The algorithm is carried out to perform the network throughput optimization, which is presented as follows:

$$\{\delta_{\mathrm{DL}}(i,j), \delta_{\mathrm{UL}}(i,j), s_{\mathrm{DL}}(i,j), s_{\mathrm{UL}}(i,j)\} = \arg\max[\Omega]. \tag{1}$$

$$\Omega = \Omega_{\mathrm{DL}} + \Omega_{\mathrm{UL}}, \tag{2}$$

In (1), we have

$$\Omega_{\mathrm{DL}} = \frac{1}{t_{\mathrm{DL}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \delta_{\mathrm{DL}}(i,j)\, s_{\mathrm{DL}}(i,j)\, R_{\mathrm{DL}}(z_{\mathrm{DL}}(i,j)), \tag{3}$$

$$\Omega_{\mathrm{UL}} = \frac{1}{t_{\mathrm{UL}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \delta_{\mathrm{UL}}(i,j)\, s_{\mathrm{UL}}(i,j)\, R_{\mathrm{UL}}(z_{\mathrm{UL}}(i,j)) \tag{4}$$

s.t.,

$$\sum_{j=1}^{J} \delta_{\mathrm{DL}}(i,j)\, s_{\mathrm{DL}}(i,j) \le S_{\mathrm{DL,max}}(i), \quad i = 1,\ldots,I, \tag{5}$$

$$\sum_{j=1}^{J} \delta_{\mathrm{UL}}(i,j)\, s_{\mathrm{UL}}(i,j) \le S_{\mathrm{UL,max}}(i), \quad i = 1,\ldots,I, \tag{6}$$

$$s_{\mathrm{DL}}(i,j)\, R_{\mathrm{DL}}(z_{\mathrm{DL}}(i,j)) \ge t_{\mathrm{DL}} R_{\mathrm{DL,max}}(j), \quad j = 1,\ldots,J, \tag{7}$$

$$s_{\mathrm{UL}}(i,j)\, R_{\mathrm{UL}}(z_{\mathrm{UL}}(i,j)) \ge t_{\mathrm{UL}} R_{\mathrm{UL,max}}(j), \quad j = 1,\ldots,J, \tag{8}$$

$$\delta_{\mathrm{DL}}(i,j)\, K(i,j) \ge m, \quad j = 1,\ldots,J, \tag{9}$$

$$\delta_{\mathrm{UL}}(i,j)\, K(i,j) \ge m, \quad j = 1,\ldots,J. \tag{10}$$

In particular, (1)–(10) are illustrated as follows.

(1) Ω is the network throughput; Ω_{DL} and Ω_{UL} are downlink and uplink throughput, respectively.

(2) $s_{\mathrm{DL}}(i,j)$ is the resource of network i that allocates to terminal j in one downlink frame.

(3) $s_{\mathrm{UL}}(i,j)$ is the resource of network i that allocates to terminal j in one uplink frame.

(4) $\delta_{\mathrm{DL}}(i,j)$ indicates the downlink of terminal j that accesses to network i. "1" means "yes," and "0" means "no."

(5) $\delta_{\mathrm{UL}}(i,j)$ indicates the uplink of terminal j that accesses to network i. "1" means "yes," and "0" means "no."

(6) $K(i,j)$ represents the user preference degree of terminal j for network i, where the value ranges from 0 to 1.

Figure 9: Network throughput changes with number of users.

Figure 10: The average users' QoE scores change with the number of users.

(7) Formulas (5) and (6) represent the resource limits in downlink and uplink, respectively.

(8) Formulas (7) and (8) represent the guarantee of QoS requirements.

(9) Formulas (9) and (10) represent the satisfaction of users' preferences for networks.

Note that the proposed load balancing algorithm is used to optimize the network throughput by selecting the access networks for different users. Since the user preferences are considered in the network optimization, the QoE of users can be improved. Using the above algorithm, we compare our proposed QoE-based load balancing strategy to the conventional SNR-based switching strategy and the QoS-based load balancing strategy [8] in terms of the overall performance.

In Figure 9, simulation results show the network throughput versus the number of users using different strategies. It can be observed that the QoE- and QoS-based strategies outperform the SNR-based ones in terms of the network throughput. Under the condition that the number of users is small, QoE- and QoS-based strategies provide similar performance. Note that when a large number of users are considered (e.g., 60–100), comparing to the performance of the QoS-based strategy, the throughput of our proposed QoE-based one is slightly lower because one has to sacrifice the network performance to meet proper QoEs for different users. However, it will be shown in Figure 10 that our proposed strategy is able to provide strongly improved performance to users in terms of the QoE comparing to that of the QoS-based strategy.

In Figure 10, the simulation results show the average users' QoE scores versus the number of users with the employments of different strategies. It can be demonstrated that the average QoE of our proposed scheme becomes much higher than that of the other two strategies. In addition, we can show that neither the SNR nor QoS strategies is able to provide reasonably good QoE performance to users. Consequently, in order to make users more satisfactory, our proposed QoE-based load balancing strategy is more desirable. In this simulation, all the strategies adopt the same network deployment and user preference knowledge. Therefore, when we average the QoE scores through many simulations, the simulation results without consideration of users' preferences are almost the same.

5. Conclusion

In this paper, we have proposed the double-cycle cognitive model and the service-oriented radio architecture for designing and implementing the cognitive wireless networks. We have first proposed the DCCM through analysis of cognitive wireless network architecture. Then, we have introduced the framework and application method of our SORA. Finally, we have employed the DCCM to design an intelligent wireless resource management system, used the SORA to implement this system for higher flexibility, and improved the QoE based on the proposed architecture.

In conclusion, using the idea of M2M technology, our proposed strategy could make the heterogeneous components of cognitive wireless network loosely coupled with each other. Thus, the whole network can be integrated and implemented more compatibly and become more intelligent.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments that have helped improving the overall quality of the paper. This work has been supported by the National Natural Science Foundation of China (Grants nos. 61231011 and 61250001) and the China National 973 project (Grant no. 2009CB320403).

References

[1] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[2] J. R. Byod, "The essence of winning and losing," 1987.

[3] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, "Cognitive networks: adaptation and learning to achieve end-to-end performance objectives," *IEEE Communications Magazine*, vol. 44, no. 12, pp. 51–57, 2006.

[4] P. Balamuralidhar and R. Prasad, "A context driven architecture for cognitive radio nodes," *Wireless Personal Communications*, vol. 45, no. 3, pp. 423–434, 2008.

[5] C. Fortuna and M. Mohorcic, "Trends in the development of communication networks: cognitive networks," *Computer Networks*, vol. 53, no. 9, pp. 1354–1376, 2009.

[6] Software CommunicationsArchitecture Specification (v2.2), http://jtnc.mil/sca/Documents/SRD_Release_V2.2.pdf.

[7] P. Pawelczak, K. Nolan, L. Doyle, S. Oh, and D. Cabric, "Cognitive radio: ten years of experimentation and development," *IEEE Communications Magazine*, vol. 49, no. 3, pp. 90–100, 2011.

[8] COST Action IC0902, http://newyork.ing.uniroma1.it/IC0902/.

[9] X. Dong, S. Q. Wei, Y. Li et al., "Service-oriented radio architecture: a novel M2M network architecture for cognitive radio systems," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 762953, 8 pages, 2012.

[10] Service Oriented Architecture, http://en.wikipedia.org/wiki/Service-oriented_architecture.

[11] Virginia Tech Cognitive Radio Open Source System, http://ossie.wireless.vt.edu/trac/wiki/Cross.

[12] P. Zhang and Z. Y. Feng, *Cognitive Radio Networks*, Science Press, 2010.

[13] S. Buljore, H. Harada, S. Filin et al., "Architecture and enablers for optimized radio resource usage in heterogeneous wireless access networks: the IEEE 1900.4 working group," *IEEE Communications Magazine*, vol. 47, pp. 122–129.

[14] M. M. Buddhikot, P. Kolodzy, S. Miller et al., "DIMSUMNet: directions in wireless networking using coordinated dynamic spectrum access," in *Proceedings of the 6th IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks*, pp. 78–85, 2005.

[15] R. Tonjes, K. Moessner, T. Lohmar et al., "OverDRiVE: spectrum efficient multicast services to vehicles," in *IST Mobile and Wireless Telecommunications Summit*, pp. 473–477, 2002.

[16] P. Leaves, K. Moessner, R. Tafazolli et al., "Dynamic spectrum allocation in composite reconfigurable wireless networks," *IEEE Communications Magazine*, vol. 42, no. 5, pp. 72–81, 2004.

[17] The 3rd Generation Partnership Project, "Transparent end-to-end packet-switched streaming service (PSS): protocols and codes (release 9), 3GPP TS 26.324 v9.2.0," 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 946931, 10 pages
http://dx.doi.org/10.1155/2013/946931

Research Article

Design and Evaluation of a Framework for Cooperative and Adaptive QoS Control of DSRC Network for Road Safety Applications

Wenyang Guan,[1] Jianhua He,[2] Zuoyin Tang,[2] and Thomas M. Chen[3]

[1] *School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China*
[2] *School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, UK*
[3] *City University London, London EC1V 0HB, UK*

Correspondence should be addressed to Jianhua He; j.he7@aston.ac.uk

Received 29 August 2013; Accepted 17 October 2013

Academic Editor: Lin Bai

Copyright © 2013 Wenyang Guan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dedicated short-range communications (DSRC) are a promising vehicle communication technique for collaborative road safety applications (CSA). However, road safety applications require highly reliable and timely wireless communications, which present big challenges to DSRC based vehicle networks on effective and robust quality of services (QoS) provisioning due to the random channel access method applied in the DSRC technique. In this paper we examine the QoS control problem for CSA in the DSRC based vehicle networks and presented an overview of the research work towards the QoS control problem. After an analysis of the system application requirements and the DSRC vehicle network features, we propose a framework for cooperative and adaptive QoS control, which is believed to be a key for the success of DSRC on supporting effective collaborative road safety applications. A core design in the proposed QoS control framework is that network feedback and cross-layer design are employed to collaboratively achieve targeted QoS. A design example of cooperative and adaptive rate control scheme is implemented and evaluated, with objective of illustrating the key ideas in the framework. Simulation results demonstrate the effectiveness of proposed rate control schemes in providing highly available and reliable channel for emergency safety messages.

1. Introduction

Road traffic safety causes huge economic and productivity loss every year. During the last decade, road safety systems have been extensively studied to actively prevent accidents or passively minimize the consequences of accidents. With significant advances in wireless communications and mobile ad hoc networking [1], vehicular network based collaborative safety applications (CSA) have been widely considered as a key solution to road safety problems. With assistance of vehicle to vehicle (V2V) and/or vehicle to infrastructure (V2I) communications collaborative safe driving may be achieved based on the obtained complex traffic and road situation information.

There could be a number of ways to support V2V communications, for example, by direct short-range radio communication or through satellites or cellular base stations. An example of vehicle network built on both direct and indirect V2V communications is shown in Figure 1, in which direct V2V communications are enabled with 802.11-like short-range radio for vehicles within a local vicinity. Vehicles equipped with short-range radio can also communicate with a roadside unit (RSU) in range. Alternatively, indirect communications can be provided by commercial communication systems, for example, wide area networks (WANs) using satellite or 2 G/3 G cellular base stations (BS) or metropolitan area networks (MANs) such as WiMAX (IEEE 802.16) [2–4]. Considering that useful safety information is usually limited to the local area around a vehicle and may be exchanged frequently, direct V2V communications are much more cost-effective compared to indirect V2V communications for large-scale vehicle safety applications. Messages exchanged by indirect V2V communications encounter unnecessary

Figure 1: An example of vehicle networks utilizing hybrid vehicle communication technologies.

and potentially excessive delays, which makes indirect V2V communications not the best candidate for real-time safety applications. On the other hand, direct V2V communications will outperform for CSA in terms of both communication performances and cost.

Among direct V2V communication technologies, dedicated short-range communications (DSRC) have been regarded as the most promising technology [5, 6]. It can complement cellular communications by providing very high data transfer rates, especially important when the applications require low communication latency and relatively small communication zones. The technology is expected to be robust and affordable for large-scale vehicle networks [5–7]. DSRC has been designed to provide both road safety (e.g., collaborative collision warning and collaborative collision avoidance) and commercial services (e.g., navigation, map and Internet access) [8–11]. The main issues related to the DSRC vehicular networks have been focused on channel model and mobility models [10], routing [12], broadcasting [13, 14], security and privacy, QoS support [15, 16], and system performance evaluation by analytical, simulation, and field test approaches [17–21]. With regard to collaborative safety applications, more concerned issues are broadcast schemes, power control and congestion control, and QoS provisioning. Compared to nonsafety applications, road safety applications are known to have strict requirements on reliable and real-time message delivery. Excessive message delay or message loss can cause unexpected negative consequences. The challenge coming from the collaborative safety applications is exaggerated by the fact that the MAC layer of DSRC technology is based on IEEE 802.11p distributed coordination function (DCF) [22]. With the random channel access specified in the IEEE 802.11 MAC it is hard to achieve deterministic QoS guarantee, even with the enhanced QoS service differentiation mechanisms in IEEE 802.11e [18, 23].

Although a lot of research efforts have been contributed to the individual research problems mentioned above for collaborative safety applications, it is noted that from the system point of view, there are still many research issues to be solved for comprehensive QoS control of vehicle networks before DSRC network based CSA can be assured to be practically useful. These research issues include but are not limited to the following list: (1) What level of QoS can be realized for DSRC in dynamic vehicle networks? (2) How to adaptively control the QoS of DSRC-based vehicle networks for CSA? (3) How effective are the QoS assurances provided by vehicle networks to CSA services? The first question has been widely studied in the literature by simulations and field tests. Those results show that the QoS assurances in DSRC vehicle networks can be a problem under stressful traffic conditions. The second and the third ones are still open issues. To the best of our knowledge, no effective solution has been reported yet for systematic QoS control of DSRC vehicle networks for CSA. In this paper, we are motivated to propose a QoS control framework for DSRC vehicle networks, aiming to provide some insights into the design of effective QoS support schemes for CSA. Next We briefly overview CSA applications and DSRC technology in Section 2. The challenges faced by QoS control in DSRC vehicle networks are discussed in Section 3. Section 4 presents a QoS control framework which exploits feedback and cross-layer design. A design example of adaptive congestion control is presented in Section 5. Preliminary results are presented and discussed in Section 6.

2. Overview of Vehicle Safety Applications and DSRC

2.1. Typical Vehicle Safety Applications and Communication Requirements. The QoS provided by DSRC networks should be viewed within the context of the vehicle safety applications under consideration. It is expected that a wide range of applications including both public safety and private operations will be deployed in order to justify the costs of installing the DSRC hardware. Recently Vehicle Safety Communications Consortium participants identified 34 safety applications that may benefit or be enabled by V2V or V2R communications. These safety applications were ranked according to the estimated benefits, effectiveness, and market penetration. Below are the applications with the highest estimated benefit opportunity for the fifth year after deployment.

(i) Near-term: traffic signal violation warning; curve speed warning; emergency electronic brake lights.

(ii) Mid-term: precrash sensing; cooperative forward collision warning; left turn assistant; lane change warning; stop sign movement assistant.

(iii) Long-term: cooperative collision warning; intersection collision warning.

In general safety applications have different communication patterns. For example, safety related message transmissions can be broadcast or unicast. The safety messages may be event-driven (triggered by an event) or periodic (sent automatically at regular intervals). Periodic safety applications (PSA) may transmit safety messages with different frequencies.

The QoS parameters required for safety applications may be varied as well. QoS parameters used widely to evaluate wireless communications for vehicle safety applications include maximum required communication ranges, minimum throughput, maximum allowable latency, and allowable message failure rate. The maximum required communication range is the effective communication distance for a transmission to support a particular application (e.g., 300 m). Maximum allowable latency is an upper bound on the tolerable delay from the instant that a message becomes available for transmission to the instant it is received at the peer safety application layer (e.g., 30 millisecond).

2.2. Multichannel Operation. Next we give a brief overview of the multichannel operation mechanism specified in the IEEE WAVE system for DSRC. The overall WAVE architecture includes IEEE 802.11p MAC and PHY standards and IEEE 1609.1 to 1609.4 standards (for resource management, security architecture, networking service, and multichannel operation, resp.). IEEE 802.11p uses essentially the same physical layer as 802.11a but operates in a 10 MHz wide channel instead of 20 MHz.

Multichannel operation is a distinctive feature of DSRC technology. In North America, the spectrum allocated to DSRC services in the 5.9 GHz licensed band is divided into seven channels, as shown in Figure 2(a). Channel 178 is the control channel (CCH) used exclusively for road safety messages and service announcements. The others are service channels (SCH), with Channel 172 dedicated for V2V communications and Channel 184 dedicated for intersections applications. SCH 174 and 176 are shared by medium-range public safety and private services, while SCH 180 and 182 are shared by short-range public safety and private services.

To enable multichannel coordination the IEEE 802.1609 standard for a WAVE system defines functional extensions to the IEEE 802.11 standard. The coexistence of safety and nonsafety applications is achieved through time division multiplexing (TDM). A synchronization procedure has been proposed for TDM channel coordination in the application level, for devices incapable of simultaneously monitoring the CCH and exchanging data on SCHs. In the procedure, a synchronization interval is defined which consists of a CCH interval, a SCH interval, and a guard interval, as shown in Figure 2(b). All WAVE devices need to monitor the CCH during the CCH interval. At the beginning of each scheduled channel interval, a guard interval is used to account for variations in the channel interval and timing inaccuracies. Upon startup a device monitors the CCH until it receives an announcement of service that utilizes an SCH, or it chooses to use the SCH based on received WAVE announcement frames. Clearly, the configuration of CCH interval and SCH interval has a significant impact on the QoS for safety applications as well as on the availability of spectrum resource for nonsafety applications.

3. QoS Control Issues for DSRC Networks

3.1. Research Challenges. The common safety messages for vehicle safety applications are exchanged mainly in the control channel. Here we focus on the QoS control for the safety applications over the control channel in DSRC networks. Besides the unreliable wireless channel and multichannel operation, there are many other challenges faced by the QoS control of DSRC networks, which are listed below.

3.1.1. Dynamic Road Traffic. A vehicle may be presented with different road traffic, urban intersections or highway, at different times. Vehicle traffic is hard to predict and can generate dynamic application traffic. The dynamic safety traffic can overwhelm the control channel in the absence of traffic controls. The DSRC protocols and parameter configurations should take the worse scenarios into account, but too conservative design and fixed configurations will make overall system performance suboptimal.

3.1.2. Distributed Operation. Most of the safety applications are based on direct peer-to-peer and broadcast communications. The communication operations are entirely distributed without centralized coordination or control. To prevent broadcast storms, road safety applications are designed to be noninteractive and scalable. The distributed, noninteractive nature of operations increases the difficulties of QoS control in a dynamic environment.

3.1.3. Multiple Safety Applications with Diverse and Strict QoS Requirements. We mentioned earlier that vehicle safety applications have diverse and strict QoS requirements. Providing multiple class QoS has been a trouble for wired networks for long time and is not to be easy for dynamic and capacity limited vehicle networks. Network protocols need to be adaptive tuned in order to maximize network throughput while ensuring QoS for safety applications, which further requires efficient feedback on the safety application performances. However, as safety messages are broadcasted, it is unlikely to receive quick feedback on the message delivery. Event-driven safety messages are generated sporadically, which makes the issue of statistic QoS performance feedback worse. Therefore, providing diverse QoS at the MAC layer is very challenging.

3.1.4. Distributed Congestion Control. Critical emergency safety messages require a control channel with high availability and low latency. However, channels are shared by emergency safety messages and periodic safety messages.

(a) DSRC frequency plan

(b) Synchronization interval

Figure 2: DSRC frequency plan and channel synchronization.

Although supported by enhanced MAC QoS mechanisms, emergency safety messages may receive unexpectedly low QoS performances under heavy traffic conditions. Periodic safety messages need to be regulated to prevent from saturating the channel but at the same time should not be controlled too aggressively because they are essential for building mutual awareness. How to fairly and efficiently control congestion in a distributed way is very challenging.

3.1.5. Dynamic Network Topology. Due to high speed movement of vehicles, connectivity of a vehicle with its neighbors can be temporary and very short. The short connectivity makes it very challenging for vehicles to maintain stable relationships and get useful feedback of message delivery performance (e.g., message reception rate and latency).

3.1.6. Random Channel Access. The simple and robust random channel access mechanism specified in the IEEE 802.11 standards is one of the reasons for its huge market success. The random channel access mechanism continues to be the foundation of DSRC networks, but it lacks the capability to support QoS guarantees for emergency safety applications. There are many interrelated factors to be considered in the MAC layer for QoS control, for example, how to determine the transmit power to ensure a certain communication range and successful message reception ratio without generating excessive interference.

3.2. Related Work. QoS control has been extensively studied for cellular networks and general mobile ad hoc networks (MANETs). The most common control mechanisms include call/connection admission control, QoS routing, and message scheduling. However, DSRC network based vehicle safety applications are mainly characterized by broadcast communications, ad hoc networking, and strict QoS requirements. These unique features make the QoS control schemes proposed for traditional cellular networks and MANETs not direct solutions to QoS control for vehicle safety applications. QoS control for DSRC based vehicle safety application is still a relatively new research area. It can be achieved through the power control, data rate control, and reliability control. Next we review some initial research efforts towards adaptive power control and congestion control for DSRC networks.

3.2.1. Feedback-Based Power Control. Guan et al. proposed a feedback-based power control algorithm for vehicular ad hoc networks [24]. The main idea is that given a target communication range designated by a vehicle safety application, the power control algorithm should select a transmission power no greater than needed for the target range. The vehicles located beyond the target range of a message selectively send feedback to the message's sender. The sender counts the number of feedbacks received for that message and compares the number to a predefined threshold. If the number is larger than the threshold, then the transmission power is incremented by a fixed step; otherwise, it is decremented by a fixed step. The idea is simple but the main focus is to reduce power consumption. The algorithm has not considered congestion control and QoS issues. An issue of instability can arise when the channel becomes congested. Under channel congestion, the number of received feedbacks can be small due to message collisions. But the transmit power will be increased according to the power control algorithm, which in turn causes heavier channel congestion.

3.2.2. Collaborative Power Control. Mittag et al. proposed a distributed power control scheme for vehicle safety applications [25]. The design aims to ensure strict fairness for the control of periodic safety messages and leave more space for safety-critical messages. The scheme consists of three main steps.

(i) Each vehicle creates a list for vehicles in the carrier sense range by exchanging periodic messages.

(ii) Each vehicle uses a theoretic model to independently determine the maximal transmit power that can satisfy the minimal channel load requirement for all the vehicles in the list.

(iii) The theoretically determined maximal transmit power is exchanged among vehicles within the transmission range. Each vehicle sets its transmit power to the minimal among the transmit powers determined theoretically by the vehicles in the transmission range.

The proposed power control scheme is collaborative and offers a high level of fairness. But a serious problem is that the maximal transmit power is determined based on theoretically computed channel load and channel load threshold. If the actual channel load is heavy, channel congestion will occur and continue on as no reaction to actual channel congestion has been considered. Another problem is that convergence

speed of the proposed congestion control method can be very low in a dynamic network environment.

3.2.3. MAC Layer Blocking. To increase the responsiveness to channel congestion and provide high channel availability for safety-critical messages, an optional mechanism for controlling transmission rate has been proposed at the MAC layer in the IEEE 802.11p specification [22]. A vehicle keeps monitoring the channel. If the sensed channel occupancy time in a monitoring window is higher than 50%, low priority PSA message transmission attempts are blocked. The drawback of this so-called MAC layer blocking mechanism is that the MAC layer supports either full or zero transmission rate. Also, the conditions to unblock PSA messages at MAC layer are not discussed. The MAC layer blocking mechanism is entirely node-centric, which can not efficiently solve the congestion control problem without vehicle collaboration.

4. A Framework for Cooperative and Adaptive QoS Control

There are many system parameters that have significant impact on the QoS performance provided to vehicle safety applications, for example, transmit power, message rate, CCH interval, and broadcast message redundancy. To date the reported QoS control for DSRC vehicle based safety applications has not been efficient. An effective framework for QoS control of DSRC vehicle networks is highly needed to bring the CSA closer to practice. That framework should allow control of system parameters systematically or individually. It should also effectively cope with dynamic vehicle scenarios and maximize the QoS performance for vehicle safety applications. We propose a QoS control framework of DSRC vehicle networks for collaborative road safety applications over the control channel [26]. It is shown in Figure 3 with the following design goals:

(i) assured high QoS for safety-critical messages;

(ii) adaptivity to traffic variations;

(iii) fair channel access;

(iv) high spectrum utilization.

The keys in the QoS control framework to achieve the goals are systematic and cross-layer design, vehicle cooperation, and QoS adaption. As single-layer based QoS control is unlikely to successfully provide QoS for CSA, we exploit systematic and cross-layer design. Vertically the QoS control functions are implemented over three layers, namely, application layer, transport layer, and MAC layer. Horizontally vehicles cooperate with their neighbors. Feedback and channel/QoS monitoring are used for QoS adaptation in order to provide required QoS and high spectrum utilization. Three main modules are proposed in the QoS control framework with each locating in the three layers to achieve the QoS control functions: adaptive safety applications (ASA) module at the application layer, cooperation and adaptation (CAA) module at transport layer, and MAC layer message rate control (MMRC) module at MAC layer. The detailed functions

Figure 3: QoS control framework.

and interfaces of the modules are explained in the following subsections.

4.1. Adaptive Safety Applications. The ASA module is responsible for three functions: feedback control, QoS monitoring, and message rate adaptation. Cooperation among the vehicles and adaptation of system parameters are key design considerations. Adaptivity is highly dependent on availability and accuracy of network and channel condition information. Through feedback and channel monitoring, a vehicle can calculate an approximate estimate of surrounding radio and network conditions. Each vehicle can easily monitor the channel load and the transmission activities of neighboring vehicles. But it still needs feedback from neighboring vehicles to make right adaptation and cooperation decisions. As safety applications are normally broadcast and noninteractive, it is not straightforward to provide feedback at MAC layer without significant modification on the 802.11 MAC layer specification. On the other hand, vehicle safety applications generate safety messages and request QoS from the underlying networks. The ASA module can easily provide the feedback on the received QoS for safety applications and is the ideal place to implement feedback control function.

With regard to feedback control, a device can provide feedback about the transmissions from an individual neighboring vehicle (e.g., statistical successful message receptions) as well as transmissions from all the surrounding vehicles (e.g., measured channel load). Feedback can be requested actively, or periodic feedback can be waited passively from other vehicles. Feedback can be sent in stand-alone messages or piggybacked on periodic safety messages. An issue is the trade-off between the overheads generated by feedback messages and the completeness and timeliness of feedback information.

The QoS monitoring function in the ASA module is responsible for measuring the actual QoS (such as latency and reception rate) seen by the vehicle safety applications. QoS monitoring should be performed for a relatively long time to make the monitoring results statistically meaningful. If the measured QoS performance is lower than required, the ASA module notifies the CAA module to take necessary actions such as traffic rate control and power control. Additionally the QoS measurement data are sent to neighboring vehicles upon request.

The message rate adaptation function in the ASA module is fairly simple. It informs the CAA module about the range of message rates for different vehicle safety applications. It is a task of the CAA module to determine proper message rates according to safety application requirements and channel congestion levels. The determined message rate is then sent back to the ASA module for use by vehicle safety applications for rate adaptation.

4.2. Cooperation and Adaptation. Among the three modules in the proposed QoS control framework, the CAA module plays a crucial role because it interacts with both ASA and MMRC modules. The CAA module is responsible for most of the network QoS control functions, for example, message rate control, transmit power control, selection of data rate, and message redundancy mode. Traffic rate, transmit power, data rate, and message redundancy can significantly affect the QoS seen by safety messages. Implementing these functions in the CAA module together can promote intelligent designs.

Interactions between the CAA module and the ASA module have been mentioned in Section 4.1, mainly on the exchange of feedback, QoS, and message rate information. The CAA module also has interactions with the MMRC module. As to be introduced in Section 4.3, the MMRC module is responsible for fast reaction to temporary channel congestion by blocking PSA messages at the MAC layer. Once the percentage of channel occupancy time is monitored to be more than a threshold (e.g., 50%), the MMRC module informs the CAA module about the temporary blocking of PSA messages at the MAC layer. The MMRC module also takes necessary actions if it is requested by the CAA module in case of low QoS being observed at the ASA module.

Message rate control and power control are mainly used to address channel congestion in the CAA module. Unlike the very simple message blocking function in the MMRC module, Both rate control and power control at the CAA module are much more dynamic and complex with multiple levels. According to the monitored QoS and channel load levels, the CAA module can cooperatively and jointly adjusts the power level, message rate, and other QoS control measures. A design example of cooperative and adaptive rate control will be presented in Section 5.

4.3. MAC Layer Message Rate Control. The MMRC module is located at the MAC layer to provide fast congestion mitigation. If a congestion event occurs, low priority PSA messages are blocked immediately at the MAC layer for a certain time. Message rate control at the MAC layer is quite simple with only two states, *unblocked* and *blocked*. In the *unblocked* state, all messages entering the MAC queue will attempt transmissions. In the *blocked* state, low priority PSA messages are transmitted at a low rate to update statuses (e.g., congestion state, position, speed, direction, etc.).

If a vehicle's MAC layer is *blocked*, a closely related issue is when the MAC *blocked* state should be returned to *unblocked*. Two MAC unblocking algorithms are suggested below.

Fixed Unblock. PSA messages are unblocked if no congestion event happens in the following $T_{\text{blk},f}$ synchronization intervals since the last congestion event happened. As the channel congestion may happen simultaneously for a cluster of vehicles, the unblocking action for those vehicles can be synchronized. For example, in a single hop DSRC vehicle network, all vehicles sense the same channel load. Therefore, MAC blocking and unblocking are fully synchronized for the vehicles, resulting in possible high message collisions and low channel utilization.

Random Unblock. A simple way to solve the synchronization problem of the *Fixed Unblock* algorithm is to randomize the waiting times for unblocking. For example, once a vehicle is blocked due to a congestion event, it chooses to wait for a random time in the range of $[1, T_{\text{blk},r}]$ before being unblocked. Here $T_{\text{blk},r}$ is a configurable parameter.

The MAC unblocking algorithms can be used alone or jointly with the transport layer algorithms for congestion control. Next we will provide a design example to evaluate the congestion control performances.

5. A Design Example

In this section, we present a design example of message rate control function in the CAA module, to illustrate some of the ideas for adaptive and cooperative congestion control described in the previous section. To facilitate the discussion, we define two channel congestion events:

(i) a direct channel congestion event (E_{dir}) happens if the channel occupancy time sensed by the vehicle is higher than 50% in a synchronization interval.

(ii) an indirect channel congestion event (E_{ind}) happens if the vehicle receives a message from a neighboring vehicle indicating an E_{dir} event.

We further define network congestion events for the purpose of rate control. We can design so-called noncooperative and cooperative rate control schemes according to their responses to the channel congestion events. In noncooperative rate control schemes, a network congestion event is said to happen only if E_{dir} events happen. On the other hand, in cooperative rate control schemes, a network congestion event is said to happen if either E_{dir} or E_{ind} events happen at least N_{guard} synchronization intervals (denoted SYN-I) after the last MAC blocking action. The configurable parameter N_{guard} is used to protect a vehicle from being forced to congestion too often by its congested neighbors.

Figure 4: Average percentage of time after PSA messages are blocked at MAC layer.

Figure 5: Standard deviation of the MAC blocking time for network with 70 nodes.

Figure 6: Average number of successfully received messages per second.

After a network congestion event occurs, congestion can be mitigated by either MAC blocking in the MMRC module or rate control in the CAA module. The CAA module can adaptively control the message rate for low priority PSA. Suppose that the message rate for PSA varies in the range of $[R_0, R_M]$. Here we use a simple additive-increase multiplicative-decrease (AIMD) algorithm for message rate adaptation [27]. When a node joins a vehicle network, it starts a *fast start mode* for its message rate. The message rate is initially set to the lowest message rate R_0 and increases by one in every N_{fast} SYN-I. This increase continues until the rate reaches the upper bound R_M or a network congestion event occurs. After the message rate reaches R_M, it remains at R_M as long as no congestion event occurs. Upon a network congestion event, half of the currently used message rate is saved as the so-called fast start threshold and fast start begins again from R_0 after PSA messages are unblocked at the MAC layer. Once the message rate reaches the fast start threshold, the rate control enters a *congestion avoidance mode* where the message rate increases by one in every N_{slow} SYN-I. By default N_{slow} is set to be larger than N_{fast}. In the *congestion avoidance mode*, the message rate for PSA remains at R_M after it reaches R_M. If a new network congestion event occurs, the process for network congestion event is repeated.

In the next section, the above proposed message rate control algorithms will be evaluated and compared to the MAC layer message blocking algorithms.

6. Simulation Results

We implemented a discrete-event driven simulator to evaluate four DSRC rate control schemes. Two schemes are configured with a fixed message rate of 4 message per second (mps) and 8 mps and *Random Unblock* in the MMRC module, which are referred to as the FRRU-4 and FRRU-8 schemes, respectively. The parameter $T_{\text{blk},r}$ was configured to 10 SYN-I for these *Random Unblock* schemes. The other two schemes are adaptive rate control with *Fixed Unblock* in the MMRC module and either noncooperative or cooperative rate control in the CAA module. The adaptive rate control schemes are denoted by N-CRFU scheme and CRFU scheme, respectively. The parameter $T_{\text{blk},r}$ was configured to 1 SYN-I for the *Fixed Unblock* schemes. The range for message rate was set to $R_0 = 1$ and $R_M = 10$ mps for the N-CRFU and CRFU schemes. Each PSA message has a fixed length of 300 bytes.

The simulated network topology represents an urban intersection with lane length of 350 m. A configurable number of nodes were randomly placed with a uniform

Figure 7: Probability of successful message reception.

distribution over the lanes. IEEE 802.11p was configured with a backoff window $W_{\min} = 2^8$ for PSA and physical data rate of 3 Mbps. Both transmission and carrier sensing range were set to 200 m. Synchronization interval parameters were set to $t_{\text{syn}} = 50$ ms and $t_{\text{cch}} = 25$ ms. Adaptive rate control parameters were configured to $N_{\text{guard}} = 20$, $N_{\text{fast}} = 1$, $N_{\text{slow}} = 15$, and $p_{\text{mb}} = 0.3$. We performed 20 simulations runs to obtain an average overall runs for each point in the following graphs. Each simulation ran for 500 seconds.

For the N-CRFU and CRFU schemes, the probability that a new MAC blocking event occurred during the time of MAC blocking was very low (smaller than 1%). Therefore, we can assume that the control channel was highly available for ESA messages. The performance measures of interest include how the vehicles were blocked at the MAC layer, the rate of successfully received messages, and probability of successful message reception. Figure 4 presents the network-wide average percentage of time of PSA messages being blocked at the MAC layer. The percentage of MAC blocking time was zero for FRRU-4 but increased sharply for FRRU-8 for large N due to a lack of adaptation to increased traffic load. MAC blocking time was very close for the N-CRFU and CRFU schemes and increased smoothly with N. Figure 5 presents the standard deviation of percentage of MAC blocking time. The standard deviation of MAC blocking time for the N-CRFU scheme was much larger than that for the CRFU scheme by up to 200%. The results demonstrate that the CRFU scheme can achieve fairer channel access and prevent some vehicles from remaining congested for a long time while at negligible cost for the other performance measures.

Figure 6 presents the average number of successfully received messages per second at each node (denoted by N_{rec}). N_{rec} for the FRRU-4 scheme increases linearly from 75 mps for network with 20 nodes to 190 mps for 50 nodes and then drops to 132 mps for 70 nodes. Clearly spectrum utilization for the FRRU-4 scheme is very low with light traffic. N_{rec} for FRRU-8 is initially 150 mps for 20 nodes but drops quickly to 96 for 70 nodes. By contrast, N_{rec} for both N-CRFU and CRFU schemes remains at a high level for all the investigated network scenarios, which demonstrates the effectiveness of the proposed rate control schemes on adapting to dynamic traffic loads. It is noted that intersection of the two curves for schemes FRRU-4 and FRRU-8 is present. The reasons can be explained as follows. When there are a small number of nodes in the network, congestion is not a major problem and the numbers of transmitted and successful messages are both larger for scheme FRRU-8. With increased number of nodes in the network, congestion becomes a serious problem for schemes with high message rate. Scheme FRRU-8 sees severe congestion and has very low number of received messages per second. In this case, the fixed rate scheme with lower message rate will perform better. Theoretically we may say that for static vehicle networks (e.g., with fixed number of nodes and fixed node locations), we could find an optimal message rate to get the highest number of received messages per second and have a centralized entity to determine the optimal message rate and control the network configuration. However, with a large scale vehicle ad hoc network, it is unlikely to have static vehicles uniformly distributed in the network. It is also difficult and expensive to deploy centralized entities for the rate adaptation purpose. Estimation of the number of nodes in the network is also challenging. Therefore, distributed rate adaptation algorithms are desirable for preventing and mitigating congestions without deployment of centralized controllers.

Figure 7 presents the average probability of successful message reception (denoted by p_{suc}) over the nodes in the network. The probability of successful message reception for a node is calculated as the average probability of its messages being successfully received by the neighbors in the node's transmission range. The overall trend for all the rate control schemes is that p_{suc} decreases with increasing number of nodes but is high even with 70 nodes. The achieved high p_{suc} can be explained by either random MAC blocking function in fixed rate schemes or rate adaptation function in N-CRFU and CRFU schemes, which successfully prevent excessive channel congestion. For the results of message success probability, it is worthy of note that there is no direct relationship between the message success probability and the message rate used in the fixed rate scheme, due to the use of MAC layer random blocking mechanism, which can stop transmission of messages once the detected congestion level is higher than a threshold. It is also observed that there is a peak with 50 nodes for scheme FRRU-4. The reason of such a peak can also be explained by the use of the random blocking mechanism in the MAC layer.

7. Conclusions

DSRC is a promising technology for road safety applications. In this paper, we studied the QoS control problem for road safety applications over large-scale DSRC-based vehicle networks. We discussed the challenging issues of

QoS control in DSRC vehicles and proposed a QoS control framework to provide cooperative and adaptive QoS for road safety applications. The proposed control framework consists of three major modules: ASA module, CAA module, and MMRC module. Vehicle cooperation and cross-layer design were design principles in the modules to achieve effective and robust QoS performance.

We presented a design example on the cooperative and adaptive rate control to illustrate the key ideas proposed in the QoS control framework. Four rate control schemes (two schemes with fixed message rate and MAC layer PSA message blocking, two schemes with rate adaptation at the CAA module) are proposed and compared by simulations. Simulation results demonstrate the effectiveness of adaptive rate control at the CAA module in terms of adapting to dynamic traffic conditions. It is also observed that cooperative rate control scheme can provide fairer channel access for PSA messages than noncooperative rate control scheme.

Acknowledgments

The work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) with Grant reference no. *EP*/1010157/1 and partially supported by the National Natural Science Foundation of China (NSFC) under Grant no. 61103177.

References

[1] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva, "A Performance comparison of multi-hopwireless Ad Hoc nework routing protocols," in *Proceedings of the 4th annual ACM/IEEE International Conference on Mobile computing and Networking (MobiCom '98)*, 1998.

[2] G. Jeffrey Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*, Pearson Education, 2007.

[3] H. Holma and A. Toskala, *LTE For UMTS: Evolution To LTEAdvanced*, John Wiley & Sons, New York, NY, USA, 2011.

[4] J. He, K. Guild, K. Yang, and H.-H. Chen, "Modeling contention based bandwidth request scheme for IEEE 802.16 networks," *IEEE Communications Letters*, vol. 11, no. 8, pp. 698–700, 2007.

[5] J. Zhu and S. Roy, "MAC for dedicated short range communications in intelligent transport system," *IEEE Communications Magazine*, vol. 41, no. 12, pp. 60–67, 2003.

[6] D. Jiang, V. Taliwal, A. Meier, W. Holfelder, and R. Herrtwich, "Design of 5.9 GHz DSRC-based vehicular safety communication," *IEEE Wireless Communications*, vol. 13, no. 5, pp. 36–43, 2006.

[7] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 74–82, 2006.

[8] M. Faezipour, M. Nourani, A. Saeed, and S. Addepalli, "Progress and challenges in intelligent vehicle area networks," *Communications of the ACM*, vol. 55, no. 2, pp. 90–100, 2012.

[9] X. Wu and S. Subramanian, "Vehicular communications using DSRC: challenges, enhancements, and evolution," *IEEE Journal on Selected Areas in Communication*, vol. 31, no. 9, 2013.

[10] S. Zeadally, R. Hunt, Y.-S. Chen, A. Irwin, and A. Hassan, "Vehicular ad hoc networks (VANETS): status, results, and challenges," *Telecommunication Systems*, vol. 50, pp. 217–241, 2012.

[11] W. Guan, J. He, C. Ma, Z. Tang, and Y. Li, "Adaptive Message Rate Control of Infrastructured DSRC Vehicle Networks for Co-existing Road Safety and Non-safety Applications," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 134238, 8 pages, 2012.

[12] M. Eiza and Q. Ni, "An evolving graph-based reliable routing schemefor VANETs," *IEEE Transactions on Vehicular Technology*, vol. 624, pp. 1493–1504, 2013.

[13] F. J. Ros, P. M. Ruiz, and I. Stojmenovic, "Acknowledgment-based broadcast protocol for reliable and efficient data dissemination in vehicular ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 33–46, 2012.

[14] X. Ma, J. Zhang, X. Yin, and K. S. Trivedi, "Design and analysis of a robust broadcast scheme for VANET safety-related services," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 46–61, 2012.

[15] K. A. Hafeez, L. Zhao, B. Ma, and J. W. Mark, "Performance analysis and enhancement of the DSRC for VANETs safety applications," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 358–363, 2013.

[16] M. I. Hassan, H. L. Vu, T. Sakurai, and L. L. H. Andrew, "Effect of retransmissions on the performance of the IEEE 802.11 MAC protocol for DSRC," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 22–34, 2012.

[17] S. I. Sou, "Modeling emergency messaging for car accident over dichotomized headway model in vehicular Ad-hoc networks," *IEEE Transactions on Communications*, vol. 61, no. 2, 2013.

[18] J. He, Z. Tang, T. O'Farrell, and T. M. Chen, "Performance analysis of DSRC priority mechanism for road safety applications in vehicular networks," *Wireless Communications and Mobile Computing*, vol. 11, no. 7, pp. 980–990, 2011.

[19] F. Martelli, M. E. Renda, G. Resta, and P. Santi, "A Measurement-based study of beaconing performance in IEEE 802.11p vehicular networks," in *Proceedings of the International Conference on Computer Communications (INFOCOM '12)*, 2012.

[20] J. Yin, T. Elbatt, G. Yeung et al., "Performance evaluation of safety applications over DSRC vehicular ad hoc networks," in *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks table of contents (VANET '04)*, pp. 1–9, 2004.

[21] N. Nasiriani, Y. P. Fallah, and H. Krishnan, "Stability analysis of congestion control schemes in vehicular Ad-hoc networks," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC '13)*, 2013.

[22] IEEE 802. 11p/d3. 05, "Draft amendment for wireless access for the vehicular environment (WAVE)," 2008.

[23] IEEE Std.802.11e, "Wireless LAN medium access control (MAC) enhancements for quality of service (QoS)," 2005.

[24] X. Guan, R. Sengupta, H. Krishnan, and F. Bai, "A feedback-based power control algorithm design for VANET," in *Proceedings of the Mobile Networking for Vehicular Environments (MOVE '07)*, pp. 67–72, 2007.

[25] J. Mittag, P. Santi, and H. Hartenstein:, "Vehicle-toVehicle communication: fair transmission power control for safety-critical information," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 7, pp. 3684–3703, 2009.

[26] W. Guan, *Adaptive QoS control of DSRC vehicle networks for collaborative vehicle safety applications [Ph.D. thesis]*, Swansea University, 2013.

[27] J. He, H.-H. Chen, T. M. Chen, and W. Cheng, "Adaptive congestion control for DSRC vehicle networks," *IEEE Communications Letters*, vol. 14, no. 2, pp. 127–129, 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 232916, 11 pages
http://dx.doi.org/10.1155/2013/232916

Research Article

An Efficient Emergency Message Broadcasting Scheme in Vehicular Ad Hoc Networks

Hayder Salman Dawood and Yuming Wang

Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Correspondence should be addressed to Yuming Wang; ymwang@hust.edu.cn

Received 17 July 2013; Accepted 17 September 2013

Academic Editor: Jianhua He

Copyright © 2013 H. S. Dawood and Y. Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Broadcasting of emergency message (EM) in the vehicular Ad hoc networks (VANETs) plays a significant role in safety applications, which require immediate broadcasting of EM to the relevant vehicles located in the risk zone (RZ). This paper proposes an efficient emergency message broadcasting (EEMB) scheme in VANETs. This scheme aims at avoiding multiple accidents and reducing traffic jams by broadcasting EM at high speed and with low overhead. When an accident takes place, the crashed vehicle (CV) will choose the best relayer(s) and broadcast the EM. The relayer will also choose the best next relayer(s) and rebroadcast the EM further. This process continues until the EM covers the whole RZ. The EEMB scheme reduces the overhead caused by beacons by using the prediction method. Our proposed scheme also overcomes the network fragmentation problem either by choosing the vehicles moving in the opposite direction as the relayer(s) if there are no vehicles in the same direction or by using the store-carry-forward (SCF) mechanism when there are no vehicles moving in both directions. The simulation results have shown that the proposed scheme has better performance in terms of the risk zone latency (RZL) and the beacons overhead (BCO).

1. Introduction

Every year, many people lose their lives or get injured on roads and huge amounts of time and fuel are exhausted because of road accidents or traffic jams. According to the report of the Ministry of Public Security of the People's Republic of China, 65225 people were killed and 254075 injured in 2010 because of road accidents in China [1]. These losses would have been avoided or minimized, if drivers had been informed of the risk ahead in advance. For this reason, both of the research community and automotive industry have paid considerable attention to the vehicular ad hoc networks (VANETs) [2]. There are two main types of VANETs applications: safety and nonsafety applications. The purpose of safety applications is to increase the safety of both the passengers and the vehicles simultaneously. This purpose can be achieved by sending emergency messages (EMs) to the vehicles located in the risk zone (RZ) [3–5]. The purpose of nonsafety applications is to improve passenger's comfort and traffic efficiency and/or to optimize the path to a destination [6]. In the safety applications of VANETs, the EM needs to be disseminated to the RZ area which is usually several kilometers behind and in the same direction of the crashed vehicle (CV). Mostly, the one-hop range is just several hundreds of meters and cannot cover the total RZ. Thus a multihop propagation is needed during the dissemination [7]. Also, most safety applications will possibly work in a broadcast fashion since EM is useful to all vehicles surrounding the sender [8]. Flooding mechanism is the first broadcasting type. In flooding, a vehicle receiving an EM including a nonduplicate sequence number (SN) from the sender will rebroadcast it directly. Obviously, this mechanism is not efficient especially in the high network density, due to the huge number of EMs flooded in the network, which leads to broadcast storm problem [9]. To solve this problem, most researchers decrease the number of redundant rebroadcast messages by allowing only some of the vehicles to rebroadcast the message instead of letting every vehicle do that [10]. In this paper, an efficient emergency message broadcasting (EEMB) scheme has been proposed. According to this scheme, the EM is made to reach the relevant vehicles located in the RZ with low overhead and at a high speed. By applying the EEMB

scheme, vehicles exchange a lower number of beacons to enable the sender to choose the best relayer(s) either in the same direction or in the opposite one and make the EM broadcast farther and faster.

The rest of this paper is organized as follows. Section 2 is a discussion of related work; Section 3 presents the operation and characteristics of the proposed scheme. In Section 4, the simulation and results are discussed. Finally, Section 5 is a conclusion.

2. Related Work

Broadcast schemes can be classified into two classes: receiver decision schemes and sender decision schemes. These classes with their main challenges will be explained in the first and second subsections. Network fragmentation issue will be identified in the last subsection.

2.1. Receiver Decision Schemes. In the receiver decision schemes [2, 11–13], the contention mechanism has been used for selecting the relayer. In this mechanism, every vehicle receives the EM and waits for a while before rebroadcasting the EM. This delay is called a waiting time (WT) and it can be calculated as a function of the distance between the vehicle and the sender. The farthest vehicle in the sender's transmission range will have the shortest WT, so it will be the relayer that will rebroadcast the EM farther. For the purpose of avoiding redundancy, the other vehicles located between this relayer and the sender will terminate their WT process when they know that the EM has been already rebroadcast. Therefore, only one vehicle will rebroadcast the EM [3, 10]. Both of urban multihop broadcast (UMB) [2] and smart broadcast (SB) [11] schemes are based on a request to broadcast (RTB)/clear to broadcast (CTB) handshake at each hop before transmitting the EM. A collision will take place when more than one vehicle sends the CTB packet simultaneously [14]. Since safety applications rely on multihop propagation, the use of the RTB/CTB handshake leads to long latency. In efficient directional broadcast (EDB) [12] and slotted 1-persistence [13] schemes, although the authors do not use the RTB/CTB handshake, they still use distance to calculate the WT. Therefore, in a sparse network, the WT is still long because the distance between the relayer and the sender is perhaps short. On the other hand, in a dense network, there will possibly be more than one vehicle that may have the same WT. In this case, the collision probability will increase resulting in the increase of latency.

2.2. Sender Decision Schemes. Contention mechanism, which is used in the receiver decision schemes, causes a long delay especially in the sparse network. This mechanism is not efficient enough to use in safety applications since the EM should be broadcast as soon as possible to the relevant vehicles located in RZ; therefore, we need to use the sender decision schemes [15–18], in which the sender makes the decision which vehicle will be the relayer and rebroadcast the EM. This eliminates the WT, but this requires that every vehicle need to know the accurate location of all its one-hop neighbors so as to choose the best relayer which is commonly behind the sender and it is the farthest one. The location of vehicles changes quickly because of high mobility. Consequently, the out-of-date information of the neighbor's location is a problem in these schemes. As a result, the sender may select a relayer out of its transmission range. Thus, the broadcasting process will fail. To solve this problem, most researchers increase the frequency of beacons which contain the vehicle's location. Anyway, this solution faces the high overhead problem, especially, when the network becomes denser.

2.3. Network Fragmentation Issue. One of the VANETs characteristics is the network fragmentation, which is a problem that prevents vehicles from broadcasting the EM to the relevant ones due to loss of direct connection between the vehicles, and this case frequently occurs in the sparse network. A scheme that does not solve this problem will fail to deliver the EM to the relevant vehicles. Some schemes are proposed to handle this problem. In [19], there is a suggestion to solve this problem by using the infrastructures. Anyhow, this solution is costly because a large number of infrastructures would be needed to cover long roads such as a freeway. Moreover, it does not match with safety applications, which usually require direct vehicle-to-vehicle ad hoc communication networks so as to be capable of delivering the information at a very low cost [20, 21]. In [22], it is suggested that the vehicle having the EM should broadcast the EM periodically until it chooses another relayer. Anyhow, this suggestion causes redundant broadcasting because many broadcasting attempts do not help the vehicle obtain any new ones. Thus, the network will suffer from the overhead. To reduce this overhead, an interval for this periodical broadcasting is recommended. This will consequently lead to a delay in the delivery of EM to the relevant vehicles.

3. An Efficient Emergency Message Broadcasting (EEMB) Scheme in VANETs

This section has two divisions. The first is concerned with explaining the EEMB scheme operation and the second is a description of the most important characteristics of the EEMB scheme.

3.1. The EEMB Scheme Operation. The EEMB scheme works in two phases: the normal phase and the emergency phase.

3.1.1. The Normal Phase. In the normal phase, no EM will be sent, so each vehicle will send a beacon to its one-hop neighbors in one of two ways: periodically or conditionally (while the difference between the predicted and real locations is greater than the threshold of error predict value). Each vehicle will be able to obtain information about its neighboring vehicles and update its neighbor list. This phase divides into two parts; namely, sending a beacon and receiving a beacon.

(1) Sending a Beacon. It is assumed that every vehicle is equipped with global positioning systems (GPSs). In addition, the vehicle can record its history of movement and

current situation. Thus, the vehicle can calculate its speed and direction using (1) and (2)

$$\text{Speed} = \frac{\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}}{t_2 - t_1}, \quad (1)$$

$$\theta = \tan^{-1} \frac{Y_2 - Y_1}{X_2 - X_1}, \quad (2)$$

where (X_2, Y_2) is the current location at the current time t_2, (X_1, Y_1) is the previous location at the previous time t_1, and θ is the moving direction of the vehicle.

So, each vehicle can broadcast a beacon containing the information shown in Figure 1.

(2) Receiving a Beacon. Each vehicle receiving a beacon will follow the steps shown below.

(a) According to the vehicle ID information, the vehicle will know if this vehicle exists in its neighbor list or not. If yes, the vehicle will update this neighbor information, and if not, it will add the vehicle to its neighbor list.

(b) Based upon the received beacon, the vehicle can predict the location of its neighbor by using (3) and (4)

$$X_p = X + \text{Speed}^* (t_2 - t_1)^* \cos(\theta), \quad (3)$$

$$Y_p = Y + \text{Speed}^* (t_2 - t_1)^* \sin(\theta), \quad (4)$$

where (X_p, Y_p) is the predicted location of the neighbor vehicle at the current time t_2, while (X, Y) and t_1 are the neighbor vehicle's location and timestamp in the last received beacon. Speed and θ are the neighbor vehicle's speed and direction from the last received beacon.

(c) The vehicle will calculate the link life time (LLT) [23] for its neighbor depending on the predicted location.

(d) The vehicle will compare its location with the predicted location of its neighbor to classify it as a member of one of the following groups: same-direction-ahead (SDA), same-direction-behind (SDB), opposite-direction-ahead (ODA), and opposite-direction-behind (ODB).

(e) The vehicle will calculate the distance (D) between it and its neighbor by using (5)

$$D = \sqrt{(X_v - X_p)^2 + (Y_v - Y_p)^2}, \quad (5)$$

where (X_v, Y_v) is the vehicle's location and (X_p, Y_p) is the predicted location for its neighbor.

The prediction process will be continued in the near future. This is based on the previous predicted value, until another beacon comes, so the neighbor's information will be updated and prediction process cycle is started again, or until the vehicle is removed from the neighbor list when its LLT value is smaller than the threshold of LLT value. Anyway, when the vehicle needs to know the location of any other vehicle in its neighbor list, it will use the predicted location value instead of the one initially informed through the beacon.

Vehicle ID	Vehicle location	Vehicle direction	Vehicle speed	Time stamp

Figure 1: Beacon message format.

First relayer ID	Second relayer ID	Sender ID	Sender position	EM SN	End of the risk zone	Time stamp

Figure 2: Emergency message format.

3.1.2. The Emergency Phase. In the emergency phase, there is an EM to be sent. So, when an accident takes place, the CV will choose the relayer(s), generate an EM, and broadcast it. When a vehicle receives the EM, it will be easy for it to know if it has to do the relay task or not. Whether the vehicle is the first relayer or the second one is decided by the field of the first relayer ID and second relayer ID in the EM format as shown in Figure 2.

In this way, the EM will warn the vehicles at the RZ not to have accidents or traffic jams.

In this phase, we have three scenarios for relayer(s) selection.

Scenario 1. As shown in Figure 3, when the CV wants to select the relayer(s) from its neighbor list, it will perform one of the following steps.

(a) If there is any vehicle behind the CV in the same direction, it will choose the farthest vehicle (FV) as a first relayer (FR), and the second farthest vehicle (SFV) as a second relayer (SR), as shown in Figure 4(a).

(b) If there is no vehicle behind the CV in the same direction, but there are vehicles behind it in the opposite direction, it will choose the FV as a FR and the SFV as a SR, as shown in Figure 4(b).

(c) If there is no vehicle behind the CV in both directions, but there are vehicles ahead of it in the opposite direction, it will choose the nearest vehicle (NV) as a FR and the second nearest vehicle (SNV) as a SR, as shown in Figure 4(c).

(d) If none of the three steps above can be achieved, as shown in Figure 4(d), the CV will store the EM until a new vehicle comes within its transmission range. And it will choose the relayer(s) according to one of the three steps above.

After the CV chooses the relayer(s), there will be two possibilities, as will be explained in Scenarios 2 and 3 below.

Scenario 2. As shown in Figure 5, when the relayer on the same direction of the CV receives the EM, it will check its location. If it is beyond the terminal point (TP) of the RZ,

Figure 3: A CV wants to select the relayer(s).

it will do nothing except broadcasting a duplicate EM as an acknowledgment. Otherwise, it selects the best next relayer(s) by performing one of the following steps.

(a) If there is any vehicle behind it in the same direction, it will choose the FV as a first next relayer (FNR) and the SFV as a second next relayer (SNR), as shown in Figure 6(a).

(b) If there is no vehicle behind it in the same direction, but there are vehicles behind it in the opposite direction, it will choose the FV as a FNR and the SFV as a SNR, as shown in Figure 6(b).

(c) If none of the two steps above can be achieved, as shown in Figure 6(c), the relayer will broadcast a duplicate EM as an acknowledgment and use the store-carry-forward (SCF) mechanism to store the EM in its buffer while checking its location. If it passes by the CV, then it will terminate the forwarding process. Otherwise, whenever it encounters another vehicle within the transmission range which satisfies one of the two steps above, the relayer will choose it as the next relayer and forward the EM to it.

Scenario 3. As shown in Figure 7 when the relayer on the opposite direction of the CV receives the EM, it will check its location. If it is beyond the TP of the RZ, it will do nothing except broadcasting a duplicate EM as an acknowledgment. Otherwise, it selects the best next relayer(s) by performing one of the following steps.

(a) If there is any vehicle ahead in its opposite direction, it will choose the FV as a FNR and the SFV as a SNR, as shown in Figure 8(a).

(b) If there is no vehicle ahead in its opposite direction, but there are vehicles ahead in the same direction, it will choose the FV as a FNR and the SFV as a SNR, as shown in Figure 8(b).

(c) If none of the two steps above can be achieved, as shown in Figure 8(c), the relayer will broadcast a duplicate EM as an acknowledgment and use the SCF mechanism while checking its location. If it passes by the TP of the RZ, it will terminate the forwarding process. Otherwise, whenever it encounters another vehicle within its transmission range that satisfies one of the two steps above, this relayer will choose that vehicle as the next relayer and forward the EM to it.

Whenever the relayer or the next relayer receives the EM, it will choose another or other next relayer(s). This selection will be based on the direction of the other next relayer(s) according to Scenario 2 or Scenario 3. But if it does not have any neighbor vehicle(s) in its neighbor list, it has to broadcast a duplicate EM to have the following benefits.

(i) Informing the sender and the second relayer that the message has been received by another vehicle. Thus, they will cancel their rebroadcast task.

(ii) The vehicles located between the relayer and the sender or between the next relayer and the relayer will receive a duplicate message. Then, if any one of these vehicles were not able to receive this message earlier, it would receive it at this moment.

3.2. The EEMB Scheme Characteristics

3.2.1. Improving the Neighbor's Localization Accuracy. The EEMB scheme reduces the overhead caused by beacons by using the prediction method [24] with some modification. More information such as location, speed, and direction are added to the beacons. This information can be used by vehicles to predict their neighbor's location in the near future, and a vehicle keeps on predicting about its neighbor's location until this neighbor is eliminated from the neighbor list. Also, when the vehicle sends its beacon which includes this information, it will use this information to determine the error between its own predicted location and the real one.

When this error is greater than the threshold of the error predict value, the vehicle itself will broadcast a new beacon that contains the current information in turn leading to updating neighbor's predictions. This method provides a more accurate localization of the vehicle's neighbors through exchanging a smaller number of beacons.

FIGURE 4: Cases of selecting the relayer(s) by the CV.

3.2.2. Efficiency Relayer Selection. The use of beacons is important not only for the localization purpose but also for updating the vehicle's neighbor list as well. Thus, a vehicle does not choose a relayer that is not available within its transmission range. So, exchanging beacons of a smaller number by means of prediction may lead to an unreliable neighbor list. A prediction method is suggested to solve this problem by calculating the distance between the vehicle and its neighbor and remove the neighbor which is out of the vehicle's transmission range from the neighbor list. Anyhow, this is not satisfactory for the maintenance of neighbor list. Even if the distance is not greater than the transmission range, but the remaining time of the neighbor in the range is not enough for a forwarding process. This is why the forwarding process fails. Therefore, we use the LLT for this purpose. LLT is the amount of time a vehicle stays within the transmission range of another vehicle.

Each vehicle will calculate the LLT value for its neighbor or neighbors depending on the prediction of location. Those neighbors whose LLT is smaller than the value which is required in the successful forwarding process will be removed from the neighbor list. So, in EEMB scheme, the sender selects the relayer(s) whose LLT value is greater than the threshold of the LLT value. Hence, the use of LLT results in efficient relayer selection and avoidance of message loss or redundancy.

3.2.3. Reliability Increment. After the sender chooses the relayer(s) and broadcasts the EM, it will set a timer (ST) to see whether the relayer(s) receive(s) the EM successfully or not. Consequently, reliability can be achieved by using one of the following two methods: ACK message method and duplicate message method. If the sender receives either of these two from the receiver before its timer gets expired, the sender will reset the timer (RT) and terminate the forwarding process. Otherwise, it will send the EM again. Since most safety applications rely on broadcast fashion, it is not possible to let each vehicle receive the message to send an ACK or a duplicate message to the sender because this will lead to a broadcast storm problem. That is why some schemes handle this problem by making a special vehicle send this ACK message or the duplicate message to the sender [2]. This vehicle is the one selected as a relayer. It should be noticed that what is meant by reliability here is the insurance of the message sent by the sender and received successfully by the relayer. Although both of the two methods mentioned above can achieve this type of reliability, EEMB scheme chooses the second one since it has got other advantages. When the sender

Figure 5: A relayer on the same direction of the CV wants to select the next relayer(s).

broadcasts its message, all the vehicles within its transmission range will receive it, but sometimes, this message may not be received successfully by all these vehicles for one reason or another. The use of the ACK method will not make the vehicles which have not received this message receive it. This method also adds an extra delay because the relayer has to send an ACK message first and rebroadcast the sender's message next. By using the second method, this extra delay is avoided, and at the same time, the probability of making some vehicles which cannot receive the message by the sender's broadcasting receive it instantly by relayer's rebroadcasting is increased.

3.2.4. Redundancy Decrement. Due to the high mobility in VANETs, the location of vehicles is continuously changing. In this case, choosing only one relayer is unsuitable for this type of network. That is why EEMB scheme chooses two relayers. If the first relayer fails to do the relay task immediately, the second one will do it after a period of time almost equal to one-hop delay. During this period of time, if the second relayer receives a duplicate message, it will cancel its forwarding process. When the relayers are located behind the sender, another benefit of choosing two relayers can be obtained: the second relayer will rebroadcast the message if the first one fails and this is better than making the sender broadcast it again since the second relayer will cover more relevant vehicles by its broadcasting than the sender's broadcasting as shown in Figure 9. Thus, the EEMB scheme will reduce the redundancy and increase reliability.

3.2.5. Overcoming the Network Fragmentation Problem. The density of vehicles varies from place to place and from time to time. Hence, a vehicle may or may not find other vehicles within its transmission range, especially if it takes into consideration only the vehicles moving in the same direction even though the freeway usually has two directions. Consequently, the EEMB scheme takes the advantage of that fact and then the vehicles moving in the opposite direction will be used as relayers if there are no vehicles in the same direction. As a result, the EEMB scheme can overcome the network fragmentation problem when it takes place in the same direction in the network and it reduces the latency caused by waiting till a vehicle in the same direction comes and becomes within the transmission range to forward the message to it. If there are no vehicles moving in both directions, the EEMB scheme will use a SCF mechanism.

4. Simulation and Results

In this section, the vehicular mobility model used in the simulation is introduced. Then, the simulation setup is described. Finally, the results are shown and discussed. A MATLAB (version R2011b) [25] is used to generate the freeway mobility model and evaluate the performance of the proposed scheme, compared with the VANET minimum connected dominating set (V-MCDS) scheme [16], which belongs to the same sender decision schemes.

In the simulation, it is assumed that both the EEMB and V-MCDS schemes work on the same wireless communication environment which is based upon the same wireless communication technologies (e.g., IEEE 802.11 series standards) and have the same lower layer communication characteristics. This simulation is just focusing on the evaluation of the performance achieved by the two schemes.

Compared with the background traffic, the safety data has the highest transmission priority, and the hop delay for safety data is very close to the theoretical value. For the safety data transmission, it looks as though no background traffic is introduced. In the EEMB and V-MCDS schemes, only the chosen relayer(s) will relay the EM. What is more, the occurrence of traffic accident is a matter of moments, and the safety data payload is very small and utilizes a very small proportion of bandwidth thus, the safety data collision is very low and will be omitted in the simulation.

4.1. Vehicular Mobility Model. We have evaluated the EEMB scheme as well as the V-MCDS scheme by comparing the two using the freeway mobility model [26], after making some modification. In this modified model, there are two directions each of which has two lanes. Vehicles are distributed randomly on the freeway lanes in both directions. Each vehicle

Figure 6: Cases of selecting the next relayer(s) by the relayer on the same direction as the CV.

chooses a random speed which is uniformly distributed in a speed interval $[S_{\min}, S_{\max}]$. This speed changes every T_s seconds, where T_s is the vehicle speed change interval (VSCI), and this speed depends on a previous one. At any time, the next speed must not be less than $S_{\min}$ or greater than $S_{\max}$. If this happens, the next speed will be equal to $S_{\min}$ or to $S_{\max}$. However, the original model has two unrealistic aspects: the first one is that the speed of the rear vehicle cannot exceed the speed of the front one, but we change this aspect as in [24], so the modified model supports the vehicle with a higher speed in crossing the vehicle with a lower speed. The second aspect is, when a vehicle moves outside the extremities of the simulation area, this vehicle is reinserted at the beginning of a random lane on the other extreme in the area. When doing this, we cannot continue to simulate the messages broadcast by this vehicle for the behavior of a vehicle near to the extreme will not be realistic, so to avoid this problem, we simulate a wide area and take a part of it (the center of the area) into consideration, and no vehicle will be in motion outside the area at the simulation time.

4.2. Simulation Setup. Two aspects of the EEMB scheme are included in the simulation. First, the vehicle moving in the opposite direction is chosen as a relayer if there are no vehicles moving in the same direction. Thus, the EM can be broadcast faster to the relevant vehicles located in the RZ. Second, the prediction method is chosen to reduce the required number of beacon exchanges while improving the neighbor's localization accuracy. To show the efficiency of this scheme, we have compared it with the V-MCDS scheme, which does not include these two aspects. In the V-MCDS scheme, only the farthest vehicle behind and in the same direction will be chosen as the relayer, and the relayer's location will not be predicted, but just gained by exchanging large number of beacons.

The parameters used in this simulation are listed in Table 1. Reference [27] shows that the theoretical number of one-hop delay varies from 1.621 ms to 1.756 ms. Thus, we have taken the average of these two numbers to represent the one-hop delay value. The threshold of error prediction value is set to 25 meters as in [24]. The minimum value allowed for the threshold of the LLT is set to (2*one-hop delay). Clearly, the beacon interval value has an effect on the performance of the schemes. When we made it too big, plenty of out-of-date problems were raised in the V-MCDS scheme, and when we made it too small, the overhead increased in the network due to the high number of beacons exchanged between vehicles. Anyhow, reducing the beacon interval is required to increase the accuracy of localization of neighbors; it is the pursuit of the V-MCDS scheme, since it is unlike EEMB scheme and has not applied the prediction mechanism. That is why, the sound choice for beacons interval value is set to 1 second and 0.5 second for the EEMB and V-MCDS schemes, respectively.

In our simulation, a vehicle will be chosen randomly as a CV at a random time, and the CV must send an EM to the vehicles located in RZ as soon as possible. This can

Figure 7: A relayer on the opposite direction of the CV wants to select the next relayer(s).

be achieved when the CV chooses the relayer(s) within its transmission range to rebroadcast the EM further. Likewise, this relayer chooses the next relayer(s) within its transmission range and forwards the EM further. As a result, the EM is disseminated from the CV to the other vehicles located in the RZ.

It is assumed that the EM life time is not limited, so in this simulation, the EEMB scheme will continue an EM disseminations until the EM reaches the extreme of the RZ or all vehicles that have been in the RZ pass by the CV. The main aim of our simulation is to make the EM cover the RZ as quickly as possible and hopefully before the vehicles located in RZ pass by the CV with the lower number of beacons exchanged.

4.3. Discussion of Results. Two important metrics have been used to show the efficiency of the EEMB scheme, namely, the risk zone latency (RZL) and the beacon overhead (BCO).

The Risk Zone Latency (RZL). It means the time interval between the EM broadcast by the CV when the accident takes place and when it reaches the extreme of the RZ, or when all vehicles in the RZ pass by the CV.

The Beacon Overhead (BCO). It is the number of beacons exchanged between the vehicles during the simulation time.

Table 1: Simulation parameters.

Parameter	Value
Length of the freeway	5 Km
Width of each lane	5 m
Vehicle density $[V_{\min}, V_{\max}]$	[30, 420] Vehicle
Vehicle speed $[S_{\min}, S_{\max}]$	[80, 120] Km/h
VSCI (T_s)	54~66 sec.
Transmission range	250 m
Risk zone size	1 Km
One-hop delay	1.6885 ms
Threshold of error predict	25 m
Threshold of LLT	3.377 ms
V-MCDS scheme beacon interval	0.5 sec.
EEMB scheme beacon interval	1 sec.
Simulation time	150 sec.

Figures 10 and 11 show the average RZL. Each result value represents the average of 120 simulation rounds. As shown in Figure 10 when the traffic density is low, the average RZL in V-MCDS scheme is very high due to the low number of vehicles on the road. Thus, the relayer may not find other vehicles within its transmission range in the same direction behind it, and then, it has to use the SCF mechanism and keeps moving for a period of time known as the disconnecting time until it encounters another vehicle within its transmission range in the same direction behind it. Then, the relayer chooses this vehicle as the next relayer and assigns the forwarding task by rebroadcasting the EM to it. The EEMB scheme takes into account that a network fragmentation problem may take place in the same direction in the network. If there is no vehicle moving in the same direction, it benefits from choosing the vehicle moving in the opposite direction as the next relayer. This approach increases the probability of finding other vehicles, whereas the disconnecting time decreases. As a result, the average RZL in the EEMB scheme is clearly less than that of the V-MCDS scheme, even in the situation of a low traffic density. However, when the traffic density increases, it is likely that the relayer finds other vehicles within its transmission range whether in the same direction or in the opposite one. Thus, the disconnecting time decreases and the average RZL decreases as well.

For this reason, the average RZL seems almost equal in both schemes, especially when the number of vehicles is larger than 270. But when taking the magnified sector of the same range as shown in Figure 11, it has been found that the EEMB scheme still has better performance than the V-MCDS scheme with regard to the average RZL, since the possibility of finding a relayer in either of the two directions in the EEMB scheme is greater than finding a relayer just in a single direction as in the V-MCDS scheme.

Figure 12 shows the comparison of the total number of beacons exchanged between the vehicles during the simulation time. According to this figure, the V-MCDS scheme has a huge number of beacons exchanged. First, a shorter beacon interval is required to reduce the out-of-date problem. Second, the V-MCDS scheme makes a vehicle send a new

Figure 8: Cases of selecting the next relayer(s) by the relayer on the opposite direction as the CV.

Figure 9: Two relayer selecting procedures.

Figure 10: Comparison of average RZL with the number of vehicles at a range from 30 to 420 vehicles in both schemes.

beacon when it changes its state (i.e., speed or direction) and any vehicle receives this beacon will in turn send a new beacon. Therefore, the BCO will increase when the number of vehicles increases (dense network), when the beacon interval decreases, and when the vehicle changes its state rapidly.

In the EEMB scheme, the BCO is smaller than that of the V-MCDS scheme, because first, through prediction, beacons exchange at a lower rate, so beacon interval increases, and second, when the vehicle changes its state, it will not send a beacon immediately, but it will send it when the difference between its predicted location and its real one becomes greater than the threshold of error predict value, and each vehicle receiving this beacon is not required to send a beacon. Thus, the number of beacons that need to be exchanged between the vehicles is reduced. As a result, the BCO of EEMB scheme during the simulation time is clearly less than half of V-MCDS scheme as shown in Figure 13.

5. Conclusions

In this paper, we have proposed an EEMB scheme to broadcast EM to the relevant vehicles located in the RZ in order to

Figure 11: Comparison of average RZL with the number of vehicles at a range from 270 to 420 vehicles in both schemes.

Figure 12: The number of beacons in both schemes.

avoid further accidents or to reduce the traffic jams. When an accident takes place on the road, the CV chooses two relayers among its neighbors and then it generates and broadcasts an EM. If the first relayer fails to rebroadcast the EM, then the second one will do that and the relayer will, in turn, choose the next relayer(s) and rebroadcast the EM until the EM covers the whole RZ. The network fragmentation problem is a real issue in VANETs, especially in the sparse network. The EEMB scheme overcomes this problem through the use of vehicles that are moving in the opposite direction to be chosen as the relayers when there are no vehicles in the same direction. An SCF mechanism has been used when there are no vehicles moving in both directions. Our scheme has also addressed the problem of beacon overhead by using the

Figure 13: Ratio of the number of EEMB scheme beacons to the number of V-MCDS scheme beacons.

prediction method which adds more information to beacon messages. Therefore, the vehicles can use this information to predict their one-hop neighbor's location in the near future, thus providing a more accurate localization of the neighbor's location by exchanging a smaller number of beacons. In the simulation, we have compared the EEMB scheme with the V-MCDS scheme through two important metrics of EM broadcasting in safety applications. These metrics are RZL and BCO. The simulation results show that the EEMB scheme has better performance than the V-MCDS scheme in terms of these two metrics.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant no. 61103177).

References

[1] 2013, http://www.mps.gov.cn/n16/n1282/n3553/2921432.html.

[2] G. Korkmaz, E. Ekici F, F. Özgüner, and Ü. Özgüner, "Urban multi-hop broadcast protocol for inter-vehicle communication systems," in *Proceedings of the 1st ACM International Workshop on Vehicular ad hoc Networks*, pp. 76–85, 2004.

[3] Z. Doukha and S. Moussaoui, "Dissemination of an emergency message in a vehicular ad hoc network," in *Proceedings of the 2011 International Conference on Communications, Computing and Control Applications (CCCA '11)*, pp. 1–6, March 2011.

[4] J. He, Z. Tang, T. O'Farrell, and T. M. Chen, "Performance analysis of DSRC priority mechanism for road safety applications in vehicular networks," *Wireless Communications and Mobile Computing*, vol. 11, no. 7, pp. 980–990, 2011.

[5] J. He, H. H. Chen, T. M. Chen, and W. Cheng, "Adaptive congestion control for DSRC vehicle networks," *IEEE Communications Letters*, vol. 14, no. 2, pp. 127–129, 2010.

[6] S. Yousefi, S. Bastani, and M. Fathy, "On the performance of safety message dissemination in vehicular ad hoc networks," in *Proceedings of the 4th European Conference on Universal*

Multiservice Networks (ECUMN '07), pp. 377–387, February 2007.

[7] Y. T. Tseng, R. H. Jan, C. Chen, C. F. Wang, and H. H. Li, "A vehicle-density-based forwarding scheme for emergency message broadcasts in VANETs," in *Proceedings of the 2010 IEEE 7th International Conference on Mobile Adhoc and Sensor Systems (MASS '10)*, pp. 703–708, November 2010.

[8] X. Ma, J. Zhang, X. Yin, and K. S. Trivedi, "Design and analysis of a robust broadcast scheme for VANET safety-related services," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 46–61, 2012.

[9] S. Y. Ni, Y. C. Tseng, Y. S. Chen, and J. P. Sheu, "The broadcast storm problem in a mobile ad hoc network," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, pp. 151–162, 1999.

[10] S. Panichpapiboon and W. Pattara-atikom, "A review of information dissemination protocols for vehicular Ad Hoc networks," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 3, pp. 784–798, 2012.

[11] E. Fasolo, A. Zanella, and M. Zorzi, "An effective broadcast scheme for alert message propagation in vehicular ad hoc networks," in *Proceedings of the 2006 IEEE International Conference on Communications (ICC '06)*, pp. 3960–3965, July 2006.

[12] D. Li, H. Huang, X. Li, M. Li, and F. Tang, "A distance-based directional broadcast protocol for urban vehicular ad hoc network," in *Proceedings of the 2007 International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '07)*, pp. 1520–1523, September 2007.

[13] N. Wisitpongphan, O. K. Tonguz, J. S. Parikh, P. Mudalige, F. Bai, and V. Sadekar, "Broadcast storm mitigation techniques in vehicular ad hoc networks," *IEEE Wireless Communications*, vol. 14, no. 6, pp. 84–94, 2007.

[14] N. Ababneh and H. Labiod, "Safety message dissemination in VANETs: flooding or trajectory-based?" in *Proceedings of the 9th IFIP Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET '10)*, pp. 1–8, June 2010.

[15] L. Bononi and M. Di Felice, "A cross layered MAC and clustering scheme for efficient broadcast in VANETs," in *Proceedings of the 2007 IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems (MASS '07)*, pp. 1–8, October 2007.

[16] L. O. Djedid, N. Lagraa, M. Yagoubi, and K. Tahari, "Adaptation of The MCDS broadcasting protocol to VANET safety applications," in *Proceedings of the 2008 International Conference on Innovations in Information Technology (IIT '08)*, pp. 534–538, December 2008.

[17] M. Durresi, A. Durresi, and L. Barolli, "Emergency broadcast protocol for inter-vehicle communications," in *Proceedings of the 11th International Conference on Parallel and Distributed Systems Workshops (ICPADS '05)*, pp. 402–406, July 2005.

[18] T. Taleb, K. Ooi, and K. Hashimoto, "An efficient collision avoidance strategy for ITS systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '08)*, pp. 2212–2217, April 2008.

[19] J. Nzoonta and C. Borcea, "STEID: a protocol for emergency information dissemination in vehicular networks," Tech. Rep., Department of Computer Science, New Jersey Institute of Technology, 2006.

[20] X. Ma and X. Chen, "Delay and broadcast reception rates of highway safety applications in vehicular ad hoc networks," in *Proceedings of the 2007 Mobile Networking for Vehicular Environments (MOVE '07)*, pp. 85–90, May 2007.

[21] S. Kuribayashi, Y. Sakumoto, H. Ohsaki, S. Hasegawa, and M. Imase, "Performance evaluation of epidemic broadcast with directional antennas in vehicular ad-hoc networks," in *Proceedings of the 11th IEEE/IPSJ International Symposium on Applications and the Internet (SAINT '11)*, pp. 260–265, July 2011.

[22] A. Benslimane, "Optimized dissemination of alarm messages in Vehicular Ad-Hoc Networks (VANET)," *High Speed Networks and Multimedia Communications*, vol. 3079, pp. 655–666, 2004.

[23] S. A. Rao, M. Pai M.M, M. Boussedjra, and J. Mouzna, "GPSR-L: greedy perimeter stateless routing with lifetime for VANETS," in *Proceedings of the 2008 8th International Conference on Intelligent Transport System Telecommunications (ITST '08)*, pp. 299–304, October 2008.

[24] A. Boukerche, C. Rezende, and R. W. Pazzi, "Improving neighbor localization in vehicular Ad Hoc networks to avoid overhead from periodic messages," in *Proceedings of the 2009 IEEE Global Telecommunications Conference (GLOBECOM '09)*, December 2009.

[25] 2013, http://www.mathworks.com/.

[26] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: a framework to systematically analyze the Impact of Mobility on Performance of RouTing protocols for Adhoc NeTworks," in *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM '03)*, IEEE Societies, pp. 825–835, 2003.

[27] A. Vindasius and S. Stanaitis, "Analysis of emergency message transmission delays in vehicular wireless mesh network," in *Proceedings of the 3rd International Conference on Advances in Mesh Networks (MESH '10)*, pp. 35–40, July 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 304601, 14 pages
http://dx.doi.org/10.1155/2013/304601

Research Article

EAP-Based Group Authentication and Key Agreement Protocol for Machine-Type Communications

Rong Jiang,[1,2] Chengzhe Lai,[2,3] Jun Luo,[1] Xiaoping Wang,[1] and Hong Wang[4]

[1] *School of Computer, National University of Defense Technology, Changsha 410073, China*
[2] *Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1*
[3] *State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China*
[4] *Xi'an Communication Institute, Xi'an 710106, China*

Correspondence should be addressed to Rong Jiang; jiangrong@nudt.edu.cn

Received 2 July 2013; Revised 23 August 2013; Accepted 23 August 2013

Academic Editor: Zhong Fan

Copyright © 2013 Rong Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine to machine (M2M) communications, also called machine-type communications (MTC), has widely been utilized in applications such as telemetry, industrial, automation, and SCADA systems. The group-based MTC, especially when MTC devices belong to non-3GPP network, will face new challenge of access authentication. In this paper, we propose a group authentication and key agreement protocol, called EG-AKA, for machine-type communications combining elliptic curve Diffie-Hellman (ECDH) based on EAP framework. Compared with conventional EAP-AKA, our protocol guarantees stronger security and provides better performance. Detailed security analysis has shown that the proposed EG-AKA protocol is secure in terms of user and group identity protection and resistance to several attacks. Furthermore, formal verification implemented in AVISPA proves that the proposed protocol is secure against various malicious attacks. Moreover, performance evaluation demonstrates its efficiency in terms of the signaling overhead, the bandwidth consumption, and the transmission cost.

1. Introduction

Machine to machine (M2M) communications [1], which is also defined as machine-type communications (MTC) [2] in release 10 of the 3rd Generation Partnership Project (3GPP), is one of the hottest issues not only in the standardization but also in the industrial circles. In M2M communications, both wireless and wired systems can communicate with other devices of the same ability. Thanks to MTC, many applications become possible [3, 4]. M2M communications uses a device, such as a sensor or meter, to capture an event (such as temperature and inventory level). Then this event is delivered through a wireless, wired, or hybrid network to an application (software program), which translates the captured event into meaningful information. For example, the event can be translated into what items need to be restocked [5]. Since MTC communications does not need direct human intervention, it is soon becoming a market-changing force for the next-generation intelligent real-time networked applications [6, 7].

Recently, most research on MTC has focused on congestion control, resource management, key management [8, 9], and so forth; however, there are few studies on security aspects. Lu et al. [10] point out that the existing challenges of M2M is energy efficiency (green), reliability, and security (GRS). Taleb and Kunz [11] present some potential challenges and solutions of MTC in 3GPP networks. Some security threats and corresponding solutions of 3GPP are discussed in [12]. Privacy preservation is also an important issue in M2M communications [13–15]. A new group message authentication protocol [16], which utilises only limited authenticated communication, combines short authenticated strings protocol with classical key agreement procedures. This SAS-based group authentication and key agreement protocol is secure against active attacks. If mobile terminals of non-3GPP short-distance wireless communication want to access the 3GPP core network, they must execute access authentication. Most access authentication protocols are based on Extensible Authentication Protocol (EAP), such as EAP-AKA [17], EAP-TTLS [18], EAP-PEAP [19], EAP-LEAP [20], and EAP-SPEKE

[21]. However, the existing access authentication protocols cannot provide enough security for MTC [22]; on the other hand, present standard has not considered the group-based access authentication. Recently, several standardization organizations start to present the concept and requirement of group authentication, but the mechanism and procedure have not yet been developed.

To the best of our knowledge, the existing network authentication systems are mainly designed for a single object, and they all need 3 or 4 rounds of interaction to realize the mutual authentication between a user and a server. In practical applications, however, there may be a large number of users with the same properties in a network, such as MTC, and user terminals can form a group when they are in the same region, or belong to the same application, or have the same behavior. In these applications, if substantial numbers of user terminals of a group access the network over a short period of time successively, the available authentication methods may suffer from network congestion by the increasing signal of the network. In order to prevent network from congesting and efficiently authenticate user terminals of a group, the concept of group authentication, which performs authentication for group units, is introduced. As a kind of network authentication technology, group authentication aims to authenticate multiple or all users over a shorter period of time. In this technology, the group is assigned a unique identifier, and user terminals are authenticated together as corporate entities. Group authentication can be fulfilled by utilizing the authentication agency or the gateway. After successful group authentication, user terminals and network side entities can share some keys.

In the current literature, a few authentication protocols of group communication have been proposed. An individual and group authentication model, which uses dynamic key cryptography and group key management for individual and group of users and services, is proposed for wireless network services [23]. Chen et al. propose G-AKA protocol for a group of mobile stations roaming from the same home network to a serving network [24]. Aboudagga et al. propose a group authentication protocol for mobile networks and design a new architecture for authentication management and an associated authentication protocol for mobile groups and individual nodes over heterogeneous domains [25]. However, there are still no appropriate group authentication methods for MTC in 3GPP. On the other hand, EAP-AKA [17] is an important authentication and key agreement protocol between 3G/LTE and non-3GPP, but EAP-AKA does not support group authentication mechanism and cannot be applied to group-based MTC. In addition, there are some vulnerabilities in EAP-AKA, such as disclosure of user identity, man-in-the-middle attack [26].

In this paper, in order to resolve group access authentication for MTC, we propose a novel group authentication and key agreement protocol based on Mun's protocol [26], named EG-AKA. Our protocol guarantees stronger security and provides better performance than the existing protocols. The main idea of our protocol is that the first MTC device of a group, which wants to access to 3GPP core network, performs a full AKA authentication procedure. In this process, the first MTC device obtains group authentication information and group temporary key (GTK) on behalf of other MTC devices of the same group. Then the authentication, authorization, and accounting server (AAA server) is enabled to carry out mutual authentication with remaining MTC devices of the group using obtained group authentication information and GTK without interacting with the home subscriber server (the HSS). The authentication delay can be decreased as a whole and the signaling overhead between the AAA server and the HSS is considerably reduced.

The remainder of this paper is organized as follows. In Section 2, we will introduce relevant background and knowledge. In Section 3, we propose our group authentication protocol. In Section 4, the authentication and other secrecy properties are verified by the model checking tools, and detailed performance evaluations are given in Section 5. Finally, we draw our conclusion and give the future work in Section 6.

2. Background

Before going to the details of the proposed protocol, we first recall the elliptic curve Diffie-Hellman technique [27], Mun's Protocol [26], which serves as the basis of the proposed EG-AKA protocol. Then, we present the abbreviations and network architecture used in this paper.

2.1. Elliptic Curve Diffie-Hellman. Elliptic curve cryptography (ECC), which is based on the algebraic structure of elliptic curves over finite fields, is a famous approach used in public-key cryptography. This cryptography was first proposed in 1985 independently by Koblitz [28] and Miller [29]. The primary advantage of ECC is that the key size is smaller while providing the same level of security, which can reduce storage and transmission requirements; that is, an elliptic curve group could provide the same level of security afforded by an RSA-based system with a large modulus and correspondingly larger key. For example, a 160 bit ECC public key should provide comparable security to a 1024 bit RSA public key. Elliptic curve Diffie-Hellman (ECDH) is an anonymous key agreement protocol that allows two parties, each having an elliptic curve public-private key pair, to establish a shared secret over an insecure channel [30]. This shared secret may be directly used as a key, or better yet, to derive another key which can then be used to encrypt subsequent communications using a symmetric key cipher. It is a variant of the Diffie-Hellman protocol using elliptic curve cryptography.

Key establishment protocol of elliptic curve Diffie-Hellman is described briefly as follows. Suppose Alice wants to establish a shared key with Bob, but the channel available for them is not secure and may be eavesdropped by the others. Initially, the domain parameters (i.e., (p, a, b, G, n, h) in the prime case or $(m, f(x), a, b, G, n, h)$ in the binary case) must be agreed upon. Also, each party must have a key pair suitable for elliptic curve cryptography, consisting of a private key d (a randomly selected integer in the interval $[1, n-1]$) and a public key Q (where $Q = dG$, that is, the result of adding

G together d times). Let Alice's key pair be (d_A, Q_A) and Bob's key pair be (d_B, Q_B). Each party must have the other party's public key (an exchange must occur). Alice computes $(x_k, y_k) = d_A Q_B$. Bob computes $(x_k, y_k) = d_B Q_A$. The shared secret is x_k (the x coordinate of the point). Most standardized protocols based on ECDH derived a symmetric key from x_k using some hash-based key derivation function. The shared secret calculated by both parties is equal, because $d_A Q_B = d_A d_B G = d_B d_A G = d_B Q_A$. The only information about her private key that Alice initially exposes is her public key. So, no party other than Alice can determine Alice's private key, unless that party can solve the elliptic curve discrete logarithm problem. Bob's private key is similarly secure. No party other than Alice or Bob can compute the shared secret, unless that party can solve the elliptic curve Diffie-Hellman problem [27].

2.2. Mun's Protocol. Mun et al. [26] propose a new authentication and key agreement protocol based on EAP-AKA designed for 3G-WLAN interworking. This protocol combines elliptic curve Diffie-Hellman (ECDH) with symmetric key cryptosystem to overcome several vulnerabilities. In addition, their protocol provides perfect forward secrecy (PFS) to guarantee stronger security, mutual authentication, and resistance to replay attack. The major advantages of their protocol can be summarized as follows:

(1) providing strong user identity protection by encrypted IMSI using shared secret key between user equipment and HSS;

(2) using ECDH to provide perfect forward secrecy between the user equipment and the AAA server;

(3) resisting against three types of man-in-the middle attack.

Mun's protocol can guarantee stronger security; however, similar to EAP-AKA, the protocol is not suitable for group-based MTC due to lack of specific mechanism. We will modify Mun's protocol to design a novel security enhanced group authentication protocol for MTC.

2.3. Network Architecture. In order to avoid confusing, we list the abbreviations used throughout the rest of this paper in Table 1.

The network architecture mainly consists of four parts: machine-type communication devices, access point, the authentication, authorization, and accounting server, and the home subscriber server, as shown in Figure 1.

Machine-Type Communication (MTC) Devices. An MTC device, which communicates through a public land mobile network (PLMN), is a device equipped for machine-type communications.

Access Point (AP). AP is a device that allows wireless devices to connect to a wired network using Wi-Fi, Bluetooth, or other related standards.

The Authentication, Authorization, and Accounting (AAA) Server. In the LTE network, the authentication, authorization, and accounting (AAA) server provides access authentication services for MTC devices on behalf of the 3GPP core network.

The Home Subscriber Server (HSS). In the LTE network, the home subscriber server (HSS) locates in 3GPP core network and provides authentication and management services for MTC devices on behalf of 3GPP core network.

Table 1: Abbreviation used in the paper.

Abbreviation	Definition
3GPP	3rd generation partnership project
3G LTE	3G long term evolution
AAA Server	The authentication, authorization, and accounting server
AKA	Authentication and Key Agreement
AP	Access point
AVISPA	Automated validation of internet security protocols and applications
CN	Core network
EAP	Extensible authentication protocol
ECC	Elliptic curve cryptography
ECDH	Elliptic curve Diffie-Hellman
GAK	Group authentication and key agreement Protocol for MTC
GK	Group Key
GTK	Group temporary key
HSS	Home subscriber server
IMSI	International mobile subscriber identification number
M2M	Machine to machine
MAC	Message authentication code
MSK	Master session key
MTC	Machine-type communications
PFS	Perfect forward secrecy
PID	Permanent identity
PLMN	Public land mobile network

Figure 1: Network architecture of MTC.

3. The Proposed Group Authentication Protocol

In this section, we give the details of the group authentication and key agreement protocol for MTC (EG-AKA) to facilitate non-3GPP MTC devices to access to 3GPP core network (CN). In order to achieve this aim, there are three phases in the proposed protocol: group initialization, authentication

data distribution, and mutual authentication and key agreement.

3.1. Group Initialization. In the group initialization phase, each MTC device has a permanent ID (PID), such as international mobile subscriber identification number (IMSI). This PID is a long-term private identity that identifies MTC device and should be installed in the MTC device by the supplier in order to allow the MTC device to register in a 3GPP network. At the same time, we assume that each MTC device has preshared a secret key with 3GPP CN, and these MTC devices form several groups based on certain principles, and then the supplier provides a group key (GK) to each group for authentication. As shown in Table 2, we create an index table to manage information of MTC devices and group; the index table contains fields of group identity, MTC device identity (PID) for each MTC device, and initial values. Table 3 is the protocol notations used in this paper.

3.2. Authentication Data Distribution. Let $MTCD_{G1-1}$ be the first MTC device initiating authentication in group 1. We assume that a secure communication channel between the AAA server and the HSS has already been established and can provide security services to the transmitted data. The authentication data distribution processes as follows.

Step 1. $MTCD_{G1-1}$ sends an access request message to the AP.

Step 2. AP sends an EAP Request/Identity message to require the identity of $MTCD_{G1-1}$.

Step 3. Upon receiving the EAP Request/Identity message sent by AP, firstly, the $MTCD_{G1-1}$ computes

$$TID_{MTCD_{G1-1}} = f^1_{K_{MTCD_{G1-1}}-HSS}\left(PID_{MTCD_{G1-1}}\right), \quad (1)$$

$$TID_{G1-1} = f^1_{K_{MTCD_{G1-1}}-HSS}\left(PID_{G1}\right), \quad (2)$$

respectively, and then $MTCD_{G1-1}$ generates $AUTH_{G1}$ as follows:

$$\begin{aligned} AUTH_{G1} = \big(&TID_{G1} \parallel TID_{MTCD_{G1-1}} \parallel \\ &R_{MTCD_{G1-1}} \parallel MAC_{MTCD_{G1-1}} \parallel \\ &ID_{HSS} \parallel ID_{AAA} \parallel ID_{AP}\big), \end{aligned} \quad (3)$$

where $MAC_{MTCD_{G1-1}}$ is calculated as

$$\begin{aligned} &MAC_{MTCD_{G1-1}} \\ &\quad = f^2_{K_{MTCD_{G1-1}}-HSS}\left(R_{MTCD_{G1-1}} \parallel ID_{AAA} \parallel ID_{AP}\right). \end{aligned} \quad (4)$$

Step 4. $MTCD_{G1-1}$ sends its $AUTH_{G1}$ to the AAA server through AP, and then the AAA server finds out corresponding HSS according ID_{HSS} and forwards $AUTH_{G1}$ and its own ID_{AAA} to the HSS by authentication data request message.

Table 2: Index table.

Group	Group ID	MTC device ID	Initial value
$G1$	PID_{G1}	PID_{G1-1}	IV_{G1-1}
		PID_{G1-2}	IV_{G1-2}
		$\vdots$	
		PID_{G1-n}	IV_{G1-n}
$G2$	PID_{G2}	PID_{G2-1}	IV_{G2-1}
		$\vdots$	

Table 3: Protocol notation.

Notation	Definition
$MTCD_{G1-1}$	The first MTC device initiating authentication in group 1
R_x	The random number generated by x
ID_x	The identity of x
PID_x	The permanent identity of x
TID_x	The temporary identity of x
K_{x-y}	The shared secret key between x and y
GK_{Gi}	The group authentication key of the ith group
GTK_{Gi}	The group temporary key of the ith group
f^1_k	Temporary identity generation function using k
f^2_k	MAC generation function using k
f^3_k	GTK generation function using k
f^4_k	Shared key generation function using k

Table 4: Temporary index table of $G1$.

Group	Group ID	MTC device ID	Initial value
$G1$	TID_{G1-1}	$TID_{MTCD_{G1-1}}$	IV_{G1-1}
		$TID_{MTCD_{G1-2}}$	IV_{G1-2}
		$\vdots$	
	TID_{G1-n}	$TID_{MTCD_{G1-n}}$	IV_{G1-n}

Step 5. When the HSS receives authentication data request message containing $MTCD_{G1-1}$'s $AUTH_{G1}$ and ID_{AAA}, it verifies the received $MAC_{MTCD_{G1-1}}$ in $AUTH_{G1}$.

If verification passes, the HSS derives $PID_{MTCD_{G1-1}}$ and PID_{G1} from $TID_{MTCD_{G1-1}}$ and TID_{G1} using $K_{MTCD_{G1-1}}$, respectively. Then HSS retrieves the corresponding group key GK_{G1} to generate a group temporary key $GTK_{G1} = f^3_{GK_{G1}}(R_{HSS} \parallel ID_{HSS})$.

Step 6. At the moment, the HSS also computes all temporary identities of the devices in group 1 and generates a temporary index table (as shown in Table 4) of group 1; then the HSS sends ID_{AP}, GTK_{G1}, R_{HSS}, and temporary index table to the AAA server by a preestablish security tunnel.

Step 7. The AAA server receives and stores ID_{AP}, GTK_{G1}, R_{HSS}, and temporary index table for future use.

3.3. Mutual Authentication and Key Agreement

Step 8. The AAA server generates R_{AAA} and computes MAC_{AAA} as follows:

$$MAC_{AAA} = f^2_{GTK_{G1}}\left(R_{AAA} \parallel R_{HSS} \parallel R_{MTCD_{G1-1}} \parallel IV_{G1-1} + i\right), \quad (5)$$

where i represents the ith run of mutual authentication with $MTCD_{G1-1}$. After that, the AAA server selects random number a and computes aP on E.

Step 9. The AAA server generates

$$AUTH_{AAA} = \left(MAC_{AAA} \parallel R_{AAA} \parallel R_{HSS}\right) \quad (6)$$

and sends $AUTH_{AAA}$ and aP to $MTCD_{G1-1}$.

Step 10. After receiving $AUTH_{AAA}$, $MTCD_{G1-1}$ verifies the received MAC_{AAA} in $AUTH_{AAA}$ as follows.

(1) Firstly, $MTCD_{G1-1}$ computes

$$GTK_{G1} = f^3_{GK_{G1}}\left(R_{HSS} \parallel ID_{HSS}\right); \quad (7)$$

(2) then, $MTCD_{G1-1}$ computes

$$MAC'_{AAA} = f^2_{GTK_{G1}}\left(R_{AAA} \parallel R_{HSS} \parallel R_{MTCD_{G1-1}} \parallel IV_{G1-1} + i\right); \quad (8)$$

(3) $MTCD_{G1-1}$ verifies whether MAC'_{AAA} equals MAC_{AAA} or not. If MAC'_{AAA} is not the same as MAC_{AAA}, the HSS or the AAA server is not valid. Therefore, the $MTCD_{G1-1}$ terminates the procedure.

Step 11. If verification is successful, $MTCD_{G1-1}$ computes bP, $K_{MTCD_{G1-1}-AAA} = f^4_{GTK_{G1}}(abP)$ and $MAC_{MTCD_{G1-1}-AAA} = f^1_{K_{MTCD_{G1-1}-AAA}}(R_{AAA} \parallel bP)$.

Step 12. $MTCD_{G1-1}$ sends $MAC_{MTCD_{G1-1}-AAA}$ and bP to the AAA server by authentication response message, at the same time, $MTCD_{G1-1}$ also calculates the MSK as EAP-AKA.

Step 13. When the AAA server receives $MAC_{MTCD_{G1-1}-AAA}$ and bP, it also computes $K_{MTCD_{G1-1}-AAA} = f^4_{GTK_{G1}}(abP)$ using bP and verifies $MAC_{MTCD_{G1-1}-AAA}$. If verification passes, AAA server also calculates the MSK as EAP-AKA.

Step 14. The AAA server sends $ID_{AP} \parallel MSK$ with EAP Success message to the AP.

Step 15. The AP verifies whether received ID_{AP} equals its own ID or not. If the result is incorrect, the AP drops the MSK and then terminates the execution. Otherwise the AP stores the MSK. Then AP encrypts ID_{AP} using the MSK and sends it with EAP Success message to $MTCD_{G1-1}$.

Step 16. Through decryption, $MTCD_{G1-1}$ recovers ID_{AP} and verifies whether or not the ID_{AP} received from the AP in Step 15 equals to the ID_{AP} used in Step 4. If the result is correct, the procedure of authentication and key agreement is successful. Consequently, $MTCD_{G1-1}$ can securely access to 3GPP CN using the MSK.

At this point, the full authentication and key agreement procedure for one MTC device is completed. The procedure is shown in Figure 2.

When other MTC device in the same group want to access the 3GPP CN, the AAA server performs mutual authentication and key agreement with $MTCD_{G1-2}$ locally using the existing GTK_{G1}. Taking the MTC device $MTCD_{G1-2}$ in the same group as an example, the full authentication and key agreement procedure for it is described as follows.

Steps 1 and 2 are similar to $MTCD_{G1-1}$s.

Step 3.* Upon receiving EAP request/identity message by AP, similarly, the $MTCD_{G1-2}$ computes $TID_{MTCD_{G1-2}} = f^1_{K_{MTCD_{G1-2}}-HSS}(PID_{MTCD_{G1-2}})$ and $TID_{G1-2} = f^1_{K_{MTCD_{G1-2}}-HSS}(PID_{G1})$, respectively, and then $MTCD_{G1-2}$ generates $AUTH_{G2}$ as follows:

$$AUTH_{G1} = \left(TID_{G1} \parallel TID_{MTCD_{G1-1}} \parallel R_{MTCD_{G1-1}} \parallel MAC_{MTCD_{G1-1}} \parallel ID_{AAA}\right), \quad (9)$$

where $MAC_{MTCD_{G1-2}}$ is calculated as

$$MAC_{MTCD_{G1-2}} = f^2_{K_{MTCD_{G1-2}}-HSS}\left(R_{MTCD_{G1-2}} \parallel ID_{AAA} \parallel ID_{AP}\right). \quad (10)$$

Step 4.* $MTCD_{G1-2}$ sends its $AUTH_{G1}$ to the AAA server through AP. Note that, the AAA server does not need to authenticate the group (G1) which $MTCD_{G1-2}$ belongs to by the HSSs assistance.

Step 5.* The AAA server begins to perform mutual authentication with $MTCD_{G1-2}$ using the temporary index table (Table 4) and GTK_{G1} received in Step 6.

The remaining steps are similar to $MTCD_{G1-1}$s.

The other MTC devices perform the authentication and key agreement procedures similar to $MTCD_{G1-2}$s until all devices complete the authentication.

4. Security Analysis

In this section, both security analysis and formal verification implemented by the AVISPA tool are conducted to show that the proposed protocol can work correctly to achieve security properties.

4.1. Security Property. In Table 5, we compare our proposed EG-AKA protocol with the other main AKA protocols: Mun's protocol [26], EAP-AKA [17], EAP-TTLS [18], EAP-PEAP

Figure 2: Authentication procedure of the first MTC device in our proposed protocol.

Table 5: Comparisons of properties among the EAP-based AKA protocols.

	Our proposed EG-AKA	Mun's protocol [26]	EAP-AKA [17]	EAP-TTLS [18]	EAP-PEAP [19]	EAP-LEAP [20]	EAP-SPEKE [21]
Type of cryptosystem	Symmetric and ECDH	Symmetric and ECDH	Symmetric	Public	Public	Public	Public
Computational overhead	Smaller	Smaller	Smallest	Large	Large	Large	Large
Protection of user identity	Yes	Yes	No	Yes	Yes	No	No
Heterogeneous network access	Yes	Yes	Yes	No	No	No	No
Secure against man-in-the middle attack	Yes	Yes	No	Yes	Yes	Yes	Yes
Secure against replay attack	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PFS	Yes	Yes	Yes	Yes	No	No	No
Support group authentication	Yes	No	No	No	No	No	No

[19], EAP-LEAP [20], and EAP-SPEKE [21]. The comparison results demonstrate that our protocol can provide the most comprehensive security performance compared to the other AKA protocols. Providing group access authentication and heterogeneous network access are the two main advantages of our protocol. In particularly, our proposed protocol meets the following security properties.

Protect User and Group Permanent Identity. In our protocol, PID cannot be got by attackers. The reason is that the MTC device generates the TID by using the $K_{MTCD_{Gi-j}}$ and then sends TID to the HSS. Therefore, the MTC device and the HSS can only retrieve user and group permanent identity included in TID through using $K_{MTCD_{Gi-j}}$. Thus, our protocol provides strong user and group identity protection.

Secure against Man-in-the Middle Attack. In our proposed protocol, only the MTC devices and HSS can obtain real ID information of the devices and the group from encrypted temporary ID information. An attacker cannot derive and modify this information. The AP receives the EAP Success message with $ID_{AP} \| MSK$ sent by the AAA server. After

that, the AP can verify whether its own ID equal to the received ID or not. If not the procedure of authentication and key agreement will fail. Furthermore, the AP will send ID_{AP} encrypted by MSK to the MTC device. The MTC device can verify whether it has accessed this AP or not. The MTC device can verity the legality of HSS by MAC_{AAA} as well. Thus, our protocol can resist against several types of man-in-the middle attack.

Secure against Replay Attack. In our protocol, random numbers $R_{MTCD_{Gi-j}}$ generated by $MTCD_{Gi-j}$, R_{HSS} generated by the HSS and R_{AAA} generated by the AAA server are temporarily used in generating challenge messages toward the opposite side, respectively. Since these random numbers used in each authentication procedure are different, even if an attacker acquires a random number in a authentication procedure, he still cannot fake challenge messages by reusing the random number in a new authentication procedure. Meanwhile, these two sites maintain an identical initial value IV_{Gi-j} to keep themselves synchronized throughout AKA processing. An out-of-sync initialization value will lead to authentication failure. Thus a node without the required random numbers and initial value cannot perform a replay attack on our system.

Resistance to Impersonate Attack. Note that, in our protocol, all the MTC devices of a group share a common GTK. If an MTC device, without loss of generality, suppose that $MTCD_{G1-1}$ intends to impersonate another MTC device in the same group, for example, $MTCD_{G1-j}$. $MTCD_{G1-1}$ may eavesdrop traffic between $MTCD_{G1-j}$ and the HSS, but $MTCD_{G1-1}$ cannot generate unique R_{G1-j} and IV_{G1-j}. Therefore, $MTCD_{G1-1}$ cannot generate a correct $MAC_{MTCD_{G1-j}-AAA}$ to impersonate $MTCD_{G1-j}$ to perform a successful authentication with the HSS. Similarly, $MTCD_{G1-1}$ cannot get the $K_{MTCD_{G1-j}-AAA}$ between $MTCD_{G1-j}$ and the AAA server. Therefore, it cannot decrypt traffic between $MTCD_{G1-j}$ and the AAA server. In summary, the 3GPP CN can easily distinguish one MTC device from another even though all MTC devices use the same GTK; at the same time, one MTC device cannot decrypt traffic between any other MTC device and the 3GPP CN.

Perfect Forward Secrecy (PFS). Our protocol utilizes ECDH to provide PFS between the MTC device and the AAA server. While generating $K_{MTCD_{Gi-j}-AAA}$, our protocol uses aP and bP that are not related with $K_{MTCD_{Gi-j}-HSS}$. Therefore, if disclosure of $K_{MTCD_{Gi-j}-HSS}$ occurs, attackers cannot get $K_{MTCD_{Gi-j}-AAA}$. In other words, guessing $K_{MTCD_{Gi-j}-AAA}$ is a computationally difficult problem.

Provide Mutual Authentication and Key Agreement. We can verify that the proposed protocol can provide a successful mutual authentication between MTC devices and the 3GPP CN by formal verification described in the Section 4.2. Key agreement includes two parts: (a) between the MTC device and the AAA server: the key agreement between the MTC device and the AAA server can achieve through ECDH with symmetric key, and the MTC device and the AAA server can share a secret key $K_{MTCD_{Gi-j}-AAA}$ by Steps 11–13; (b) between the MTC device and the AP: the key agreement between the MTC device and the AP is the same as EAP-AKA [7], and the MTC device and AP can securely communicate with other by the MSK.

```
goal
secrecy_of kma
authentication_on mtcd_aaa
authentication_on aaa_mtcd
end goal
```

Figure 3: Analysis goals of the model.

4.2. Formal Verification. The primary goal of our proposed protocol is to provide mutual authentication and key agreement services between MTC devices and the 3GPP CN. We tested our protocol using formal security verification tool known as the "Automated Validation of Internet Security Protocols and Applications" (AVISPA) [31]. The AVISPA project aims at developing a push-button, industrial-strength technology for the analysis of large-scale Internet security-sensitive protocols and applications. This technology will speed up the development of the next generation of network protocols, improve their security, and therefore increase the public acceptance of advanced, distributed IT applications based on them. AVISPA will achieve this by advancing specification and deduction technology to the point where industry protocols can be specified and automatically analyzed. A central aim of the project is then to integrate this technology into a robust automated tool, tuned on practical, large-scale problems, and migrated to standardization bodies, whose protocol designers are in dire need of such tools. In the AVISPA tool, protocols are specified using the High Level Protocol Specification Language (HLPSL for short). Then, the HLPSL specification is translated into an Intermediate Format which is used by the various verification tools embedded in AVISPA. We use On-the-fly-Model-Checker (OFMC) and SAT-based model checker (SATMC) to text our EG-AKA protocol. The authentication goals that we need to verify are shown in Figure 3. The output of the model checking results are shown in Figures 4 and 5. We can conclude that the proposed protocol can accomplish the goal of mutual authentication, and it can resist those malicious attacks such as replay attacks, MitM attacks, and secrecy attacks under the test of AVISPA using the OFMC back-end and SATMC back-end.

5. Performance Evaluation

In this section, we give a detailed performance evaluation of the proposed protocol from the signaling overhead and the transmission cost point of view.

5.1. Signaling Overhead. In order to evaluate the signaling overhead, we consider the following scenario: the number

```
% OFMC
SUMMARY
SAFE
DETAILS
BOUNDED_NUMBER_OF_SESSIONS
PROTOCOL
D:\SPAN\testsuite\results\EG-AKA.if
GOAL
as_specified
BACKEND
OFMC
COMMENTS
STATISTICS
parseTime: 0.00 s
searchTime: 0.01 s
visitedNodes: 24 nodes
depth: 4 plies
```

Figure 4: Results reported by the OFMC back-end.

```
SUMMARY
SAFE
DETAILS
STRONGLY_TYPED_MODEL
BOUNDED_NUMBER_OF_SESSIONS
BOUNDED_SEARCH_DEPTH
BOUNDED_MESSAGE_DEPTH
GOAL
as_specified
BACKEND
  SATMC

COMMENTS

STATISTICS
  attackFound                  false      boolean
  upperBoundReached            true       boolean
  graphLeveledOff              3          steps
  satSolver                    zchaff     solver
  maxStepsNumber               11         steps
  stepsNumber                  3          steps
  atomsNumber                  0          atoms
  clausesNumber                0          clauses
  encodingTime                 0.09       seconds
  solvingTime                  0          seconds
  if2sateCompilationTime       0.66       seconds

ATTACK TRACE
no attacks have been found..
```

Figure 5: Results reported by the SATMC back-end.

of MTC device is n, and the number of group is m. Suppose that each MTC device launches x (re)authentications. For EAP-AKA, authentication procedures performed by an MTC device require the total number of signaling messages which grows linearly with x. In EAP-AKA protocol, there are 12 signaling messages for one complete authentication procedure. Thus, the number of signaling message of a MTC device is $12x$ and the total number of signal message is $12nx$. In Mun's protocol, the MTC device runs a full authentication using 8 messages at one time, a total of $8x$ messages is required. Similarly, when n MTC devices belonging to m group perform authentication, there are a total of $12nx$ messages for EAP-AKA, and a total of $8nx$ messages for Mun's protocol. In the proposed protocol, the first MTC device initiating authentication in the group complete the whole procedure of authentication and the number of signaling

Figure 6: Comparison of the number of signaling messages of several EAP-based protocols.

message is 8. The rest devices of the group only need 6 signaling messages. In this scenario, the number of the rest devices is $n - m$ and the total number of signaling message is $8m + 6(n - m)$. If each device executes another $x - 1$ re-authentications, then the total number of signaling message is $8m + 6(n - m) + 6n(x - 1)$. Figure 6 illustrates the number of signaling messages of the proposed procedure over the existing authentication protocols for several different cases. It can be seen that signaling messages of several AKA protocols are increasing as the number of MTC devices increases. Among three AKA protocols, our EG-AKA outperforms other protocols. This is because our protocol shifts the impact of the number of MTC devices on network to the impact of that of the number of MTC device groups on network; our EG-AKA can reduce both authentication delay and signaling overhead within the core network.

5.2. Bandwidth Consumption. In order to analyze the bandwidth consumption, we assume that x AVs are transmitted every time the HSS successfully authenticates one ME, and there are n MTCDs forming m group. Without loss of generality, Table 6 shows the setting of parameters for evaluating bandwidth consumption.

The bandwidth consumption of AKA protocols are as follows, where bw_{first} represents the bandwidth consumption of the authentication of the first MTCD.

(1) Bandwidth analysis of EAP-AKA: the sizes of authentication messages are calculated as follows:

$$bw_{first} = \sum_{i=1}^{5} |Message_i| = 704 + 608x \text{ bits}. \quad (11)$$

Table 6: Setting of parameters.

Parameters	Value (bits)
ID/TID	128
GTK	128
MAC	64
Random number (RN)	128
ECDH key	192
Timestamp	32

The overall bandwidth consumption for n devices is calculated as $n \times (704 + 608x)$.

(2) Bandwidth analysis of Mun's scheme: the sizes of authentication messages are calculated as follows:

$$bw_{first} = \sum_{i=1}^{7} |Message_i| = 2432 \text{ bits.} \quad (12)$$

(i) $Message_1 = 2|ID| + |Timestamp| + |MAC| = 352$ bits.

(ii) $Message_2 = 2|ID| + |Timestamp| + |MAC| = 352$ bits.

(iii) $Message_3 = |RN| + |MAC| + |TK| + |ID| = 448$ bits.

(iv) $Message_4 = 2|RN| + |MAC| + |ECDH\ key| = 512$ bits.

(v) $Message_5 = |ID| + |ECDH\ key| = 256$ bits.

(vi) $Message_6 = |ID| + |MSK| = 256$ bits.

(vii) $Message_7 = |ID| + |MSK| = 256$ bits.

The overall bandwidth consumption for n devices is calculated as $n \times 2432$.

(3) Bandwidth analysis of EG-AKA: the sizes of authentication messages are calculated as follows:

$$bw_{first} = \sum_{i=1}^{6} |Message_i| = 2688 \text{ bits.} \quad (13)$$

(i) $Message_1 = 3|ID| + |RN| + |MAC| = 576$ bits.

(ii) $Message_2 = 4|ID| + |RN| + |MAC| = 704$ bits.

(iii) $Message_3 = |RN| + |GTK| + |ID| = 384$ bits.

(iv) $Message_4 = 2|RN| + |MAC| + |ECDH\ key| = 512$ bits.

(v) $Message_5 = |MAC| + |ECDH\ key| = 256$ bits.

(vi) $Message_6 = |ID| + |MSK| = 256$ bits.

Consider

$$bw_{remaining} = \sum_{i=1}^{3} |Message_i| = 1024 \text{ bits,} \quad (14)$$

where $bw_{remaining}$ represents the bandwidth consumption of authentication of each remaining ME.

(i) $Message_1 = 2|RN|+|MAC|+|ECDH\ key| = 512$ bits.

(ii) $Message_2 = |MAC| + |ECDH\ key| = 256$ bits.

(iii) $Message_3 = |ID| + |MSK| = 256$ bits.

The overall bandwidth consumption for n devices is calculated as $m * 2688 + (n - m) \times 1024$.

Figure 7 shows the bandwidth consumption of several AKA protocols, when the number of the MEs is different. From Figures 7(a) to 7(d), we can see that the bandwidth consumption of our EG-AKA protocol is much better than that of EPS-AKA and Mun's scheme. Meanwhile, our EG-AKA protocol can provide much better security compared to the other protocols.

5.3. Transmission Cost. In order to evaluate the transmission cost, assume that energy dissipated during 1-message transmission between MTC device and HSS is 1 unit, the energy dissipated during 1-message transmission between MTC device and AAA server is a unit ($a < 1$), and energy dissipated during 1-message transmission between AAA server and HSS is b unit ($b < 1$). Assume that the number of devices in a group is y.

Since the other EAP-AKA based protocols only enhance the security aspect and the procedure of signaling mode is the same as the traditional EAP-AKA protocol, we only compare our proposed protocol with the traditional EAP-AKA protocol. We consider the following two case as shown in Figure 2 in our proposed protocol:

(a) the AAA server has to fetch the fresh authentication vector form the HSS;

(b) the AAA server already has the fresh authentication vector.

In case (a), there are 4 messages between the MTC device and the AAA server, and there are 2 messages between the AAA server and HSS during one authentication procedure. The communication cost of our proposed protocol in this case is

$$C_{prop1} = (4a + 2b) + (y - 1) \times 4a. \quad (15)$$

In case (b), since the AAA server already has the fresh authentication vector, it does not need to communicate with the HSS anymore. Thus, the communication cost of our proposed protocol in this case is

$$C_{prop2} = 4ay. \quad (16)$$

Similarly, in the EAP-AKA protocol, there are 8 messages between the MTC device and the AAA server, and there are 2 messages between the AAA server and HSS during one authentication procedure. Therefore, the communication cost of the EAP-AKA protocol in case (a) is

$$C_{EAP-AKA1} = (8a + 2b)\, y \quad (17)$$

and in case (b) is

$$C_{EAP-AKA2} = 8ay. \quad (18)$$

Figure 7: Comparison of the bandwidth consumption of several EAP-based protocols.

Suppose that the AAA server fetches n authentication vectors during the authentication procedure. The average communication cost of the proposed protocol is

$$C_{prop} = \frac{1}{n}C_{prop1} + \frac{n-1}{n}C_{prop2} = \frac{4any + 2b}{n}. \quad (19)$$

The average communication cost of the EAP-AKA protocol is

$$C_{EAP-AKA} = \frac{1}{n}C_{EAP-AKA1} + \frac{n-1}{n}C_{EAP-AKA2} = \frac{8any + 2by}{n}. \quad (20)$$

(a) $n = 1$ (b) $n = 3$

(c) $n = 5$ (d) $n = 10$

Figure 8: Comparison of the *IR*.

We define a improvement rate *IF* to evaluate the improvement of our proposed protocol compared to the EAP-AKA protocol. The definition of improvement rate *IR* is:

$$IR = \frac{C_{EAP-AKA} - C_{prop}}{C_{EAP-AKA}} = \frac{4any + 2by - 2b}{8any + 2by}. \quad (21)$$

From the definition of *IR*, we know that the bigger the *IR* is, the smaller the transmission cost of our proposed protocol is. Figure 8 plots the improvement rate *IR* varying with the number of devices, the number of fetched authentication vectors, and the energy dissipated during 1-message transmission between the MTC device and AAA server. From the figures, we can easily see that the more the number of MTC devices in the group is, the bigger the *IR* is. The reason is that in our proposed protocol we only need one communication between the AAA server and the HSS for the whole group authentication. While in the EAP-AKA protocol each MTC device has to execute a complete authentication. Furthermore, the more number of authentication vector the AAA server fetches from the HSS, the bigger the *IR* is. The reason is that our proposed protocol only needs one authentication vector for the whole group. The communication cost can be reduced dramatically.

6. Conclusion and Future Work

In this paper, we propose a group authentication and key agreement protocol for MTC device under the EAP framework, named EG-AKA. To the best of our knowledge, there is no protocol in the current literature that handles specific group access authentication for non-3GPP MTC. The proposed EG-AKA protocol not only enhances security on the basis of Mun's protocol, but also design specific group authentication mechanism for MTC. Formal verification and security analysis show that the proposed protocol is secure and fulfill its design goals. Detailed evaluations of performance illustrate that the proposed protocol achieves better performance in terms of transmission and signaling overhead compared with several existing protocols. In our future work, we will consider more practical group authentication protocol based on symmetric cryptography for resource-constrained devices in heterogeneous networks.

```
role mtcd (MTCD, AAA: agent, GK:
    symmetric_key, F1, F2, F3, F4: hash_func,
    SND, RCV: channel (dy))
played_by MTCD def =
local

        State: nat,
        Rmtcd, Rhss, Raaa, IDhss, Key: text,
        Kma: message

init
State := 0
transition
(1)     State = 0/\RCV (start) = |>
        State' := 2/\Rmtcd' := new ()
                        /\SND (Rmtcd')
(2)     State = 2/\RCV ({F2(Raaa'. Rhss'.
        Rmtcd)}_({F3(IDhss'. Rhss')}_GK).
        Raaa'. Rhss')
/\witness (MTCD, AAA, aaa_mtcd, Raaa', Rhss')
        = |>
        State' := 4/\Kma' := {F4(Key')}_({F3(
            IDhss'. Rhss')}_GK)/\secret (Kma',
            kma, {AAA, MTCD})
(3)     State = 4/\SND ({F1(Raaa. Key')}_Kma)/\
          wrequest (MTCD, AAA, mtcd_aaa, Raaa) = |>
State' := 6
end role % the role of MTCD

role aaa (MTCD, AAA: agent, GK:
        symmetric_key, F1, F2, F3, F4: hash_func,
        SND, RCV: channel (dy))
played_by AAA def =
local

        State: nat,
        Rmtcd, Rhss, Raaa, IDhss, Key: text,
        Kma: message

init
State := 1
transition
(1)     State = 1/\RCV (Rmtcd') = |>
        State' := 3/\Raaa' := new ()
        /\SND ({F2(Raaa'. Rhss'. Rmtcd')}_ ({
            F3(IDhss'. Rhss')}_GK). Raaa'.
            Rhss')
        /\wrequest (AAA, MTCD, aaa_mtcd, Raaa',
            Rhss')
(2) State = 3/\RCV ({F1(Raaa. Key')}_Kma')
        /\witness (AAA, MTCD, mtcd_aaa, Raaa)
        = |>
      State' := 5/\Kma' := {F4(Key')}_({F3(
            IDhss'. Rhss')}_GK)
        /\secret (Kma', kma, {AAA, MTCD})
end role % the role of AAA
```

Figure 9: The formal security verification program.

Appendix

For more details see Figure 9.

Acknowledgments

This work is supported by China Scholarship Council and the National Natural Science Foundation of China under Grant no. 61170261.

References

[1] T. Bourgeau, H. Chaouchi, and P. Kirci, "Machine-to-machine communications," in *Next-Generation Wireless Technologies*, pp. 221–241, Springer, New York, NY, USA, 2013.

[2] 3GPP TR 23. 888 V1. 4. 0. System Improvements for Machine-Type Communications, 2011.

[3] R. Deng, J. Chen, C. Yuen, P. Cheng, and Y. Sun, "Energy-efficient cooperative spectrum sensing by optimal scheduling in sensor-aided cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 716–725, 2012.

[4] P. Cheng, R. Deng, and J. Chen, "Energy-efficient cooperative spectrum sensing in sensor-aided cognitive radio networks," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 100–105, 2012.

[5] B. Emmerson, "M2M: the Internet of 50 billion devices," *WinWin Magazine*, pp. 19–22, 2010.

[6] C. Lai, H. Li, X. Li, and J. Cao, "A novel group access authentication and key agreement protocol for machine-type communication," *Transactions on Emerging Telecommunications Technologies*. In press.

[7] C. Lai, H. Li, Y. Zhang, and J. Cao, "Security issues on machine to machine communications," *KSII Transaction on Internet and Information Systems*, vol. 6, no. 2, pp. 498–514, 2012.

[8] M. Wen, Y.-F. Zheng, W.-J. Ye, K.-F. Chen, and W.-D. Qiu, "A key management protocol with robust continuity for sensor networks," *Computer Standards and Interfaces*, vol. 31, no. 4, pp. 642–647, 2009.

[9] R. Jiang, J. Luo, F. Tu, and J. Zhong, "LEP: a lightweight key management scheme based on ebs and polynomial for wireless sensor networks," in *Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC '11)*, pp. 1–5, Xi'an, China, September 2011.

[10] R. Lu, X. Li, X. Liang, X. Shen, and X. Lin, "GRS: the green, reliability, and security of emerging machine to machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 28–35, 2011.

[11] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: potential, challenges, and solutions," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 178–184, 2012.

[12] 3GPP TR 33. 868 V0. 5. 0, Security aspects of Machine-Type Communications, 2011.

[13] B. Wang, B. Li, and H. Li, "Knox: privacy-preserving auditing for shared data with large groups in the cloud," in *Applied Cryptography and Network Security*, pp. 507–525, Springer, New York, NY, USA, 2012.

[14] R. Jiang, J. Luo, and X. Wang, "An attack tree based risk assessment for location privacy in wireless sensor networks," in *Proceedings of the 8th IEEE International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4, 2012.

[15] B. Wang, B. Li, and H. Li, "Oruta: privacy-preserving public auditing for shared data in the cloud," in *Proceedings of the IEEE 5th International Conference on Cloud Computing (CLOUD '12)*, pp. 295–302, 2012.

[16] S. Laur and S. Pasini, "Sas-based group authentication and key agreement protocols," in *Public Key Cryptography-PKC*, pp. 197–213, Springer, 2008.

[17] J. Arkko and H. Haverinen, "Extensible authentication protocol method for 3rd generation authentication and key agreement (EAP-AKA)," 2006.

[18] P. Funk and S. Blake-Wilson, "EAP Tunneled TLS Authentication protocol version 1 (EAP-TTLS v1)," http://tools.ietf.org/html/draft-funk-eap-ttls-v1-00.

[19] Microsoft, "Securing Wireless LANs with PEAP and Passwords, Introduction: Choosing a Strategy for Wireless LAN Security," http://technet.microsoft.com/en-us/library/dd162271.aspx.

[20] O. George, "Ultimate wireless security guide: an introduction to LEAP authentication," Tech. Rep., 2007.

[21] D. P. Jablon, "Strong password-only authenticated key exchange," *ACM SIGCOMM Computer Communication Review*, vol. 26, no. 5, pp. 5–26, 1996.

[22] C. Lai, H. Li, R. Lu, X. Shen, and J. Cao, "A unified end-to-end security scheme for machine-type communication in lte networks," in *Proceedings of the 2nd IEEE/CIC International Conference on Communications in China (ICCC '13)*, pp. 1–6, 2013.

[23] H. H. Ngo, X. Wu, P. D. Le, and B. Srinivasan, "An individual and group authentication model for wireless network services," *Journal of Convergence Information Technology*, vol. 5, no. 1, pp. 82–94, 2010.

[24] Y.-W. Chen, J.-T. Wang, K.-H. Chi, and C.-C. Tseng, "Group-based authentication and key agreement," *Wireless Personal Communications*, vol. 62, no. 4, pp. 965–979, 2012.

[25] N. Aboudagga, J.-J. Quisquater, and M. Eltoweissy, "Group authentication protocol for mobile networks," in *Proceedings of the 3rd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '07)*, White Plains, NY, USA, October 2007.

[26] H. Mun, K. Han, and K. Kim, "3G-WLAN interworking: security analysis and new authentication and key agreement based on EAP-AKA," in *Proceedings of the Wireless Telecommunications Symposium (WTS '09)*, pp. 1–8, Prague, Czech Republic, April 2009.

[27] Mathcam, "PlanetMath-Elliptic Curve Diffie-Hellman key exchange," http://planetmath.org/DiffieHellmanKeyExchange.

[28] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation*, vol. 48, no. 177, pp. 203–209, 1987.

[29] V. S. Miller, "Use of elliptic curves in cryptography," in *Proceedings of the Advances in Cryptology (CRYPTO '85)*, pp. 417–426, Springer, 1986.

[30] E. B. Barker, D. Johnson, and M. E. Smid, "SP 800-56A. Recommendation for Pair-Wise Key Establishment Schemes Using Discrete Logarithm Cryptography (Revised)," 2007.

[31] T. A. Team, "AVISPA v1. 1 User Manual 2006," http://avispa-project.org/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 598946, 15 pages
http://dx.doi.org/10.1155/2013/598946

Research Article

Connectivity Analysis for Free-Flow Traffic in VANETs: A Statistical Approach

Chen Chen, Xiaobo Du, Qingqi Pei, and Yanan Jin

State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China

Correspondence should be addressed to Chen Chen; cc2000@mail.xidian.edu.cn

Received 8 July 2013; Accepted 29 August 2013

Academic Editor: Lin Bai

Copyright © 2013 Chen Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Intervehicle communication gives vehicles opportunities to exchange packets within the limited radio range and self-organized in Ad Hoc manner into VANETs (Vehicular Ad Hoc Networks). However, due to issues such as the high mobility, insufficient market penetration ratio, and lacking of roadside units, connectivity is actually a big problem in VANETs. In addition, relying only on the direct connectivity in most of the previous works, say broadcasting which provides one-hop connections between nodes is far from the continuously growing application demands in VANETs, such as inter-vehicle entertainments, cooperative collision avoidances, and inter-vehicle emergency notifications. Therefore, the indirect connectivity from multihop forwarding is also a necessary complement especially for the case where direct connection is hardly achieved. In this paper, we define a new metric, that is, available connectivity, to consider both direct and indirect connectivity. After analyzing the statistical properties of direct and indirect connectivity in vehicular environment, the available connectivity is proposed and quantified for practical usage. Numerical results show that our available connectivity could provide correct and useful references for protocols design and performance improvements of different applications.

1. Introduction

Vehicular Ad hoc Networks [1–3] (VANETs) are distributed, self-organized communication networks composed of moving vehicles and are thus characterized by very high nodal mobility and limited degrees of freedom in the mobility patterns. The discussed IEEE 1609 Wireless Access in Vehicular Environments [4] (WAVE) draft is being developed for VANETs applications mainly considering safety-related scenarios, such as cooperative forward collision warning [5, 6] (CCW) systems, traffic signal violation warning [7] systems, and lane change warning [8] systems, and at the same time taking account of some requirements from nonsafety-related applications [9]. Indeed, all the above applications greatly rely on the packets exchange on reliable links and robust connections between vehicles. From this perspective, network connectivity becomes a fundamental and crucial issue to any practical application in VANETs. Regarding the node mobility, vehicles can not only depend on the direct one-hop broadcasting; they also need packets to be indirectly forwarded to the destinations through store-carry or multi-hop relays. However, it is worth noting that both direct and indirect connections are sometime hardly achieved especially during sparse communication environments such as the suburb highway or cases with low market penetration ratio [10]. As a result, the limited connectivity directly affects the possible speed at which information can be transported over a VANET and hence directly influence the up-to-datedness of wanted notifications which may result in accidents.

In summary, analysis to the connectivity is very necessary in VANETs. Actually, a number of models have been proposed in previous works such as stochastic process [11] based, Gaussian unitary ensemble (GUE) of random-matrix theory [12] based, and queuing theory [13] based and almost all their works discussed the connectivity with the famous three-phase traffic theory in which the fundamental relationship between traffic parameters on a road segment is given by the following equation:

$$F = S \times K, \tag{1}$$

Figure 1: Relationship between basic parameters in the vehicle traffic theory.

Figure 2: CCW application scenario.

Figure 3: Traffic jams avoidance.

where F, S, and K are the traffic flow, average speed, and traffic density, respectively. The influences on the connectivity from the above parameters can be disclosed from Figure 1. Before the state transforming point, at which packets begin to be queued in transceivers' sending buffers when the density of vehicles increases, the connectivity is intermittent and thus needs further studying to improve the performance. For the stage after the state transforming point, which is named the congestion flow state and depicted by dashed line, the connectivity could be basically guaranteed regarding the larger density. In that case, road is often jammed by vehicles very close to each other whereby the direct connectivity can be readily obtained. Accordingly, our research will focus on the free-flow state depicted by the solid line throughout our paper.

The rest of the paper is organized as follows. Section 2 provides our research motivations and discusses the possible usages of the available connectivity in VANETs. Section 3 presents previous related works. Section 4 introduces our definition of the available connectivity, assumptions, and the statistical analysis of direct and indirect connectivity. Numerical results and performance evaluations are presented in Section 5. Our work is concluded in Section 6 followed by acknowledgments and cited references.

2. Motivations

"Connectivity" in Ad Hoc networks actually has a mature body of research but still absorbed lots of research interests in VANETs recently. With VANETs gradually stepping into our daily life, the connectivity has played an important role in many road applications to ensure driving safety and increase comfortableness. For instance, in CCW scenario, a good connectivity could help to avoid the chain collisions through disseminating warning notifications farther and quicker. In Figure 2, the red car crashed with the blue car and sent a CCW message. Then, car 1 could receive the CCW by direct connection and take brake in time to avoid crash. However, without the help of the indirect connection, that is, multi-hop forwarding in this case, car 2, and 3 may collide with the front car if the intervehicle distances are approaching the unsafe range [14]. In Figure 3, vehicles 2~5 could detect the front traffic jam by the messages store-carry-forwarded by vehicle 1 from the red sender. Therefore, owing to the indirect connectivity, they can beforehand enter into the auxiliary road to avoid the congestion and save time. In summary, the connectivity, direct or indirect, could provide more opportunities for vehicles to make their decisions wisely and safely and to achieve better driving experiences. Hence, a good definition of the connectivity or the metric to measure the connectivity is essential in VANETs.

In our work, we define the connectivity as the probability of connections which could be achieved directly or indirectly. Numerical results show that our definition can explore the potential transmission opportunities for vehicles especially in the safety-related applications context. The connection possibility is discussed through statistical analysis in terms of either connected durations or inter-vehicle distances. To reflect the high dynamics of the connectivity due to different mobility patterns, we also introduce the influence of velocity into our work.

Indeed, a wise definition of the available connectivity can practically make many otherwise complex problems easy. For example, to design an admission control strategy in VANETs, the available connectivity could provide important references for the connectivity improvement by giving the specific time or location at which the vehicles may be admitted into. In addition, to evaluate the performance of emergency-related applications, the available connectivity could roughly work out the delivery ratio of emergency notifications from given vehicles. On the other hand, to make the emergency notifications spread faster and further, the available connectivity could also be taken as a better criterion for relays selection. Besides, to route the packets based on the available connectivity, the throughput can be improved and the backup multiple paths may be simultaneously figured

out whereby transmission robustness could be guaranteed. In a dense network, available connectivity may also be able to offer a reference for the threshold of packet generation rate to maintain sufficient connectivity but not make the overall network congested. In summary, along with the available connectivity, a plethora of works now can be implemented to improve the safety, comfortableness, and efficiency in VANETs.

3. Related Works

Although connectivity analysis is a classical issue in wireless communication networks, it is now still a hot topic in VANETs regarding the recently increasing research activities [15–17]. In my point of view, the connectivity research in VANETs can be generally classified into two categories: one for investigating the lifetime properties of connected links or paths; another via the measurement of inter-vehicle distances.

Researches on lifetime properties mainly focus on the statistical analysis of the connection duration. In [11], the authors discussed the connection duration expression in detail considering the velocity vectors with yaw angles. Their numerical results found that high relative velocities impose a hard task on some cooperative maneuvers including underlying routing protocols. However, even the short connection duration is still enough for emergency notification scenarios. In [18], the connection duration between two adjacent vehicles was figured out, given their speeds, directions, and the radio ranges. According to the calculated connection period expression, an admission control strategy is proposed to determine whether the next vehicle is allowed to be injected into the current traffic flow without interrupting the ongoing connections. In [19], metrics for evaluating nodal connectivity in VANETs had been presented considering the different nodal mobility patterns. The connection period distribution was characterized by average duration of the k-hop path existing between any two nodes. They also showed by simulation that multi-hop paths have much poorer connectivity performance than the single-hop one in VANETs.

Researches on the connectivity through inter-vehicle distances attempt to disclose the relationship between the connectivity and the distance which is defined as the length of the connected path. In [11], the distance distribution had been expressed by a tuple (h,v,ρ), where h denotes the inter-vehicle distance, v indicates the velocity, and ρ is the market penetration ratio, respectively. By taking specific propagation model and the radio range into consideration, the connectivity formula was given. Without loss of generality, the expression was then further extended to the connectivity of a node with degree n. In [13], the authors introduced an equivalent $M/D/\infty$ queuing model and applied Laplace transform to the tail probability of the connection distance to find an explicit expression for the expected value of this distance. Their numerical results showed that the increasing of traffic flow and the vehicles' radio range correspondingly leads to the increasing of the discussed connectivity in terms of platoon size and connection distance. In [20], the paper showed that, under free-flow state, the connectivity increases with either the density or the number of lanes growing. Especially, they found that the inter-vehicle distances had a major impact on connectivity when it is within 3-4 times larger than the radio range. Beyond this distance, the connectivity declines slowly. In [21], a simple geometric analysis to the connectivity was proposed in vehicular networks under a distance-based radio communication model. The authors presented the notion of the connectivity robustness based on inter-vehicle distances, whereby a local computable function could be obtained to provide a sufficient condition for the connectedness of the network. For the above-mentioned works, although, most of them could deduce out or express the connectivity with vehicles density [22, 23], speed or equivalent speed [24], radio ranges [22, 25], and so forth, their methods actually can all be equivalently extended to the approach using statistical distributions of connection durations or inter-vehicle distances. Generally speaking, it is difficult to give a globally uniform probability density function (PDF) or cumulative distribution function (CDF) expression of the inter-vehicle distances or connection durations in different application scenarios such as overtaking, platoon, and collisions. But, when the traffics are in the free-flow state, the distribution function and probability of the minimum node degree can be exactly calculated whereby available connectivity could be figured out.

Figure 4: The analytical model used for highway and the vehicles traffic follows Poisson arrival process.

4. The Connectivity Analysis in VANETs

4.1. The Analytical Model and Definitions. As we have stated before, the connectivity in the free-flow state is worth studying in view of its uncertainty and scarcity. Thus, in this section, we introduce a highway as our discussing objective where vehicles form a free-flow state traffic with lower distribution density. In our analysis, an equipped vehicle is taken as a node for simplicity. The discussed scenario is plotted in Figure 4, where all cars drive on a single lane represented by a line. We assume that X_0 requests to send packets to cars on its right. X_0 here is the source node at the origin of the one-dimensional coordinate system and we also take it as the reference point later for our connectivity analysis. The arrival stochastic process of these vehicles is supposed to be a Poisson process with intensity F. Let ε denote the average density of equipped vehicles in our discussed road segment. Thus, the arrival rate λ is equal to $F\varepsilon$.

Definition 1. Define the connectivity of two nodes as $C(x)$, where $C(x)$ is the probability that node x could reach the reference node X_0.

Figure 5: The relationship between $M/D/\infty$ system and highway model.

Definition 2. Let i be the farthest node which can communicate with X_0. i in later discussion is also named as the end node with radio range R_i.

X is a variable representing the distance between X_0 and the end node. The PDF and CDF of X can be expressed as $f(x) = \Pr(X = x)$ and $F(x) = \Pr(X \le x)$, respectively. If we suppose that X is within $[x, x + \Delta x]$, it means the following (i) There is at least one node in $[x, x + \Delta x]$. (ii) The end node could connect with X_0 by single-hop broadcasting. (iii) There is no node within $(x+\Delta x, x+\Delta x+R_i)$. If and only if the above three conditions are all met, we say the node i is the end node [26].

We suppose that two nodes are connected if they are falling into the radio range of each other and the disconnection due to collisions or asymmetrical channel conditions is not taken into account.

4.2. The Statistical Model with Constant Radio Range Setting. In this section, we discuss the connectivity when all nodes choose the same constant radio range, that is, R. We model the one-dimension highway as an $M/D/\infty$ system [27] as shown in Figure 5.

We investigate $C(x)$ of any equipped vehicle x in the highway. In free-flow state, vehicles move independently and their velocities follow the normal distribution [28] with PDF $f_v(v)$ as

$$f_v(v) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(v-\mu)^2/2\sigma^2}, \tag{2}$$

where μ is the average speed and σ is the standard deviation of velocity. We also introduce a metric, that is, speed factor, to reflect the impact of velocity on the inter-vehicle distances and the resulted connectivity.

Definition 3. A is defined as the speed factor with unit hour/km as

$$A = \int_{v_{\min}}^{v_{\max}} \frac{\widehat{f}_v(v)}{v} dv, \tag{3}$$

where

$$\widehat{f}_v(v) = \frac{f_v(v)}{\int_{v_{\min}}^{v_{\max}} f_v(s)\, ds}. \tag{4}$$

$v_{\min}$ and $v_{\max}$ in (4) indicate the permitted minimum and maximum velocity on highway, respectively, and they satisfy the following equation. Consider

$$2\int_{v_{\min}}^{v_{\max}} f_v(s)\, ds = \operatorname{erf}\left(\frac{v_{\max}-\mu}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{v_{\min}-\mu}{\sigma\sqrt{2}}\right). \tag{5}$$

For a PDF curve of a normal distribution, the region corresponding to the x-axis coordinate ranging from $[\mu - 3\sigma, \mu + 3\sigma]$ covers about 99.7% of the probability area; thus, we reasonably choose $\mu - 3\sigma$ and $\mu + 3\sigma$ as the typical values of $v_{\min}$ and $v_{\max}$, respectively. Indeed, vehicles in highway can travel neither too fast for safety consideration nor too slow for traffic efficiency. Although speed limits are set, overtaking is still allowed to reflect the practical case.

With the above definitions and assumptions, the expression of connectivity with constant radio range setting could be given as follows.

Theorem 4. *For the case where vehicles all have the same constant radio range, the connectivity could be deduced as*

$$C(x) = \begin{cases} 1, & x \in [0, R), \\ 1 - e^{-\lambda AR} - (x-R)\lambda A e^{-\lambda AR}, & x \in [R, 2R), \\ C(nR) - w(x), & x \in [nR, (n+1)R), \end{cases} \tag{6}$$

where $n \ge 2$, and

$$C(nR) = \lim_{x \to nR} C(x), \quad x \in [(n-1)R, nR),$$

$$w(x) = \lambda A e^{-\lambda AR} \int_{nR}^{x} \left[\sum_{k=0}^{n-1} Q(k)\right] dt,$$

$$Q(k) = \frac{[t-(k+1)R]^k - e^{-\lambda AR} \cdot [t-(k+2)R]_+^k}{k!} \cdot \left(-\lambda A e^{-\lambda AR}\right)^k, \tag{7}$$

$$x_+^k = \begin{cases} x^k, & x \ge 0 \\ 0, & x < 0. \end{cases}$$

Proof. Let $f_d(t)$ and $F_d(t)$ be the PDF and the CDF of the maximum radio range of X_0. With Figure 5 and based on [27], we have

$$f(x) = f_d(x + R), \tag{8}$$

where $f_d(t)$ corresponds to the PDF of busy period in $M/D/\infty$ system, and its Laplace transform is given by

$$L(f_d(t)) = f_d(s) = \frac{e^{-sR}(1 - F_H(R))}{1 - \int_0^R e^{-sh} dF_H(h)}, \tag{9}$$

where $F_H(h)$ is the CDF of inter-vehicle distances. Note that inter-vehicle distances in $M/D/\infty$ system are independent and identically distributed (IID) [29].

Therefore, we have the following equations:

$$f(x-R) = f_d(x),$$
$$L(F(x)) = F(s) = \frac{1}{s}e^{sR} f_d(s), \tag{10}$$

where $f_d(s)$ and $F(s)$ correspond to the Laplace transform of $f_d(x)$ and $F(x)$, respectively.

Based on Definition 2 for an end node, the connectivity of any node at x is

$$C(x) = 1 - F(x-R), \tag{11}$$

and its Laplace transform is

$$L(C(x)) = C(s) = \frac{1}{s} - e^{-sR}F(s). \tag{12}$$

With above equations, (12) can be rewritten as

$$C(s) = \frac{1}{s} - \frac{1}{s}\frac{e^{-sR}(1-F_H(R))}{1-\int_0^R e^{-sh}dF_H(h)}. \tag{13}$$

Actually, in [13], the CDF expression of the inter-vehicle distance is given by

$$F_H(h) = 1 - e^{-\lambda A h}, \tag{14}$$

and the average inter-vehicle distance is written as

$$E[H] = \frac{1}{\lambda A}. \tag{15}$$

Thus, with (12) and (14), we have

$$C(s) = \frac{1}{s} - \frac{1}{s}\frac{s+\lambda}{se^{(s+\lambda)R}+\lambda}. \tag{16}$$

Applying an inverse Laplace transforming [30] to (16), we have the expression of $C(x)$ as (6). □

Remark 5. $C(x)$ is equal to 1 when x is within the range $[0, R)$; that is, a node falling into tother radio range can connect to it if it wishes.

Remark 6. If a node is exactly located at $x = R$, then $1-C(R)$ is the probability that it cannot connect to the reference node. The probability that there is no node within range $(0, R]$ is $e^{-\lambda AR}$ for a Poisson arrival process.

Remark 7. $C(x)$ decreases as the relative distance x increases. In the extreme case, $C(x) \to 0$ when $x \to \infty$.

The above remarks also validate Theorem 4. To understand this theorem in depth, we also explore some connectivity-related statistical properties including the probability that a given node is isolated, the area that a sent message may cover, and the probability of the number of connected nodes. We also classify the connected nodes into two categories by relaying hops between them and analyzing the corresponding connectivity.

(1) Probability that a Given Node Is Isolated. The probability that a given node is isolated, denoted by P_I, can be determined by the following theorem.

Theorem 8. *The isolation probability is given by*

$$P_I = (1 - C(R))^2 = e^{-2\lambda AR}. \tag{17}$$

For a node at x, P_I is equal to the probability that there is no node within the range $[x-R, x+R]$.

(2) The Area that a Sent Message May Cover. We discuss the area that a sent message from the reference node can cover. The single and multi-hop cases are both considered. Generally, this characteristic can also be evaluated with coverage probability. Here, we use the expectation to show the average value in general case. Because the system is initially assumed not empty, the mathematic expectation could be given by the following theorem.

Theorem 9. *The expectation of the coverage area for a message sent from the reference node is*

$$E(\mathrm{Cov}_+) = -\frac{\partial}{\partial s}(1 - s\cdot C(s))\big|_{s=0} + E[H], \tag{18}$$

where Cov_+ *means the area corresponding to the right side of a given node.*

Proof. Based on [27], we can obtain the following equations for different system initial conditions. Consider

$$E[\mathrm{Cov}_+] = -\frac{\partial f_d(s)}{\partial s}\big|_{s=0} + E[H]. \tag{19}$$

Then, with (10) and (12), (18) could be derived out. □

Corollary 10. *If both sides of the discussed node are considered,* (18) *could be rewritten as*

$$E[\mathrm{Cov}] = E[\mathrm{Cov}_+ + \mathrm{Cov}_-] = \frac{2e^{\lambda AR}}{\lambda A}, \tag{20}$$

where Cov_- *corresponds to the area on the left.*

(3) Probability of the Number of Connected Nodes. The node which can reach the reference node is called a connected node. Here we discuss the possible number of the connected nodes, that is, CN. Similarly, we define CN_+ as the number on the right and CN_- on the left.

Theorem 11. *The Probability Mass Function (PMF) of CN is*

$$p(CN = m) = (m+1)(1 - F_H(R))^2 F_H(R)^m, \tag{21}$$

where $m = 0, 1, \ldots$.

Proof. We first focus on CN_+ and then extend the result to CN_-.

The PMF of CN_+ for $k = 0, 1, \ldots$ can be expressed as

$$p(\mathrm{CN}_+ = k) = (1 - F_H(R))F_H(R)^k. \tag{22}$$

The result for CN_- is proved in the same way. Then, the PMF for both sides is derived as follows. Consider

$$\begin{aligned} p(CN = m) &= \sum_{k=0}^{m} p(CN = m, CN_+ = k) \\ &= \sum_{k=0}^{m} p(CN_- = m-k, CN_+ = k) \\ &= \sum_{k=0}^{m} p(CN = m-k)\, p(CN_+ = k). \end{aligned} \tag{23}$$

Simplify the above equation and then the result can be rewritten as (21). □

Theorem 12. *The tail probability of CN is*

$$p(CN \geq m) = [m(1 - F_H(R)) + 1]\, F_H(R)^m. \tag{24}$$

Proof. According to (21), we have

$$\begin{aligned} p(CN \geq m) &= \sum_{k=m}^{\infty} p(CN = k) \\ &= \sum_{k=m}^{\infty} (k+1)(1 - F_H(R))^2 F_H(R)^k \\ &= [m(1 - F_H(R)) + 1]\, F_H(R)^m. \end{aligned} \tag{25}$$

□

In addition, the probability that at most k nodes are connected is

$$p(CN_+ \leq k) = 1 - F_H(R)^{k+1}. \tag{26}$$

The probability generating function (PGF) of CN is

$$PGF_{CN_+}(z) = \frac{1 - F_H(R)}{1 - zF_H(R)}. \tag{27}$$

And the expected value of CN_+ is

$$E[CN_+] = \lim_{z \to 1} \frac{\partial PGF_{CN_+}(z)}{\partial z} = \frac{F_H(R)}{1 - F_H(R)}. \tag{28}$$

Substituting (14) into (28), we obtain

$$E[CN_+] = e^{\lambda AR} - 1. \tag{29}$$

Corollary 13. *The expectation of CN is*

$$E[CN] = 2e^{\lambda AR} - 2. \tag{30}$$

Let Q be the probability lower bound when the number of connected nodes is CN_{re+} and it can be written as

$$Q = \left(1 - e^{-\lambda AR}\right)^{CN_{re}}. \tag{31}$$

For a given Q and CN_{re+}, we may obtain a critical radio range, which will be explained later in Corollary 14 as

$$R_c = -\frac{1}{\lambda A} \ln\left(1 - Q^{1/CN_{re}}\right). \tag{32}$$

Actually, the radio range is together determined by the transmission power P_{tx}, the additive Gaussian noise power P_{noise}, the signal-to-noise ratio (SNR) ψ, and the path loss factor α. The relationship of them is described by (33) *where K is a constant depending on the system setting.*

$$R = \left(\frac{KP_{tx}}{\psi P_{noise}}\right)^{1/\alpha}. \tag{33}$$

Similarly, the equation for Q considering both sides is

$$Q = (CN_{re} + 1)\, e^{-2\lambda AR}\left(1 - e^{-\lambda AR}\right)^{CN_{re}}. \tag{34}$$

Corollary 14. *The critical range R_c could be determined by the intersection of following curves:*

$$\begin{aligned} y_1 &= \frac{Q}{CN_{re} + 1} e^{2\lambda Ar}, \\ y_2 &= \left(1 - e^{-\lambda Ar}\right)^{CN_{re}}. \end{aligned} \tag{35}$$

According to (32) *and* (33)*, the connectivity could be flexibly controlled via radio range adjustment.*

(4) Two Kind of Connected Nodes

Definition 15. The connected nodes can be classified into two categories: (1) nodes within the radio range of the reference node, which is denoted by IRN; (2) nodes out of its radio range expressed by ORN.

An IRN node is the one which can directly communicate with the reference node. We also separately discussed the left and right side cases of IRN. Notice that IRN_+ follows Poisson distribution with parameter λAR and its PMF is

$$p(IRN_+ = k) = \frac{(\lambda AR)^k}{k!} e^{-\lambda AR}. \tag{36}$$

In the same way, the IRN considering both sides could be derived as

$$p(IRN = m) = \frac{(2\lambda AR)^m}{m!} e^{-2\lambda AR}. \tag{37}$$

And the expectation of IRN is given by

$$E[IRN] = E[IRN_+] + E[IRN_-] = 2\lambda AR. \tag{38}$$

The ORN indicates the number of nodes beyond the radio range of the reference node and connecting to X_0 through multiple hops.

Theorem 16. *The PMF of ORN_+ is*

$$p(ORN_+ = i) = \begin{cases} e^{-\lambda AR}(1 + \lambda AR) & i = 0, \\ e^{-2\lambda AR}\left(1 - e^{-\lambda AR}\right)^{i-1} \\ \quad \times \left[e^{\lambda AR} - 1 - \lambda AR\right] & i \geq 1. \end{cases} \tag{39}$$

Proof. From (22), (26), and (36), for $i \geq 1$, we have

$$\begin{aligned} p(\mathrm{ORN}_+ = i) &= \sum_{k=0}^{\infty} p(\mathrm{ORN}_+ = i, \mathrm{IRN}_+ = k) \\ &= \sum_{k=1}^{\infty} p(\mathrm{ORN}_+ = i, \mathrm{IRN}_+ = k) \\ &= \sum_{k=1}^{\infty} p(\mathrm{CN}_+ = k+i) \\ &\quad \times p(\tau_k \leq R, \tau_{k+1} > R \mid \mathrm{CN}_+ = k+i), \end{aligned} \tag{40}$$

where τ_k, τ_{k+1} represent the sum of the first k and $k+1$ inter-vehicle distances, respectively. Let $H_1, H_2, \ldots$ be the inter-vehicle distances and they are IID to each other; that is,

$$\begin{aligned} \tau_k &= H_1 + H_2 + \cdots + H_k, \\ \tau_{k+1} &= H_1 + H_2 + \cdots + H_k + H_{k+1}, \\ \tau_{k+1} &= \tau_k + H_{k+1}. \end{aligned} \tag{41}$$

When $\mathrm{CN}_+ = k+i$, the following inequality can be deduced:

$$H_p \mid \mathrm{CN}_+ \leq R, \quad p = 1, 2, \ldots, k+i. \tag{42}$$

And the PDF of $H_p \mid \mathrm{CN}_+$ is

$$f_{H_p|\mathrm{CN}_+}(h) = \frac{\lambda A e^{-\lambda A h}}{1 - e^{-\lambda A R}}, \quad h \in [0, R]. \tag{43}$$

Taking Laplace transform to the above equation, we have

$$f_{H_p|\mathrm{CN}_+}(s) = \frac{\lambda A}{1 - e^{-\lambda A R}} \frac{1 - e^{-(\lambda A + s)R}}{s + \lambda A}. \tag{44}$$

According to the independence of inter-vehicle distances, the following equation can be derived as

$$f_{\tau_k|\mathrm{CN}_+}(s) = \left(\frac{\lambda A}{1 - e^{-\lambda A R}}\right)^k \left(\frac{1 - e^{-(\lambda A + s)R}}{s + \lambda A}\right)^k. \tag{45}$$

Then the PDF of $\tau_k \mid \mathrm{CN}_+$ is written as

$$\begin{aligned} f_{\tau_k|\mathrm{CN}_+}(t) &= \left(\frac{\lambda A}{1 - e^{-\lambda A R}}\right)^k \\ &\quad \cdot \sum_{j=0}^{k} C_k^j (-1)^j \frac{(t - jR)^{k-1}}{(k-1)!} e^{-\lambda A (t - jR)} U(t - jR), \end{aligned} \tag{46}$$

where $U(\cdot)$ is a step function.

Meanwhile, the following conditional probability can be obtained as

$$\begin{aligned} &p(\tau_{k+1} > R, \tau_k \leq R \mid \mathrm{CN}_+ = k+i) \\ &\quad = \int_0^R p(H_{k+1} > R - t \mid \mathrm{CN}_+ = k+i, \tau_k = t)\, dF_{H|\mathrm{CN}_+}(t) \\ &\quad = \int_0^R f_{\tau_k|\mathrm{CN}_+}(t)\, dt \int_{R-t}^{R} f_{H_{k+1}|\mathrm{CN}_+}(h)\, dh. \end{aligned} \tag{47}$$

With (22), (43), (46), and (47), we have

$$\begin{aligned} &p(\mathrm{ORN}_+ = i) \\ &\quad = \sum_{k=1}^{\infty} p(\tau_k \leq R, \tau_{k+1} > R \mid \mathrm{CN}_+ = k+i) \\ &\quad\quad \cdot p(\mathrm{CN}_+ = k+i) \\ &\quad = e^{-2\lambda A R}\left(1 - e^{-\lambda A R}\right)^{i-1}\left[e^{\lambda A R} - 1 - \lambda A R\right]. \end{aligned} \tag{48}$$

When $i = 0$, we have

$$\begin{aligned} p(\mathrm{ORN}_+ = 0) &= 1 - \sum_{i=1}^{\infty} p(\mathrm{ORN}_+ = i) \\ &= e^{-\lambda A R}(1 + \lambda A R). \end{aligned} \tag{49}$$

Then, the PMF of ORN_+ is rewritten as (39). □

Corollary 17. *The PMF of ORN is expressed as*

$$p(ORN = m) = \begin{cases} e^{-2\lambda A R}(1 + \lambda A R)^2 & m = 0, \\ 2e^{-2\lambda A R}(1 + \lambda A R)\left[1 - e^{-\lambda A R} - \lambda A R e^{-\lambda A R}\right] & m = 1, \\ (m+1)e^{-3\lambda A R}\left(1 - e^{-\lambda A R}\right)^{m-2} \\ \quad \times \left(e^{-\lambda A R} - 1 - \lambda A R\right) \\ \quad \cdot \left(e^{-\lambda A R} - 1 - \lambda A R + \dfrac{2\lambda A R e^{\lambda A R}}{m+1}\right) & m \geq 2. \end{cases} \tag{50}$$

Proof. Through (51), we have

$$p(\mathrm{ORN} = m) = \sum_{i=0}^{m} p(\mathrm{ORN}_- = m - i, \mathrm{ORN}_+ = i). \tag{51}$$

□

Corollary 18. *The expectations of ORN_+ and ORN are*

$$\begin{aligned} E[ORN_+] &= e^{\lambda A R} - \lambda A R - 1, \\ E[ORN] &= 2e^{\lambda A R} - 2\lambda A R - 2. \end{aligned} \tag{52}$$

We define $Hops_+$ as a variable indicating the number of needed hops from the reference node to a connected node on X_0's right side and the following inequalities holds

$$\begin{aligned} Hops_+ &\leq CN_+, \\ E[Hops_+] &\leq E[CN_+] = e^{\lambda A R} - 1, \\ P(Hops_+ \geq h) &\leq P(CN_+ \geq h) = \left(1 - e^{-\lambda A R}\right)^h. \end{aligned} \tag{53}$$

4.3. Statistical Model Considering Random Radio Range. If the channel fading is taken into account, the reachable radio range should be reconsidered. Thereupon, the connectivity

will also need reexamination. For notion simplicity, we name such variable "radio range" as ERR (equivalent radio range). The $M/G/\infty$ queuing model is introduced to check the connectivity in this case. The traffic still follows a Poisson arrival process. Let $F_R(r)$ be the CDF of the radio range corresponding to the service time in $M/G/\infty$.

The PDF of the busy period is [27]

$$B(d) = 1 - (\lambda A)^{-1} \sum_{n=1}^{\infty} \delta^{(n)}(d), \tag{54}$$

where $\delta^{(n)}(d)$ is the nth convolution of $\delta(d)$, and

$$\delta(d) = \lambda A \left[1 - F_R(d)\right] \exp\left\{-\lambda A \int_0^d \left[1 - F_R(r)\right] dr\right\}. \tag{55}$$

Note that, in our work, route selection is based on the shortest path algorithm. And (54) corresponds to the case that nodes at $[0, d]$ can receive the messages sent from X_0. Indeed, such nodes may not communicate with each other in some cases. Therefore, $B(d)$ cannot be directly taken as the connectivity.

Hereinafter, we define $C^F(x)$ as the instantaneous connectivity with a random fading channel. Similarly, the PDF and CDF of X are rewritten as $f^F(t)$, $F^F(x)$ under this case. The distances between adjacent vehicles are denoted by H_i $(i = 0, 1, \ldots)$ and they are also IID to each other. Here, vehicles may have different ERRs $\{R_i\}$ experiencing different channel fadings.

Theorem 19. *The connectivity for the random radio range case can be expressed as*

$$C^F(x) = 1 - \sum_{m=1}^{\infty} \int_0^{x-a} \cdots \int_0^{x-a-x_0-\cdots-x_{m-2}} \int_0^{\infty} \cdot \prod_{i=0}^{m-1} \left[1 - F_R(x_i)\right]^2 F_R(x_m)\, dF_H(x_m) \cdots dF_H(x_0), \tag{56}$$

where a denotes the distance between the end node and its nearest relay.

Proof. We assume that there are m connected nodes on X_0's right side. It means that the end node is the mth one. Thus, we have

$$H_i \le \min(R_i, R_{i+1}), \quad i = 0, \ldots, m-1, \qquad H_m > R_m. \tag{57}$$

According to the definition of X, then

$$\begin{aligned}
F^F(x) &= P(X \le x) \\
&= \sum_{m=1}^{\infty} p\left(\sum_{j=0}^{m-1} H_j \le x, H_m > R_m, H_i \le \min(R_i, R_{i+1})\right) \\
&= \sum_{m=1}^{\infty} \int_0^x p\left\{\sum_{j=0}^{m-1} H_j \le x, H_i \le \min(R_i, R_{i+1}), H_m > R_m \mid H_0 = x_0\right\} dF_H(x_0) \\
&\cdots \\
&= \sum_{m=1}^{\infty} \int_0^x \cdots \int_0^{x-x_0-\cdots-x_{m-2}} \int_0^{\infty} p\left\{H_i \le \min(R_i, R_{i+1}), H_m > R_m\right\} \cdot dF_H(x_m) \cdots dF_H(x_0) \\
&= \sum_{m=1}^{\infty} \int_0^x \cdots \int_0^{x-x_0-\cdots-x_{m-2}} \int_0^{\infty} dF_H(x_m) \cdots dF_H(x_0) \\
&\quad \cdot \prod_{i=0}^{m-2} \left[1 - F_R(x_i)\right]^2 \left[1 - F_R(x_{m-1})\right] \\
&\quad \times \left[F_R(x_m) - F_R(x_{m-1})\right].
\end{aligned} \tag{58}$$

The connectivity between the reference node and a given node with $X = x$ is

$$C(x) = 1 - F^F(x - a). \tag{59}$$

□

To reflect the general fading types in vehicular environment, the Rayleigh and lognormal shadowing are introduced in our work. According to [31–33], the corresponding CDF and expectations of the radio range for the above fading models are as follows.

Lemma 20. *The PDF and average ERR of Rayleigh fading are*

$$F_R(r) = 1 - e^{-r^{\alpha} \psi P_{noise}/KP_{tx}}, \tag{60}$$

$$E[R] = \frac{\Gamma\left(\alpha^{-1}\right)}{\alpha} \left(\frac{KP_{tx}}{\psi P_{noise}}\right)^{1/\alpha}. \tag{61}$$

Lemma 21. *The PDF and average ERR for lognormal shadowing are*

$$F_R(r) = \Phi\left(\frac{\ln\left(r^{\alpha} \psi P_{noise}/KP_{tx}\right)}{\sigma_s}\right), \tag{62}$$

$$E[R] = \left(\frac{KP_{tx}}{\psi P_{noise}}\right)^{1/\alpha} e^{\sigma_s^2/2\alpha^2}, \tag{63}$$

where σ_s is the standard deviation, and

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \tag{64}$$

Note that one just discusses the cases under free-flow state, say λ is generally very small. Then, in a Rayleigh fading channel, the pass loss factor could be taken as a constant during the transmission of a frame. Indeed, when λ becomes very large, the above assumption may fail in some cases.

Similar to the analyzing procedure when the radio range is constant, one also investigates some statistical characteristics for random radio range case, that is, probability that a given node is isolated, the area that a message may cover, the probability of the number of connected nodes, and its value for IRN and ORN, respectively.

(1) The Probability That a Given Node Is Isolated. For a fading channel, the number of nodes within R is Poisson distributed with parameter $2\lambda AE[R]$ and can be expressed as

$$P_I^F = e^{-2\lambda AE[R]}. \tag{65}$$

(2) The Area that a Transmitted Message May Cover. Based on [30], we know, the computational complexity of (56) is very high. However, by [27], the expectation value of the coverage area for a sent message is

$$E[\mathrm{Cov}_+] = \frac{1}{\lambda A P_0}, \tag{66}$$

where

$$\begin{gathered} P_0(t) = \exp\left\{-\lambda A \int_0^t (1 - F_R(r))\, dr\right\}, \\ P_0 = \lim_{t\to\infty} P_0(t). \end{gathered} \tag{67}$$

The expectations for the area on the right and both sides are, respectively,

$$\begin{gathered} E[\mathrm{Cov}_+] = \frac{e^{\lambda AE[R]}}{\lambda A}, \\ E[\mathrm{Cov}] = \frac{2e^{\lambda AE[R]}}{\lambda A}. \end{gathered} \tag{68}$$

(3) The Probability of the Number of Connected Nodes. As stated in the corresponding parts for constant radio rage case, CN^F, CN_+^F, and CN_-^F here are the number of connected nodes on both sides, the right and the left, respectively. Then, the probability of CN_+^F given radio range of each car is

$$\begin{aligned} &p\left(\mathrm{CN}_+^F = m \mid \{R_i\}\right) \\ &\quad = p\left(H_i \le \min(R_i, R_{i+1}), H_m > \min(R_m, R_{m+1}) \mid \{R_i\}\right) \\ &\quad = e^{-\lambda A \min(R_m, R_{m+1})} \prod_{i=1}^{m-1} \left(1 - e^{-\lambda A \min(R_i, R_{i+1})}\right). \end{aligned} \tag{69}$$

Because the radio range is a stochastic variable under a random channel, by (69), the number of connected nodes is mainly determined by the nodes with smaller ranges.

If we focus on the nodes which are within the radio range of X_0, the number of them is named HN_+ and its PGF [27] is

$$\begin{aligned} g(s) = \sum_{k=0}^{\infty} s^k \int_0^{\infty} \cdots \int_0^{\infty} \prod_{i=0}^{k-1} F_R(h_k) \cdot \lambda A e^{-\lambda A h_k} \\ \cdot \left\{ \lambda A e^{-\lambda A h_i} \left[F_R\left(\sum_{j=i}^{k} h_j\right) - F_R(h_i) \right] dh_i \right\} dh_k. \end{aligned} \tag{70}$$

With (70), the PMF, CDF, and expectation expressions could be derived in theory. However, the procedures are actually very computationally complex. Owing to the works by [34, 35], the above statistics could be deduced in a simple way. Suppose that the vehicular arrival process is a Markovian arrival process (MAP) and it is stationary, ergodic, and nonempty. Without loss of generality, we assume that the arrival traffic flow size in MAP is one [27] and nodes enter into this system one by one.

According to [36], HN_+ follows $\mathrm{DPH}(\alpha_0, \mathbf{I} - \int_0^{\infty} e^{-\lambda A r \mathbf{I}} dF_R(r))$, where $\mathbf{I}$ is the identity matrix and $\alpha_0 \cdot \mathbf{1} = 1$. $\mathbf{1}$ is an all-1 matrix. Thus, the lower bound of HN_+ can be derived as

$$\mathrm{LB}[E(HN_+)] = \frac{1}{\int_0^{\infty} e^{-\lambda A r} dF_R(r)}, \tag{71}$$

and the result considering both sides is

$$\mathrm{LB}[E(HN)] = \frac{2}{\int_0^{\infty} e^{-\lambda A r} dF_R(r)}. \tag{72}$$

(4) Two Kinds of Connected Nodes. Let IRN^F be the number of nodes within the radio range of X_0. For $m = 0, 1, \ldots$, the PMF and its expectation, respectively, are

$$\begin{gathered} p\left(\mathrm{IRN}^F = m\right) = \frac{(2\lambda AE[R])^m}{m!} e^{-2\lambda AE[R]}, \\ E\left[\mathrm{IRN}^F\right] = \int_0^{\infty} E[\mathrm{IRN}(r)]\, dF_R(r) = 2\lambda AE[R]. \end{gathered} \tag{73}$$

For the number of nodes out of the radio range of X_0, we have

$$\begin{gathered} E\left[\mathrm{ORN}^F\right] = E\left[\mathrm{CN}^F\right] - E\left[\mathrm{IRN}^F\right], \\ E\left[\mathrm{hops}^F\right] \le E\left[\mathrm{CN}^F\right], \\ P\left(\mathrm{hops}^F \ge h\right) \le P\left(\mathrm{CN}^F \ge h\right), \end{gathered} \tag{74}$$

where hops^F is the needed number of hops to reach X_0 for a sent message.

4.4. Available Connectivity. The connectivity $C(x)$ stated before attempts to evaluate the connection probability between two nodes, either for single-hop broadcasting or for multi-hop relay. However, to completely reflect the connection possibility involving the potential chances, we introduce a new metric, that is, available connectivity, for a node to estimate its communication capability with X_0.

Definition 22. Available connectivity is defined as a metric to measure the connection possibility for a given node considering the contributions on connections from others.

Let i be an equipped vehicle running on a one-dimensional highway. $\mathrm{AC}(i,t)$, say the available connectivity, is an instantaneous variable at time t. However, considering the high dynamics and frequently changing topology of VANETs, it is difficult to calculate the available connectivity every second. In addition, to get the available connectivity in a large-scale VANET with hundreds of nodes is still impractical. Thereupon, we introduce two limitations on $\mathrm{AC}(i,t)$:

Limitation I.

$$C(x) \geq C_{\mathrm{th}}. \tag{75}$$

Limitation II.

$$p(\mathrm{hops} \leq n) \geq P_{\mathrm{th}}, \tag{76}$$

where C_{th}, P_{th} denote the lower bound of connectivity and the upper bound of the number of hops.

To make our work practical, we discretize the time continuous variable $\mathrm{AC}(i,t)$. Let θ be the sampling period, say available connectivity is measured every θ seconds. T is an element of the set $\{0,\theta,2\theta,\ldots\}$. Then, the available connectivity for any given node i can be rewritten as $\mathrm{AC}(i,T)$. In general, a larger value of $\mathrm{AC}(i,T)$ corresponds to a bigger available connectivity. This means there are many potential nodes that could be used as forwarders to reach the destination node. It is worth noting that, instead of considering the connectivity between a given node and the reference node for analysis simplicity, we now focus on any communication pair in the network.

Definition 23. Available connectivity includes two kinds of dissemination ways, say the direct and indirect. Namely, available connectivity is equal to the arithmetic sum of the direct connectivity and indirect connectivity.

(i) Direct connectivity: the connection probabilities between a given node and its IRNs.

(ii) Indirect connectivity: the connection probabilities between a given node and its IRNs.

Theorem 24. *The available connectivity of node i at time T is*

$$\mathrm{AC}(i,T) = \sum_{x_{k,T}} \sum_{n} p\left(hops = n, i \longleftrightarrow x_{k,T}\right), \tag{77}$$

where $x_{k,T}$ are the nodes satisfying Limitations I and II, and $i \leftrightarrow x_{k,T}$ represents that i and $x_{k,T}$ can communicate to each other at time T.

Similarly, we have

$$DirCon(i,T) = \sum_{x_{k,T}} p\left(hops = 1, i \longleftrightarrow x_{k,T}\right), \tag{78}$$

where $DirCon(i,T)$ indicates the direct connectivity for i at T. If hops between any two communication pair are larger than 2, we have

$$IndCon(i,T) = \sum_{x_{k,T}} \sum_{n} p\left(hops = n, i \longleftrightarrow x_{k,T}\right), \tag{79}$$

where $IndCon(i,T)$ denotes the indirectly connectivity for i at T.

With (77) and the definition of our available connectivity, we know that the number of hops influences the connection performance significantly. For different applications, the requirements of hop bounds are variously depending on the delay limitation, information generation frequency, energy efficiency, and so forth. Considering the high dynamics in VANETs, larger hops may correspondingly incur intolerable delays, thus resulting in the service failure especially for emergent events. Therefore, a small hop bound is acceptable in VANETs.

When using the shortest path route selection algorithm, the following two corollaries hold.

Corollary 25. *For the case that the radio range is a constant, Limitation I could be rewritten as*

$$\left(1 - e^{-\lambda A R}\right)^{hops} \geq P_{th}, \tag{80}$$

thus

$$hops \leq \frac{\ln P_{th}}{\ln\left(1 - e^{-\lambda A R}\right)}. \tag{81}$$

Corollary 26. *Similarly, we can derive out the corresponding expression for the case where the radio range is a random variable depending on channel fadings. Consider*

$$\prod_{m=1}^{hops} \left(1 - e^{-\lambda A \min(R_{m-1}, R_m)}\right) \leq P_{th}. \tag{82}$$

With (82), the probability upper bound of the number of hops could be deduced.

5. Numerical Results

In this section, we will evaluate the performance of our proposed metric and investigate its feasibility and effectiveness with different simulation settings.

5.1. The Relationship between the Connectivity and R, λ, and A. According to (6) and (56), it is noted that the radio range R, the equipped vehicle arrival rate λ, and the speed factor A all have influences on the connectivity. In [37], the author provided some typical values for the expectation and standard deviation for vehicular speed on highway which follow the normal distribution as listed in Table 1. We will also use these settings in our later simulations.

Table 1: Normal-speed statistics.

μ (km/h)	σ (km/h)	A (h/km)
70	21	0.016105
90	27	0.012526
110	33	0.010249
130	39	0.008672
150	45	0.007516

Figure 6: The relationship between speed factor A (hour/km) and the expectation μ (km/hour), the standard deviation σ (km/hour).

The speed factor A is actually a function of μ and σ. Thus, we will firstly analyze the impact of μ, σ on A. The result is plotted in Figure 6. It is noted that the speed factor decreases when the mean of the velocity increases. In addition, the standard deviation has only a very limited influence on A. For σ, a great increasing of it could only slightly augment the speed factor.

The connectivity performance when the radio ranges of all nodes are the same constant is showed in Figure 7. In this simulation scenario, we placed 31 vehicles along a one-dimensional highway and investigated the resulted connectivity for a given node at x with different R (m), λ (vehicles/hour), and A (hour/km).

From Figure 7, it is noted that when $x > R$, $C(x)$ decreases with the inter-vehicle distance to the reference node increases. However when $x < R$, the connectivity will reasonably be 1. Therefore, it can be noticed that all the curves in Figure 7 show a sharp change at $x = R$ where the amplitude of this variation is determined by the values of λ, A, and R. Indeed, any rise of λ or A leads to an increase of $C(x)$. Meanwhile, a larger radio range improves the connectivity performance greatly as shown in Figure 7. In addition, because A is determined by the mean and standard deviation of the speed, a vehicle with a large A will correspond to a smaller mean and standard deviation. Therefore, with other parameters unchanged, a bigger A will

Figure 7: Connectivity versus x (m) with the radio range R (m), vehicle arrival rate λ (vehicles/hour), and speed factor A (hour/km) changing when all the nodes use the same constant radio range.

increase the connection probability considering moderate topology change.

As a result, it is noted that the variation of the radio range R will mainly influence the connectivity performance compared to other factors. With this conclusion, a critical radio range could be calculated for transmitting power saving consideration and applied to scenarios with determined connectivity requirements.

5.2. The Isolation Probability. The probability that a given node is isolated plays an important role in VANETs. Usually, it can be used for algorithms concerning clustering forming, relay selection, routing metric, and so forth. To reflect its influence on connectivity, the isolation probability is shown in Figures 8 and 9 based on (17) and (65), where Figure 8 corresponds to the constant radio range case and Figure 9 for the random radio range case. It is noted that in Figure 8, R is still the main influential factor to P_I. Speed factor A is in inverse proportion to the isolation probability where a large A will decrease the isolated possibility for a given node. The reason is the same with Figure 7 that a larger A means a less dynamic network, thus resulting in a relatively stable topology. Additionally, the increase of λ will reduce P_I due to more neighbors available.

For random radio range case, we compare the isolation probability under the deterministic, Rayleigh, and shadowing fading channel, respectively, as shown in Figure 9. From this figure and (33), (61), and (63), we notice that the Rayleigh shows the worst performance of P_I while lognormal shadowing seems be suitable to reduce isolations between nodes and the deterministic model has a decent performance in between the other two.

Figure 8: The isolation probability considering nodes on both sides for the constant radio range case.

Figure 9: The isolation probability P_I considering nodes on both sides under different channels where $A = 0.012526$ hour/km, $KP_{tx}/(\psi P_{\text{noise}}) = 3 \times 10^4$; (1) deterministic channel, $R = 108$ m; (2) Rayleigh channel, $\alpha = 2.2$; (3) shadowing fading channel, $\alpha = 2.2$, $\sigma_s = 2$.

5.3. The Number of Connected Nodes. Based on Theorem 16 and (14), the average tail probability of the number of connected nodes is plotted in Figure 10 for the constant radio range case. It is noted that the speed factor, arrival rate, and radio range all greatly influence the average tail probability. A smaller A, which corresponds to a larger average speed, will decrease the probability of vehicles forming a bigger cluster.

Figure 10: The average tail probability of the number of connected nodes for the constant radio range case.

The reason is that high mobility will make the network more dynamic and links unstable. As for the arrival rate λ, it seems to be in positive proportion to the average tail probability. In addition, among the three influential factors, radio range still contributes the most, which is consistent with the results from Figures 7 and 8.

Figure 11 depicts the corresponding average tail probability of the random radio range case. We set $KP_{tx}/(\psi P_{\text{noise}})$ to 1.2×10^5, the path loss factor α to 2.2, and the standard deviation for the lognormal shadowing to 2. Curves in Figure 11 show that a large arrival rate and low average velocity both increase the number of connected nodes. Compared to the curves from the lognormal shadowing in Figure 11, the results corresponding to the Rayleigh fading channel all show worse performance when other parameters are the same.

5.4. Available Connectivity. In this subsection, we will investigate the performance of our introduced available connectivity metric. For simplicity, the threshold of connectivity and P_{th} are set to 0.1 and 0.5, respectively. The sampling period to investigate the available connectivity is every 4 seconds.

The performance for the constant radio range case is shown in Figure 12. When setting the arrival rate λ to 500 vehicles/hour, speed factor A to 0.008672 hour/km (corresponding to $\mu = 130$, $\sigma = 39$ km/hour), and the radio range R to 200 m, the available connectivity of the reference node within 80 seconds is almost zero. After increasing A to 0.016105, the available connectivity begins to improve. It means that low average velocities lead to high available connectivity. In Figure 12, the radio range and vehicle arrival rate both greatly influence the available connectivity and the radio range is still the main factor, which is consistent with the previous results. A little rise of the radio range leads to

Figure 11: The average tail probability of the number of connected nodes on the right of the reference node for the random radio range case.

Figure 12: Available connectivity versus t (second) within 80 seconds for the case of constant radio range.

a great increasing of the available connectivity. Thus, when traffic flow is sparse, that is, in free-flow state, the better way to improve the available connectivity is increasing the radio range. However, when traffic is dense, growth of radio range may not be a better choice for obtaining available connectivity considering the introduced interferences.

Figure 13: Available connectivity versus t (seconds) within 80 seconds for the random radio range case.

The available connectivity in random fading channels are plotted in Figures 13 and 14, where we set $KP_{tx}/(\psi P_{\text{noise}})$ to 1.2×10^5, the path loss factor α to 2.2, and the standard deviation for lognormal shadowing to 2. The available connectivity of the reference node within 80 seconds in Rayleigh fading and lognormal shadowing channel is described in Figure 13. Similar to the results in Figure 12, the curves are approximately straight lines. Besides, high equipped vehicle arrival rate leads to a high available connectivity. For example, the available connectivity is about 9.5 when λ is equal to 500 vehicles/hour, and is about 13 when λ is equal to 700 vehicles/hour when the lognormal shadowing channel is used. Additionally, it can be noticed that a lower average velocity may pose positive impact on the available connectivity. Further, because the average radio range in shadowing channel is greater than that in Rayleigh channel, the available connectivity is higher under the shadowing fading channel.

To show the impact of hop bounds on the available connectivity, the performance from different hop bounds and channel fadings is plotted in Figure 14. Note that the upper bound of hops in Figure 13 is adaptively adjusted according to the average radio range. Because the hop bound is decided by application requirements, we introduce different settings including 5 (correspond to emergency case), 10, and 20 (corresponding to the delay tolerable services) in Rayleigh fading channel and lognormal shadowing channel. As shown in Figure 14, it is noticed that the different hop bounds settings greatly influence the available connectivity and a larger hop bound means a higher available connectivity. Thereupon, hop bounds should also be taken into consideration when designing different applications in VANETs.

Figure 14: The impact of hop bounds on the available connectivity under random fading channels.

6. Conclusion

In this paper, a useful metric, that is, available connectivity has been introduced in VANETs. Our proposed available connectivity comes not only from direct connections from neighbors, but also from the indirect links by multi-hop relay. The elaborate investigations of the related statistical properties of the connectivity have been given to reflect the influential factors and give the important references for applications design. Numerical results show that our proposed available connectivity has many potential relationships with network parameters and may provide important references for future protocols design in VANETs.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61201133, 61172055, and 61072067), Xian Municipal Technology Transfer Promotion Project (CX12178(6)), the Fundamental Research Funds for the Central Universities (K5051301011), the Postdoctoral Science Foundation of China (20100481323), the Program for New Century Excellent Talents (NCET-11-0691), the "111 Project" of China (B08038), and the Foundation of Guangxi Key Lab of Wireless Wideband Communication & Signal Processing (11105).

References

[1] L. Bononi and M. Di Felice, "A cross layered MAC and clustering scheme for efficient broadcast in VANETs," in *Proceedings of the IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems (MASS '07)*, pp. 1–8, Pisa, Italy, October 2007.

[2] L. Wang, S. K. Zhang, X. N. Wang, and Y. Zhang, "Research on vehicle automatically tracking mechanism in VANET," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 592129, 10 pages, 2013.

[3] D. Wu, Y. Ling, H. S. Zhu, and J. Liang, "The RSU access problem based on evolutionary game theory for VANET," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 143024, 7 pages, 2013.

[4] S. Eichler, "Performance evaluation of the IEEE 802.11p WAVE communication standard," in *Proceedings of the IEEE 66th Vehicular Technology Conference (VTC '07)*, pp. 2199–2203, Baltimore, Md, USA, September 2007.

[5] T. ElBatt, S. K. Goel, G. Holland, H. Krishnan, and J. Parikh, "Cooperative collision warning using dedicated short range wireless communications," in *Proceedings of the 3rd international workshop on Vehicular ad hoc networks (VANET '06)*, pp. 1–9, York, NY, USA, September 2006.

[6] C. Chen, L. Liu, Q. Q. Pei, X. B. Du, and X. M. Zhao, "Improving driving safety based on safe distance design in vehicular sensor networks," *IEEE International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 469067, 13 pages, 2012.

[7] A. Carter, "The status of vehicle-to-vehicle communication as a means of improving crash prevention performance," Tech. Rep. 05-0264, National Highway Traffic Safety Administration, Washington, DC, USA, 2005.

[8] E. Schoch, F. Kargl, M. Weber, and T. Leinmüller, "Communication patterns in VANETs," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 119–125, 2008.

[9] H. Hartenstein and K. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, 2008.

[10] M. Torrent-Moreno, M. Killat, and H. Hartenstein, "The challenges of robust inter-vehicle communications," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '05)*, vol. 1, pp. 319–323, 2005.

[11] R. Nagel, "The effect of vehicular distance distributions and mobility on VANET communications," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '10)*, pp. 1190–1194, San Diego, Calif, USA, June 2010.

[12] A. Y. Abul-Magd, "Modeling highway-traffic headway distributions using superstatistics," *Physical Review E*, vol. 76, no. 5, Article ID 057101, 4 pages, 2007.

[13] S. Yousefi, E. Altman, R. El-Azouzi, and M. Fathy, "Analytical model for connectivity in vehicular Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3341–3356, 2008.

[14] C. Young, B. R. Chang, J. Lin, and R. Fang, "Cooperative collision warning based highway vehicle accident reconstruction," in *Proceedings of the 8th International Conference on Intelligent Systems Design and Applications (ISDA '08)*, vol. 1, pp. 561–565, Kaohsiung, Taiwan, November 2008.

[15] C. Chen, L. Liu, X. B. Du, X. L. Wei, and C. X. Pei, "Available connectivity analysis under free flow state in VANETs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, article 270, 19 pages, 2012.

[16] Z. Yan, H. Jiang, Z. Shen, Y. Chang, and L. Huang, "K-connectivity analysis of one-dimensional linear VANETs," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 426–433, 2012.

[17] L. Cheng and S. Panichpapiboon, "Effects of intervehicle spacing distributions on connectivity of VANET: a case study from measured highway traffic," *IEEE Communications Magazine*, vol. 50, no. 10, pp. 90–97, 2012.

[18] A. Cardote, S. Sargento, and P. Steenkiste, "On the connection availability between relay nodes in a VANET," in *Proceedings of the IEEE Globecom Workshops (GC '10)*, pp. 181–185, Miami, Fla, USA, December 2010.

[19] I. W. H. Ho and K. K. Leung, "Node connectivity in vehicular ad hoc networks with structured mobility," in *Proceedings of the 32nd IEEE Conference on Local Computer Networks (LCN '07)*, pp. 635–642, Dublin, Ireland, October 2007.

[20] M. M. Artimy, W. Robertson, and W. J. Phillips, "Connectivity in inter-vehicle ad hoc networks," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering*, vol. 1, pp. 293–298, May 2004.

[21] D. P. Spanos and R. M. Murray, "Robust connectivity of networked vehicles," in *Proceedings of the 43rd IEEE Conference on Decision and Control (CDC '04)*, vol. 3, pp. 2893–2898, December 2004.

[22] M. M. Artimy, W. J. Phillips, and W. Robertson, "Connectivity with static transmission range in vehicular ad hoc networks," in *Proceedings of the 3rd Annual Communication Networks and Services Research Conference*, pp. 237–242, May 2005.

[23] M. Kafsi, P. Papadimitratos, O. Dousse, T. Alpcan, and J. P. Hubaux, "Vanet connectivity analysis," http://arxiv.org/abs/0912.5527.

[24] S. Durrani, X. Zhou, and A. Chandra, "Effect of vehicle mobility on connectivity of vehicular ad hoc networks," in *Proceedings of the IEEE 72nd Vehicular Technology Conference Fall (VTC '10)*, pp. 1–5, Ottawa, Canada, September 2010.

[25] M. Khabazian and M. K. M. Ali, "A performance modeling of connectivity in vehicular Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2440–2450, 2008.

[26] W. L. Jin and B. Wang, *Connectivity of Vehicular Ad Hoc Networks with Continuous Node Distribution Patterns*, Faculty Research, University of California Transportation Center, Berkeley, Calif, USA, 2010.

[27] L. Liu and D. H. Shi, "Busy Period in GIX/G/∞," *Journal of Applied Probability*, vol. 33, pp. 815–829, 1996.

[28] R. P. Roess, E. S. Prassas, and W. R. McShane, *Traffic Engineering*, Pearson/Prentice Hall, New Jersey, NJ, USA, 2004.

[29] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, John Wiley & Sons, New York, NY, USA, 1950.

[30] W. Stadje, "The busy period of the queueing system M/G/∞," *Journal of Applied Probability*, vol. 22, no. 3, pp. 697–704, 1985.

[31] L. Hanzo, C. H. Wong, and M. S. Yee, *Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalised and Space-Time Coded TDMA, CDMA and OFDM Systems*, John Wiley and Sons, New York, NY, USA, 2002.

[32] D. Miorandi and E. Altman, "Connectivity in one-dimensional ad hoc networks: a queueing theoretical approach," *Wireless Networks*, vol. 12, no. 5, pp. 573–587, 2006.

[33] D. Miorandi and E. Altman, "Coverage and connectivity of ad hoc networks in presence of channel randomness," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, pp. 491–502, March 2005.

[34] M. Boegsted, "Connectivity analysis of one-dimensional ad-hoc networks," *Wireless Networks*, vol. 17, no. 1, pp. 87–101, 2011.

[35] M. Telek and G. Horvath, "A minimal representation of Markov arrival processes and a moments matching method," *Performance Evaluation*, vol. 64, no. 9-12, pp. 1153–1168, 2007.

[36] M. F. Neuts, *Algorithmic Probability: A Collection of Problems*, Chapman and Hall, London, UK, 1995.

[37] M. Rudack, M. Meincke, and M. Lott, "On the dynamics of ad hoc networks for inter-vehicles communications (IVC)," in *Proceedings of the International Conference on Wireless Networks (ICWN '02)*, pp. 40–48, 2002.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 360267, 10 pages
http://dx.doi.org/10.1155/2013/360267

Research Article

Mixed Cooperation MAC Protocol with Sleep Mechanism for Data Acquisition in Wireless Machine-to-Machine Networks

Yulei Zhao, Bing Du, and Ning Ge

State Key Laboratory on Microwave and Digital Communications, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Yulei Zhao; sunstone2222@163.com

Received 18 July 2013; Revised 3 September 2013; Accepted 15 September 2013

Academic Editor: Lin Bai

Copyright © 2013 Yulei Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

By exploring the idle state of each sensor node, a mixed sleep-cooperative time division multiple access (TDMA) media access control (MAC) protocol (MS-CTDMA) is proposed for the wireless data gathering network in wireless machine-to-machine (M2M) networks. The basic idea is that in the idle state, each sensor node dynamically goes into sleep state or cooperative state to maximize the network lifetime. A single-hop network delay model of MS-CTDMA is established by the $M^X/G/1$ N-policy queue with multiple exhaustive vacations, which provides a way to balance energy consumption in network and the delay of each node with the delay sensitive traffic constraint. Furthermore, based on some reasonable assumptions, the proposed MS-CTDMA analysis model can be extended to the whole M2M network. Consequently, we propose an optimal source node selection strategy from the perspective of the relay node during its idle time, regarding the traffic load, residual energy, and channel state. Numerical results reveal that with proper tradeoff between delays and energy conservations, MS-CTDMA can significantly prolong the network lifetime in Rayleigh fading channels compared to pure sleep scheme.

1. Introduction

Machine-to-Machine (M2M) technology is widely used to gather information, and the machines are small and low-power. Cooperative communications enhance the link quality of the wireless network in a distributed fashion. The peer terminals in wireless M2M networks can achieve the diversity gain via mutual cooperation realized by the prevalent cooperative relay protocols, such as amplify-and-forward (AF) and decode-and-forward (DF) in [1]. In fact, the source and the relay form the virtual antenna array which brings benefits of the spatial diversity. For data acquisition system in wireless networks, a node receives transmission signals sent by its neighbors due to the wireless broadcast nature. Reference [2] proposed iterative collaborative relay beamforming strategies based on AF to maximize the received signal-to-interference-and-noise ratio (SINR) in wireless M2M networks. Time division multiple address (TDMA) is adopted to guarantee the quality of service (QoS) of the traffic data. More important, TDMA system can save energy by permitting the nodes to work with sleep mode in wireless data acquisition network [3]. Recently, several cooperative retransmission media access control (MAC) protocols have been investigated for TDMA systems. In [4], a Cooperative TDMA (C-TDMA) protocol was proposed to improve the throughput of the wireless network. Reference [5] extended the previous work to the dynamic slot assignment scheme for TDMA systems.

However, cooperations introduce delay to the system. In order to investigate the delay character of the cooperative relay network, the researchers first established a delay model of the single source and single-relay cooperative automatic-repeat request (ARQ) protocols for failure transmission [6], in which the Poisson arrival model of frames is assumed to estimate the delay caused by the cooperative relay when error frame triggers the retransmission.

For the consideration of saving energy in wireless sensor networks, [7] demonstrated that the network lifetime of wireless sensor network is more important for data collection network. Reference [8] established the energy model of single-hop wireless sensor networks. The transceiver of each sensor node uses energy when sending, receiving, or listening, and the ratio energy consumption is about 1.5 : 1 : 1 [9, 10]. Energy

efficiency and power allocation of cooperation transmission in wireless data gathering network were investigated in [11]. Reference [11] adopted the minimum energy consumption (MIE) policy and maximum residual energy (MARE) policy to allocate transmit power between the relay and the source node. Based on the control message of RTS-CTS, the proposed transmission scheme could decrease the signaling overhead. Reference [12] extended MARE policy to pricing strategy to maximize the network lifetime. The proposed energy pricing strategy could balance each node's energy consumption. Besides, [13] proposed the improved power allocation scheme based on communication distance for each group to maximize the lifetime of the whole sensor network. The priority queuing model combined with the vacation queuing model [14] is used to analyze the energy-aware MAC for differentiated services in wireless packet networks.

Only in sleep state, the energy consumption is very low. The microprocessor embedded in sensor node can go to the sleep state automatically when there is no task. So it is essential to introduce sleep mode to wireless sensor MAC protocol to save power [9]. Reference [15] adopted the vacation queue model, depicted extensively in [16], to study the sleep mode in the 802.16e system, which increased the energy efficiency and prolonged the lifetime of the wireless network.

However, it seems that the above cooperative transmission protocols cannot consider the incurred overhead extensively, which includes energy consumption of the relay node, additional delay of the relay traffic, and signaling for cooperating nodes selection [17]. Furthermore, if a terminal is in its idle state (here, we define, idle state is that the incoming buffer length of one node is less than the predefined threshold), it can become a relay node or sleep node according to its residual energy, traffic load, and received signal strength. When the node selects cooperative target during its idle state, many-to-one cooperative communication topology is established, and [18, 19] used game theory to analyze power allocation problem for this topology. When the node selects sleep scheme during its idle state, intuitively it saves power for itself, but if adopting random or independent sleep scheme, the network lifetime of sensor network cannot be maximized, because the network lifetime is attributed to the shortest duration time of the sensor node. Thus, in idle state we combined the sleep mode with the cooperative transmission, which is able to average energy cost over several sensor nodes. Both cooperative scheme and sleep scheme incur the transmit delay.

To deal with both energy and delay challenges, we propose a mixed cooperation MAC protocol with sleep mechanism (MS-CTDMA) for data acquisition system. And a queue model for MS-CTDMA protocol is given to measure the pros and cons of the proposed transmission. The contributions of this research are mainly at the following three points.

(1) MS-CTDMA protocol is introduced by fully utilizing the idle time of each node to save power. During idle state each node can select sleep scheme or cooperative transmission.

(2) We establish a $M^X/G/1$ N-policy vacation queue model for MS-CTDMA to carry out the performance analysis. The vacation time in $M^X/G/1$ is equivalent to the idle time in MS-CTDMA.

(3) Considering many-to-one cooperation communication topology, we give the optimal node selection strategy from the perspective of the relay naming source node selection, which is more efficient to average energy over the whole network.

The rest of the paper is organized as follows. Section 2 describes MS-CTDMA protocol elaborately. In Section 3, we establish the $M^X/G/1$ N-policy multiple vacation queue model for MS-CTDMA. Section 4 derives the system performance extensively. The optimal source node selection strategy is also depicted in this section. Numerical results and discussions are given in Section 5. And finally Section 6 draws the conclusions and the future work.

2. System Model with MS-CTDMA Protocol

2.1. Network Model. The research in this paper considers the wireless data acquisition network with sensor nodes and a sink node. In this network, each node only transfers collected data to the sink node. We focus on single-hop wireless network which means that all nodes are within one-hop transmit range. Moreover, each node can obtain spatial diversity through 2-hop cooperative transmission. This wireless packet system adopts TDMA scheduling protocol. Each node has its own allocated time slot and only has the chance to access the channel in its time slot. All nodes are synchronized and the MAC frames have equal length. Without loss of generality, this study concentrates on single relay cooperative scheme which incurs less overhead. As depicted in Figure 1, this data acquisition system contains M sensor nodes and one sink node. Each node alternatively acts as source node during its busy state or as relay node within its idle state. A frame time is divided into two fields: a control time slot (CT) and a data time slot (DT). CT conveys the scheduling information for each node. DT contains the time slots reserved for data transmission. Each node coordinates the cooperative transmission mechanism through CT.

An Rayleigh fading channel [20] is considered between two nodes, and the reciprocal channel is symmetry. $h_{i,j}$ denotes the complex channel gain for communication link from node i to node j. This research supposes that each sensor has the independent additive white Gaussian noise (AWGN) with the identical power spectral density (PSD) N_0. The instantaneous signal-to-noise ratio (SNR) of i to j can be expressed as [12]

$$\gamma_{i,j} = \frac{P_i \left| h_{i,j} \right|^2}{N_0 W}, \tag{1}$$

where P_i denotes the transmit power of node i and W is the transmit bandwidth. All sensor nodes are assumed to have the constant modulation and coding, so they have the same packet data transmit rate I_p. Without assistance from the relay

Figure 1: Typical data acquisition system and time slot assignment.

node, the transmit power P_i^d of node i satisfies (2) to achieve the packet data transmit rate [11]:

$$P_i^d \geq \frac{1}{\lambda_{i,j}} \left(2^{I_p/W} - 1\right). \tag{2}$$

For presentation brevity, $\lambda_{i,j}$ is defined as $\lambda_{i,j} = |h_{i,j}|^2/(N_0 W)$.

2.2. MS-CTDMA MAC Protocol. To maximize the sensor network lifetime, which is defined as the network working time until one sensor node uses up its energy, we proposed MS-CTDMA protocol. Introducing sleep mechanism into sensor node's idle state can reach low network energy consumption, but it also prolongs average packet delay. And the cooperative transmission could average power consumption over sensor nodes to lengthen the network lifetime. For example, as shown in Figure 1, $d_{M-1,D}$ is greater than $d_{1,D}$ and $d_{i,j}$ denotes the distance between node i and j. So node $M-1$ drains up its energy faster than node 1 under the same traffic load. Through combining cooperative relay strategy with sleep mode, the source node gets benefit of lowering its transmit power with the help of the relay node. The key idea of our proposed protocol is to efficiently utilize each node's transmit power in its idle state by choosing to sleep or to cooperate, so as to maximize network lifetime.

Each node is in one of the following states: init state, busy state, setup state, closedown state, sleep state or cooperation state. The state transition diagram is depicted in Figure 2. A node is said to be in sleep state when it turns off the radio transreceiver to save its battery energy. In a sensor node's cooperation state, the idle node acts as the relay node to save the source node's energy consumption by cooperative transmission based on source node selection, not relay node selection.

Initially each node starts out with the initial state. In this state the sensor node completes initialization and waits for the packet arrival. If the sensor has data to transmit, it transfers to the busy state. During the busy state, this node acts as the source terminal S. S is listening to the cooperative relay request (CREQ) from other idle sensor nodes by signaling packet. Then S sends the negotiation (CNEG) message back to the request relay node. Once a request sensor node selects S as its cooperative target and replies to S by cooperative acknowledgement (CACK), S and this relay node establish the cooperative relay link to help S to transmit. After all

Busy state: service its own data traffic
Setup state: cooperative negotiation stage
Cooperation state: cooperative transmission stage
Sleep state: enter the sleep mode
Closedown state: release the cooperative transmission

Figure 2: State transition diagram of each sensor node.

packets in data buffer have been sent out, the node changes back to the idle state, where it can be a relay node R or go into sleep mode according to the following steps. If it were a relay node, first, R enters the setup state to select the optimal S by broadcasting CREQ. If R successfully finds the source node from other nodes, it transfers to the cooperative state. Otherwise, R fails to find an appropriate S and turns off the radio to go into the sleep state. After sleep or cooperation, R reexamines the queue length. If there is not enough traffic coming and the data buffer size is less than N which denotes the predefined threshold, R maintains the cooperative or sleep status. On the contrary, if R's data buffer size grows larger than N, it goes to the closedown state and releases the cooperative relay link or sleep mode. The node turns to be a source node to transmit its own data traffic until the next idle state is satisfied.

The criterion to choose optimal source node for R is vital for the mixed transmission strategy. The mixed strategy tries to give a tradeoff between the network lifetime and packet delay determined by considering the multiple important elements: the residual battery energy, the channel state information at the physical layer, the incoming traffic load at the MAC layer, and the packet delay for a different traffic type. During the same constant time interval, if the predefined threshold N is increasing, the number of entering setup state is decreasing. In other words, the overhead for sleep or cooperation is reducing. For the same reason, the long-sleep time T_s or cooperative time T_c leads to low overhead. But these energy saving behaviors also incur lengthy packet delay.

3. Vacation Queue Model for MS-CTDMA

The $M^X/G/1$ vacation queue model with N-policy multiple exhaustive vacations is explored to analyze the performance of MS-CTDMA protocol in this section. The access mechanism of MS-CTDMA is coincident with the N-policy vacation queue model inherently. The idle state in MS-CTDMA can be modeled by the vacation state in $M^X/G/1$ vacation queue model.

Here, we assume that the traffic data buffer size is unlimited, which is reasonable when analyzing the average

performance of the MS-CTDMA systems. The incoming traffic of each terminal is assumed to be compound Poisson process. The batch arrival rate is λ, and the number of packet in each batch is denoted by stochastic variable X. Batch Markovian arrival process (BMAP) is suited to the bursty nature of packet traffic for wireless data acquisition scenario [21]. The data acquisition moment is triggered by outside environment changing which coincides with the model of Poisson arrival. At this moment, a new batch of sensor data with random packets is generated into the transmit buffer. In this paper, all sensor nodes are equal and have the same traffic load.

We give some general assumptions first. The packet length keeps constant, that is, L_d for data packet and L_s for signaling packet. The number of arrival packets obeys discrete random distribution which is determined by a different sensor type. The discrete probability density function is depicted by $p\{X = x_i\} = g_i$, where $i = 1, 2, \ldots, E$ denotes random acquisition type. Over each transmission link, the physical layer adopts capacity-achieving code and adjusts the transmit power to assure the required bit error rate (BER) [11], so the retransmission mechanism is not considered in this paper. Thus the service time is only determined by the packet transmission time. I_i denotes the packet transmission rate of the terminal i, $(i = 1, 2, \ldots, M)$. Coding and modulation of the transmission keep constant without adaptation, so I_i does not vary with channel state information. Without loss of generality, each sensor node has the same processing ability, that is, $I_i = I_p$ for each i. The corresponding data packets are served as customers on first-come-first-serve (FCFS) basis until the system becomes empty. S_i, $(i = 1, 2, \ldots, M)$ denotes the service time of each terminal, also known as the transmission time in physical layer. Based on the above assumption, S_i follows the deterministic distribution, and the mean of S_i can be expressed as

$$E\left[S_i\right] = \frac{L_d}{I_i}. \tag{3}$$

3.1. State Transition for MS-CTDMA Protocol. First, we illustrate the state transition for MS-CTDMA protocol. In proposed MS-CTDMA protocol with $M^X/G/1$ N-policy vacation queue model, the setup time and closedown time are constant values, denoted by T_u and T_d, respectively, which is determined by the S's negotiation mechanism. Cooperative time T_c is independent identical distributed (i.i.d) random variable with a general distribution function, denoted by $F_c(t)$. Sleep time T_s is also i.i.d random variable depicted by $F_s(t)$. Figure 3 shows the state transitions for each node's operation cycle in the established model [22]. When the incoming traffic length of one node is larger than N, this node transfers from the init state to the busy state. And it services its own data traffic until the queue length is empty. When data length in the queue is less than N, according to MS-CTDMA protocol, it goes into the setup state to determine whether to sleep or to cooperate during its idle state or vacation state. After the vacation time, if the node's queue length is growing larger than N, the state transfers into closedown state and ends from vacation state.

The vacation time in $M^X/G/1$ N-policy vacation queue model can be regarded as the cooperative time or sleep time, which is determined by optimal source selection strategy in MS-CTDMA. And in this research, the vacation time follows the deterministic cumulative distribution function $F_v(t)$, which is

$$F_v(t) = \begin{cases} 0, & t < T_{VE} \\ 1, & t \geq T_{VE}. \end{cases} \tag{4}$$

$F_v(t)$ denotes its probability density function. The mean value of the vacation time is $E[T_v] = T_{VE}$, which is the integral time of the packet transmission time. The long T_{VE} leads to low overhead and vice versa. In this paper, T_c and T_s have identical probability distribution with T_v.

3.2. The Average Queue Length. We establish the embedded Markov chain to calculate the average queue length in this subsection. The queue length at the packet departure instants can be mapped into the state space of the embedded Markov chain. Let D_n, $n = 1, 2, \ldots$ denote the state space or the embedded Markov chain, where $n = 1, 2, \ldots$ is the packet departure instants. They satisfy

$$D_{n+1} = \begin{cases} D_n - 1 + A, & \text{for } D_n \geq N \\ D_n + A, & \text{for } D_n < N, \end{cases} \tag{5}$$

where A denotes the arriving packets during the vacation time and is determined by batch traffic arrival rate λ. From [16] we know that the rmbedded Markov chain is positive recurrent when the utilization load $\rho < 1$. The queue length D_q in this $M^X/G/1$ vacation model can be decomposed into the sum of the two independent random variables:

$$D_q = D + D_v, \tag{6}$$

where D is the mean queue length of an ordinary $M^X/G/1$ queue without vacations and D_v is the additional mean queue length due to the vacation effect.

3.3. The Mean Waiting Time. The mean packet queue waiting time W_q of $M^X/G/1$ multiple vacation models combined with N-policy is used to measure the quality of service (QoS) incurred by packet delay. According to tht Poisson arrivals (see time average (PASTA)) [16], the average queue waiting time of the embedded Markov chain is equal to steady state waiting time. From [23], we can get

$$W_q = \left(D_q + E\left[X_R\right]\right) E\left[S\right] + P(B)\, E\left[S_R\right] + P(I)\, \frac{E\left[N_{I_R}\right]}{\lambda E\left[X\right]}, \tag{7}$$

where $P(B)$ and $P(I)$ indicate the probability of busy time and idle time of each sensor node. They can be expressed as $P(B) = \rho$ and $P(I) = 1 - \rho$. From Little's theorem [16], $D_q = \lambda E[X] W_q$. So (7) can be formulated as follows:

$$W_q = \frac{E\left[S\right]}{1-\rho} E\left[X_R\right] + \frac{\rho}{1-\rho} E\left[S_R\right] + \frac{E\left[N_{I_R}\right]}{\lambda E\left[X\right]}, \tag{8}$$

Figure 3: The time distribution with the state transition.

where $E[X_R]$ denotes the former waiting packet in the same batch and $E[S_R]$ is residual service time. From [23] we have

$$E[S_R] = \frac{1}{2}\frac{E[S_i^2]}{E[S_i]}, \tag{9}$$

$$E[X_R] = \frac{1}{2}\left[\frac{E[X^2]}{E[X]} - 1\right]. \tag{10}$$

For N-policy $M^X/G/1$ system with multiple vacations, the last component of (8) can be expressed as

$$\frac{E[N_{I_R}]}{\lambda E[X]} = E[V_R] + \frac{1}{\lambda E[X]}\frac{\sum_{n=0}^{N-1} n\beta_n}{\sum_{n=0}^{N-1}\beta_n}, \tag{11}$$

where $E[N_{I_R}]$ means the additional packets generated by sensor devices during the idle sate. $E[V_R]$ is the expected duration of the residual vacation and can be expressed as

$$E[V_R] = \frac{E[V^2]}{2E[V]}. \tag{12}$$

Through (9)–(12), we could obtain the mean waiting time W_q from (8).

3.4. The Average Cooperative Probability. Each node selects its working scheme within its vacation time or idle state. So the probability of the event P_c that each node takes action to sleep or cooperate for the sink node equals the ratio of the vacation time to one cycle total serving time, which is defined as the interval between two consecutive busy period ending instants, denoted by T_{NN}. And the N_{th} traffic frame accumulation period is defined by T_N. The forward recurrence time of the vacation time is delegated by T_R. The time that the relay node R processes the traffic queue is denoted by T_A. The relationship among T_{NN}, T_N, T_R, and T_A is shown in Figure 3. Equation (13) can be derived from the embedded Markov analysis in Section 3.2 [16] as follows:

$$E[T_{NN}] = \frac{1}{1-\rho}\left(\frac{N}{\lambda_i} + E(T_R)\right). \tag{13}$$

Hence, the cooperative time T_{VAC} is

$$E[T_{\text{VAC}}] = \frac{N}{\lambda_i} + E(T_R). \tag{14}$$

From (13) and (14), we can get the cooperative relay probability $P_c = E[T_{\text{VAC}}]/E[T_{NN}] = 1-\rho$.

4. Performance Metric and Optimal Source Selection Strategy

In this section, we first present the total energy consumption during one operation cycle. Then the delay sensitive problem is formulated. The analysis queue model of MS-CTDMA protocol can be extended to the whole network based on energy consumption target. The optimal source node selection strategy is depicted extensively when MS-CTDMA protocol is running on the one-hop network.

4.1. Total Energy Consumption. The energy consumption within one operation cycle T_{NN} can be expressed as

$$E_{T_{NN}} = E_{T_{\text{VAC}}} + E_{T_A}, \tag{15}$$

where $E_{T_{\text{VAC}}}$ and E_{T_A} denote the energy consumption during vacation state and busy state, respectively. T_{VAC} and T_A are random variables, so it is necessary to calculate the expected value of total energy consumption $E[E_{T_{NN}}]$. During busy state, node i serves its traffic data with transmit power P_{iS} and the expected energy consumption $E[E_{T_A}] = P_{iS}E[T_A]$.

During the vacation time, node i goes into sleep or cooperation state. The transmit power is denoted by P_{iR}, and the power value for receiving data from source node is equal to P_{iR}, when $P_{iR} = 0$ node i selects the sleep mode. The setup state at the beginning of vacation and the closedown state at the end of the vacation also use energy. From (14), when traffic load λ is constant, $E[T_{\text{VAC}}]$ is attributed to predefined threshold N and the probability distribution of vacation time. Each operation cycle needs one setup stage and closedown

stage. The expected vacation energy consumption can be depicted as

$$E\left[E_{T_{\mathrm{VAC}}}\right] = \frac{E\left[T_v\right]}{E\left[T_u + T_v + T_d\right]} P_{iR} E\left[T_{\mathrm{VAC}}\right] + E\left[T_u + T_d\right] P_{iR}. \tag{16}$$

The last component is the energy overhead.

4.2. Delay Tolerant Constraint. In data collection scenarios, the traffic load is always low and nonrealtime. In other words, the collecting data needs to be sent to the sink node more reliably, while the time requirement is relatively relax. However, if the delay is longer than the deadline time D_T, QoS of the traffic decreases rapidly, and the sink node is not able to process the usable data. Thus, mean waiting time must satisfy $W_q \geq D_T$. From (16), prolonging T_v and N can minimize the energy overhead, which also incurs longer average packet delay. We can formulate the problem as

$$\begin{aligned} &\min_{N,T_v} \frac{E\left[T_{\mathrm{VAC}}\right]}{E\left[T_B\right]} E\left[T_u + T_d\right] P_{iR} \\ &\text{s.t.}\ \ W_q \leq D_T, \end{aligned} \tag{17}$$

where T_B indicates the minimum vacation time. This optimization is a tradeoff between the additional delay for node i and decreasing energy overhead incurred by state transition. And this solution is independent with optimal source selection strategy during the vacation time.

4.3. Extend the Results to the Whole Network. M-node data collection system is an open network, each node has its independent BMAP traffic with the batch arrival rate λ_i, $i = 1, \ldots, M$. Particularly, in a single-hop TDMA system, each node transfers its traffic in its assigned time slot and receives other nodes' traffic during its idle state with the cooperative probability $P_c(i)$, $i = 1, \ldots, M$. During the busy state, each node has the probability $P_s(i)$, $i = 1, \ldots M$ to be the source node with a relay. So, the network scenario satisfies the definition of the open Jackson network from [16].

According to the Jackson network, an M-node network can be divided into several single nodes to analyze the network performance. For simplicity of deduction, we assume that there are always multiple available relays between the source and the destination, and one source node only chooses one cooperative relay node. Based on the independent characteristic of each node, P_c and P_s can be derived as

$$\begin{aligned} P_c &= P(I)\left(1 - P(I)^{(M-1)}\right), \\ P_s &= P(B)\left(1 - P(B)^{(M-1)}\right). \end{aligned} \tag{18}$$

4.4. Optimal Source Selection Strategy. In this paper, the cost function [12] is established to evaluate operation mode of each node. At ending instants of the busy period, each node determines whether to sleep or to be as a cooperative relay for one source node. When the node acts as a source node, it also has the opportunity to be selected as the cooperative target by a relay node. During T_{VAC}, each node can choose to sleep to save battery energy or to be a relay node and find the optimal cooperative source node to minimize the cost. Therefore, within T_A, each node may have the chance to be assisted by one node. The whole cost function can be formulated as

$$\begin{aligned} C_i &= P_c T_{\mathrm{VAC}}\left(\omega_i P_{iR} + \omega_k P_{kS}\right) \\ &\quad + P_s T_A\left(\omega_i P_{iS} + \omega_j P_{jR}\right) + \omega_i \frac{T_B}{T_{\mathrm{VAC}}} E_{TS}. \end{aligned} \tag{19}$$

The first component denotes the cost value during T_{VAC}, which is the stochastic random variable. P_{kS} depicts the transmit power of source node k selected by relay node i during its vacation or idle period. The second component is the cost within T_A, which is equal to $T_{NN} - T_{\mathrm{VAC}}$. P_{jR} represents the power allocation of relay node j, which selects the node i as its source node. The last component indicates that setup and release energy consumption cost of each node. And $E_{TS} = E[T_u + T_d]P_{iR}$. The expected value of C_i is expressed as

$$\begin{aligned} E\left[C_i\right] &= E\left[T_{VAC}\right]\left(\omega_i P_{iR} + \omega_k P_{kS}\right) \\ &\quad + E\left[T_A\right]\left(\omega_i P_{iS} + \omega_j P_{jR}\right) \\ &\quad + \omega_i \frac{E\left[T_B\right]}{E\left[T_{VAC}\right]} E_{TS}. \end{aligned} \tag{20}$$

Compared with (19), we incorporate P_c and P_s into the source node selection strategy. The first component and second component are determined by working strategy during vacation time and power allocation within one operation cycle. Minimization of the last component is discussed in Section 4.2. The optimal sleep or cooperation selection and power allocation are to minimize (20).

4.4.1. The Energy Cost Factor. The energy cost factor is critical for the optimal source node selection. In this paper, we take into account the residual battery energy, the initial battery energy, and the traffic load to calculate the cost factor as follows:

$$\omega_i = \omega_0 \left(\frac{E_{i,\mathrm{ini}}}{E_{i,\mathrm{rem}}}\right)^l \left(\frac{\rho_i}{\rho_B}\right)^m, \tag{21}$$

where ω_0 denotes the initial cost factor of each node in the network, which reflects priority level of each node. The proportion of $E_{i,\mathrm{ini}}$ to $E_{i,\mathrm{rem}}$ indicates the cost factor of per unit energy for each node. The node with low proportion is likely to assist other nodes in its vacation or idle state. ρ_B is the basic traffic measurement, and the ratio of ρ_i to ρ_B implies the normalized traffic load for each node. The node with traffic load is reasonable to sleep in its idle state, since it has already consumed a lot of energy. l and m are exponent components of energy and traffic, respectively, to adjust the cost between residual energy and traffic load. For the sake of simplicity, l and m take value 1 or 2.

4.4.2. The Optimal Source Selection Strategy. According to (20), the cost value can be decomposed into three independent targets. Within T_{VAC}, the optimal source node selection can be expressed as a linear programming problem:

$$\begin{aligned} &\min_{P_{iR},P_{kS},k} C_{i1} = \omega_i P_{iR} + \omega_k P_{kS} \\ &\qquad \lambda_{k,D} P_{kS} + \lambda_{i,D} P_{iR} \geq 2^{2I_p/W} - 1 \\ &\text{s.t.} \quad P_{S,\text{low}} \leq P_{kS} \leq P_{\max} \\ &\qquad 0 \leq P_{iR} \leq P_{\max}. \end{aligned} \tag{22}$$

The first constrained condition is depicted extensively in [24]. $P_{S,\text{low}}$ is determined by instantaneous SNR between nodes k and D denoted by $\gamma_{k,D}$ and between nodes k and i depicted by $\gamma_{k,i}$. $P_{S,\text{low}}$ guarantees that the achieved rate of k to D is I_p and that the transmit rate of k to i is $2I_p$. $P_{iR} = 0$ means that the node selects to sleep during its idle state. $P_{\max}$ is a constant value which is constrained by sensor device. So, it can be expressed as

$$\begin{aligned} P_{S,\text{low}} &= \min\{P_{S,\text{low}A}, P_{S,\text{low}B}\} \\ P_{S,\text{low}A} &= \frac{1}{\lambda_{k,D}}\left(2^{I_p/W} - 1\right) \\ P_{S,\text{low}B} &= \frac{1}{\lambda_{k,i}}\left(2^{2I_p/W} - 1\right). \end{aligned} \tag{23}$$

Similarly, the problem that each node becomes the optimal target source assisted by other relay nodes can also be formulated as the following linear programming problem:

$$\begin{aligned} &\min_{P_{iS},P_{jS},j} C_{i1} = \omega_i P_{iS} + \omega_j P_{jR} \\ &\qquad \lambda_{k,D} P_{iS} + \lambda_{j,R} P_{iR} \geq 2^{2I_p/W} - 1 \\ &\text{s.t.} \quad P_{S,\text{low}} \leq P_{iS} \leq P_{\max} \\ &\qquad 0 \leq P_{jR} \leq P_{\max}. \end{aligned} \tag{24}$$

The energy consumption incurred by state transition is a minimized base on the precondition that the average packet delay must be satisfied. This optimal programming problem is discussed in Section 4.2.

4.4.3. Solutions to Optimal Power Allocation. The linear programming problem (22) can be solved by geometric methods effectively introduced by [12]. From (23), when $P_{S,\text{low}A} \leq P_{S,\text{low}B}$, there is no need for node i to assist node k, since the cost of i and k would both increase under the cooperation relay scheme. But other than that, i and k will always get benefit from the cooperation transmission, when $P_{iR} = 0$, in other words, no power is allocated to i, and i goes to sleep state during its vacation time.

As shown in Figure 4, $Y1$, $Y2$, and $Y3$ represent the constrained condition $\lambda_{k,D} P_{kS} + \lambda_{i,R} P_{iR} = 2^{2I_p/W} - 1$ in the plane when I_p is set to different values. According to constraint values of $P_{S,\text{low}}$ and $P_m ax$, the valid power allocation region is the intersection of $ABEF$ and one of the upper half plans of $Y1$, $Y2$, or $Y3$. This optimal power allocation problem is convex optimization, so the optimal solutions are intersection point, such as A or C, which is determined by the slope of objective line and condition line. When $\omega_k/\omega_i \geq \lambda_{k,D}/\lambda_{i,R}$, that is, $C1$, the left point A is the solution point. On the contrary, when $\omega_k/\omega_i < \lambda_{k,D}/\lambda_{i,R}$, that is, $C2$, the right point F is the optimal point. The rectangle $ABEF$ is divided into three regions ABC, $ACDF$, and DEF.

FIGURE 4: Geometric solutions for different regions.

The research in this paper focuses on sleep state, $P_{iR} = 0$. So our study below is in region DEF, where I_p satisfies

$$\begin{aligned} 2^{2I_b/W} - 1 &> \lambda_{k,D} P_{S,\text{low}} \\ 2^{2I_b/W} - 1 &\leq \lambda_{k,D} P_{\max}. \end{aligned} \tag{25}$$

When $\omega_k/\omega_i \geq \lambda_{k,D}/\lambda_{i,R}$, the solutions is

$$\begin{aligned} P_{kS}^* &= P_{S,\text{low}} \\ P_{iR}^* &= \frac{1}{\lambda_{i,R}}\left(2^{2I_p/W} - 1 - \lambda_{k,D} P_{S,\text{low}}\right). \end{aligned} \tag{26}$$

In this case, the cost of per unit energy for source node k is much greater than relay node i. The cooperation scheme is able to average power consumption over i and k, which maximizes the network operation time.

When $\omega_k/\omega_i < \lambda_{k,D}/\lambda_{i,R}$, the solutions is

$$\begin{aligned} P_{kS}^* &= \frac{1}{\lambda_{k,S}}\left(2^{2I_p/W} - 1\right) \\ P_{iR}^* &= 0. \end{aligned} \tag{27}$$

In this case, node i selects going into sleep state to save power other than cooperative transmission. At the beginning of each vacation time of node i, (27) is solved. Likewise, the solutions for region, $ACDF$ and DEF can be solved. Due to space limitation, we omit the detailed results.

Figure 5: The average energy consumption and average waiting time by changing vacation time.

Figure 6: The energy consumption and average waiting time by changing predefined N.

4.4.4. The Implementation of Optimal Source Selection. The optimal relay selection is the usual way to perform cooperation. But in this paper, we implement the optimal source selection strategy within setup state during the vacation time. Once node i reaches setup state, it transmits CREQ control packet during CT. In the same frame time, when source node k receives CREQ signaling, it responses to node i a CNEG control packet containing the received signal strength indication (RSSI), traffic load, and residual battery energy. When node i gathers all the valid source candidates information, it makes decision whether to sleep or to cooperate during vacation time due to problem (22) and (24) solutions. Then, i sends the CACK packet data in the next frame time.

5. Numerical Results

In order to evaluate the performance of MS-CTDMA, we consider the scenario depicted in Figure 1. The sink node D is in the center of this region. In the research analytical evaluations, the signaling packet length L_c is 32 bits and data packet length L_d is 1000 bits. The received signal strength is lessening with proportion of square transmission distance. The fading is assumed to be constant in each vacation time. In addition, it is assumed that the average number of each batch is that $E[X] = 8$. Other parameters are set as $N_0 = 10^{-14}$ W/Hz, $W = 1$ MHz, $I_p = 1$ Mbps, and $P_{\max} = 0.1$ W, respectively.

First, we explore the relationship between average packet delay and average energy consumption. Three different batch arrival rates are assigned to each node, and the node adopts the identical pure sleeping strategy. In this evaluation, the cooperative mechanism is not considered, because it does not affect vacation overhead.

Figure 5 shows a plot of the average packet delay and average energy consumption by varying the vacation time T_v. We can save energy by increasing the vacation time or sleep time, but it also causes the packet delay. The longer T_v leads to low energy overhead for state transition. This benefit is obvious, and the energy overhead of state transition is stable at the end. Thus, ultimate energy conservation is determined by the traffic load, and the appropriate value of T_v is selected according to the system requirements. From Figure 5, the heavy traffic load can lead to less average energy consumption due to the reduced state transition number when serving the same number of packets. In this scenario, average waiting time is dominated by the vacation time.

Figure 6 shows several curves of the average energy consumption versus delays by varying the predefined threshold N. The increasing line demonstrates that the additional delay of each node introduced by N and the decreasing line demonstrates the average energy consumption of the same node. We clearly concluded that the tradeoff between average energy consumption and average delay can be adjusted by N. Compared to Figure 5, we discover that the average waiting time is insensitive to mean arrival time by changing T_v, but low traffic load can lead to heavy delay by varying N. So it is reasonable to adjust T_v to reach the delay and energy requirements.

Figure 7 shows the network lifetime of senor network using sleep scheme and our MS-CTDMA protocol by changing the relay node traffic load, when the predefined N has two values. In this simulation, three nodes have the same channel condition to the destination D, such as distance and fading condition. But the traffic load of relay node is lower than the other nodes. From Figure 7, we see that MS-CTDMA protocol considers the impact of different traffic load, and the nodes which have longer idle state are responsible for assisting other heavy traffic nodes to maximize the network

Figure 7: Network lifetime by changing average arrival time.

Figure 8: Network lifetime by changing distance to sink node.

lifetime. When the network operates with pure sleep scheme, the network lifetime is up to the shortest lifetime node. But with MS-CTDMA protocol, the node with more idle time or less traffic load can lend its energy to other nodes.

Figure 8 shows the network lifetime of the senor network by changing distance to sink node. Here, we assume that three nodes have the same traffic load. The relay node is near to the sink node so the channel condition is different. In this condition, the performance of MS-CTDMA is also superior to the pure sleep mode. Comparing Figure 8 with Figure 7, we find that the impact of network lifetime with relay traffic load

Figure 9: Network lifetime by changing average arrival time and distance.

is more distinct than that with relay distance to sink node by the predefined threshold N.

In Figure 9, we compare the network lifetime by changing traffic load and distance to sink node simultaneously. It is obvious that the maximization point is reached when the distance is nearest and the mean arrival time is smallest.

6. Conclusion

In this paper, we propose a novel mixed cooperative TDMA MAC protocol with sleep mode, namely, MS-CTDMA, for wireless data acquisition networks. When traffic queue length of a terminal is less than the predefined threshold N, it actively devotes itself to be a cooperative relay for its single-hop neighbor or go to the sleep state. And the $M^X/G/1$ N-policy vacation queue model is developed to analyze the delay performance of MS-CTDMA to obtain the best tradeoff between the energy consumption in network and the additional delay of the node. In addition, we find that the achievements of the proposed relaying model can be extended to the network under the Jackson network theorem's conditions. Then, an optimum source node selection strategy is given for the maximization of the network lifetime. The analysis and numerical results demonstrate that MS-CTDMA can improve the operation time of data collection significantly.

As a future work, we would like to investigate the multi-hop TDMA systems with multiple relays based on dynamic slot assignment scheme.

Conflict of Interests

The authors do not have any conflict of interests with the content of the paper.

Acknowledgments

This research is supported by the National Major Project no. 2011ZX03004-001-01 and National Natural Science Foundation of China no. 61201183 and no. 61132002.

References

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[2] D. Wang, L. Bai, X. Zhang, W. Guan, and C. Chen, "Collaborative relay beamforming strategies for multiple destinations with guaranteed QoS in wireless machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 525640, 10 pages, 2012.

[3] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[4] Z. Yang, Y.-D. Yao, X. Li, and D. Zheng, "A TDMA-based MAC protocol with cooperative diversity," *IEEE Communications Letters*, vol. 14, no. 6, pp. 542–544, 2010.

[5] J. K. Lee, H. J. Noh, and J. Lim, "Dynamic cooperative retransmission scheme for TDMA systems," *IEEE Communications Letters*, vol. 16, no. 12, pp. 2000–2003, 2012.

[6] I. Cerutti, A. Fumagalli, and P. Gupta, "Delay models of single-source single-relay cooperative ARQ protocols in slotted radio networks with poisson frame arrivals," *IEEE/ACM Transactions on Networking*, vol. 16, no. 2, pp. 371–382, 2008.

[7] W.-J. Huang, Y.-W. Peter Hong, and C.-C. Jay Kuo, "Lifetime maximization for amplify-and-forward cooperative networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1800–1805, 2008.

[8] D. Zhang and Z. Chen, "Energy-efficiency of cooperative communication with guaranteed E2E reliability in WSNs," *International Journal of Distributed Sensor Networks*, vol. 2013, p. 11, 2013.

[9] Q. Gao, K. J. Blow, D. J. Holding, and I. Marshall, "Analysis of energy conservation in sensor networks," *Wireless Networks*, vol. 11, no. 6, pp. 787–794, 2005.

[10] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom '01)*, pp. 166–179, Rome, Italy, July 2001.

[11] Z. Zhou, S. Zhou, J.-H. Cui, and S. Cui, "Energy-efficient cooperative communication based on power control and selective single-relay in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 8, pp. 3066–3079, 2008.

[12] F. Ke, S. Feng, and H. Zhuang, "Relay selection and power allocation for cooperative network based on energy pricing," *IEEE Communications Letters*, vol. 14, no. 5, pp. 396–398, 2010.

[13] J. J. Liaw, C. W. Chou, and C. Y. Dai, "The Lifetime Extension of Wireless Sensor Networks Using Adaptive Energy Allocation by Distance," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 982573, p. 8, 2013.

[14] A. Fallahi and E. Hossain, "Distributed and energy-aware MAC for differentiated services wireless packet networks: a general queuing analytical framework," *IEEE Transactions on Mobile Computing*, vol. 6, no. 4, pp. 381–394, 2007.

[15] E. Chu, B. H. Jung, D. K. Sung, and Y. Park, "Modeling and analysis of an energy-efficient sleep-mode operation in IEEE 802.16e system," in *Proceedings of the Communications IEEE International Conference on (ICC '09)*, June 2009.

[16] N. Tian and Z. J. Zhang, *Vacation Queue Models: Theory and Application*, Springer, New York, NY, USA, 2005.

[17] W. Zhuang and M. Ismail, "Cooperation in wireless communication networks," *IEEE Wireless Communications*, vol. 19, no. 2, pp. 10–20, 2012.

[18] M. W. Baidas and A. B. Mackenzie, "Auction-based power allocation for many-to-one cooperative wireless networks," in *Proceedings of the 7th International Wireless Communications and Mobile Computing Conference (IWCMC '11)*, pp. 1677–1682, July 2011.

[19] D. Yang, X. Fang, and G. Xue, "Game theory in cooperative communications," *IEEE Wireless Communications*, vol. 19, no. 2, pp. 44–49, 2012.

[20] J. G. Proakis and M. Salehi, *Digital Communications*, MCGraw-Hill, New York, NY, USA, 2007.

[21] Y. Wu, T. Shu, Z. Niu, and J. Zheng, "A vacation model with setup and close-down times for transmitter buffer of ARQ schemes," in *Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '03)*, pp. 2054–2058, September 2003.

[22] H. W. Lee, S. S. Lee, J. O. Park, and K. C. Chae, "Analysis of the $M^X/G/1$ with N-policy and mutiple vacations," *Journal of Applied Probability*, vol. 31, pp. 476–496, 1994.

[23] K. C. Chae and H. W. Lee, "$M^X/G/1$ vacation models with N-policy: heuristic interpretation of the mean waiting time," *Journal of the Operational Research Society*, vol. 46, no. 2, pp. 258–264, 1995.

[24] T. Wang, A. Cano, G. B. Giannakis, and J. N. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," *IEEE Transactions on Communications*, vol. 55, no. 7, pp. 1427–1438, 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 796549, 8 pages
http://dx.doi.org/10.1155/2013/796549

Research Article

An Energy-Heterogeneous Clustering Scheme to Avoid Energy Holes in Wireless Sensor Networks

Gong Bencan, Jiang Tingyao, Xu Shouzhi, and Chen Peng

College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China

Correspondence should be addressed to Gong Bencan; gonbc@sina.com

Received 11 July 2013; Accepted 2 September 2013

Academic Editor: Jianhua He

Copyright © 2013 Gong Bencan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Clustering techniques can reduce energy consumption of nodes and increase the scalability of the network. However, in uniformly deployed clustering wireless sensor networks, the uneven distribution of communication loads often causes the energy-hole problem, which means that the energy of the nodes in the hole region will be used up sooner than the nodes in other regions. In order to solve the problem, this paper theoretically analyzes the energy consumption of nodes in different network areas, gives the expression of the optimal energy distribution of heterogeneous nodes, and designs an energy-efficient clustering routing protocol. The goal of our work is to propose an energy-heterogeneous clustering scheme (EHCS) that allows the initial energy of the sensor nodes to be varied with the distance to sink. The nearer nodes from sink have more energy. Simulation results show that the method can balance the energy consumption among sensor nodes, achieve an obvious improvement on the network lifetime, and effectively avoid the energy-hole problem.

1. Introduction

Wireless sensor networks are composed of a large number of sensor nodes with limited energy resources. In many applications, the sensor nodes are usually inaccessible to the user after being deployed, and thus, replacement of the energy resource is not feasible. Hence, energy efficiency becomes a key design issue in order to improve the life span of the entire network. In multihop wireless sensor networks characterized by many-to-one traffic pattern, the cluster heads close to sink have to forward data for cluster heads away from sink and thus will tend to die earlier, resulting in network being partitioned and network lifetime being shortened. This is called the energy-hole problem. Simulations results [1] have showed that the energy expended by a sensor node drops significantly as they move away from the sink, and the power expended by a node in the 6th corona is less than 10% of the energy expended by a node in the first corona.

In order to prolong the network lifetime, this paper proposes an energy-heterogeneous clustering scheme where sensor nodes with larger data forwarding loads are given more initial energy. EHCS is only suitable for periodic data gathering applications where in a certain time, the energy consumption of nodes in different network areas is determined and can be calculated theoretically.

The remainder of this paper is organized as follows. Section 2 introduces the literature review. Section 3 defines the network model. Section 4 analyzes the optimal initial energy of nodes. Section 5 describes the clustering algorithm. Section 6 evaluates the performance of the protocol via simulation experiments. Finally, Section 7 reaches the conclusions.

2. Related Work

In order to improve the utilization of network energy, many efficient algorithms have been presented. The first type of algorithms is to adjust transmission ranges of sensor nodes. Tran-Quang et al. [2] have formulated the transmission range adjustment optimization problem as a 0-1 multiple choice knapsack problem and presented a dynamic programming method which allows sensor nodes to dynamically set their individual transmission levels according to their residual energy. Song et al. [3] have proposed an improved corona model with levels for analyzing sensors with adjustable transmission ranges in a WSN with a circular multihop

deployment. They have proved that searching optimal transmission ranges of sensors among all coronas is a multi-objective optimization problem (MOP) and have designed a centralized algorithm and a distributed algorithm for assigning the transmission ranges of sensors in each corona for different node distributions. Liu et al. [4] have presented the particle swarm optimization algorithm (PSO) to solve the routing problem of avoiding the energy hole. The algorithm redefines the particle of the PSO, the operation of particle, and the "flying" rules. Liu and Guo [5] have proposed an analytical model to characterize the energy-hole problem and have described an iterative process to determine the optimum value of the model parameters. Zeng et al. [6] have analyzed the characteristics of data distribution in WSN and have obtained some results including the distribution of energy consumptions, the lifetime of each node in different localities, and the data transferring delay. They also proposed energy-hole avoidance for WSN based on adjust transmission power.

The second type of algorithms is to use nonuniform node distribution. Wu et al. [7] have investigated the theoretical aspects of the nonuniform node distribution strategy to avoid the energy hole around the sink and have proved that in a circular sensor network with a uniform node distribution and constant data reporting, the unbalanced energy depletion among the nodes in the whole network is unavoidable. If the ratio between the node densities of the adjacent $(i + 1)$th corona and the ith corona is equal to $(2i-1)/q(2i+1)$, a suboptimal energy efficiency among the inner parts of the network can be obtained, where q is the geometric proportion. Li et al. [8] have proposed a load-similar node distribution strategy to balance energy consumption and solve the energy-hole problem, based on the analysis of traffic load distribution in the continuous space of the network. Sensor nodes are deployed according to the load distribution. More nodes will be deployed in the range where the average load is higher. Halder et al. [9] have derived the principle of nonuniform node distribution that ensures energy balancing. They also developed a nonuniform, location-wise predetermined node deployment strategy based on this principle leading to an increase in network lifetime.

The third type of algorithms is to use mobile sinks or mobile relays to prolong the lifetime of wireless sensor networks. Marta and Cardei [10] have proposed solution to use mobile sinks that change their location when the nearby sensors' energy becomes low. In this way, the sensors located near sinks change over time. In deciding a new location, a sink searches for zones with richer sensor energy. Luo and Hubaux [11] have proved that the best mobility strategy consists in following the periphery of the network (assume that the sensors are deployed within a circle). Gao et al. [12] have proposed a data collection scheme called maximum amount shortest path (MASP) to optimize the mapping between members and subsinks. MASP has been formulated as an integer linear programming problem which is solved by a genetic algorithm. A communication protocol has been designed to implement MASP, which is also applicable in sensor networks with low density and multiple sinks. Cheng et al. [13] have proposed the architecture of multiple mobile sinks sparse wireless sensor network and an opportunistic transmission scheduling algorithm.

The above literature discuss solutions on the problem of the hot spots in flat networks. In the hierarchical network, clustering techniques can increase the scalability of wireless sensor networks and enable the efficient utilization of the limited energy resources. Many efficient energy-aware clustering routing protocols are proposed, for example, LEACH [14], LEACH_C [15], HEED [16], EADEEG [17], and PEGASIS [18].

LEACH (low-energy adaptive clustering hierarchy) utilizes randomized rotation of local cluster-heads to evenly distribute the energy load among the sensors in the network. LEACH uses localized coordination to enable scalability and robustness for dynamic networks and incorporates data fusion into the routing protocol to reduce the amount of information that must be transmitted to sink.

While LEACH has advantage to using distributed cluster formation algorithm, each node makes autonomous decisions. Then, this protocol offers no guarantee about the placement and number of cluster-head nodes. LEACH-C (LEACH-Centralized) is a protocol that uses a centralized clustering algorithm to form the clusters. During the set-up phase, each node sends information about its current location and energy level to sink. Sink runs an optimization algorithm to determine the clusters for that round. Sink may produce better clusters than those formed using the distributed algorithm by dispersing the cluster-head nodes throughout the network.

HEED (hybrid energy-efficient distributed clustering) is a novel distributed clustering approach, which does not make any assumptions about the presence of infrastructure or about node capabilities, other than the availability of multiple power levels in sensor nodes. It periodically selects cluster heads according to a hybrid of the node residual energy and a secondary parameter, such as node proximity to its neighbors or node degree. HEED terminates in $O(1)$ iterations, incurs low message overhead, and achieves fairly uniform cluster head distribution across the network.

EADEEG (an energy-aware data gathering protocol for wireless sensor networks) adopts a new clustering parameter for cluster head election, which can better handle the heterogeneous energy capacities. Furthermore, it also adopts a simple but efficient approach, namely, intracluster coverage to cope with the fractional area coverage problem. EADEEG achieves a good performance in terms of lifetime by minimizing energy consumption for communications and balancing the energy load among all nodes.

In order to reduce the number of nodes communicating directly with sink, PEGASIS (power-efficient gathering in sensor information systems) organizes all sensor nodes into a near-optimal chain according to nodes' location by greedy algorithm. Each node transmits data to the closest possible neighbor. The data is passed along the chain from one node to another node and is fused. Finally, one designated cluster head transmits the combined data to sink. All network nodes take turns as the cluster head to balance the energy consumption of nodes. It is better than LEACH by about 100–300% in terms of network lifetime for different network

topologies, but it has the greater delay and needs overall location information to construct the chain.

The above clustering protocols use randomized rotation of cluster heads to balance energy dissipation among sensor nodes. They can achieve local energy balance, but cannot avoid the energy-hole problem. From a global view, the nodes close to sink still deplete the more energy than other nodes. In order to solve the problem, UCS [19], EECS [20], and EEUC [21] use different methods.

UCS (unequal clustering size) assumes that the network is heterogeneous and cluster head nodes have more energy and predetermined locations. It organizes the network into heterogeneous clusters, where some more powerful nodes take on the cluster head role to control network operation, it is important to ensure that energy dissipation of these cluster head nodes is balanced.

EECS (energy-efficient clustering scheme) elects cluster heads with more residual energy through local radio communication while achieving well cluster head distribution; furthermore, it introduces a novel method to balance the load among the cluster heads. EECS is fully distributed and more energy efficient. Simulation results show that EECS outperforms LEACH significantly with prolonging the network lifetime over 35%.

EEUC introduces an unequal clustering mechanism to balance the energy consumption among cluster heads. Cluster heads closer to sink have smaller sizes than those heads father away from sink, thus cluster heads closer to sink can preserve some energy for the purpose of intercluster data forwarding.

3. Network Model

A sensor network consisting of N sensor nodes randomly deployed within a circle field to continuously monitor the surrounding environment. The ith sensor node is denoted by n_i, and the corresponding sensor node set is denoted by $S = \{n_1, n_2, \ldots, n_N\}$. We assume that the sensor network has the following properties.

(i) All sensor nodes and sink are stationary after deployment. There is a base station (i.e., sink) which is located in the center of the sensing field and has unlimited energy.

(ii) Sensor nodes are heterogeneous, and nodes in different rings have different initial energy. All nodes are location-unaware, that is, not equipped with GPS-capable antennae. The energy of nodes cannot be renewed.

(iii) All nodes are assigned a unique identifier and have the function of data aggregation. The intracluster sensed information is highly correlated; thus, the cluster head can aggregate N data packets from its members into a single length-fixed packet. But intercluster data is not correlated and cannot be fused.

(iv) Nodes can figure out the distance to the sender according to RSSI (received signal strength indication) and can use power control to tune the amount of transmission power by the transmission distance to receiver. Such as the Berkeley Mote, it has 100 transmission power levels.

Figure 1: Network structure model.

The network structure model is illustrated in Figure 1.

The whole network region is divided into multilayer rings, and each ring has the same width w. Assuming that the number of rings is k, C_i denotes the ith ring from the inside to the outside. In clusters, the communication radius of nodes is $r_c = w$. In order to guarantee connectivity among the different cluster heads, the maximal communication radius of cluster heads is set to $r_h = 2r_c$. Therefore, the cluster heads in rings C_1 and C_2 can directly send data to sink, but the cluster heads in ring C_i $(i \in 3, \ldots, k)$ can only transfer data to the cluster heads in ring C_{i-1} $(i \in 3, \ldots, k)$. The relay load steps up from the outside to the inside, and the energy of nodes also increases.

4. Energy Analysis

4.1. Energy Consumption Model. This paper uses energy consumption model in [15]. The energy dissipation of transmission depends on the distance between the transmitter and receiver. When the distance is relatively far, the multiple-path fading channel model (d^4 power loss) is used. Otherwise, the free space model (d^2 power loss) is used. The energy consumption for transmission of an l-bit packet over distance d is

$$E_s(l, d) = \begin{cases} l \cdot E_{\text{elec}} + l \cdot \varepsilon_{\text{fs}} \cdot d^2, & d < d_0, \\ l \cdot E_{\text{elec}} + l \cdot \varepsilon_{\text{mp}} \cdot d^4, & d \geq d_0. \end{cases} \quad (1)$$

The radio dissipation E_{elec} is to run the transmitter or receiver circuitry, which depends on many factors, such as the digital coding and modulation. The amplifier energies ε_{fs} and ε_{mp} are, respectively, the energy required by power amplification to achieve an acceptable bit-error rate in the two models. The distance $d_0 = \sqrt{\varepsilon_{\text{fs}}/\varepsilon_{\text{mp}}}$ is a threshold value.

To receive this message, the radio dissipates energy is

$$E_r(l) = l \cdot E_{\text{elec}}. \quad (2)$$

Figure 2: Maximal cluster heads.

Figure 3: Minimal cluster heads.

The energy expended for data merging is

$$E_m(m,l) = m \cdot l \cdot E_{\mathrm{DA}}, \tag{3}$$

where m is the number of messages, and E_{DA} is the energy dissipation per bit for fusion.

4.2. Theory Analysis

Theorem 1. *Assume that the area of the ring C_i is S_i, and the intracluster communication radius is r_c, then in the ring C_i, the expected value of the number of cluster heads is*

$$H_i = \left\lceil \frac{4 \cdot S_i}{3\sqrt{3} \cdot r_c^2} \right\rceil. \tag{4}$$

Proof. The wireless sensor network is a network of high-density, and clusters need to cover all the nodes in the entire network. Then, the overlap between clusters exists, and only one cluster head node exists in any cluster. When the overlap among clusters is most, the number of the cluster heads is maximal. The figure of maximal cluster heads is showed in Figure 2.

In Figure 2, the cluster head A is a regular hexagon with the edge length $r_c/\sqrt{3}$, and the area is $\sqrt{3} \cdot r_c^2/2$. So, in the ring C_i, the maximal number of the cluster heads is

$$H_{\max} = \left\lceil \frac{2 \cdot S_i}{\sqrt{3} \cdot r_c^2} \right\rceil. \tag{5}$$

On the contrary, when the overlap among clusters is least, the number of the cluster heads is minimal. The figure of minimal cluster heads is showed in Figure 3.

In Figure 3, the cluster head A is a regular hexagon. So in the ring C_i, the minimal number of the cluster heads is

$$H_{\min} = \left\lceil \frac{2 \cdot S_i}{3\sqrt{3} \cdot r_c^2} \right\rceil. \tag{6}$$

Therefore, in the ring C_i, the expected value of the number of cluster heads is

$$H_i = \left\lceil \frac{4 \cdot S_i}{3\sqrt{3} \cdot r_c^2} \right\rceil. \tag{7}$$

□

Theorem 2. *Assume that the area of the ring C_i is S_i, the intracluster communication radius is r_c, and the region covered by the cluster is approximately thought as a circle with radius r, then the expected value of the squared distance from the cluster members to the cluster head is*

$$d_{ch}^2 = \frac{9 \cdot r_c^2}{8\sqrt{3} \cdot \pi}. \tag{8}$$

Proof. The expected value of the squared distance from the members to the cluster head (assumed to be at the center of mass of the cluster) is given by

$$\begin{aligned} d_{\mathrm{ch}}^2 &= \iint \left(x^2 + y^2\right) \rho(x,y)\, dx\, dy \\ &= \iint r^2 \rho(r,\theta)\, r\, dr\, d\theta, \end{aligned} \tag{9}$$

where ρ is the density of nodes,

$$\rho = \frac{4}{3\sqrt{3} \cdot r_c^2}. \tag{10}$$

The cluster radius r can be computed by the following formula:

$$\begin{aligned} \pi \cdot r^2 &= \frac{S_i}{H_i} = \frac{3\sqrt{3} \cdot r_c^2}{4}, \\ r &= \frac{\sqrt{3} \cdot \sqrt[4]{3} \cdot r_c}{2\sqrt{\pi}}, \\ d_{\mathrm{ch}}^2 &= \rho \int_{\theta=0}^{2\pi} \int_{r=0}^{\sqrt{3} \cdot \sqrt[4]{3} \cdot r_c / 2\sqrt{\pi}} r^3\, dr\, d\theta. \end{aligned} \tag{11}$$

Therefore, the expected value of the squared distance from the cluster members to the cluster head is

$$d_{\mathrm{ch}}^2 = \frac{3\sqrt{3} \cdot r_c^2}{8\pi}. \tag{12}$$

□

Theorem 3. *Assume that the area of the ring C_i is S_i and the intracluster communication radius is r_c, then the expected*

value of the distance between two neighboring cluster heads is

$$d_{hh} = \frac{\sqrt{3\sqrt{3}} \cdot r_c}{\sqrt{\pi}}. \tag{13}$$

Proof. The expected value of the distance between two neighboring cluster heads is given by

$$\pi\left(\frac{d_{\text{hh}}}{2}\right)^2 = \frac{S_i}{H_i}. \tag{14}$$

Therefore,

$$d_{\text{hh}} = 2\sqrt{\frac{S_i}{H_i \cdot \pi}} = \frac{\sqrt{3\sqrt{3}} \cdot r_c}{\sqrt{\pi}}. \tag{15}$$

□

Theorem 4. *Assume that the area of ring C_i is S_i, the number of nodes in ring C_i is N_i, the density of nodes is ρ, the distance from the cluster member to the cluster head is d_{ch} ($d_{ch} \le d_0$), the distance between two neighboring cluster heads is d_{hh}, and the maximal radius of ring C_i is R_i, then the initial energy of each node is as follows.*

In ring C_i, ($i \in [2, \ldots, k-1]$), there will be two conditions for consideration:

(1) *when $d_{ch} \le d_0$ and $d_{hh} \le d_0$,*

$$\begin{aligned} e_i = {} & \frac{lE_{elec}\left(2\sum_{j=i+1}^{k} H_j + 2N_i - H_i\right)}{N_i} \\ & + \frac{l\varepsilon_{fs} d_{hh}^2 \left(\sum_{j=i+1}^{k} H_j + H_i\right)}{N_i} \\ & + \frac{l\varepsilon_{fs} d_{ch}^2 (N_i - H_i) + lE_{DA} N_i}{N_i}; \end{aligned} \tag{16}$$

(2) *when $d_{ch} \le d_0$ and $d_{hh} > d_0$,*

$$\begin{aligned} e_i = {} & \frac{lE_{elec}\left(2\sum_{j=i+1}^{k} H_j + 2N_i - H_i\right)}{N_i} \\ & + \frac{l\varepsilon_{mp} d_{hh}^4 \left(\sum_{j=i+1}^{k} H_j + H_i\right)}{N_i} \\ & + \frac{l\varepsilon_{fs} d_{ch}^2 (N_i - H_i) + lE_{DA} N_i}{N_i}, \end{aligned} \tag{17}$$

where $N_i = S_i\rho = \pi(R_i^2 - R_{i-1}^2)\rho$, $2 \le i \le k$.

Proof. In ring C_i ($i \in [2, \ldots, k-1]$), the energy of cluster heads is dissipated in the following four aspects: (1) receiving the data from cluster members; (2) fusing the data from cluster members; (3) sending the data in cluster; (4) forwarding the data from other cluster heads.

(1) For receiving data from cluster members, the energy dissipation of a cluster head in ring C_i is

$$E_r = lE_{\text{elec}}\left(\frac{N_i}{H_i} - 1\right). \tag{18}$$

(2) For merging data from cluster members, the energy dissipation of a cluster head in ring C_i is

$$E_m = lE_{\text{DA}}\frac{N_i}{H_i}. \tag{19}$$

(3) For sending one frame data, the energy dissipation of a cluster head in ring C_i is

$$E_s = lE_{\text{elec}} + l\varepsilon_{\text{fs}} d_{\text{hh}}^2. \tag{20}$$

(4) For collecting one frame data, the total energy dissipation of a cluster head in ring C_i is

$$\begin{aligned} E_{\text{ch}} &= E_r + E_m + E_s \\ &= lE_{\text{elec}}\left(\frac{N_i}{H_i} - 1\right) + lE_{\text{DA}}\frac{N_i}{H_i} + lE_{\text{elec}} + l\varepsilon_{\text{fs}} d_{\text{hh}}^2. \end{aligned} \tag{21}$$

(5) The energy dissipation of a cluster member in ring C_i is

$$E_{\text{mem}} = lE_{\text{elec}} + l\varepsilon_{\text{fs}} d_{\text{ch}}^2. \tag{22}$$

(6) For collecting one frame data, in ring C_i, the total energy dissipation in a cluster is

$$E_{\text{cluster}} = E_{\text{ch}} + E_{\text{mem}}\left(\frac{N_i}{H_i} - 1\right). \tag{23}$$

(7) In ring C_i, the amount of data that nodes needs to forward is

$$S = \sum_{j=i+1}^{k} H_j. \tag{24}$$

(8) For forwarding data from the outer cluster heads, the energy dissipation of nodes in ring C_i is

$$E_f = 2lE_{\text{elec}}\left(\sum_{j=i+1}^{k} H_j\right) + l\varepsilon_{\text{fs}} d_{\text{hh}}^2\left(\sum_{j=i+1}^{k} H_j\right). \tag{25}$$

(9) In ring C_i, the total energy dissipation of all nodes is

$$\begin{aligned} E_i &= H_i E_{\text{cluster}} + E_f \\ &= 2lE_{\text{elec}} N_i + lE_{\text{DA}} N_i + l\varepsilon_{\text{fs}} d_{\text{hh}}^2 H_i \\ &\quad + l\varepsilon_{\text{fs}} N_i d_{\text{ch}}^2 - lE_{\text{elec}} H_i - l\varepsilon_{\text{fs}} H_i d_{\text{ch}}^2 \\ &\quad + 2lE_{\text{elec}}\left(\sum_{j=i+1}^{k} H_j\right) + l\varepsilon_{\text{fs}} d_{\text{hh}}^2\left(\sum_{j=i+1}^{k} H_j\right). \end{aligned} \tag{26}$$

(10) When $d_{\mathrm{ch}} \leq d_0$ and $d_{\mathrm{hh}} \leq d_0$, in ring C_i, the initial energy of each node is

$$\begin{aligned} e_i &= \frac{E_i}{N_i} \\ &= \frac{lE_{\mathrm{elec}}\left(2\sum_{j=i+1}^{k} H_j + 2N_i - H_i\right)}{N_i} \\ &\quad + \frac{l\varepsilon_{\mathrm{fs}} d_{\mathrm{hh}}^2 \left(\sum_{j=i+1}^{k} H_j + H_i\right)}{N_i} \\ &\quad + \frac{l\varepsilon_{\mathrm{fs}} d_{\mathrm{ch}}^2 \left(N_i - H_i\right) + lE_{\mathrm{DA}} N_i}{N_i}. \end{aligned} \tag{27}$$

(11) When $d_{\mathrm{ch}} \leq d_0$ and $d_{\mathrm{hh}} > d_0$, the proving process is similar, so we omit it.

□

5. Clustering Algorithm

After the network deployment, sink broadcasts a *Sink_ADV* message to start all nodes to work, and each node computes out the approximate distanceto sink by the received signal strength of the *Sink_ADV* message.

5.1. Cluster Setup. In EHCS, in order to reduce the energy cost for the cluster head competition, a small number of nodes are chosen as candidate cluster heads to compete for final cluster heads by the predefined probability. Each candidate cluster head broadcasts *Hello* message in radius r_c and collects the *Hello* messages from the adjacent candidate cluster heads to establish the neighbor information table which includes the node number *ID*, the residual energy of the node, and the state of the node (candidate node or common node). The node with the largest residual energy is elected as the final cluster head. After the election of the cluster head ends, each cluster head broadcasts *Head_ADV* message to announce its own cluster head identity. Common nodes join the nearest cluster by the received signal strength of the *Head_ADV* message and send a *Join_Msg* message to the cluster head. Finally, each cluster head broadcasts *TDMA* messages to its cluster member to assign the time slot of the data collection. During the cluster setup, a common node will passively become a cluster head if it is not within the communication range of all cluster heads and cannot join any cluster.

5.2. Formation of Intercluster Forwarding Tree. The cluster head broadcasts a *TREE_ADV* message which includes the node number *ID*, the residual energy of the node, and the distance to sink. After collecting the *TREE_ADV* messages from the adjacent cluster heads, each cluster head establishes the neighbor cluster head information table which includes the number *ID* of the adjacent cluster head, the residual energy E of the adjacent cluster head, the distance from the adjacent cluster head to sink, and the distance from it to the adjacent cluster head.

In order to setup the intercluster forwarding tree, each cluster head chooses a suitable neighbor cluster head as its parent node. For the cluster head i, if $S(i)$ denotes the set of its neighbor cluster heads, its parent node $p(i)$ is selected by the following equation:

$$p(i) = \begin{cases} \text{Sink}, & \text{if } d(i, \text{Sink}) < r_h, \\ j, & \text{otherwise}, \end{cases} \tag{28}$$

$$\begin{aligned} j = \arg\min \{&\cos t(k), k \in S(i) \\ &\text{and } d(k, \text{Sink}) < d(i, \text{Sink})\}, \\ \cos t(k) = w * &\frac{d^2(i,k) + d^2(k, \text{Sink})}{d_{\max}} \\ &+ (1-w) * \frac{E_{\max} - E(k)}{E_{\max}}, \\ &0 \leq w \leq 1, \\ d_{\max} = \max &\left\{d^2(i,k) + d^2(k, \text{Sink})\right\}, \\ E_{\max} = \max &\{E(k)\}, \end{aligned} \tag{29}$$

where d denotes the distance, and E is the residual energy of nodes. If sink is in the intercluster communication range, the cluster head directly passes data to sink; otherwise, it chooses the minimum cost node as the parent node from the neighbor cluster heads closer to sink. The cost is computed by the relay distance and the residual energy of nodes to reduce and balance the network energy consumption. After all cluster heads have found a parent node, an intercluster tree rooted at sink is created.

5.3. Data Collection. The cluster members collect the data and send it to the cluster head. Then, the cluster head fuses the intracluster data and transmits the compressed data to its parent node along the forwarding tree. The parent node forwards the received data to sink. After collecting the predetermined number of frames, the network enters into the next round and reclustering.

6. Simulations Results

In order to evaluate the performance of the protocol, EHCS is compared with multihop HEED and LEACH-C via simulations. Simulation parameters are listed in Table 1.

Figure 4 shows the node distribution in the three protocols.

Figure 5 shows the clustering structure in EHCS. "∗" denotes the cluster heads, and "•" denotes the cluster members. Each cluster head chooses a neighbor cluster head closer to sink as its relay node.

Table 2 shows the death time of nodes in the three protocols and the improvement rate of EHCS over LEACH_C and multihop HEED. If the death time of the first node is used as a comparison standard, EHCS is better than LEACH_C by

Table 1: Simulation parameters.

Parameter	Value
Radius of network	200 m
Number of nodes N	300
Location of sink	Regional center
d_0	87 m
E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/m^2
ε_{mp}	0.0013 pJ/bit/m^4
E_{DA}	5 nJ/bit/signal
Packet size l	4000 bits
Frames/round	20
Intracluster communication radius r_c	50 m
Intercluster communication radius r_h	100 m
Width of the ring w	50 m

Figure 4: Distribution of nodes.

287.5% and multihop HEED by 20.8%. So, EHCS can better balance the energy consumption of nodes and prolong the lifetime of the network.

7. Conclusions

Clustering is one of the key technologies of wireless sensor networks and has a significant impact on the network performance, but in clustering networks, many-to-one traffic pattern easily leads to the energy-hole problem because of the unbalanced energy dissipation among sensor nodes. We design an energy-heterogeneous clustering scheme where nodes closer to sink have more initial energy to provide the enough energy for forwarding data from outer cluster heads. Simulation results show that the proposed protocol is better than LEACH_C by 287.5% and multihop HEED by 20.8%, if the death time of the first node is used as a comparison standard.

Figure 5: Clustering structure in EHCS.

Table 2: Performance comparison among three protocols.

Percent of dead nodes	Death time (rounds)			Improvement rate (%)	
	LEACH_C	HEED	EHCS	LEACH_C	HEED
First node	24	77	93	287.5%	20.8%
10%	43	93	100	132.6%	7.5%
20%	53	98	102	92.5%	4.1%
30%	73	102	104	42.5%	2.0%
40%	89	105	106	19.1%	1.0%
50%	98	107	108	10.2%	0.9%

Acknowledgment

This work is supported by the National Natural Science Foundation of China (nos. 61272236, 61174177, and 41172298).

References

[1] A. Wadaa, S. Olariu, L. Wilson, M. Eltoweissy, and K. Jones, "Training a wireless sensor network," *Mobile Networks and Applications*, vol. 10, no. 1, pp. 151–168, 2005.

[2] V. Tran-Quang, P. N. Huu, and T. Miyoshi, "A transmission range optimization algorithm to avoid energy holes in wireless sensor networks," *IEICE Transactions on Communications B*, vol. 94, no. 11, pp. 3026–3036, 2011.

[3] C. Song, M. Liu, J. Cao, Y. Zheng, H. Gong, and G. Chen, "Maximizing network lifetime based on transmission range adjustment in wireless sensor networks," *Computer Communications*, vol. 32, no. 11, pp. 1316–1325, 2009.

[4] A. Liu, X. Wu, and Z. Chen, "Energy-hole avoidance routing algorithm for WSN based on PSO," *Journal of Computer Research and Development*, vol. 46, no. 4, pp. 575–582, 2009.

[5] Y. Liu and W. Guo, "A balanced energy depletion strategy for the energy hole problem in wireless sensor networks," *Chinese Journal of Electronics*, vol. 18, no. 3, pp. 535–538, 2009.

[6] Z. Zeng, Z. Chen, and A. Liu, "Energy-hole avoidance for WSN based on adjust transmission power," *Chinese Journal of Computers*, vol. 33, no. 1, pp. 12–22, 2010.

[7] X. Wu, G. Chen, and S. K. Das, "Avoiding energy holes in wireless sensor networks with nonuniform node distribution,"

IEEE Transactions on Parallel and Distributed Systems, vol. 19, no. 5, pp. 710–720, 2008.

[8] Q. Li, M. Liu, M. Yang, and G. Chen, "Load-similar node distribution for solving energy hole problem in wireless sensor networks," *Chinese Journal of Software*, vol. 22, no. 3, pp. 451–465, 2011.

[9] S. Halder, A. Ghosal, and S. D. Bit, "A pre-determined node deployment strategy to prolong network lifetime in wireless sensor network," *Computer Communications*, vol. 34, no. 11, pp. 1294–1306, 2011.

[10] M. Marta and M. Cardei, "Improved sensor network lifetime with multiple mobile sinks," *Pervasive and Mobile Computing*, vol. 5, no. 5, pp. 542–555, 2009.

[11] J. Luo and J. Hubaux, "Joint mobility and routing for lifetime elongation in wireless sensor networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '05)*, pp. 1735–1746, March 2005.

[12] S. Gao, H. Zhang, and H. Xu, "Efficient data gathering approach in sensor networks with path-fixed sinks," *Chinese Journal of Software*, vol. 21, no. 1, pp. 147–162, 2010.

[13] L. Cheng, C. Chen, and J. Ma, "Selection scheme of mobile sinks in wireless sensor networks," *Chinese Journal on Communications*, vol. 29, no. 11, pp. 12–19, 2008.

[14] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Siences (HICSS '00)*, pp. 3005–3014, IEEE Computer Society, Maui, Hawaii, USA, January 2000.

[15] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[16] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 660–669, 2004.

[17] M. Liu, J. Cao, G. Chen, L. Chen, X. Wang, and H. Gong, "EADEEG: an energy-aware data gathering protocol for wireless sensor networks," *Chinese Journal of Software*, vol. 18, no. 5, pp. 1092–1109, 2007.

[18] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proceedings of the IEEE Aerospace Conference*, pp. 1125–1130, IEEE Aerospace and Electronic Systems Society, Big Sky, Mont, USA, March 2002.

[19] S. Soro and W. B. Heinzelman, "Prolonging the lifetime of wireless sensor networks via unequal clustering," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS '05)*, pp. 236–242, IEEE Computer Society, Denver, Colo, USA, April 2005.

[20] M. Ye, C. Li, G. Chen, and J. Wu, "EECS: an energy efficient clustering scheme in wireless sensor networks 10a.2," in *Proceedings of the 24th IEEE International Performance, Computing, and Communications Conference (IPCCC '05)*, pp. 535–540, April 2005.

[21] A. Liu, X. Wu, Z. Chen, and W. Gui, "Research on the energy hole problem based on unequal cluster-radius for wireless sensor networks," *Computer Communications*, vol. 33, no. 3, pp. 302–321, 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 373849, 11 pages
http://dx.doi.org/10.1155/2013/373849

Research Article

Outage Probability of Dual-Hop Relay System with Interference and Feedback Delay in Satellite M2M Networks

Yuan Lin,[1,2] Yuanzhi He,[2] Qinfei Huang,[2] Hao Yin,[2] and Shaodong Feng[3]

[1] *Institute of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, China*
[2] *Institute of China Electronic Equipment System Engineering Company, Beijing 100141, China*
[3] *PLA 96610 Troop, Beijing 102208, China*

Correspondence should be addressed to Yuan Lin; utopia86@126.com

Received 10 July 2013; Accepted 3 September 2013

Academic Editor: Lin Bai

Copyright © 2013 Yuan Lin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The machine-to-machine (M2M) communication enables information exchange between machine devices, which can be carried out without or with minimal human interaction. In this paper, we analyse the outage probability of satellite M2M communication system using amplify-and-forward (AF) and decode-and-forward (DF) relaying schemes. Both the channel feedback delay and multiple interferers at the relay and destination are taken into consideration. The channels of the two hops are modeled as Rayleigh and Rician, respectively. The closed-form outage probability expressions for both AF fixed-gain and DF relaying system are obtained, and the validness of our analytical results are verified by the Monte Carlo simulations. It is shown that the channel delay affects the outage probability performance more significantly compared with the interferences. It is also worthy to note that the DF relaying system provides better performance than AF fixed-gain relaying system, especially when the feedback delay is short. Moreover, it can be found that the performance gaps of the two relaying schemes become trivial as the feedback delay increases.

1. Introduction

Wireless communication systems can benefit from dual-hop relaying, which is well known as an efficient way to extend coverage area, improve throughputs, obtain spatial diversity, and overcome channel impairments such as fading, shadowing, and path loss [1–4]. Dual-hop relaying can be employed in many different communication scenarios. In the wireless M2M networks, information has to been exchanged between machine devices. Because the machine devices are small and power-limited, so it is hard to guarantee the outage probability performance requirements of the destination devices due to the propagation loss between the source devices and the destination device. Recently, many increasing researches showed interest in the relay devices to deal with this problem [5, 6]. In dual-hop wireless M2M relay networks, signals are transmitted by relay node while they might be attenuated beyond detection by propagation loss if directly targeting the destination nodes, due to the low-power machine devices. Also, with the help of multiple antennas in source and destination, the spatial diversity of MIMO relay can help to mitigate the effect of channel fading. Among various relaying schemes, amplify-and-forward (AF) and decode-and-forward (DF) are two commonly used relaying protocols. If the relay only amplifies the received signal at the relay and forwards it to the destination, it is called amplify-and-forward (AF) relaying system. Whereas, if the relay decodes the received signal, reencodes it, and forwards it to the destination, it is called decode-and-forward (DF) relaying system.

The performance analysis of dual-hop AF and DF relaying systems has been extensively investigated. Some have analyzed the performance of single-antenna AF or DF relaying networks in [7–9]. The significant benefits of deploying multiple antennas in relaying network were demonstrated in [10]. For beamforming being an important multiantenna technique, the dual-hop relaying systems employing beamforming were studied in [11–14]. Among various beamforming techniques, the combination of maximal ratio transmission (MRT) and maximal ratio combining (MRC) is widely used in two-hop multiple input and multiple output

(MIMO) relaying networks due to the efficiency and simplicity [11, 13, 14]. In [11, 13], the author studied the end-to-end outage probability and error performance of dual-hop AF and DF relaying system over Rayleigh and Nakagami-m fading channels, respectively, and the outage probability of beamforming with antenna correlation in a two-hop AF MIMO relay network was analyzed in [14]. When multiple antennas are deployed at the source, relay, and destination, the outage probability of MIMO AF relay networks with optimal beamforming was derived in [12].

However, most of literatures assumed perfect channel state information (CSI). In practical scenarios, CSI is feedback from the receiver to the transmitter, where the transmitting delay leads to a mismatch between the available CSI and the actual CSI. The effect of feedback delay on AF relaying network was studied in [15]. It was shown that feedback delay highly degrades the performance of the system since no diversity advantage can be obtained. Along with the feedback delay, the system performance is also highly degraded by the interference [16–18]. In [17, 18], the author investigated the outage probability of AF and opportunistic DF relaying system with interference over Nakagami-m fading channels, respectively. As far as we know, few works have considered the joint effect of feedback delay and interference [19, 20]. In [19], for AF dual-hop relaying system, the author studied the joint effect of feedback delay and interference over Rayleigh fading channels, but it only considered the scenario where each node is equipped with a single antenna. In [20], the authors studied the multiantenna AF relaying networks with feedback delay and interference, but it assumed that the interference only exists at the relay. Furthermore, the above works assumed the dual hops obey the same fading statistics. Nowadays, the support of M2M communication services for the vehicular market using satellite network represents a good opportunity. These services can make use of the backward satellite link capacity for the transmission of telemetry data, requiring only a small fraction of capacity in the forward direction. In the scenario of backward satellite link (from satellite mobile terminal to earth station) of the satellite system, assume satellite mobile terminal and earth station both have multiantennas, the satellite services as relay, then the channel of the satellite mobile terminal to the satellite is often modeled as Rayleigh due to large scattering, while the channel between the satellite and the earth station could be Rician fading because of the composite effect of line of sight component and scattering. To handle this scenario, our paper not only considers the effect of feedback delay but also the interferers at the relay and destination. Furthermore, we also consider the mixed Rayleigh-Rician channel. To the best of the authors' knowledge, no results with the above assumptions have been reported in public.

In this paper, we investigate the joint effect of channel feedback delay and interference for AF fixed-gain and DF dual-hop relaying systems over mixed Rayleigh-Rician channels. We firstly derive the end-to-end equivalent signal-to-interference and noise ratio (SINR), then the expressions for the outage probability of the two systems are given in closed form. Numerical results verify the validness of our theoretical analysis. Simulation results also show the outage probability gaps of AF fixed-gain and DF dual-hop relaying systems with different parameters.

This paper is organized as follows. In Section 2, we describe the system model with feedback delay and interferers at both the relay and destination. In Section 3, we obtain the closed-form expression for the outage probability of the two relaying schemes with channel feedback delay and interferences. In Section 4, the validity of the performance analysis is verified by Monte Carlo simulations. Finally, conclusions are drawn in Section 5.

2. System Model

We consider a dual-hop half-duplex relay system in wireless M2M networks as shown in Figure 1, where a source machine device equipped with N_s antennas communicates with a destination machine device equipped with N_d antennas through a single-antenna relay.

The source employs MRT and the destination employs MRC. Due to the feedback delay, the weight vectors for the MRT are computed based on the delayed CSI. We assume a mixed channel that the source-to-relay link as Rayleigh fading and the relay-to-destination link as Rician fading. We also assume that the relay and destination operate in an interference-limited environment; therefore, the received signal at relay and destination are impaired by N_1 and N_2 cochannel interferences, respectively. In the first phase, the source transmits the signal to the relay, and then the relay receives the signal and the N_1 interfering signals. At the relay, the received signal is given by

$$y_r(t) = \sqrt{P_s}\mathbf{w}_s^{\mathbf{H}}(t)\,\mathbf{h}_{rs}(t)\,s_1(t) + \sum_{i=1}^{N_1}\sqrt{P_{ri}}h_{ri}(t)\,s_{ri}(t) + n_r(t), \tag{1}$$

where N_1 is the number of interferers at R. $s_1(t)$, $s_{ri}(t)$, and $n_r(t)$ are the transmitted signal, the ith interferer, and zero mean additive white Gaussian noise (AWGN) at relay, satisfying $E\{|s_1(t)|^2\} = 1$, $E\{|s_{ri}(t)|^2\} = 1$ and $E\{|n_r(t)|^2\} = \sigma_r^2$, respectively. P_s and P_{ri} are the transmitted power of source and the ith interferer at relay, respectively. $\mathbf{h}_{rs}(t) = [h_{rs}^1(t), \ldots, h_{rs}^{N_s}(t)]^{\mathbf{T}}$ is the Gaussian channel vector from source to relay satisfying $E\{\mathbf{h}_{rs}(t)\mathbf{h}_{rs}^{\mathbf{H}}(t)\} = \mathbf{I}_{N_s}$. $h_{ri}(t)$ ($i = 1, \ldots, N_1$) is the ith Gaussian channel coefficient between the ith interferer and R with $E\{|h_{ri}(t)|^2\} = 1$. $\mathbf{w}_s(t)$ is the beamforming weight vector at the source, satisfying $\|\mathbf{w}_s(t)\|_F^2 = 1$; the weight vector is computed based on the outdated CSI, which is given by

$$\mathbf{w}_s(t) = \frac{\mathbf{h}_{rs}(t - T_d)}{\|\mathbf{h}_{rs}(t - T_d)\|_F}, \tag{2}$$

where $\mathbf{h}_{rs}(t-T_d)$ is the delayed channel vector by a time delay T_d of the perfect source-relay channel $\mathbf{h}_{rs}(t)$. The relationship between $\mathbf{h}_{rs}(t-T_d)$ and $\mathbf{h}_{rs}(t)$ is given by the following widely adopted time-varying channel model [21]:

$$\mathbf{h}_{rs}(t) = \rho_d\mathbf{h}_{rs}(t - T_d) + \sqrt{1 - |\rho_d|^2}\,\mathbf{e}(t), \tag{3}$$

Figure 1: System model.

where $\rho_d = J_0(2\pi f_d T_d)$ is the normalized correlation coefficient between $h_{rs}^i(t)$ and $h_{rs}^i(t-T_d)$. $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind [22], f_d is the Doppler shift, and $\mathbf{e}(t)$ is the zero mean Gaussian error vector with $E\{\mathbf{e}(t)\mathbf{e}^H(t)\} = \mathbf{I}_{N_s}$.

For a dual-hop AF relaying system, in the second phase, the received scalar signal $y_r(t)$ at the relay is multiplied by a gain G and forwarded to the destination, and the destination receives the amplified signal from the relay and the N_2 interfering signals. Then, the destination transformed the received vector signal to a scalar with a weight vector $\mathbf{w}_d(t)$, satisfying $\|\mathbf{w}_d(t)\|_F^2 = 1$, and the received signal is given by

$$y_d(t) = \mathbf{w}_d^{\mathbf{H}}(t)\left(Gy_r(t)\mathbf{h}_{dr}(t) + \sum_{j=1}^{N_2}\sqrt{P_{dj}}h_{dj}(t)s_{dj}(t) + \mathbf{n}_d(t)\right), \tag{4}$$

where $\mathbf{h}_{dr}(t) = (h_{dr}^1(t), \ldots, h_{dr}^{N_d}(t))^{\mathbf{T}}$ is the Gaussian channel vector from relay to destination. $\mathbf{h}_{dj}(t)$ $(j = 1, \ldots N_2)$ is the jth Gaussian channel vector between the jth interferer and destination with $E\{\mathbf{h}_{dj}(t)\mathbf{h}_{dj}^{\mathbf{H}}(t)\} = \mathbf{I}_{N_d}$. P_{dj} is the transmitted power of the jth interferer. $s_{dj}(t)$ and $n_d(t)$ are the jth interferer and zero mean additive white Gaussian noise (AWGN) at destination, satisfying $E\{|s_{dj}(t)|^2\} = 1$ and $E\{\mathbf{n}_d(t)\mathbf{n}_d^{\mathbf{H}}(t)\} = \sigma_d^2\mathbf{I}_{N_d}$, respectively. According to the principles of MRC, the weight vector $\mathbf{w}_d(t)$ is chosen to be

$$\mathbf{w}_d(t) = \frac{\mathbf{h}_{dr}(t)}{\|\mathbf{h}_{dr}(t)\|_F}. \tag{5}$$

For the fixed-gain relay system, G is given by

$$G = \sqrt{\frac{P_r}{P_s E\left\{\left|\mathbf{w}_s^{\mathbf{H}}\mathbf{h}_{rs}(t)\right|^2\right\} + \sum_{i=1}^{N_1} P_{ri} + \sigma_r^2}}, \tag{6}$$

where P_r is the transmit power of relay. After some manipulations, the instantaneous end-to-end SINR at destination can be expressed by

$$\begin{aligned}\gamma^{\mathrm{AF}} = {} & G^2 P_s\left|\mathbf{w}_d^{\mathbf{H}}(t)\mathbf{h}_{dr}(t)\right|^2\left|\mathbf{w}_s^{\mathbf{H}}(t)\mathbf{h}_{rs}(t)\right|^2 \\ & \times\left(G^2\left|\mathbf{w}_d^{\mathbf{H}}(t)\mathbf{h}_{dr}(t)\right|^2\sum_{i=1}^{N_1}P_{ri}\left|h_{ri}(t)\right|^2\right. \\ & + \sum_{j=1}^{N_2}P_{dj}\left|\mathbf{w}_d^{\mathbf{H}}(t)\,\mathbf{h}_{dj}(t)\right|^2 \\ & \left. + G^2\left|\mathbf{w}_d^{\mathbf{H}}(t)\,\mathbf{h}_{dr}(t)\right|^2\sigma_r^2 + \sigma_d^2\right)^{-1}.\end{aligned} \tag{7}$$

Substituting (6) into (7) and with the assumption that relay and destination are interference-limited (i.e., $n_r(t)$ and $\mathbf{n}_d(t)$ can be neglected, the SINR is equivalent to signal-to-interference ratio (SIR)), γ^{AF} can be further written as

$$\gamma^{\mathrm{AF}} = \frac{\gamma_1\gamma_2}{\gamma_2\gamma_R + C\gamma_D}, \tag{8}$$

where $\gamma_1 = \overline{\gamma_1}|\mathbf{w}_s^{\mathbf{H}}(t)\mathbf{h}_{rs}(t)|^2$, $\gamma_2 = \overline{\gamma_2}|\mathbf{w}_d^{\mathbf{H}}(t)\mathbf{h}_{dr}(t)|^2$, $\gamma_R = \sum_{i=1}^{N_1}(P_{ri}/\sigma_r^2)|h_{ri}(t)|^2$, $\gamma_D = \sum_{j=1}^{N_2}(P_{dj}/\sigma_d^2)|\mathbf{w}_d^{\mathbf{H}}(t)\mathbf{h}_{dj}(t)|^2$, and $C = (P_s/\sigma_r^2)E\{|\mathbf{w}_s^{\mathbf{H}}(t)\mathbf{h}_{rs}(t)|^2\} + \sum_{i=1}^{N_1}(P_{ri}/\sigma_r^2) + 1$ with $\overline{\gamma_1} = P_s/\sigma_r^2$, $\overline{\gamma_2} = P_r/\sigma_d^2$.

For the second phase of the dual-hop DF relaying system, the relay decodes the source data $s_1(t)$, reencodes, and transmits it to the destination, and the destination receives the reencoded signal $s_2(t)$ with $E\{|s_2(t)|^2\} = 1$ and N_2 interfering signals. Then, the destination utilizes a weight

vector $\mathbf{w}_d(t)$ to transform the vector signal to a scalar signal, satisfying $\|\mathbf{w}_d(t)\|_F^2 = 1$, and the received signal is given by

$$y_d^{\mathrm{DF}}(t) = \mathbf{w}_d^{\mathbf{H}}(t)\left(\mathbf{h}_{dr}(t)\, s_2(t) + \sum_{j=1}^{N_2}\sqrt{P_{dj}}\mathbf{h}_{dj}(t)\, s_{dj}(t) + \mathbf{n}_d(t)\right). \tag{9}$$

From (1) and (9), the instantaneous equivalent SIR at relay and the destination can be expressed as

$$\gamma_1^{\mathrm{DF}} = \frac{P_s\left|\mathbf{w}_s^{\mathbf{H}}(t)\mathbf{h}_{rs}(t)\right|^2}{\sum_{i=1}^{N_1} P_{ri}\left|h_{ri}(t)\right|^2} \triangleq \frac{\gamma_1}{\gamma_R}, \tag{10}$$

$$\gamma_2^{\mathrm{DF}} = \frac{P_r\left|\mathbf{w}_d^{\mathbf{H}}(t)\,\mathbf{h}_{dr}(t)\right|^2}{\sum_{j=1}^{N_2} P_{dj}\left|\mathbf{w}_d^{\mathbf{H}}(t)\,\mathbf{h}_{dj}(t)\right|^2} \triangleq \frac{\gamma_2}{\gamma_D}, \tag{11}$$

where γ_1, γ_2, γ_R and γ_D are defined as above.

3. Outage Probability

Outage probability is an important system performance metric in wireless communication. It is defined as the end-to-end instantaneous equivalent SIR falling below a given threshold γ_{th}.

3.1. Amplify-and-Forward Relaying System. For AF relaying system, it can be expressed by

$$P_{\mathrm{out}}^{\mathrm{AF}}(\gamma_{\mathrm{th}}) = P\left(\gamma^{\mathrm{AF}} < \gamma_{\mathrm{th}}\right) = P\left(\frac{\gamma_1\gamma_2}{\gamma_2\gamma_R + C\gamma_D} < \gamma_{\mathrm{th}}\right). \tag{12}$$

In order to calculate (12), we firstly evaluate the cumulative distribution function (CDF) of γ_1, the probability distribution function (PDF) of γ_2, γ_R, and γ_D, and also the value of C, respectively. According to [21], the CDF of γ_1 can be given by

$$F_{\gamma_1}(\gamma) = 1 - \sum_{m=0}^{N_s-1}\sum_{n=0}^{N_s-m-1}\binom{N_s-1}{m}\left(\left|\rho_d\right|^2\right)^{N_s-m-1} \times\left(1-\left|\rho_d\right|^2\right)^m \frac{\gamma^n e^{-\gamma/\overline{\gamma_1}}}{n!\overline{\gamma_1}}. \tag{13}$$

The PDF of γ_1 can also be obtained in [21] by

$$f_{\gamma_1}(\gamma) = \frac{1}{\overline{\gamma_1}^{N_s}}\sum_{m=0}^{N_s-1}\binom{N_s-1}{m}\left(\left|\rho_d\right|^2\right)^{N_s-m-1} \times\left(\overline{\gamma_1}\left(1-\left|\rho_d\right|^2\right)\right)^m \gamma^{N_s-m-1}e^{-\gamma/\overline{\gamma_1}} \times\left((N_s-m-1)!\right)^{-1}. \tag{14}$$

With the help of [22, Eq. (2-1-118)], The PDF of γ_2 is given by

$$f_{\gamma_2}(\gamma) = \frac{1}{2\sigma^2\overline{\gamma_2}}\left(\frac{\gamma}{\overline{\gamma_2}S^2}\right)^{(n-2)/4} \times e^{-(S^2+(\gamma/\overline{\gamma_2}))/2\sigma^2} I_{n/2-1}\left(\sqrt{\frac{\gamma}{\overline{\gamma_2}}}\frac{S}{\sigma^2}\right), \tag{15}$$

where $I_a(x)$ is the ath-order modified Bessel function of the first kind [23, Eq. (8.431.1)], which can be expressed by

$$I_a(x) = \sum_{k=0}^{\infty}\frac{(x/2)^{a+2k}}{k!\Gamma(a+k+1)} \quad (x \geq 0). \tag{16}$$

Substituting (16) and $n = 2N_d$ into (15), The PDF of γ_2 can be further given by

$$f_{\gamma_2}(\gamma) = \sum_{k=0}^{\infty}\frac{\gamma^{N_d+k-1}S^{2k}e^{-(S^2+(\gamma/\overline{\gamma_2}))/2\sigma^2}}{\overline{\gamma_2}^{N_d+k}\left(2\sigma^2\right)^{N_d+2k}k!\Gamma(N_d+k)}, \tag{17}$$

where we consider γ_2 as the sum of squares of $2N_d$ statistic independence Gaussian random variables which have the same variance $\sigma^2 = \overline{\gamma_2}^2$ and the mean values $\overline{\gamma_2}m_i$ ($i = 1, 2, \ldots, 2N_d$); here, we define $S^2 = \overline{\gamma_2}^2\sum_{i=1}^{2N_d} m_i^2$ as the sum of squares of $\overline{\gamma_2}m_i$ and Rice factor $K = S^2/2\sigma^2$, accordingly.

As with Theorem 2 in [24], the PDF of γ_R is given by

$$f_{\gamma_R}(\gamma) = \sum_{i=1}^{\rho(A)}\sum_{j=1}^{\tau_i(A)}\chi_{ij}(A)\frac{P_{r[i]}^{-j}}{(j-1)!}\gamma^{j-1}\exp\left(-\frac{\gamma}{P_{R[i]}}\right), \tag{18}$$

where $A = \mathrm{diag}(P_{r1}/\sigma_r^2, P_{r2}/\sigma_r^2, \ldots, P_{rN_1}/\sigma_r^2)$, $\rho(A)$ is the number of distinct diagonal elements of A, $P_{r[1]} > P_{r[2]} > \cdots > P_{r[\rho(A)]}$ are the distinct diagonal elements in decreasing order, $\tau_i(A)$ is the multiplicity of $P_{r[i]}$, and $\chi_{ij}(A)$ is the (i, j)th characteristic coefficient of A [24]. Similarly, the PDF of γ_D is given by

$$f_{\gamma_D}(\gamma) = \sum_{u=1}^{\rho(B)}\sum_{v=1}^{\tau_u(B)}\chi_{uv}(B)\frac{P_{d[u]}^{-v}}{(v-1)!}\gamma^{v-1}\exp\left(-\frac{\gamma}{P_{d[u]}}\right), \tag{19}$$

where $B = \mathrm{diag}(P_{d1}/\sigma_d^2, P_{d2}/\sigma_d^2, \ldots, P_{dN_2}/\sigma_d^2)$, $\rho(B)$, $P_{d[u]}$, and $\chi_{uv}(B)$ are defined as (18) similarly.

With the help of (14), C can be calculated as

$$\begin{aligned} C &= \frac{P_s}{\sigma_r^2}E\left\{\left|\mathbf{w}_s^{\mathbf{H}}(t)\,\mathbf{h}_{rs}(t)\right|^2\right\} + \sum_{i=1}^{N_1}\frac{P_{ri}}{\sigma_r^2} + 1 \\ &= \frac{P_s}{\sigma_r^2\overline{\gamma_1}}E\{\gamma_1\} + \sum_{i=1}^{N_1}\frac{P_{ri}}{\sigma_r^2} + 1 \\ &= \frac{P_s}{\sigma_r^2\overline{\gamma_1}^{N_s+1}} \\ &\quad\times\int_0^{\infty}\sum_{m=0}^{N_s-1}\frac{\binom{N_s-1}{m}\left(\left|\rho_d\right|^2\right)^{N_s-m-1}\left(\overline{\gamma_1}\left(1-\left|\rho_d\right|^2\right)\right)^m}{(N_s-m-1)!} \\ &\quad\times\gamma^{N_s-m}e^{-\gamma/\overline{\gamma_1}}d\gamma + \sum_{i=1}^{N_1}\frac{P_{ri}}{\sigma_r^2} + 1. \end{aligned} \tag{20}$$

With the help of [23, Eq. (3.351.3)],

$$\int_0^\infty x^n e^{-\mu x} dx = n!\mu^{-n-1} \quad (\mathrm{Re}\,\mu > 0). \tag{21}$$

C can be further obtained as

$$C = \overline{\gamma_1} \sum_{i=1}^{N_s-1} \binom{N_s-1}{i} \left(|\rho_d|^2\right)^{N_s-i-1} \left(1-|\rho_d|^2\right)^i (N_s-i) + \sum_{i=1}^{N_1} \frac{P_{ri}}{\sigma_r^2} + 1. \tag{22}$$

With the help of (13), (17), (18), and (19) and also the binomial theorem, (12) can be further formulated as

$$\begin{aligned} P_{\mathrm{out}}^{\mathrm{AF}}(\gamma_{\mathrm{th}}) &= P\left(\gamma^{\mathrm{AF}} < \gamma_{\mathrm{th}}\right) \\ &= P\left(\gamma_1 < \frac{(\gamma_2\gamma_R + C\gamma_D)\gamma_{\mathrm{th}}}{\gamma_2}\right) \\ &= \iiint_0^\infty F_{\gamma_1}\left(\frac{(\gamma_2\gamma_R + C\gamma_D)\gamma_{\mathrm{th}}}{\gamma_2}\right) \\ &\quad \times f_{\gamma_R}(\gamma_R) f_{\gamma_2}(\gamma_2) f_{\gamma_D}(\gamma_D)\, d\gamma_R\, d\gamma_2\, d\gamma_D \\ &= 1 - \int_0^\infty \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-m-1} \sum_{p=0}^{n} \binom{N_s-1}{m} \left(|\rho_d|^2\right)^{N_s-m-1} \\ &\quad \times \left(1-|\rho_d|^2\right)^m \\ &\quad \times \left(\frac{\gamma_{\mathrm{th}}^n \gamma_R^p e^{-\gamma_{\mathrm{th}}\gamma_R/\overline{\gamma_1}}}{n!\overline{\gamma_1}^n}\right) \\ &\quad \times f_{\gamma_R}(\gamma_R)\, d\gamma_R \\ &\quad \times \iiint_0^\infty \left(\frac{C\gamma_D}{\gamma_2}\right)^{n-p} e^{-\gamma_{\mathrm{th}} C\gamma_D/\overline{\gamma_1}\gamma_2} f_{\gamma_2}(\gamma_2) f_{\gamma_D}(\gamma_D)\, d\gamma_2\, d\gamma_D \\ &= 1 - I_1 \times I_2, \end{aligned} \tag{23}$$

where

$$\begin{aligned} I_1 &= \int_0^\infty \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-m-1} \sum_{p=0}^{n} \binom{N_s-1}{m} \left(|\rho_d|^2\right)^{N_s-m-1} \\ &\quad \times \left(1-|\rho_d|^2\right)^m \\ &\quad \times \frac{\gamma_{\mathrm{th}}^n \gamma_R^p e^{-\gamma_{\mathrm{th}}\gamma_R/\overline{\gamma_1}}}{n!\overline{\gamma_1}^n} f_{\gamma_R}(\gamma_R)\, d\gamma_R, \\ I_2 &= \iiint_0^\infty \left(\frac{C\gamma_D}{\gamma_2}\right)^{n-p} \\ &\quad \times e^{-\gamma_{\mathrm{th}} C\gamma_D/\overline{\gamma_1}\gamma_2} f_{\gamma_2}(\gamma_2) f_{\gamma_D}(\gamma_D)\, d\gamma_2\, d\gamma_D. \end{aligned} \tag{24}$$

Substituting (18) into I_1, I_1 can be derived as

$$\begin{aligned} I_1 = \int_0^\infty \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-m-1} \sum_{p=0}^{n} \sum_{i=1}^{\rho(A)} \sum_{j=1}^{\tau_i(A)} &\left(\chi_{ij}(A) \binom{N_s-1}{m}\right. \\ &\times \binom{n}{p} \left(|\rho_d|^2\right)^{N_s-m-1} \\ &\times \left(1-|\rho_d|^2\right)^m \gamma_{\mathrm{th}}^n \\ &\times \left.\left(n!\overline{\gamma_1}^n (j-1)! P_{r[i]}^j\right)^{-1}\right) \\ &\times \gamma_R^{p+j-1} e^{-(\gamma_{\mathrm{th}}/\overline{\gamma_1} + 1/P_{r[i]})\gamma_R} d\gamma_R. \end{aligned} \tag{25}$$

With the help of [23, Eq. (3.351.3)], I_1 can be further obtained as

$$\begin{aligned} I_1 = \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-m-1} \sum_{p=0}^{n} \sum_{i=1}^{\rho(A)} \sum_{j=1}^{\tau_i(A)} &\chi_{ij}(A) \binom{N_s-1}{m} \\ &\times \binom{n}{p} \left(|\rho_d|^2\right)^{N_s-m-1} \\ &\times \left(1-|\rho_d|^2\right)^m \gamma_{\mathrm{th}}^n (p+j-1)! \\ &\times \left(n!\overline{\gamma_1}^n (j-1)! P_{r[i]}^j \right. \\ &\times \left.\left(\frac{\gamma_{\mathrm{th}}}{\overline{\gamma_1}} + \frac{1}{P_{r[i]}}\right)^{p+j}\right)^{-1}. \end{aligned} \tag{26}$$

Substituting (17) into I_2, I_2 can be derived as

$$\begin{aligned} I_2 = \iint_0^\infty \sum_{k=0}^{\infty} &\left(S^{2k} (c\gamma_D)^{n-p} e^{-S^2/2\sigma^2}\right. \\ &\times \left(\left(2\sigma^2\right)^{Nd+2k} \overline{\gamma_2}^{Nd+k}\right. \\ &\times \left.\left. k!\Gamma(N_d+k)\right)^{-1}\right) \\ &\times \gamma_2^{Nd+k+p-n-1} \\ &\times e^{-(\gamma_{\mathrm{th}} C\gamma_D/\overline{\gamma_1}\gamma_2 + \gamma_2/2\sigma^2\overline{\gamma_2})} f_{\gamma_D}(\gamma_D)\, d\gamma_2\, d\gamma_D. \end{aligned} \tag{27}$$

With the help of [23, Eq. (3.471.9)],

$$\int_0^\infty x^{\nu-1} e^{-(\beta/x)-\gamma x} dx = 2\left(\frac{\beta}{\gamma}\right)^{\nu/2} K_\nu\left(2\sqrt{\beta\gamma}\right) \quad (\mathrm{Re}\,\beta > 0, \mathrm{Re}\,\gamma > 0). \tag{28}$$

I_2 can be further obtained as

$$I_2 = \int_0^\infty \sum_{k=0}^{\infty} \Big(2^{(p-N_d-3k-n+2)/2} S^{2k} (c\gamma_D)^{(N_d+k+n-p)/2} \times e^{-S^2/2\sigma^2} \sigma^{p-n+2-N_d-3k} \gamma_{\text{th}}^{(N_d+k+p-n)/2} \times \left(\overline{\gamma_2}^{(N_d+k+n-p)/2} \overline{\gamma_1}^{(N_d+k+p-n)/2} k! \Gamma(N_d+k) \right)^{-1} \Big) \times K_{N_d+k+p-n} \left(2\sqrt{\frac{\gamma_{\text{th}} C \gamma_D}{2\overline{\gamma_1}\overline{\gamma_2}\sigma^2}} \right) f_{\gamma_D}(\gamma_D)\, d\gamma_D, \tag{29}$$

where $K_n(x)$ is the nth-order modified Bessel function of the second kind [23, Eq. (8.432.1)]. Substituting (19) into (28) and after some algebraic manipulations, I_2 can be given by

$$I_2 = \int_0^\infty \sum_{u=1}^{\rho(B)} \sum_{v=1}^{\tau_u(B)} \sum_{k=0}^{\infty} \Big(\chi_{uv}(B)\, 2^{(p-N_d-3k-n+2)/2} s^{2k} \times (c\gamma_D)^{(N_d+k+n-p)/2} e^{-S^2/2\sigma^2} \times \sigma^{p-n+2-N_d-3k} \gamma_{\text{th}}^{(N_d+k+p-n)/2} \times \gamma_D^{v-1} e^{-\gamma_D/P_{d[u]}} \times \Big(P_{d[u]}^{v} \overline{\gamma_2}^{(N_d+k+n-p)/2} \times \overline{\gamma_1}^{(N_d+k+p-n)/2} \times (v-1)! k! \Gamma(N_d+k) \Big)^{-1} \Big) \times K_{N_d+k+p-n} \left(2\sqrt{\frac{\gamma_{\text{th}} C \gamma_D}{2\overline{\gamma_1}\overline{\gamma_2}\sigma^2}} \right) d\gamma_D. \tag{30}$$

With the help of [23, Eq. (6.631.3)],

$$\int_0^\infty x^\mu e^{-\alpha x^2} K_v(\beta x)\, dx = \frac{1}{2} \alpha^{-(1/2)\mu} \beta^{-1} \Gamma\left(\frac{1+v+\mu}{2} \right) \times \Gamma\left(\frac{1-v+\mu}{2} \right) e^{\beta^2/8\alpha} W_{-(1/2)\mu,(1/2)v} \left(\frac{\beta^2}{4\alpha} \right) \quad (\operatorname{Re}\mu > |\operatorname{Re} v| - 1). \tag{31}$$

After further calculations, I_2 can be finally obtained as

$$I_2 = \sum_{u=1}^{\rho(B)} \sum_{v=1}^{\tau_u(B)} \sum_{k=0}^{\infty} \chi_{uv}(B) \Big(2S^{2k} C^{(N_d+k+n-p)/2} \times e^{(\gamma_{\text{th}} C P_{d[u]}/4\overline{\gamma_1}\overline{\gamma_2}\sigma^2 - S^2/2\sigma^2)} \times \left(2\sigma^2\right)^{(p-n-N_d-3k)/2} \times \gamma_{\text{th}}^{(N_d+k+p-n)/2} \times P_{d[u]}^{(Nd+k+n-p-1)/2} \times \Gamma(N_d+k+v) \times \Gamma(v-p+n) \times \Big(\overline{\gamma_1}^{(N_d+k+p-n)/2} \times \overline{\gamma_2}^{(N_d+k+n-p)/2} \times k!\,(v-1)! \times \Gamma(N_d+k) \Big)^{-1} \Big) \times W_{-(N_d+k+n-p+2v-1)/2,(N_d+k+p-n)/2} \left(\frac{\gamma_{\text{th}} C P_{d[u]}}{2\overline{\gamma_1}\overline{\gamma_2}\sigma^2} \right), \tag{32}$$

where $W_{ij}(x)$ is the Whittaker function defined in [23, Eq. (9.222.1)]. Now substituting (14) and (32) into (23), the closed-form expression for the outage probability can be obtained in

$$P_{\text{out}}^{\text{AF}}(\gamma_{\text{th}}) = 1 - \Bigg[\sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-m-1} \sum_{p=0}^{n} \sum_{i=1}^{\rho(A)} \sum_{j=1}^{\tau_i(A)} \binom{N_s-1}{m} \times \binom{n}{p} \left(|\rho_d|^2\right)^{N_s-m-1} \times \left(1-|\rho_d|^2\right)^m \times \Big(\chi_{ij}(A)\, \gamma_{\text{th}}^n (p+j-1)! \times \Big(n! \overline{\gamma_1}^n (j-1)! \times \left(\frac{\gamma_{\text{th}}}{\overline{\gamma_1}} + \frac{1}{P_{r[i]}} \right)^{p+j} \times P_{r[i]}^{j} \Big)^{-1} \Big)$$

$$\times \sum_{u=1}^{\rho(B)} \sum_{v=1}^{\tau_u(B)} \sum_{k=0}^{\infty} \chi_{uv}(B) \Big(2S^{2k} C^{(N_d+k+n-p)/2} \times e^{(\gamma_{\text{th}} C P_{d[u]}/4\overline{\gamma_1}\overline{\gamma_2}\sigma^2 - S^2/2\sigma^2)} \times \left(2\sigma^2\right)^{(p-n-N_d-3k)/2} \times \gamma_{\text{th}}^{(N_d+k+p-n)/2} \times P_{d[i]}^{(Nd+k+n-p-1)/2} \times \Gamma\left(N_d+k+v\right) \times \Gamma\left(v-p+n\right) \times \Big(\overline{\gamma_1}^{(N_d+k+p-n)/2} \times \overline{\gamma_2}^{(N_d+k+n-p)/2} \times k!\,(v-1)! \times \Gamma\left(N_d+k\right) \Big)^{-1} \Big) \times W_{-(N_d+k+n-p+2v-1)/2,(N_d+k+p-n)/2} \left(\frac{\gamma_{\text{th}} C P_{d[u]}}{2\overline{\gamma_1}\overline{\gamma_2}\sigma^2} \right) \Bigg]. \tag{33}$$

3.2. Decode-and-Forward Relaying System. For the dual-hop DF relaying system, the outage probability can be defined as the probability that the minimum of the single-hop SIRs falls below a given threshold γ_{th}. Therefore, it can be expressed by

$$\begin{aligned} P_{\text{out}}^{\text{DF}}(\gamma_{\text{th}}) &= P\left(\min\left(\gamma_1^{\text{DF}}, \gamma_2^{\text{DF}}\right) < \gamma_{\text{th}}\right) \\ &= 1 - P\left(\gamma_1^{\text{DF}} > \gamma_{\text{th}}, \gamma_2^{\text{DF}} > \gamma_{\text{th}}\right) \\ &= 1 - \left(1 - F_{\gamma_1^{\text{DF}}}(\gamma_{\text{th}})\right)\left(1 - F_{\gamma_2^{\text{DF}}}(\gamma_{\text{th}})\right). \end{aligned} \tag{34}$$

By using the CDF of γ_1 in (13), the outage probability of the γ_1^{DF} in (10) can be calculated as

$$\begin{aligned} &F_{\gamma_1^{\text{DF}}}(\gamma_{\text{th}}) \\ &= P\left(\frac{\gamma_1}{\gamma_R} < \gamma_{\text{th}}\right) = P\left(\gamma_1 < \gamma_{\text{th}}\gamma_R\right) \\ &= \int_0^{\infty} F_{\gamma_1}(\gamma_{\text{th}}\gamma_R) f_{\gamma_R}(\gamma_R)\, d\gamma_R \\ &= 1 - \int_0^{\infty} \sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-m-1} \sum_{i=1}^{\rho(A)} \sum_{j=1}^{\tau_i(A)} \Big(\chi_{ij}(A) \binom{N_s-1}{m} \times \left(|\rho_d|^2\right)^{N_s-m-1} \times \left(1-|\rho_d|^2\right)^m \gamma_{\text{th}}^n \times \left(n!\overline{\gamma_1}^n (j-1)! P_{r[i]}^j\right)^{-1} \Big) \times \gamma_R^{n+j-1} e^{-(\gamma_{\text{th}}/\overline{\gamma_1}+1/P_{r[i]})\gamma_R} d\gamma_R. \end{aligned} \tag{35}$$

With the help of [23, Eq. (3.351.3)], the outage probability of γ_1^{DF} can be further obtained as

$$\begin{aligned} &F_{\gamma_1^{DF}}(\gamma_{\text{th}}) \\ &= 1 - \Bigg[\sum_{m=0}^{N_s-1} \sum_{n=0}^{N_s-m-1} \sum_{i=1}^{\rho(A)} \sum_{j=1}^{\tau_i(A)} \binom{N_s-1}{m} \left(|\rho_d|^2\right)^{N_s-m-1} \times \left(1-|\rho_d|^2\right)^m \\ &\quad \times \Big(\chi_{ij}(A)\, \gamma_{\text{th}}^n \left(\frac{\gamma_{\text{th}}}{\overline{\gamma_1}} + \frac{1}{P_{r[i]}} \right)^{-n-j} (n+j-1)! \times \left(n!\overline{\gamma_1}^n (j-1)! P_{r[i]}^j\right)^{-1} \Big) \Bigg]. \end{aligned} \tag{36}$$

In order to calculate the CDF of γ_2^{DF}, we firstly evaluate the CDF of γ_D. Given the PDF of γ_D in (19), the CDF of γ_2^{DF} can be calculated as

$$\begin{aligned} F_{\gamma_D}(\gamma) &= \int_0^{\gamma} f_{\gamma_D}(x)\, dx \\ &= \int_0^{\gamma} \sum_{u=1}^{\rho(B)} \sum_{v=1}^{\tau_u(B)} \chi_{uv}(B) \frac{P_{d[u]}^{-v}}{(v-1)!} x^{v-1} \exp\left(-\frac{x}{P_{d[u]}}\right) dx. \end{aligned} \tag{37}$$

With the help of [23, Eq. (3.351.1)],

$$\int_0^u x^n e^{-\mu x} dx = \frac{n!}{\mu^{n+1}} - e^{-u\mu} \sum_{k=0}^{n} \frac{n!}{k!} \frac{u^k}{\mu^{n-k+1}} \quad (u > 0, \operatorname{Re}\mu > 0, n = 0, 1, 2, \ldots). \tag{38}$$

The CDF of γ_D can be given by

$$\begin{aligned} F_{\gamma_D}(\gamma) &= \sum_{u=1}^{\rho(B)} \sum_{v=1}^{\tau_u(B)} \chi_{uv}(B) - \sum_{u=1}^{\rho(B)} \sum_{v=1}^{\tau_u(B)} \sum_{l=0}^{v-1} \chi_{uv}(B) \frac{\gamma^l e^{-\gamma/P_{d[u]}}}{l! P_{d[u]}^l} \\ &= 1 - \sum_{u=1}^{\rho(B)} \sum_{v=1}^{\tau_u(B)} \sum_{l=0}^{v-1} \chi_{uv}(B) \frac{\gamma^l e^{-\gamma/P_{d[u]}}}{l! P_{d[u]}^l}, \end{aligned} \tag{39}$$

where the fact $\sum_{u=1}^{\rho(B)}\sum_{v=1}^{\tau_u(B)}\chi_{uv}(B) = 1$ is utilized. Using the CDF given in (39), the outage probability of γ_2^{DF} in (11) can be derived as

$$\begin{aligned}
F_{\gamma_2^{\mathrm{DF}}}(\gamma_{\mathrm{th}}) &= P\left(\frac{\gamma_2}{\gamma_D} < \gamma_{\mathrm{th}}\right) = P\left(\gamma_D > \frac{\gamma_2}{\gamma_{\mathrm{th}}}\right) \\
&= \int_0^\infty \left(1 - F_{\gamma_D}\left(\frac{\gamma_2}{\gamma_{\mathrm{th}}}\right)\right) f_{\gamma_2}(\gamma_2)\, d\gamma_2 \\
&= \int_0^\infty \sum_{u=1}^{\rho(B)}\sum_{v=1}^{\tau_u(B)}\sum_{l=0}^{v-1}\sum_{k=0}^{\infty}\Big(\chi_{uv}(B)\, S^{2k} e^{-s^2/2\sigma^2} \\
&\qquad \times \Big(\gamma_{\mathrm{th}}{}^{l}\, l!\, k!\, \Gamma(N_d + k)\, \overline{\gamma_2}^{N_d+k} \\
&\qquad \times (2\sigma^2)^{N_d+2k} P_{d[u]}^{l}\Big)^{-1}\Big) \\
&\qquad \times \gamma_2^{N_d+k+l-1} \\
&\qquad \times e^{-(1/\gamma_{\mathrm{th}}P_{d[u]} + 1/2\sigma^2\overline{\gamma_2})\gamma_2}\, d\gamma_2.
\end{aligned} \tag{40}$$

With the help of [23, Eq. (3.351.3)], the outage probability of γ_2^{DF} can be further obtained as

$$\begin{aligned}
&F_{\gamma_2^{\mathrm{DF}}}(\gamma_{\mathrm{th}}) \\
&\quad = \sum_{u=1}^{\rho(B)}\sum_{v=1}^{\tau_u(B)}\sum_{l=0}^{v-1}\sum_{k=0}^{\infty}\chi_{uv}(B)\, S^{2k} e^{-s^2/2\sigma^2} \\
&\qquad \times (N_d + k + l - 1)! \\
&\qquad \times \left(\frac{1}{\gamma_{\mathrm{th}}P_{d[u]}} + \frac{1}{2\sigma^2\overline{\gamma_2}}\right)^{-N_d-k-l} \\
&\qquad \times \Big(\gamma_{\mathrm{th}}^{l}\, l!\, k!\, \Gamma(N_d + k)\, \overline{\gamma_2}^{N_d+k} \\
&\qquad \times (2\sigma^2)^{N_d+2k} P_{d[u]}^{l}\Big)^{-1}.
\end{aligned} \tag{41}$$

Substituting (36) and (41) into (34), the outage probability can be further derived as

$$\begin{aligned}
&P_{\mathrm{out}}^{\mathrm{DF}}(\gamma_{\mathrm{th}}) \\
&\quad = 1 - \Bigg[\sum_{m=0}^{N_s-1}\sum_{n=0}^{N_s-m-1}\sum_{i=1}^{\rho(A)}\sum_{j=1}^{\tau_i(A)}\binom{N_s-1}{m}\left(|\rho_d|^2\right)^{N_s-m-1} \\
&\qquad \times \left(1 - |\rho_d|^2\right)^m \\
&\qquad \times \Big(\chi_{ij}(A)\, \gamma_{\mathrm{th}}^{n}\left(\frac{\gamma_{\mathrm{th}}}{\overline{\gamma_1}} + \frac{1}{P_{r[i]}}\right)^{-n-j} \\
&\qquad \times (n + j - 1)! \\
&\qquad \times \left(n!\,\overline{\gamma_1}^{n}\,(j-1)!\,P_{r[i]}^{j}\right)^{-1}\Big) \\
&\qquad \times \Bigg(1 - \sum_{u=1}^{\rho(B)}\sum_{v=1}^{\tau_u(B)}\sum_{l=0}^{v-1}\sum_{k=0}^{\infty}\chi_{uv}(B)\, S^{2k} e^{-s^2/2\sigma^2} \\
&\qquad \times (N_d + k + l - 1)! \\
&\qquad \times \left(\frac{1}{\gamma_{\mathrm{th}}P_{d[u]}} + \frac{1}{2\sigma^2\overline{\gamma_2}}\right)^{-N_d-k-l} \\
&\qquad \times \Big(\gamma_{\mathrm{th}}^{l}\, l!\, k!\, \Gamma(N_d + k)\, \overline{\gamma_2}^{N_d+k} \\
&\qquad \times (2\sigma^2)^{N_d+2k} P_{d[u]}^{l}\Big)^{-1}\Bigg)\Bigg].
\end{aligned} \tag{42}$$

4. Numerical Results

In this section, we verify the theoretical analysis via Monte Carlo simulations. We assume that $\overline{\gamma_1} = \overline{\gamma_2}$, the value of C is defined in (22). Here, we consider two interferers as $N_1 = N_2 = 2$. The power distribution factor of interference is defined as $m = P_{r[1]}/P_{r[2]} = P_{d[1]}/P_{d[2]} = 2$ and SIR is defined as $\mathrm{SIR} = \overline{\gamma_1}/(P_{r[1]} + P_{r[2]}) = \overline{\gamma_2}/(P_{d[1]} + P_{d[2]})$. With the help of [24] and the references therein, for $m > 1$, it can be obtained that $\rho(A) = \rho(B) = 2$, $\tau_1(A) = \tau_2(A) = \tau_1(B) = \tau_2(B) = 1$, $\chi_{11}(A) = P_{r[1]}/(P_{r[1]} - P_{r[2]})$, $\chi_{11}(B) = P_{D[1]}/(P_{D[1]} - P_{D[2]})$, $\chi_{21}(A) = P_{r[2]}/(P_{r[2]} - P_{r[1]})$, $\chi_{21}(B) = P_{d[2]}/(P_{d[2]} - P_{d[1]})$.

Figure 2 shows the outage probability versus average SNR per hop for the dual-hop AF and DF relaying systems. The system parameters are set to be $N_s = N_d = 3$ and $P_{r[1]} = P_{d[1]} = 5\,\mathrm{dB}$, and the SIR threshold is selected as $\gamma_{\mathrm{th}} = 5\,\mathrm{dB}$. Figure 3 shows the outage probability with average SIR chosen from 5 dB to 30 dB, and the average SNR is selected as $\overline{\gamma_1} = \overline{\gamma_2} = 10\,\mathrm{dB}$. Here, we assume Jake's fading spectrum, and the feedback delay parameters are set as $\rho_d = 1$, $\rho_d = 0.64$ and $\rho_d = 0.29$ corresponding to $f_dT_d = 0$, $f_dT_d = 0.2$, and $f_dT_d = 0.3$, respectively. It can be found that the simulation results and analytical results are perfectly matched, which demonstrates the effectiveness of our performance analysis. It can be seen that long feedback delay, that is, ρ_d, significantly degrades the outage probability performance. We can also find that the outage probability decreases as the SNR per hop or SIR per hop increases. It is also worthy to note that DF relaying system performances are better than AF relaying system when feedback delay is short, while the performance gap becomes small as the feedback delay increases.

Figures 4 and 5 show the outage probability versus Rician factor K for the two systems. We choose SIR threshold and average SNR per hop as $\gamma_{\mathrm{th}} = 10\,\mathrm{dB}$ and $\overline{\gamma_1} = \overline{\gamma_2} = 10\,\mathrm{dB}$, respectively. Here, we consider four different scenarios, that is, channel delay and channel interference (CD + CI), channel delay only (CD only), channel interference only (CI only), and without channel delay and channel interference (No CD,

Figure 2: Outage probability of AF and DF relaying system for different ρ_d with $P_{r[1]} = P_{d[1]} = 5$ dB, $N_s = N_d = 3$, and $\gamma_{\text{th}} = 5$ dB.

Figure 3: Outage probability of AF and DF versus different SIR with $N_s = N_d = 3$, $\overline{\gamma_1} = \overline{\gamma_2} = 10$ dB, and $\gamma_{\text{th}} = 5$ dB.

Figure 4: Outage probability of AF versus different K with $\gamma_{\text{th}} = 10$ dB, $N_s = N_d = 3$.

Figure 5: Outage probability of DF versus different K with $\gamma_{\text{th}} = 10$ dB, $N_s = N_d = 3$.

No CI). For the scenarios with CD, we choose $\rho_d = 0.29$ and $P_{r[1]} = P_{d[1]} = 2$ dB. As can clearly be seen from Figures 4 and 5, the outage probability decreases as the K increases for both AF and DF relaying systems, but the improvement is slow and unobvious, which means that the outage probability is not decided by the second hop. It is shown that the simulation results and the analytical results are in perfect agreement. It also can be noticed that the best performance can be obtained without channel delay and interference. Moreover, we can find that channel delay has larger influence on the outage probability compared with the interference.

Figures 6 and 7 show the performance of outage probability for different antenna configurations. The system parameters are set as $\gamma_{\text{th}} = 5$ dB, $f_d T_d = 0.2$, and $P_{r[1]} = P_{d[1]} = 5$ dB. It can be seen clearly that the analytical results exactly agree with the simulation results. We can find that the outage probability is improved as the number of antenna increases for both AF and DF relaying system. It is also worthy to note that the number of transmit antennas at source significantly influences the outage probability performance, which means

Figure 6: Outage probability of AF versus different N_s and N_d with $\gamma_{\text{th}} = 5$ dB, $f_d T_d = 0.2$.

Figure 7: Outage probability of DF versus different N_s and N_d with $\gamma_{\text{th}} = 5$ dB, $f_d T_d = 0.2$.

Figure 8: Outage probability of AF versus different scenarios with $\gamma_{\text{th}} = 5$ dB, $N_s = N_d = 3$.

Figure 9: Outage probability of DF versus different scenarios with $\gamma_{\text{th}} = 5$ dB, $N_s = N_d = 3$.

that the channel delay in the first hop is the main cause that affects the outage probability performance.

Figures 8 and 9 show the outage probability versus average SNR (dB) per hop for the two systems. Here, we also consider four different scenarios, that is, CD + CI, CD only, CI only, and No CD, No CI. It is shown that the simulation results and the analytical results are in perfect agreement. Similar to Figures 4 and 5, same conclusion can be reached that, without channel delay and interference, the best performance can be obtained, while channel delay degrades the outage probability performance more significantly compared with the interference.

5. Conclusions

In this paper, we investigated the effects of channel feedback delay and cochannel interference on the performance of outage probability in dual-hop relaying system. We considered a mixed Rayleigh and Rician fading channel model and obtained closed-form outage probability expressions for both AF fixed-gain and DF relaying system. The validness of our analytical results was verified by Monte Carlo simulations.

Our analysis is useful to the system design engineer for performance evaluation of satellite M2M networks.

Acknowledgment

This work was supported by the China National Natural Science Foundation under Grant 61231011.

References

[1] H. Yanikomeroglu, "Fixed and mobile relaying technologies for cellular networks," in *Proceedings of the 2nd Workshop Applications and Services in Wireless Networks (ASWN '02)*, pp. 75–81, Paris, France, July 2002.

[2] R. Pabst, B. H. Walke, D. C. Schultz et al., "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, 2004.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—Part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[4] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[5] D. Wang, L. Bai, X. Zhang, W. Guan, and C. Chen, "Collaborative relay beamforming strategies for multiple destinations with guaranteed QoS in wireless machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 525640, 10 pages, 2012.

[6] C. Chen, L. Bai, M. Feng, M. Huang, and T. Tian, "Distributed beamforming for relay assisted multiuser machine-to-machine networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 213727, 9 pages, 2012.

[7] Z. Fang, L. Li, and Z. Wang, "Asymptotic performance analysis of multihop relayed transmissions over Nakagami-m fading channels," *IEICE Transactions on Communications B*, vol. E91, no. 12, pp. 4081–4084, 2008.

[8] X. Qu, X. Xu, H. Li, X. Tao, and H. Tian, "Analysis of outage capacity for DF dual-hop relay and optimal power allocation," in *Proceedings of the IET International Conference on Wireless Sensor Network 2010 (IET-WSN '10)*, pp. 194–197, November 2010.

[9] D. Senaratne and C. Tellambura, "Unified exact performance analysis of two-hop amplify-and-forward relaying in nakagami fading," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1529–1534, 2010.

[10] Y. Fan and J. Thompson, "MIMO configurations for relay channels: theory and practice," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1774–1786, 2007.

[11] M. O. Hasna and M.-S. Alouini, "End-to-end performance of transmission systems with relays over Rayleigh-fading channels," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1126–1131, 2003.

[12] M. Li, M. Lin, Q. Yu, W. P. Zhu, and L. Dong, "Optimal beamformer design for dual-hop MIMO AF relay networks over Rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 8, pp. 1402–1414, 2012.

[13] M. O. Hasna and M.-S. Alouini, "Harmonic mean and end-to-end performance of transmission systems with relays," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 130–135, 2004.

[14] R. H. Y. Louie, Y. Li, H. A. Suraweera, and B. Vucetic, "Performance analysis of beamforming in two hop amplify and forward relay networks with antenna correlation," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 3132–3141, 2009.

[15] H. A. Suraweera, T. A. Tsiftsis, G. K. Karagiannidis, and A. Nallanathan, "Effect of feedback delay on amplify-and-forward relay networks with beamforming," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1265–1271, 2011.

[16] C. Zhong, S. Jin, and K.-K. Wong, "Dual-hop systems with noisy relay and interference-limited destination," *IEEE Transactions on Communications*, vol. 58, no. 3, pp. 764–768, 2010.

[17] F. S. Al-Qahtani, T. Q. Duong, C. Zhong, K. A. Qaraqe, and H. Alnuweiri, "Performance analysis of dual-hop AF systems with interference in nakagami-m fading channels," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 454–457, 2011.

[18] A. M. Salhab, F. Al-Qagtani, S. A. Zummo, and H. Alnuweiri, "Exact outage probability of opportunistic DF relay systems with interference at both the relay and the destination over Nakagami-m fading channels," *GLOBECOM*, pp. 4560–4565, 2012.

[19] F. S. Al-Qahtani, C. Zhong, H. Alnuweiri, and K. A. Qaraqe, "Performance analysis of partial relay selection with feedback delay in the presence of interference," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference: Energizing Global Communications (GLOBECOM '11)*, p. 15, December 2011.

[20] H. Phan, T. Q. Duong, M. Elkashlan, and H.-J. Zepernick, "Beamforming amplify-and-forward relay networks with feedback delay and interference," *IEEE Signal Processing Letters*, vol. 19, no. 1, pp. 16–19, 2012.

[21] Y. Ma, D. Zhang, A. Leith, and Z. Wang, "Error performance of transmit beamforming with delayed and limited feedback," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1164–1170, 2009.

[22] J. G. Proakis, *Digital Communication*, McGraw-Hill Companies Press, 4th edition, 2009.

[23] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, 7th edition, 2007.

[24] A. Bletsas, H. Shin, and M. Z. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3450–3460, 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 392083, 17 pages
http://dx.doi.org/10.1155/2013/392083

Research Article

A Smart Electronic Tagging System Based on Context Awareness and Machine-to-Machine Interworking

Gil-Yong Lee,[1] Nam-Yeol Yun,[1] Sung-Chul Lee,[2] and Soo-Hyun Park[1]

[1] *Ubiquitous System Lab., Graduate School of BIT, Kookmin University, Jeongneung-Ro 77, Seoungbuk-Gu, Seoul 136-702, Republic of Korea*
[2] *Convergent SoC Research Center, Korea Electronics Technology Institute, Seongnam 463-816, Republic of Korea*

Correspondence should be addressed to Soo-Hyun Park; shpark21@kookmin.ac.kr

Received 16 April 2013; Accepted 9 August 2013

Academic Editor: Jianhua He

Copyright © 2013 Gil-Yong Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The existing electronic tagging system traces the location of a sex offender using triangulation by communicating with Global Positioning System (GPS) satellites and mobile phone base stations. The acquired location information is used to prevent the offenders from perpetrating repeat crimes. However, the battery resources of such a system are inadequate as it has to trace the location of the moving target in real time and consumes a large amount of battery power while communicating with GPS satellites and mobile phone base stations. In addition, the systems cannot infer the mental state of the targets or detect their alcohol consumption levels, which may be necessary for the prevention of a repeat crime. The purpose of this study is to connect the Ubiquitous System Network (USN), which consumes little electricity, and Android mobile platforms, which are commonly used for machine-to-machine communication. Thus, this system will legally facilitate the protection of minors by providing information only about the target's approach to certain facilities. In addition, the system uses an Android platform to process data measured by the USN's sensors, which can also detect alcohol intake and infer the mental state of the target, and then initiates the corresponding real-time context-awareness services.

1. Introduction

Electronic tagging is a service that monitors the location or condition of a person wearing a tagging device using Global Positioning System (GPS) satellites and electronic devices. The tagging device is attached to a person who may commit repeat crimes, such as kidnapping, homicide, or sexual assault, and thus the service is used to help protect society from these offenders.

In Korea, SK Telecom, Elastic Network, and Samsung Data System have released various electronic tagging product services since 2008 in which a Ubiquitous Sensor Network (USN) is connected using Code Division Multiple Access (CDMA) [1]. In addition, studies have been conducted on inserting a microelectronic chip in the body of a sexual offender or using a monitoring target in which bioelectronics technology is implemented [2].

Users of the system can be seen as regards ex-convicts as potential criminals and thus are open to charges of violating the principles of equality and proportionality of punishment, infringing human rights, imposing a double punishment, and so forth [3]. Nevertheless, the research on electronic tagging for the purpose of preventing major offences such as kidnapping, homicide, and sexual assault continues briskly. However, the existing electronic tagging technology has several problems in terms of its ability to facilitate the prevention of sexual assaults and other crimes: low electrical capacity of batteries, difficulty in detecting an emotion or physical change in the target and determining the target's exact location, and so forth.

The legal constraints in Korea are such that the list of people who are required to wear an electronic tagging device is managed solely by the Ministry of Justice, and detailed information about them is not open to the police or ordinary citizens. In Korea, the real-time location of a person wearing an electronic tagging is not publicly available in order to protect the human rights of the ex-convict. Thus, it is impossible to detect the approach of such persons

(a) (b) (c)

Figure 1: Existing electronic tagging system. (a) Attached device, (b) tracking devices, and (c) home supervision device.

to facilities for children, such as nurseries or kindergarten schools, making minors vulnerable to being the victim of repeating offenders. As it is illegal in Korea to locate the exact position of a person who is wearing an electronic tagging device [4], the Smart Electronic Tagging (SET) system, based on the Machine-to-Machine (M2M) method, which notifies, in real time, only the approach of a person wearing an electronic tagging to facilities for children, is developed.

The remainder of this paper is organized as follows. In Section 2, an introduction to the existing electronic tagging system, with which the proposed SET system will be compared, is given for. In addition, methods for saving energy, location tracing, and inferring the mental state of the target using sensors in M2M and the existing sensor network are described. In Section 3, the architecture, flow of data in each device, principles of operation of the proposed SET system, and a context-awareness service using sensors are explained. In Section 4, the details of the development of the hardware and software for each device in the SET system are given, and the components of the realization of the test bed and the data results of the tests are shown. Finally, in Section 5, we present conclusions and the direction of future studies.

2. Related Works

2.1. The Existing Electronic Tagging System. Three types of systems are used for monitoring people wearing an electronic tag: passive, active, and tracking. The method used in Korea belongs to the tracking type [3].

The electronic tagging system used in Korea is composed of the attached device (electronic tag), tracking device (portable terminal), and home supervision device, as shown in Figure 1. When the electronic tag and portable terminal are located within 5 m of each other and communicate normally, location tracing is accomplished through communication between the portable terminal and GPS satellites [5]. The satellites transmit the information to a central location tracing control center in the Seoul Probation Office for the real-time 24-hour tracking of targets [6]. If the portable terminal of a target is in an area where it cannot easily communicate with GPS satellites, its location is traced to within a 100 m radius using the Location Based Service (LBS) of mobile communication networks or Wireless-Fidelity (Wi-Fi) using the CDMA signal based on Time Division Multiple Access (TDMA) and Observed Time Difference Of Arrival (OTDOA) received by the portable terminal every 10 s [7].

However, the method of communicating with GPS satellites or CDMA base stations allows a time gap of around 45 s between the previous and actual location of the person who is wearing the electronic tag [6, 7]. As this method shows only the radius within which the target is located, it does not show on which floor of a building the person is located. Furthermore, it does not give detailed information about the emotion or physical state of the target. As a result, the ability of the existing electronic tagging system to help prevent a crime if it is committed in the basement of the apartment building where the target lives, or if a repeat crime is planned in nearby facilities for children, is limited.

2.2. Machine-to-Machine. M2M is defined slightly differently in different countries. In Korea, M2M is defined as a "next-generation fusion infrastructure which provides intelligent services using various media instead of the services based on single media by expanding the communication between man for the usage of information into the communication between man and machine and machine and machine" [8].

The M2M architecture defined by European Telecommunications Standards Institute (ETSI) is shown in Figure 2 [9]. All the architectures of the SET system, such as the collection of information by the tail machine, communication between the tail machine (i.e., the sensor node) and the gateway, processing of the information sent by the tail machine, and provision of services to users using application programs that do not require manual operation, are structured in the same way.

SET was established as an M2M system that uses USN and the Android platform. The sensor node, which is the tail machine of the SET system, autonomously collects the sensing data of the target using the heart rate, acceleration, and body temperature sensors, and the collected information

Figure 2: Machine-to-machine architecture.

is transmitted to the gateway within the scope of communication through USN. The sensor information is processed in the Android platform and transmitted to the server through Wi-Fi. The server adjusts the urgency level given to the processed information according to the degree of danger and sends notices of impending danger to the different components.

2.3. Energy Saving. In the SET system, the lifetime of the sensor node in the electronic tag significantly affects lifetime of the overall system, which uses the wireless sensor network. That is, the lifetime of the wireless sensor network is affected by how efficiently the sensor node uses the limited energy. Therefore, it is essential to minimize the waste of energy in an SET system that uses sensor nodes.

The main factors leading tor energy waste in wireless sensor networks are collision, overhearing, and control packet overhead. Thus, the adaptive idle listening method should be applied to the SET system, which uses a wireless sensor network, to reduce the energy used by the sensor nodes. In addition, it is also necessary to examine the research on saving energy based on the Media Access Control (MAC) protocol of the SET system.

The asynchronous MAC protocol [10] is a method that transmits data when the communication channel is idle. The checking of the physical link of the channel relies heavily on the Radio Frequency (RF) chip and has the shortcoming of increased energy waste due to the transmission node overhearing the preamble. Moreover, the asynchronous MAC protocol consumes unnecessary energy since data collisions can be found even after all preambles are transmitted.

Still, the asynchronous MAC protocol is easily implemented because it does not have to be synchronized, and it can maximize energy saving since it consumes a small amount of energy when there is little competition and information in the wireless network. Moreover, the asynchronous MAC protocol is appropriate for the proposed SET system as it can reduce idle listening.

The weakness of the asynchronous Berkeley MAC (B-MAC) protocol lies in the increase in transmission delay, the collision of data, and large overhead due to long preambles. Energy also continues to be consumed because of the continuous disagreement in communication time between the sensor nodes and the coordinator due to the fixed length of the duty cycle.

However, in the SET system proposed in this study, the coordinator and the sensor nodes communicate only within the communication scope, causing little data collision due to competition. The phenomenon of energy dissipation due to the increased delay in data can be resolved by using the adaptive duty cycle method [11], and the proper wakeup and idle listening can be predicted by the duration field included in the ACK.

As previously mentioned, if competition intensifies, the possibility of data collision increases in the B-MAC protocol, which has a long preamble. Clearly, the problem can be solved using the result of a previous research on Low Power Listening (LPL), but as the CC2420 RF chip in this system does not support LPL, it is impossible to use the technique to occupy the channel with the preamble before data transmission.

Therefore, the CC2420 chip checks the condition of the channel in various modes using the Channel Clear Assessment (CCA) for data transmission and communication and approaches the channel using the back-off method if the communication channel is in use. The CCA consumes less

energy than LPL due to the short time required to distinguish noise from the signal. In addition, most IEEE 802.15.4 RF chip manufacturing companies define the Listening Interval Time (LIT) to conduct functions similar to those of LPL as longer than the time of the completion of the preamble transmission. Through these methods, the energy attrition due to a fixed wakeup time can be overcome, and the same functions as LPL can be accomplished [12–14]. Using CC2420, as many as 16 bytes of preamble can be supported, which can be adjusted by the user. The channel-occupying time is only 0.512 ms. If LPL is operated here, the energy consumed for turning the RF switch transmitter on and off increases, reducing efficiency.

An additional problem pertaining to LPL is that data transmitted after the preamble are transmitted for a time longer than the length of the data transmission in the process of communication between the sensor node of the electronic tag and the coordinator. Although the coordinator knows that the sensor node is transmitting data, it does not know the exact timing of the data transmission, and thus, it has to receive all the preambles.

If 64 bytes of data are transmitted consecutively based on CC2420, as in this study, the interval time between packets will be 8 ms with about 2 ms of channel occupying time as

$$\frac{\mathrm{TX}_{\text{data length}}}{\mathrm{PPDU}_{\text{data length}}} = \frac{64\,(\text{byte}) * 8\,(\text{bit})}{250,000\,(\text{bit/sec})} = 2.048\,(\text{ms})\,. \quad (1)$$

The sensor node of the electronic tag checks the channel in every preamble cycle. CC2420 takes longer time to transmit than the channel occupying time (2 ms), and the channel is checked for a time shorter than the interval between imaginary preambles (10 ms). If the channel is to be used in idle condition by checking the channel for a time shorter than 10 ms, the beacon should be received or the back-off must be conducted after transmitting the data. More energy may be consumed when using CC2420 than when using LPL, as the CCA should be repeated during the time of the imaginary preamble packet. However, the sensor node of the electronic tag can reduce energy consumption to one-fifth of that used for LPL in the B-MAC protocol by transmitting the preamble packet for just 2 ms at 10 ms intervals.

2.4. Location Tracking. The location tracking of an electronic tag is calculated by the triangulation method using the distance measured from the GPS satellite during communication.

The GPS satellite communication method is used more frequently outdoors than indoors. The GPS chip is expensive and much more additional energy is consumed by GPS satellite communication [15, 16]. Therefore, various methods that use less energy than GPS satellite communication to acquire of location information have been researched [17].

However, it is not necessary for the SET system to know the exact location of the target being monitored, because it is necessary to know only when the target approaches a certain area in real time. Since the above methods consume a large amount of energy, they will not be applied to the SET system. Instead, the SET system will apply a cell-based location tracking method, which transmits details of the target's approach, in order to reduce consumption of energy used in location tracing.

Figure 3: Cell-based location tracking.

In the SET system, the operating distance for communication between the sensor node and the gateway is a 40 m radius from the gateway. Within this distance, the sensor node and the gateway can communicate even in the worst conditions. In other words, a gateway can communicate with all the sensor nodes within an 80 m distance. This is explained in detail in Section 4. Therefore, as shown in Figure 3, a SET system can be implemented by installing several gateway cells to cover all facilities for children, such as a nurseries and kindergarten or elementary schools.

Therefore, in the cell-based location tracking method for the SET system, the sensor node of the electronic tag does not consume battery resources for communication with GPS satellites and exact location tracing, and thus it is possible to monitor the target for a longer time than that permitted by the existing electronic tagging systems without changing the battery.

2.5. Infer a Psychological State Using Sensor. In Korea, the number of sexual assaults has more than doubled over the past 10 years, and the rate of repeat offenders is about 60%, which is very high. The repeat rate of offenders who have sexually assaulted children aged 13 and younger is higher than that for other crimes by up to 10% [18]. The prevention of such repeat crimes would be facilitated by gathering data for changing the emotional state of the offenders and analyzing it to predict a repeat offense.

While it is realistically impossible to gather data directly for changing emotions, it is possible to analyze the physical changes caused by emotional changes and gather data related to them. A good example of such technology is the polygraph.

However, the patterns and profiles of sexual assaults found in real-time sensor data have yet to be organized, and there is no exact real-time biometric information for each sexual offender. In addition, it is prohibited to distribute detailed biometric profiles of sexual offenders to the public.

Therefore, in this study, the psychological condition of the target is inferred using the data measured by heart rate, body temperature, and acceleration sensors. The heart rate sensor measures the number of cardiac impulses by gauging the main artery in the leg of the target, and the body temperature sensor and accelerometer measure the body temperature

Table 1: The recommended exercise intensity.

Intensity of exercise	Low	Medium	High
Energy consumption (Cal/min)	2.0–4.9	7.5–9.9	12.5
Maximum intake of oxygen (VO2 max, %)	30–49	50–74	75 and over
Percentage of heart rate per each intensity (HR, %)	35–59	60–79	80 and over

Table 2: Maximum and 80% heart rate according to age.

Age	20	30	40	50	60	70+
Maximum heart rate (bpm)	200	192	183	175	166	158
80% of maximum heart rate (bpm)	160	153	146	140	133	126

and the acceleration of the x, y, and z axes of the target's ankle, respectively. The psychological condition of the target is inferred in real time using the sensor data. These inferred data will be used to interrupt and prevent repeated sexual assaults.

The standard for the heart rate to be applied in the SET system is referred to in general medical literature [19–26]. The heart rate is the number of heart beats per a unit of time, and it is usually expressed in Beats Per Minute (BPM). The heart rate varies according to the body's demand for oxygen intake and carbon dioxide emission while exercising or sleeping. The normal heart rate range is 60–100 bpm [21–24]. The heart rate is also related to age and gender [27].

The reasons for an increase in heart rate are as follows. First, the heart rate can increase due to a rise in blood pressure and respiration rate caused by mental excitement [28]. Second, the heart rate can increase due to the transmission of catecholamines to the sympathetic nerve due to stress [29]. Third, if the body temperature increases for various reasons, the blood vessels on the surface of the skin expand to reduce the body temperature and as a result the amount of circulating blood decreases. However, oxygen and glucose, which are carried by the blood, have to be supplied constantly, and thus the heart rate increases to maintain the total cardiac output [30].

A standard for heart rate is required in order to infer the mental state of the target being monitored by sensors using the SET system. Therefore, the standard for the heart rate that signals the danger that the target will commit a repeat crime was set to 160 bpm, according to the following information.

The heart rate of ordinary people varies widely, with 60–100 bpm being considered the normal range. However, when an ordinary person exercises, the heart rate rises sharply, and different persons will have different heart rates for different exercise intensities. Doctors recommend doing moderate exercise and refraining from over exercising to maintain health. They suggest exercises that maintain about a heart rate that is 80% of the maximum (HR_{max}), following the guidelines of the American College of Sports Medicine shown in Table 1 [31].

The formula for the calculation of the maximum heart rate was established by Dr. William Haskell, a sport physiologist, and Samuel Fox, a heart medical scientist, during the late 1960s [32]. A revised formula was published by Gellish et al., Londeree, and Moeschberger [33, 34], which is written as

$$HR_{max} = 217 - (0.85 * \text{age}). \quad (2)$$

The maximum heart rate is calculated for each age using (2), and 80% of the maximum heart rate is calculated using (3). About 80% of the maximum heart rates calculated for different ages using (3) is shown in Table 2 as

$$80\% \text{ of } HR_{max} = [217 - (0.85 * \text{age})] * 0.80. \quad (3)$$

The body of an ordinary habitual sex offender secretes dopamine due to the sensitive reaction of the autonomic nervous system and the endocrine system as soon as he decides to commit a repeat crime. This dopamine excites the α, $\beta 1$ adrenaline receptor, and the sympathetic nerve is excited by the secretion of adrenaline, increasing the heart rate, and the heart's contractile power. The heart rate at this time rises to a point that cannot be reached by normal people even after exercising [35–37]. Therefore, the standard value for the heart rate sensor was set at 160 bpm, which is higher than that which can be reached by normal people after correct exercising.

According to a report of the Korean Institute of Criminology, about 12% of the rapists in Korea committed rape after drinking alcohol [38, 39], but as shown in Table 3 almost 40% of repeating sexual offenders committed sexual assaults after drinking alcohol. In addition, most assailants who committed sexual assaults on minors and women confessed that they committed the crime after drinking alcohol [40].

Thus, when identifying the context for a sexual assault, it is most important to determine whether the target is under the influence of alcohol, because in that case the probability of the sexual offender committing a repeat crime is almost 40%. If a target is found near facilities for children after consuming alcohol, the situation becomes very dangerous.

For ordinary people, the heart rate after drinking alcohol is measured mostly in the 60–100 bpm range [41]. Therefore, it is not easy to decide whether someone has consumed alcohol using only a heart rate sensor; however, alcohol consumption can be inferred on the basis of data that show changes in the heart rate pattern similar to those seen after drinking alcohol [41–43]. In ordinary people, the heart rate rises by about 10 bpm from the average heart rate within 2 h of drinking alcohol. Subsequently, the difference between the highest heart rate and the lowest heart rate increases to more than 20 bpm within 10 h. In addition, the body temperature remains low after alcohol consumption [44, 45]. Such characteristics of the values measured by sensor are used to infer the identification of the context.

Table 3: The number of sex offenders and repeat offenders under the influence of alcohol in Korea.

Crime	2006	2007	2008	2009	2010
Total rapes (person)	7,255	6,441	8,957	9,351	11,402
Rapes under influence of alcohol (persons)	907	843	1,154	1,170	1,375
Percentage (%)	12.5	13.1	12.9	12.5	12.1
Repeat offenders (person)	757	686	1,024	1,107	4,693
Rapes under influence of alcohol among repeat offenders (person)	315	306	489	413	1,825
Percentage (%)	41.6	44.6	47.8	37.3	38.9

3. The Proposed System

3.1. The Smart Electronic Tagging System. The SET system is based on M2M. It is composed of two parts, the USN and the Android platform; an overview of its system architecture is shown in Figures 4 and 5.

The constituents of the SET system are the USN-based Security Sensor Node (SSN), which is attached to the monitored target as an electronic tag; the Android platform-based Security Sensor Gateway (SSG), which collects information; the Central Crime Management Center (CCMC), which manages and controls the overall system, being linked to the gateway, and provides context-awareness service; the smart mobile terminals, which receive information about the target's approach.

Since the information about the target must be protected from network intervention, hacking, and so forth, the information is collected through a USN-based network and encoded in the security wireless system. Then, the information is received, decrypted, and transmitted through Wi-Fi and the Internet by the unsecured mobile management system, separately.

3.2. Security Sensor Node. The SSN is the electronic tag of the target. When it is within the communication sphere of the SSG, which plays the role of coordinator, it transmits the sensed packet information to the SSG periodically. If it is not within the communication sphere of the SSG, as aforementioned, a channel condition checking method through CCA using the CC2420 RF chip and LIT is applied. In addition, the energy-saving phase is performed in the SSN by predicting wakeup and idle listening using the duration field included in the ACK. Therefore, the SSN does not transmit the sensed packet information moving as in the adaptive sleep mode.

The SSG requests information about the approaching SSN through the broadcast beacon message in about 1,000 ms units. Thus, the SSN node prevents the loss of energy caused by periodical transmission and induces the connection of the cluster unit. If there are several SSNs in the communication sphere of the SSG, the SSN forms a star topology network structure to conduct individual communication with the SSG.

The SSN is composed of three parts: the Sensor Subsystem (SS), which measures and collects the sensor information, the Communication Subsystem (CS), which communicates with the SSG and the device that collects fixed information, and the Main Module Subsystem (MMS), which is responsible for processing the collected sensing data, encoding, and minimizing power usage.

The primary data in the SET system are created as shown in Table 4 using the data value generated through the body temperature gauge, accelerometer, and heart rate sensor hardware of the SS by the TinyOS v2.0.2 based MMS in the SSN. The MMS is composed of software for low power processing and information transmission that is appropriate for the proposed SET system, on the basis of the TinyOS. In order to avoid the risk of electronic tagging information being duplicated and falsified and to protect the data, the data measured at the SS are encoded by the Advanced Encryption Standard (AES) 128 bit encryption algorithm [46] using the TinyOS option.

If the broadcast beacon message is received from the SSG or fixed node through the CS, these encoded data are transmitted to the SSG or fixed node through the CC2420 RF chip.

3.3. Security Sensor Gateway. The SSG is the center of the M2M system, which connects the USN and the Android platform in the SET system. The hardware of the SSG is composed of a coordinator, USB-to-serial board, and Android Gateway Device (AGD).

The SSG is linked to the USB interface through the GPIO expansion of the coordinator, and it communicates with the wireless sensor network. When the mobile node of the electronic tag (SSN) receives the SSG's broadcast beacon message within a 40 m radius of the SSG, the SSN transmits the collected sensor data to the SSG.

At this time, the coordinator broadcasts a request through the beacon to the SSN to transmit the AES 128 bit encoded data. The encoded sensor data that the coordinator receives from the SSN are restored to the initial 64 byte data through an AES 128 bit decryption algorithm in its own TinyOS option. The decrypted information is transmitted to the AGD through the USB-to-serial interface.

Then, the AGD converts the decrypted 64 byte data to 44 byte data using the Detection Application Program (DAP) as shown in Table 5, transmits the data to the CCMC through the Wi-Fi communication network connected to the AGD, and displays this information on the AGD screen through the

Figure 4: The architecture overview in the SET system.

Figure 5: The data communication overview in the SET system.

Android Graphical User Interface (GUI) layout developed in Extensible Markup Language (XML).

3.4. Central Crime Management Center. The CCMC is a 24/7 server that is connected through the Internet. The CCMC communicates data using many AGDs set in a certain Internet Protocol (IP) and port; it receives data from the AGD through Wi-Fi communication and saves it in the Data Base (DB).

At the same time, the CCMC analyzes the data and provides the context-awareness service. To provide the context-awareness service, the SET system sets up danger levels

Table 4: The packet structure in SSN.

Frame attribute	Data Length (byte)	Attribute
Am type	1	The code to identify the unique packet
Destination address	2	Address of reception node
Source address	2	Address of transmission node
Payload length	1	Size of payload packet
Group ID	1	Sensor node group ID
AM type	1	The code to identify the unique packet
STX	1	Start bit
Node ID	2	Sensor node ID
Unique ID	6	Unique ID
RTC	6	Year/month/day/hour/minute/second
Temperature	8	Body temperature sensor (highest, lowest, average, current value)
Accelerometer	24	Acceleration sensor (highest, lowest, average, current value)
Heart rate	8	Heart rate sensor (highest, lowest, average, current value)
ETX	1	End bit

Table 5: The packet structure converted into AGD.

Frame Attribute	Data Length (byte)	Attribute
STX	1	Start bit
AGD_no	1	Android gateway device number
SN_no	1	Sensor node number
Temperature	8	Body temperature sensor (highest, lowest, average, current value)
Accelerometer	24	Acceleration sensor (highest, lowest, average, current value)
Heart rate	8	Heart rate sensor (highest, lowest, average, current value)
ETX	1	End bit

Figure 6: The flow chart for the context-awareness service.

in advance by classifying the risk grades of the monitored targets, as shown in Table 6.

The M2M-based SET system processes the complicated data obtained from the heterogeneous USN and Android platform to offer the various types of context-awareness services. A context-awareness service is a smart service that is applied appropriately to real-time dangerous situations based on the information measured by the sensors.

The SSN of the monitored target is given a unique number in advance for each danger level according to the type of the target's previous crime, as shown in Table 6. The CCMC extracts and analyzes the information in the context-awareness data according to the received SSN unique number and provides the appropriate information in real time based on the analysis of the context.

The heart rate, body temperature, and 3-axis acceleration sensors are attached to the SSN. In addition to the SSN's unique number information, the CCMC uses the sensor measurement values of these sensors to infer the mental state of the target and provides status information according to the flow chart of context-awareness services, as shown in Figure 6, to help preempt crimes.

The procedures of the context-awareness service are as follows. From the SSN's heart rate, body temperature, and acceleration sensors, the CCMC receives the accumulated highest, lowest, average, and current values of each sensor. First, the CCMC checks whether the monitored target has consumed alcohol. First, CCMC checks whether the current

Table 6: The danger level in the SET system.

Danger level	SSN unique number	Explain	Service scope
Urgent	1	A target who killed a child or a handicapped person is approaching	SMS to police department and smart mobile Location, warning data transmission to neighborhood AGD and CCMC
Emergency	2	A target who kidnapped a child or a handicapped person is approaching	SMS to police department and smart mobile Location, warning data transmission to neighborhood AGD and CCMC
Hazard	3	A target who molested a child or a handicapped person is approaching	SMS to smart mobile Warning data transmission to neighborhood AGD
Peril	4	A target who was convicted of habitual robbery is approaching	SMS to smart mobile Warning data transmission to neighborhood AGD
Caution	5	A target who has a habitual drunkenness problem is approaching	SMS to smart mobile
Attention	6	Another dangerous person is approaching.	SMS to smart mobile

Table 7: The context-awareness algorithm.

Query	Alcohol consumption	Heart rate change	Body temperature change	Acceleration condition	Context awareness
1	Yes	Any	Any	Any	Sexual molestation intention
2	No	No	Any	Running/walking	Wandering
3	No	No	Any	Stop	Sexual molestation intention
4	No	Yes	Yes	Running	Exercising
5	No	Yes	Yes	Walking	Tension due to stress or pain
6	No	Yes	Yes	Stop	Sexual molestation intention
7	No	Yes	No	Any	Sexual molestation intention

value of the measured heart rate surpasses the recorded highest value and sees whether the value of heart rate exceeds the highest heart rate by more than 10 bpm compared to the average value in order to confirm that a pattern of alcohol consumption is present. As the body temperature decreases after drinking alcohol, the CCMC checks whether the current value of the measured body temperature is lower than the average value and confirms through the accelerometer that the target is not running. The 3-axis acceleration sensor attached to the ankle of the target receives the data value of the x, y, and z axes about every 1000 ms and it has the highest, lowest, average, and current values. If there is no motion, the data values of the three axes in the acceleration sensor change minutely. If the target stops, walks, runs, or takes the elevator, the data values of the three axes will change dramatically, and the changing patterns for each behavior shown on the graphs make it possible to distinguish between them [47–50]. This will be explained when the values actually measured by the three-axis acceleration sensor in the acceleration sensor measurement experiment in Section 4 are given. If all these conditions are satisfied, the CCMC will decide that the monitored target has consumed alcohol and intends to make a sexual assault, as shown in Table 7. The danger level is then set as the highest "urgent" level to initiate the context-awareness service.

The CCMC also checks whether the current value of the measured heart rate has surpassed the lowest value, and whether the value of the heart rate exceeds the lowest heart rate by more than 20 bpm compared to the highest value in order to confirm whether a pattern of alcohol consumption is present and confirms through the acceleration sensor that the target is not running. If these conditions are satisfied, the CCMC will decide that the monitored target consumed alcohol and the danger level is set at "urgent" level. If it is inferred that the target consumed alcohol, as in the two above examples, the CCMC sets the danger level of the target as the "urgent" level for 24 hours, and the context-awareness service is provided. In a normal person, alcohol elimination from the body takes approximately 24 hours [51]. Therefore, the dangerous elements of the target can last for 24 hours. Except for the two examples described above, the target is regarded as not having consumed alcohol.

Second, the CCMC determines whether the average value of the heart rate recorded by the sensor exceeds 160 bpm. It was explained in Section 2 that this is the standard set for a heart rate that signals the danger that an offender is committing a repeat crime. If the heart rate is 160 bpm or lower and the value of the acceleration sensor indicates running or walking, the monitored target is wandering around a facility for children. Thus, the context-awareness service is implemented according to the danger level shown in Table 6.

Third, if the data value of the acceleration sensor in case 2 indicates that the target has stopped, the target has infiltrated

into the facility for children and is not moving. In this case, it is inferred that the target is watching the children attentively, and they are vulnerable to being an object for a repeat crime. Thus, the danger level is set as the highest level and the context-awareness service begins to be implemented.

Fourth, if the heart rate exceeds 160 bpm, the target's body temperature is measured immediately. The CCMC checks whether the current value of the body temperature according to the sensor has increased or decreased within 0.5°C of the average value. Usually, the body temperature of ordinary people aged 18–40 is 36.4–37.2°C. The body temperature changes naturally according to the time of the day; the normal range of change is about 0.5°C [52]. Therefore, if the body temperature changes by more than 0.5°C, the CCMC checks whether it is higher than the average value and whether the target is running according to the value measured by the acceleration sensor. If all these conditions are satisfied, it is inferred that the target is exercising, and the normal context-awareness service is offered.

Fifth, if the heart rate of the target exceeds 160 bpm and the body temperature is more than 0.5°C above or below the average value and it is confirmed through the acceleration sensor that the target is walking, it is inferred that the target is in tension due to stress or pain, and the normal service is provided

Sixth, if the heart rate of the target exceeds 160 bpm and the body temperature is more than 0.5°C above or below the average value and it is confirmed through the acceleration sensor that the target is stationary, it is inferred that the target is mentally excited by a plan to commit a repeat crime or is imagining an intended sexual assault or molestation and the minors in the relevant area are very likely to be in danger. In this case, as the minors can be the potential victims of a repeat crime of the monitored target, the highest danger level is set, and the corresponding context-awareness service is provided.

Seventh, if the heart rate of the target exceeds 160 bpm and the body temperature changes within a 0.5°C range, it is inferred, regardless of the change in the acceleration sensor, that the target is mentally excited by a plan to commit a repeat crime or is imagining an intended sexual assault or molestation. In this case, the highest danger level, "urgent," is set and the corresponding context-awareness service is provided to protect the minors.

This context-awareness service is accompanied by the AGD's unique number information in the packet data. In this case, the location of the monitored target can be found via cell-based location tracking without GPS information, and as a result both the context-awareness and location-awareness service can be activated. When the location awareness service is provided using GPS, the packet data of the GPS is composed of about 26–64 bytes, depending on the number of GPS communication satellites [53]. However, in the SET system, it is possible to provide the location awareness service with 1 byte of information using the AGD's unique number.

The CCMC sets the procedures of the context-awareness service according to the danger level of the approaching monitored target and implements services appropriate for each level. Notification of the highest danger level, the urgent level, is sent by Short Message Service (SMS) to multiple sets of smart mobiles registered in advance at the police station and the CCMC. The communication module of the server includes the connection-waiting function for the smart mobile. It sends the current data in real time whenever there is a request for communication from multiple sets of smart mobiles. In addition, the information on danger is notified to the AGD and the CCMC adjacent to the AGD that is being approached by the target. Moreover, the data are saved in the DB and the information is shown on the screen of the persons using the server through the GUI-based monitoring program developed in XML.

3.5. Smart Mobile Device. The mobile phone used is an Android smart phone in which the Reception Application (RA) is installed. When the specific IP and port of the server are input using the RA, the smart phone receives the data from the server in real time in about 1000 ms units. A text banner, sound, or vibration alerts the user to the incoming data. These data include the danger level of the target and the measured time.

In addition, if the smart mobile is not able to receive data due to lack of reception, the server of the CCMC saves the information on the IP and the MAC Address of the smart mobile and transmits the information to the smart mobile later, when it is connected. The multitasking function of the Android OS is used to save the information in the external memory in a real time log even if the smart phone is not activated all the time, and this saved information can be viewed later.

4. Implementation

4.1. The Test Bed Configuration. The SET system is composed of the USN and the Android platform based on M2M, as shown in Figure 4. The SET system shown in Figure 7 constitutes a test bed.

All the programs in the SET system were developed in two computer languages. The programs for transmitting data from the SSN to the coordinator were developed using the nesC language, and the programs for transmitting from the AGD to the CCMC and the smart mobile device were developed using the Java language. To show the entity relationships of the data in each device in the SET system in a unified modeling language for object-oriented analysis and design, a class diagram is presented in Figure 8.

The SSN and SSG were produced and implemented directly; their detailed specifications are shown in Table 8.

The CCMC is contained in an ordinary x86-based laptop computer with a Wi-Fi wireless Internet network; its OS is Windows 7. If a target approaches a facility for children or a certain facility, the CCMC indicates the information on the screen of the laptop computer for user in the GUI-based monitoring program using the context-awareness service.

The smart mobile device is an ordinary Android smart phone. When it receives data concerning the approach of a target from the CCMC through the Internet, it alerts the user through vibration, sound, and a text banner; the data appear on the screen through an Android application, as shown in

Table 8: The detail specification of SSN and SSG.

	Security sensor node	Security sensor gateway
CPU	Texas Instruments MSP430 F1611 16 Bit RISC	ARM Cortex A8 1 GHz
Memory	FLASH 1 MB, EEPROM 2 KB	DDR SDRAM 512 MB, T-Flash (Micro SD) 2ea
RF Chip	Chipcon Wireless Transceiver CC2420 (IEEE 802.15.4, 250 Kbps, 2.4 GHz)	High speed IEEE802.11 b/g
Sensors	Body temperature, heart rate, 3-axes acceleration	3-axes geomagnetic, 3-axes accelerometer
Power	AAA 1.5 V battery 2 pieces	Li Polymer 2,700 mAh or AC Input 220 V 60 Hz, DC Output 5 V/3 A
Software	TinyOS v2.0.2	Linux Kernel 2.6.32, Android Froyo 2.2
Etc.	SPI interface GPIO expansion	5 inch WVGA (800 × 480) HD TFT-LCD

Figure 7: The test bed configuration of the SET system.

Figure 9(a). In addition, the data received from the beginning of the communication with the CCMC are saved in text files for each day to be viewed at any time using the "Show History Log" button of the application, as shown in Figure 9(b).

4.2. Experimental Results. The heart rate, accelerometer, and body temperature sensors are attached to the SSN. The accelerometer is a device that can measure the amount of change in acceleration in 3D space: the x, y, and z axes. Analysis of the data of the 3-axis accelerometer allows patterns of stop, walking, running, climbing steps, and so forth to be determined sequentially in a 3D space; they are simultaneously divided into action units. Action patterns according to movement measured by an accelerometer have been researched using various techniques [54–56]; the Hidden Markov Model (HMM) is considered an appropriate model of the characteristics of time series data [57].

However, the application of patterns based on the HMM increases the load on the SET system, and the response speed in the case of uncertainty is slow. Therefore, in the SET system, the axis of time and the axis of the measured value of the sensor were separated into individual sections. The SSN was attached to the ankle of a subject to test the accelerometer. The subject performed movements of six kinds sequentially for about 67 m, and the data values were measured, as shown in Figure 10.

The lateral axis is the value of the time per second, and the longitudinal axis is the value of the 3-axis acceleration of the subject obtained through the accelerometer. The subject walked for about 17 minutes in area A, ran for about 17 m in area B, climbed steps for about 17 m in area C, was in a descending elevator for about 4 m in area D-1, was in an ascending elevator for about 4 m in area D-2, and remained standing for about 8 m in area E on the 6th floor of the building. Using the values measured in the test, the subject's condition of walking, running, or remaining at rest could be inferred with the 3-axis measured values of the accelerometer alone. The data on the movement in the x, y, and z axes were applied to the context-awareness service.

The heart rate sensor and the body temperature sensor are also attached to the SSN. The heart rate sensor is a device that measures the heart rate using an Infrared-Light Emitting Diode (IR-LED) light that penetrates a body's tissue. The frequency of the change in light reflected from the tissue represents the heart beat. The body temperature sensor is a device that measures body temperature by sensing the value of the electrical resistance of the body, which varies according to its temperature.

Figure 8: Class diagram of the entity relationships.

The SSN was attached to the ankle of the subject to test the heart rate and body temperature sensors. The subject performed six kinds of movements sequentially for a total of about 67 m in the same conditions as those used to test the accelerometer. The tests were conducted twelve times and the data values were measured, as shown in Figure 11.

The lateral axis is the value of the time per second, and the longitudinal axis on the left is the subject's heart rate data values obtained through the heart rate sensor. The longitudinal axis on the right is the subject's body temperature data values obtained through the body temperature sensor. The subject walked for about 17 m in area A. The value of the heart rate and the body temperature rose from the initial value when the subject was at rest, but it maintained a certain value while walking. In area B, the subject ran for about 17 m. The value of the heart rate and the body temperature rose rapidly as the subject began to run, and thereafter changed slightly according to the speed of the subject's movement. In area C, the subject climbed steps for about 17 m. The heart rate and body temperature values remained the same as

(a) (b)

Figure 9: The smart mobile device. (a) The screen that shows the banner information and (b) the screen that shows the received data.

Figure 10: The data values measured with the 3-axis acceleration sensor.

Figure 11: The data values measured with a heart rate and body temperature sensor.

those measured immediately after the subject ceased running but continued to decrease gradually; however, the values were maintained at a higher level than those for the walking condition in area A. The subject was also in a descending elevator for about 4 m in area D-1 and in an ascending elevator for about 4 m in area D-2 and remained standing for about 8 m in area E on the 6th floor of the building. During these three activities, the subject simply stood still, and thus the value of the heart rate and the body temperature kept decreasing gradually until returning to the normal value.

Thus, the data values for the subject's walking, running, and rest conditions were obtained through measurement of the heart rate and the body temperature. These data can be applied to the context-awareness service.

The sensor data are generated in the SSN of the monitored target and arrive at the smart mobile after passing the SSG and CCMC of the SET system. USN communication is transmitted from the SSN to the SSG. A test was performed in this environment to find the time required for data transmission according to the distance of the SSN. The measurement values obtained are shown in Figure 12. In addition, at this time, the CPU usage in the AGD Android platform was also measured.

The lateral axis is the value of the distance per meter between the SSN and SSG, and the longitudinal axis on the left is the value of the time per second taken by the

Figure 12: The data transmission time in the SET system and the measured amount of CPU usage in the AGD.

Figure 13: The power consumption in SSN and the electronic tag.

data generated in the SSN, as shown in Table 4, to arrive at the smart mobile device. The longitudinal axis on the right represents the amount of CPU used by the data generated in the SSN to pass the AGD after being converted as shown in Table 5.

USN communication is used to transmit the data generated by the SSN to the SSG. According to the test, the data information of the SSN is received by the SSG within a distance of up to 50 m. However, if the distance between the SSN and the SSG is more than 40 m, 2 out of 12 pieces of data (17%) transmitted for about 12000 ms are not received. It is thought that this phenomenon is caused by environmental noise. As the SSN plays the role of the electronic tag, 100% of the information of the SSN must be received within the range of the SSG. Therefore, the maximum range for the communication of SSG in the SET system has been determined to be a 40 m radius.

In addition, there was no significant change in the amount of usage of the CPU in the AGD in relation to distance when only one SSN was used for the test. However, when two SSNs were used at the same time, the amount of usage of the CPU increased gradually as the distance increased, and at 50 m the amount of usage of the CPU reached 50%. It is inferred that as the CPU of the AGD is a single Application Processor (AP), it uses multithreading when overlapped data packets are received according to the distance, which increased the amount of usage of the CPU. If the AP of the AGD had been a dual or quad core, the load would not have been so large. In addition, it is inferred that with the current, single AP, the time for data transmission is delayed, as much of the data processing slows down when four or more SSNs connect with the SSG at the same time.

The SSN in the SET system had a limited power supply. It collected the information on the monitored target in about 1000 ms units using the built-in sensors. When the SSN received the SSG's broadcast beacon message, it transmitted the collected sensor data to the SSG. However, if the SSN was not within the communication sphere of the SSG, it went to sleep mode for energy saving.

In the existing electronic tagging system, the electronic tag had a limited power supply. It communicated with portable terminal through 400 MHz Industrial, Scientific, and Medical (ISM) two-way radio bands. It communicated with the portable terminal 24 hours a day in about 10000 ms units. If the two-way communication was successful, the electronic tag went to sleep mode to save energy until the next communication.

The SSN's power switch was changed from the OFF to the ON state for measuring power consumption within the communication sphere of the SSG. However, the electronic tag had no power switch, and therefore the power consumption was measured assuming that the tag was in the ON state while being within communication range of the portable terminal. The power consumption of the SSN and the electronic tag was measured for 1000 ms in the same environment. The measurement results are shown in Figure 13. The horizontal axis is the value of the time in ms and the vertical axis is the value of the current in mA that was consumed in communication between the SSN and SSG, that is, the electronic tag and the portable terminal, respectively.

The SSN consumed a maximum current of 18 mA during approximately 25 ms in order to initialize the device and consumed a maximum current of 13 mA during approximately 40 ms in order to collect the values of the sensor data. In addition, when the SSN communicated with SSG, it consumed a maximum current of 26 mA during approximately 30 ms when receiving and a maximum of 38 mA current during approximately 120 ms when transmitting. It consumed less than 1 mA in sleep mode.

It was difficult to measure the amount of current used for initializing the electronic tag, because there was no power switch. When the electronic tag communicated with the portable terminal, it consumed a maximum current of 31 mA during approximately 150 ms when receiving and a maximum of 46 mA during approximately 100 ms when transmitting. It consumed less than 2 mA current in sleep mode.

The SSN used 4500 mA while communicating with the SSG, whereas the electronic tag used approximately 8000 mA

while communicating with the portable terminal for 1000 ms in the same environment. These results show that the SSN can save approximately 56% of the energy used by the electronic tag. Therefore, the SET system is better than the existing electronic tagging system in terms of energy efficiency.

5. Conclusion

It is difficult to change batteries in the existing electronic tagging system, and thus it is essential to use as little power as possible. However, the existing electronic tagging system has problems in accomplishing low power consumption as it conducts location tracking by using a triangulation method to communicate with the GPS satellites and mobile phone base stations. Exact location tracking is also difficult to achieve in this way as the maximum error is 2 km, due to buildings, forests, or low reception regions in downtown areas. In addition, while sexual assaults are increasing every year and sexual crimes against minors when the offender is under the influence of alcohol are becoming more ferocious, the existing electronic tagging does make it possible to know whether the monitored target only consumed alcohol. It has become much more difficult to protect minors as location tracking of a target wearing an electronic tag by unauthorized people is illegal under the Korean law.

USN technology is used because it is lightweight, cheap, and simple to operate, in addition to consuming very little electricity. Additionally, the usage of smart phones has become common in Korea and the US, with about 50% of the population using such a device. Thus, this thesis has proposed the SET system, which uses the M2M method based on the USN and Android platforms, utilizing only their strong points. This paper proposed M2M and context-awareness service methodology as a system that will notify the approach of a monitored target. In our study, a test bed was established to verify the validity of the SET system.

The equipment actually developed and produced to implement this system includes an SSN, coordinator, USB-to-serial board, and an AGD. The Java language was used for the programming of the SSG, CCMC, and smart mobile device. The performance of the equipment in the system was confirmed through the establishment and testing of the test bed. In the future, this study will be expanded by linking the electronic tagging system with the SET system for the purpose of monitoring sexual assaults.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2012461).

References

[1] W. Y. Kim, *Improve Productivity Using M2M*, Services Science National Forum Union Seminar, Seoul, Republic of Korea, 2010.

[2] P. Liu, S. H. I. Yeung, K. A. Crenshaw, C. A. Crouse, J. R. Scherer, and R. A. Mathies, "Real-time forensic DNA analysis at a crime scene using a portable microchip analyzer," *Forensic Science International: Genetics*, vol. 2, no. 4, pp. 301–309, 2008.

[3] Y. C. Yoon, "Critical study on the electronic surveillance system in Korea," *Korean Criminological Review*, vol. 19, no. 3, pp. 201–228, 2008.

[4] "Act on electronic monitoring device attachment, etc, for specific crimes offenders," 9765 §16-2, 2009.

[5] R. Cotter and W. D. Lint, "GPS-electronic monitoring and contemporary penology: a case study of US GPS-electronic monitoring programmes," *The Howard Journal of Criminal Justice*, vol. 48, no. 1, pp. 76–87, 2009.

[6] H. S. Kang and H. G. Moon, "A study on the effect and suggestions of GPS electronic monitoring program," *Korean Academy of Probation and Parole Services Review*, vol. 10, no. 2, pp. 39–135, 2010.

[7] S. K. Oh, "A study on the scheme for development of the electronic monitoring system," *Seoul Law Review*, vol. 20, no. 1, pp. 311–343, 2012.

[8] B. A. Kim, "The policy direction of machine-to-machine," *Telecommunications Technology Association Journal*, vol. 134, pp. 34–40, 2011.

[9] D. Boswarthick, "M2M activities in ETSI," in *Proceedings of the Society for Modeling & Simulation International Conference*, Sophia Antipolis, France, 2009.

[10] W. H. Liao and H. H. Wang, "An asynchronous MAC protocol for wireless sensor networks," *Journal of Network and Computer Applications*, vol. 31, no. 4, pp. 807–820, 2008.

[11] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 307–320, November 2006.

[12] Texas Instruments Inc., "True system-on-chip solu-tion for 2.4 GHz IEEE 802.15.4/ZigBee," CC2430 Data Sheet (Revision 2.1) SWRS036F, Texas Instruments Inc., Dallas, Tex, USA, 2012.

[13] Atmel Inc., "AT86RF233-ultra low power 2.4 GHz transceiver for IEEE 802.15.4, ZigBee, and ISM applications," Revision 8351B-MCU Wireless, 2012.

[14] Atmel Inc., *Evaluation Platform and Reference Design for the Atmel AT86RF233 Radio Transceiver SDK Guide*, Revision A, 2012.

[15] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone," in *Proceedings of the Advanced Computing Systems Association Annual Technical Conference*, Boston, Mass, USA, 2010.

[16] USGlobalSat, Inc., *EM-406a GPS Receiver Engine Board Specification*, Version 1.4.1, 2012.

[17] S. B. Kotwal, R. Abrol, S. Nigam, S. Verma, and G. S. Tomar, "Analysis of one anchor RSSI and AOA based iterative localization algorithm for wireless sensor networks," *International Journal of Simulation: Systems, Science and Technology*, vol. 11, no. 2, pp. 35–41, 2010.

[18] C. G. Park, "A study on the recent legislation for prevention of major crime against juvenile," *Korean Juvenile Protection Review*, vol. 14, pp. 51–86, 2010.

[19] S. H. Baek, J. S. Park, J. K. Lim, J. H. Choi, J. Y. Bahk, and J. S. Hyun, "The effects of sexual excitement induced by visual erotic stimulation on serum dopamine and testosterone," *Korean Journal of Andrology*, vol. 18, no. 2, pp. 125–129, 2000.

[20] R. K. Hanson and K. E. Morton-Bourgon, "The characteristics of persistent sexual offenders: a meta-analysis of recidivism studies," *Journal of Consulting and Clinical Psychology*, vol. 73, no. 6, pp. 1154–1163, 2005.

[21] D. Y. Oh, *Psychological characteristics of child sexual molesters [M.S. thesis]*, Hanyang University, Seoul, Republic of Korea, 2007.

[22] W. J. Kim and M. H. Hwang, "The effects of alcohol administration on heart rate and catecholamine response after submaximal exercise," *The Korean Journal of Sports Science*, vol. 8, no. 1, pp. 453–462, 1999.

[23] H. S. Cho, D. G. Shin, J. H. Bae et al., "Optimal resting heart rate in adult population: factors related to the heart rate," *The Korean Journal of Medicine*, vol. 72, no. 2, pp. 172–180, 2007.

[24] Y. A. Shin, M. J. Kim, and J. H. Ahn, "The comparison of heart rate variability and blood pressure according to posture and exercise type," *Exercise Science*, vol. 20, no. 3, pp. 249–260, 2011.

[25] Y. Y. Jin, E. S. Ahn, and H. S. Kang, "The test of validity using an accelerometer for the assessment of physical activity-related energy expenditure," *The Korean Journal of Obesity*, vol. 20, no. 4, pp. 185–192, 2011.

[26] I. J. Kim, *Correlation of heart rate variability with severity of depression and anxiety in normal adults [M.S. thesis]*, Hallym University, Kangwon, Republic of Korea, 2006.

[27] H. L. Kim, D. I. Park, T. D. Kwon, and S. K. Hwang, "Changes of heart rate and blood pressure after exercise in athletes of various age groups," *Korean Alliance for Health, Physical Education, Recreation and Dance Review*, vol. 24, no. 1, pp. 1129–1141, 1985.

[28] J. Archer and B. Lloyd, *Sex and Gender*, Cambridge University Press, New York, NY, USA, 2002.

[29] E. Boenisch and C. M. Haney, *The Stress Owner's Manual: Meaning, Balance & Health in Your Life*, Impact Publishers, 2nd edition, 2003.

[30] S. S. Kety and C. F. Schmidt, "The effects of active and passive hyperventilation on cerebral blood flow, cerebral oxygen consumption, cardiac output, and blood pressure of normal young men," *Journal of Clinical Investigation*, vol. 25, no. 1, pp. 107–119, 1946.

[31] S. Haufe, S. Engeli, P. Budziarek et al., "Determinants of exercise-induced fat oxidation in obese women and men," *Hormone and Metabolic Research*, vol. 42, no. 3, pp. 215–221, 2010.

[32] W. C. Miller, J. P. Wallace, and K. E. Eggert, "Predicting max HR and the HR-V-O2 relationship for exercise prescription in obesity," *Medicine and Science in Sports and Exercise*, vol. 25, no. 9, pp. 1077–1081, 1993.

[33] R. L. Gellish, B. R. Goslin, R. E. Olson, A. McDonald, G. D. Russi, and V. K. Moudgil, "Longitudinal modeling of the relationship between age and maximal heart rate," *Medicine and Science in Sports and Exercise*, vol. 39, no. 5, pp. 822–829, 2007.

[34] B. R. Londeree and M. L. Moeschberger, "Effect of age and other factors on maximal heart rate," *Research Quarterly for Exercise and Sport*, vol. 53, no. 4, pp. 297–304, 1982.

[35] J. C. Blader and W. L. Marshall, "The relationship between cognitive and erectile measures of sexual arousal in non-rapist males as a function of depicted aggression," *Behaviour Research and Therapy*, vol. 22, no. 6, pp. 623–630, 1984.

[36] M. Zuckerman, "Physiological measures of sexual arousal in the human," *Psychological Bulletin*, vol. 75, no. 5, pp. 297–329, 1971.

[37] A. Hill, P. Briken, C. Kraus, K. Strohm, and W. Berner, "Differential pharmacological treatment of paraphilias and sex offenders," *International Journal of Offender Therapy and Comparative Criminology*, vol. 47, no. 4, pp. 407–421, 2003.

[38] T. M. Kim, "A critical study on the status and countermeasures of sexual crimes," *Korean Criminological Review*, vol. 22, no. 3, pp. 5–44, 2011.

[39] Crime and Criminal Justice Statistics, *Korean Crime Victimization Survey: KCVS*, Korean Institute of Criminology, 2012.

[40] G. H. Cho, "The influence that the alcohol affects to the human body," *Korean Association of Addiction Crime Review*, vol. 2, no. 1, pp. 25–38, 2011.

[41] P. F. D. Bau, R. S. Moraes, C. H. D. Bau, E. L. Ferlin, G. A. Rosito, and F. D. Fuchs, "Acute ingestion of alcohol and cardiac autonomic modulation in healthy volunteers," *Alcohol*, vol. 45, no. 2, pp. 123–129, 2011.

[42] H. Abe, Y. Kawano, S. Kojima et al., "Biphasic effects of repeated alcohol intake on 24-hour blood pressure in hypertensive patients," *Circulation*, vol. 89, no. 6, pp. 2626–2633, 1994.

[43] J. Rossinen, M. Viitasalo, J. Partanen, P. Koskinen, M. Kupari, and M. S. Nieminen, "Effects of acute alcohol ingestion on heart rate variability in patients with documented coronary artery disease and stable angina pectoris," *The American Journal of Cardiology*, vol. 79, no. 4, pp. 487–491, 1997.

[44] R. D. Myers, "Alcohol's effect on body temperature: hypothermia, hyperthermia or poikilothermia?" *Brain Research Bulletin*, vol. 7, no. 2, pp. 209–220, 1981.

[45] T. Danel, C. Libersa, and Y. Touitou, "The effect of alcohol consumption on the circadian control of human core body temperature is time dependent," *The American Journal of Physiology—Regulatory Integrative and Comparative Physiology*, vol. 281, no. 1, pp. R52–R55, 2001.

[46] S. Slijepcevic, M. Potkonjak, V. Tsiatsis, S. Zimbeck, and M. B. Srivastava, "On communication security in wireless Ad-hoc sensor networks," in *Proceedings of the 11th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pp. 139–144, Pittsburgh, PA, USA, 2002.

[47] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2010, Proceedings of the 4th International Workshop on Knowledge Discovery from Sensor Data, Washington, DC, USA.

[48] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.

[49] T. W. Park, J. W. Lee, I. S. Hwang, C. K. Yoo, L. Nachman, and J. H. Song, "E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems (SenSys '11)*, pp. 260–273, Seattle, Wash, USA, November 2011.

[50] Q. An, Y. Ishikawa, J. Nakagawa et al., "Evaluation of wearable gyroscope and accelerometer sensor (Pocket-IMU2) during walking and sit-to-stand motions," in *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication*, pp. 731–736, Paris, France, 2012.

[51] R. Swift and D. Davidson, "Alcohol hangover: mechanisms and mediators," *Alcohol Health and Research World*, vol. 22, no. 1, pp. 54–60, 1998.

[52] T. H. Goo, H. S. Goo, and M. Y. Park, "The characteristic of the body surface temperature change of the normal adult male according to the sleep," *The Korea Association of Mental Science Review*, vol. 19, no. 5, pp. 97–104, 2003.

[53] Etek Navigation Inc., *MTK NMEA Packet User Manual*, Revision 0.3, Etek Navigation Inc., Taiwan, China, 2006.

[54] S. K. Kim, G. H. Park, S. H. Jeon et al., "HMM-based motion recognition with 3-D acceleration signal," *Journal of Korea Institute of Information Scientists and Engineers: Computing Practices and Letters*, vol. 15, no. 3, pp. 216–220, 2009.

[55] B. Choe, J. H. Hong, and S. B. Cho, "Accumulative dynamic time warping for 3-axis accelerometer based gesture recognition," *Proceedings of the 36th Korea Computer Congress*, vol. 36, no. 1, pp. 562–565, 2009.

[56] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uWave: accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, vol. 5, no. 6, pp. 657–675, 2009.

[57] J. Wu, G. Pan, D. Zhang, G. Qi, and S. Li, "Gesture recognition with a 3-D accelerometer," in *Ubiquitous Intelligence and Computing*, vol. 5585 of *Lecture Notes in Computer Science*, pp. 25–38, Springer, New York, NY, USA, 2009.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 318407, 7 pages
http://dx.doi.org/10.1155/2013/318407

Research Article

A Utility-Based Rate Allocation of M2M Service in Heterogeneous Wireless Environments

Yao Huang,[1] Hui Tian,[1] Jie Zhang,[2] Cheng Qin,[1] and Zhibo Wang[1]

[1] *State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, P.O. Box 92, 10 Xi Tu Cheng Road, Haidian District, Beijing 100876, China*
[2] *The Communications Group, Department of Electronic and Electrical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK*

Correspondence should be addressed to Yao Huang; bupthuangyao@163.com

Received 4 June 2013; Revised 8 August 2013; Accepted 24 August 2013

Academic Editor: Lin Bai

Copyright © 2013 Yao Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a dynamic adaptive machine-to-machine (M2M) service rate allocation scheme for optimum traffic distribution in heterogeneous wireless environments (HWEs). According to the M2M service characteristics, it proposes the utility function, which forms a convex optimization problem that maximizes the utility of M2M service and can be solved by the Lagrange multiplier. The simulation results show the proposed method convergence, and it can achieve load balance with maximized throughput and minimized cost.

1. Introduction

Machine-type communication (MTC) is regarded as one of the next frontiers in wireless communications. Recently, there are growing researches about MTC due to the fact it can be widely used in many aspects, such as smart building, smart grid, and environment monitoring.

In 3GPP, it is proposed that each MTC device attaches to the existing cellular infrastructure (e.g., LTE-Advanced) [1, 2], by which higher layers connections between the MTC controller and MTC devices are provided. Thus, the subsequent problems lie in the access management on the air interface [3]. Furthermore, there are also many systems architecture difficulties about MTC [4–6]. Reference [4] identifies potential issues on the air interface. Reference [5] explores systems architecture problems that are associated with the evolving home M2M network. Reference [6] addresses the new requirement to use the M2M equipment with high configuration and flexibility to open up exciting new use cases, services, and applications.

Although MTC has been developed for years, it still faces many challenges due to the differences from the conventional human-type communications, such as low mobility, time controlled [1].

A main MTC characteristic is that there could be a large number of M2M devices existent in the network. It is possible that congestion will happen when massive M2M devices simultaneously connect to the network, as wireless network resources are limited. In order to avoid congestions, some improved measures are proposed. Access class baring [7–9] is effectively to solve plenty of M2M devices access, but the cost is not involved in [7–9]. Meanwhile, because massive M2M devices can produce a great deal of M2M service data, it is necessary to reduce the transmission cost.

The M2M service characteristics are also different for various MTC applications [10]. For example, some applications require hard timing constraints. It will be therefore dangerous when timing constraints of security monitoring are violated. Some applications demand soft timing constraints, such as metering water and gas or electricity. These M2M services can tolerate soft time delay. Therefore, the characteristics of M2M service need to be fully considered in M2M communications.

In addition, it is well known that in the future multiple radio access technologies, such as 3G/4G, WiMax, and WLAN, will coexist and form heterogeneous wireless environments (HWEs), where users have options to access the best wireless network that fits their services requirements, and network operators can increase revenue with a more

probability usage of radio resources [11–13]. Reference [11] proposes a rate allocation algorithm over different interfaces with an objective to optimize multiuser performance; the target of this algorithm is to achieve maximum throughput. An explicit adaptive traffic allocation scheme [12] is based on the cooperation of wireless wide area networks and wireless personal area network. This method can achieve internetworking load balance and minimize the whole transmission delay.

For M2M service, it also can be transmitted in HWEs. In this paper, we propose an effective rate allocation scheme for M2M service, which depends on the utility function. The utility function takes both the throughput and the cost into account. It aims to maximize the throughput and minimize the cost according to the characteristics of M2M service.

The rest of this paper is organized as follows. The M2M communications scenario is described in Section 2. Section 3 provides an optimal solution for dynamic traffic allocation and describes our simulation results, and Section 4 concludes the paper.

2. M2M Communications

2.1. M2M Communications Scenario. In a cell, there are massive M2M devices, for example, real-time video surveillance of intelligent transportation M2M systems and video replay transmission of environment monitoring M2M systems. It will result in network congestion when massive M2M devices simultaneously access the same network. If M2M service can be transmitted in multiple networks, it is known then that network congestion will be decreased. However, if an M2M device can access different radio access networks (RANs), its price will be increased, and it will consume more energy for listening to more RANs to access. In fact, it is unnecessary for each M2M device to access different RANs. Meanwhile, M2M devices do not move, move infrequently, or move only within a certain region. For certain management purposes, many M2M devices can be grouped as clusters [2].

Based on the discussion above, it is assumed that different application M2M devices can be divided into different groups, and all of the M2M devices data are sent to the access node. Thus, the access node can operate groups, M2M devices, and there will be large M2M service data on the access node.

Between M2M devices and the access node communications, many kinds of wireless technologies can be involved, such as, Wi-Fi, ZigBee, RFID, and Bluetooth. And many measures can be taken to guarantee the access of these different M2M devices groups. For example, in the back off scheme, different M2M devices groups can take different back off windows, which can effectively decrease congestion [4]. In this paper, the communications from M2M devices to the access node are not considered. And it is assumed that the access node can access different RANs and operate massive M2M devices. This assumption can reduce the M2M device price and access collisions among M2M devices.

In addition, if an M2M device data are large, it also can be viewed as the access node.

In Figure 1, An M2M communication scenario is illustrated. Node A is the access node. It means that other groups

Figure 1: An M2M communication scenario in HWEs.

of MTC devices transmit data to the node A. Therefore, the data in the noda A are large. And the node A can access multiple RANs.

For different RANs, some RANs can offer shorter data transmission delay, and some RANs can provide lower service expense. Meanwhile, different M2M services have different requirements, such as delay.

Hence, how to allocate the rate in the different RANs should be solved. For M2M service, two important attributes are observed in this paper, the throughput and the cost. $T(\lambda)$ and $C(\lambda)$ represent them, respectively. When the remaining resources of RANs are known, the proposed scheme can distribute the rate in the different RANs based on the M2M service characteristics.

Firstly, some of variables are explained as follows.

R bps is the total data rate.

L_p bit is the packet length.

λ ($\lambda = R/L_p$) is the packet arrival rate and follows the Poisson distribution.

λ_i is the allocation rate in the ith RAN, where $i = 1, \ldots, N$.

N is the number of RANs.

R_i^a bps is the available resources of RANs.

μ_i ($\mu_i = R_i^a/L_p$) is the serving rate and follows the Poisson distribution.

$T(\lambda)$ is the total gain function.

$C(\lambda)$ is the total cost function.

The data transmission through RANs is modeled as $M/M/1$ queue. According to [12], the steady-state probabilities of m packets in ith RAN can be written as

$$P_m^{\mathrm{RAN}i} = \left(\frac{\lambda_i}{\mu_i}\right)^m \left(1 - \frac{\lambda_i}{\mu_i}\right). \quad (1)$$

Consequently, the average packets L_i transmitted by the ith RAN are deduced from the following formula [12]:

$$L_i = \sum_m m\left(\frac{\lambda_i}{\mu_i}\right)^m \left(1 - \frac{\lambda_i}{\mu_i}\right) = \frac{\lambda_i}{\mu_i - \lambda_i}. \quad (2)$$

Hence, the average delay $T(\lambda_i)$ in the ith RAN can be calculated as

$$T(\lambda_i) = \frac{L_i}{\lambda_i} = \frac{1}{\mu_i - \lambda_i}. \quad (3)$$

To some extent, minimization of the transmission delay can proportionally maximize the throughput; thus, the total gain function $T(\lambda)$ is given as

$$T(\lambda) = \sum_{i=1}^{N} \frac{1}{\mu_i - \lambda_i}. \quad (4)$$

Additionally, for the ith RAN, b_i is the transmission cost, which means that the M2M service will cost b_i when a certain number of packets are transmitted by the ith RAN, and the number is assigned as the unit quantity. If the M2M packet length and the M2M service expense are known, then b_i can be obtained. For example, if the packet length is 10 bits and the service expense is 0.01 Yuan per bit, b_i can be calculated as $10 * 0.01 = 0.1$ Yuan per packet. In other words, b_i is the known information in the RAN. Meanwhile, L_i is the number of the average transmitted packets in the ith RAN. The cost in the ith RAN can be written as

$$C(\lambda_i) = L_i b_i = \frac{\lambda_i b_i}{\mu_i - \lambda_i}. \quad (5)$$

The objective $C(\lambda_i)$ is the increasing function of λ_i. Then, the total cost function $C(\lambda)$ is given as

$$C(\lambda) = \sum_{i=1}^{N} \frac{\lambda_i b_i}{\mu_i - \lambda_i}. \quad (6)$$

Let U_k represent the utility function of the M2M service k. It is calculated as

$$U_k = \beta_1 T(\lambda) + \beta_2 C(\lambda), \quad (7)$$

where β_1, β_2 are weighting factors. If the M2M service requires hard timing constraints, then $\beta_1 > \beta_2$. When $\beta_1 = 1$, $\beta_2 = 0$, the throughout is only considered in this situation. If the M2M service requires non real-time and low cost, then $\beta_1 < \beta_2$, and even $\beta_1 = 0$, $\beta_2 = 1$. In this situation, the cost is only considered. The values of both β_1 and β_2 depend on the M2M service characteristics.

Because $T(\lambda)$ means the overall delay and $C(\lambda)$ means the total cost, smaller values of both $T(\lambda)$ and $C(\lambda)$ will be better obtained in the actual system. Hence, minimizing U_k can get better utility for M2M service k.

2.2. Optimal Solution. When the utility function U_k is obtained by (7), there are also some constraints in U_k optimization.

2.2.1. Allocation Constraint 1. The delay of M2M service k is determined by the maximum delay D_k, which is the time from the MTC device to the MTC sever in all RANs. D_i is the other delay of uplink traffic towards the destination in the ith RAN. It is defined as $1/(\mu_i - \lambda_i) < D_k - D_i$.

2.2.2. Allocation Constraint 2. The total rate of M2M service distributed in N RANs is calculated by $\lambda = \sum_{i=1}^{N} \lambda_i$.

2.2.3. Allocation Constraint 3. In an $M/M/1$ queue with arrival rate of λ_i and serving rate of μ_i, it is known that $0 \le \lambda_i < \mu_i$.

The rate allocation of M2M service k should minimize (7) by satisfying these constraints. Thus, the optimization problem can be modeled as an objective with constraints as follows:

$$\min U_k$$
$$\text{s.t.} \begin{cases} 0 \le \lambda_i < \mu_i \\ \lambda = \sum_{i=1}^{N} \lambda_i \\ \dfrac{1}{\mu_i - \lambda_i} < D_k - D_i, \end{cases} \quad (8)$$

where $i = 1, 2, \ldots, N$.

To show the concavity of the objective function U_k, the following general form is observed:

$$\begin{aligned} U_k &= \beta_1 T(\lambda) + \beta_2 C(\lambda) \\ &= \beta_1 \sum_{i=1}^{N} \frac{1}{\mu_i - \lambda_i} + \beta_2 \sum_{i=1}^{N} \frac{\lambda_i b_i}{\mu_i - \lambda_i}. \end{aligned} \quad (9)$$

The first derivatives of $T(\lambda)$ and $C(\lambda)$ are

$$\begin{aligned} \frac{\partial T(\lambda)}{\partial \lambda_i} &= \frac{1}{(\mu_i - \lambda_i)^2}, \\ \frac{\partial C(\lambda)}{\partial \lambda_i} &= \frac{b_i \mu_i}{(\mu_i - \lambda_i)^2}. \end{aligned} \quad (10)$$

The second derivatives of $D(\lambda)$ and $C(\lambda)$ are

$$\begin{aligned} \frac{\partial^2 T(\lambda)}{\partial {\lambda_i}^2} &= \frac{2}{(\mu_i - \lambda_i)^3}, \\ \frac{\partial^2 C(\lambda)}{\partial \lambda_i} &= \frac{2 b_i \mu_i}{(\mu_i - \lambda_i)^3}. \end{aligned} \quad (11)$$

As long as $\mu_i > \lambda_i$, $\partial T(\lambda)/\partial \lambda_i > 0$, $\partial C(\lambda)/\partial \lambda_i > 0$, $\partial^2 T(\lambda)/\partial {\lambda_i}^2 > 0$, $\partial^2 C(\lambda)/\partial \lambda_i > 0$, then $\partial U_k/\partial \lambda_i > 0$, $\partial^2 U_k/\partial \lambda_i^2 > 0$. Therefore $U_k = \beta_1 T(\lambda) + \beta_2 C(\lambda)$ is a convex

function in λ_i. The Lagrange function can be written as follows:

$$L(\lambda_i,\theta,\nu_i) = \beta_1\sum_{i=1}^{N}\frac{1}{\mu_i-\lambda_i} + \beta_2\sum_{i=1}^{N}\frac{\lambda_i b_i}{\mu_i-\lambda_i} - \theta\left(\sum_{i=i}^{N}\lambda_i - \lambda\right) - \sum_{i=1}^{N}\nu_i\left(\mu_i - \frac{1}{D_k-D_i} - \lambda_i\right) - \sum_{i=1}^{N}w_i(\mu_i-\lambda_i). \tag{12}$$

According to $1/(\mu_i-\lambda_i) < D_k-D_i$, it can be calculated that $\lambda_i < \mu_i-(1/(D_k-D_i))(1/(D_k-D_i) > 0)$. Meanwhile, $\lambda_i < \mu_i$. That is to say that, if $\lambda_i < \mu_i-(1/(D_k-D_i))$, it is known that $\lambda_i < \mu_i$. Hence, the Lagrange function can be expressed as

$$L(\lambda_i,\theta,\nu_i) = \beta_1\sum_{i=1}^{N}\frac{1}{\mu_i-\lambda_i} + \beta_2\sum_{i=1}^{N}\frac{\lambda_i b_i}{\mu_i-\lambda_i} - \theta\left(\sum_{i=i}^{N}\lambda_i - \lambda\right) - \sum_{i=1}^{N}\nu_i\left(\mu_i - \frac{1}{D_k-D_i} - \lambda_i\right). \tag{13}$$

In order to deal with the Lagrange function, the Karush-Kuhn-Tucker (KKT) conditions [14] are used. Then, (14), (15), (16), and (17) are obtained as follows:

$$\frac{\beta_1}{(\mu_i-\lambda_i)^2} + \frac{\beta_2 b_i\mu_i}{(\mu_i-\lambda_i)^2} - \theta + \nu_i = 0, \tag{14}$$

$$\mu_i - \frac{1}{D_k-D_i} - \lambda_i > 0, \tag{15}$$

$$\nu_i\left(\mu_i - \frac{1}{D_k-D_i} - \lambda_i\right) = 0, \tag{16}$$

$$\sum_{i=i}^{N}\lambda_i - \lambda = 0, \tag{17}$$

where $i = 1,\ldots,N$. Because $\mu_i - 1/(D_k-D_i) - \lambda_i > 0$, we can get $\nu_i = 0$. Then (14) can be expressed as

$$\frac{\beta_1}{(\mu_i-\lambda_i)^2} + \frac{\beta_2 b_i\mu_i}{(\mu_i-\lambda_i)^2} - \theta = 0, \quad \text{thus } \theta > 0. \tag{18}$$

The first derivative of $L(\lambda_i,\theta,\nu_i)$ about λ_i is

$$\frac{\partial L}{\partial\lambda_i} = \frac{\beta_1}{(\mu_i-\lambda_i)^2} + \frac{\beta_2 b_i\mu_i}{(\mu_i-\lambda_i)^2} - \theta. \tag{19}$$

The second derivative of $L(\lambda_i,\theta,\nu_i)$ about λ_i is

$$\frac{\partial^2 L}{\partial\lambda_i^2} = \frac{2\beta_1}{(\mu_i-\lambda_i)^3} + \frac{2\beta_2 b_i\mu_i}{(\mu_i-\lambda_i)^3}. \tag{20}$$

Let $f(\lambda_i^k)$ and $f'(\lambda_i^k)$ represent $\partial L/\partial\lambda_i^k$ and $\partial^2 L/\partial(\lambda_i^k)^2$ respectively, where k represents kth iteration. It has been proved that the objective function (9) is concave and that $f(\lambda_i^k)$ is increasing in λ_i with fixed $\beta_1,\beta_2,b_i,\mu_i$ for $f'(\lambda_i^k) > 0$.

(1) Collect the available resources information from each RAN.
(2) If $k = 0$, then
(3) Initialize $\lambda_i^0 = 0$ and $\theta^0 = 0$
(4) Else
(5) Calculate using Newton's method.
$$\lambda_i^{k+1} = \lambda_i^k - \frac{f(\lambda_i^k)}{f'(\lambda_i^k)}$$
(6) Updated θ^{k+1} according to (22)
$$\theta^{k+1} = \left[\theta^k + \varepsilon\left(\sum_{i=1}^{N}\lambda_i^k - \lambda\right)\right]^+$$
(7) $k = k + 1$. Go to step (5) until the difference between λ_i^{k+1} and λ_i^k values are less than a minimum, such as 0.000001.
(8) End if
(9) After the algorithm converges the rate is allocated according to λ_i^{k+1}.

Algorithm 1

Meanwhile, $f''(\lambda_i^k) > 0$. Therefore, we can use the Newton iteration method [15] to get the converged optimal value of λ_i. Regardless of initial values λ_i^0, the last convergence value λ_i^k will satisfy (8). Then, a better approximation λ_i^{k+1} can be written as

$$\lambda_i^{k+1} = \lambda_i^k - \frac{f(\lambda_i^k)}{f'(\lambda_i^k)}. \tag{21}$$

The updated θ^k value for rate allocation is calculated by

$$\theta^{k+1} = \left[\theta^k + \varepsilon\left(\sum_{i=1}^{n}\lambda_i^k - \lambda\right)\right]^+, \tag{22}$$

where $\varepsilon\,(\varepsilon > 0)$ is a constant step size and $[\theta^k]^+ = \max(\theta^k, 0)$.

The iteration will continue until the difference between λ_i^{k+1} and λ_i^k is less than a minimal value, which can be predefined by the algorithm. Because λ_i is the convergence value, λ_i^k will more approach the actual value if the number of iterations is more. In other words, the minimal value controls the accuracy. The smaller the minimal value, the higher the accuracy.

Then, the allocation rate λ_i among RANs is obtained based on (21) and (22), and it is the global optimal results of the convex optimization problem for (8).

2.3. Algorithm Flow. In conclusion, we use an iterative algorithm to find the solution of optimization problem, and it is shown in Algorithm 1.

3. Simulation Results

In this section, the performances of the proposed traffic allocation scheme are evaluated. In the numerical simulation, two different M2M services are investigated. One M2M service requires hard timing constraints; $\beta_1 = 0.9$ and

Figure 2: Two M2M services in HWEs with $b = [1 \quad 1 \quad 1]$.

Figure 3: Two M2M services in HWEs with $b = [1 \quad 4 \quad 8]$.

$\beta_2 = 0.1$ are defined. The other M2M service requires soft timing constraints and low cost; $\beta_1 = 0.1$ and $\beta_2 = 0.9$ are defined. The data rate R is 6 Mbps for both M2M services.

Meanwhile, there are three RANs. It is assumed that the available resources are 4 Mbps for each RAN. For comparison, two different cost combinations are defined for these three RANs. The first cost combination is $b = [1 \quad 1 \quad 1]$ for RANs 1, 2, and 3, which means that the M2M service will consume 1 when 1M packets are, respectively, transmitted by RANs 1, 2, and 3. The second combination is $b = [1 \quad 4 \quad 8]$ for RANs 1, 2, and 3.

In addition, the radio channel environment is good to evaluate the performance of the proposed algorithm.

3.1. Two M2M Services in HWEs. Figures 2 and 3 show two M2M services in the same HWEs, respectively. Figures 4 and 5 show the same M2M service in the different HWEs, respectively. In Figure 2, the allocated rates are almost equal in all RANs because the cost of each RAN is 1. This is also consistent with the actual situation.

As mentioned above, the hard timing constraints M2M service pays more attention to the delay, and the soft timing constraints M2M service focuses on the cost. Hence, the allocated rates are different when the cost of RANs is charged in Figures 3, 4, and 5. Meanwhile, for the soft timing constraints M2M service, because 1 is the lowest cost of RANs, the allocated rate is the highest in the three RANs.

Based on Figures 2, 3, 4, and 5, because the two different M2M services have different characteristics, the convergence values are different in the three RANs by the Newton method. And these figures indicated that the proposed scheme is effective in rate allocation for M2M service.

3.2. The Cost and Delay Comparison. In Figure 6, the maximum delay of soft timing constraints M2M service is higher

Figure 4: Soft timing constraints in HWEs with $b = [1 \quad 1 \quad 1]$ and $b = [1 \quad 4 \quad 8]$.

than that of hard timing constraints M2M service. For hard timing constraints M2M service, because it pays more attention to the delay, the delays are equal in HWEs with the different cost.

In Figure 7, the sum cost of soft timing constraints M2M service is lower than that of hard timing constraints M2M service. Figures 6 and 7 indicate that the hard timing constraints M2M service requires lower delay, while the soft timing constraints M2M service focuses on lower cost. In addition, because $b = [1 \quad 1 \quad 1]$, the cost of hard timing constraints

Figure 5: Hard timing constraints in HWEs with $b = [1\ 1\ 1]$ and $b = [1\ 4\ 8]$.

Figure 6: The delay comparison.

Figure 7: The cost comparison.

Figure 8: The utility comparison between two schemes.

M2M service is almost the same with the cost of soft timing constraints M2M service.

3.3. Algorithm Comparison. In order to validate the proposed rate allocation scheme, the commonly used load balancing scheme [12] is briefly given as comparison as shown in Figure 8.

It should be noted that lower value means better utility. As is shown from Figure 8, the proposed scheme can get better performance than the compared scheme, when the M2M service requires longer delay and lower cost. Meanwhile, the values of the utility are almost equal when $b = [1\ 1\ 1]$ and the M2M service requires hard timing constraints, because the distributed rates are the similar values as depicted in Figure 2.

4. Conclusions

In the future M2M communications, M2M service may be transmitted in HWEs. The rate allocation for M2M service among multi-RANs is an important and challenging research. In this paper, an adaptive rate allocation scheme is proposed based on the utility function. According to the characteristics of M2M service, the utility function may be different. And the rate is distributed based on the utility function in HWEs to achieve higher throughput and lower cost. The simulation shows that it is an effective one compared with traditional method.

Acknowledgments

This work was sponsored by the Projects 61121001 and 60971125 supported by the Natural Science Foundation

of China, the Project IRT1049 supported by the Program for Changjiang Scholars and Innovative Research Team in University, Chinese Highway Institution Project 2012ZX03005010-003, and China Scholarship Council.

References

[1] "Service requirements for machine-type communications," 3GPP TS 22. 368 V11. 4. 0, 2012.

[2] "System improvement for machine-type communications," 3GPP TR 23. 888 V1. 6. 1, 2012.

[3] S.-Y. Lien and K.-C. Chen, "Massive access management for QoS guarantees in 3GPP machine-to-machine communications," *IEEE Communications Letters*, vol. 15, no. 3, pp. 311–313, 2011.

[4] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, 2011.

[5] M. Starisinic, "System architecture challenges in the home M2M network," in *Proceedings of the Applications and Technology Conference (LISAT '10)*, pp. 1–7, IEEE press, Farmingdale, NY, USA, May 2010.

[6] I. Cha, Y. Shah, A. U. Schmidt, A. Leicher, and M. Meyerstein, "Addressing new security threats," *IEEE Vehicular Technology Magazine*, vol. 4, no. 3, pp. 69–75, 2009.

[7] CATT, "R2-100182: access control of MTC devices," 3GPP TSG RAN WG2 Meeting 68bis, 2010.

[8] ZTE, "R2-104662: MTC simulation results with specific solutions," 3GPP TSG RAN WG2 Meeting 71, 2010.

[9] S.-Y. Lien, T.-H. Liau, C.-Y. Kao, and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 27–32, 2012.

[10] "Service requirements for machine-type communications," 3GPP TS 22. 368 V10. 1. 0, 2010.

[11] W. Fu and D. P. Agrawal, "Multi-connection and rate allocation in heterogeneous wireless networks," in *Proceedings of the IEEE Global Communications Conference Workshops (GLOBECOM '10)*, pp. 1157–1161, Miami, Fla, USA, December 2010.

[12] L. Sun, H. Tian, Q.-Y. Sun, D.-M. Shen, and P. Zhang, "Traffic allocation scheme with cooperation of WWAN and WPAN," *IEEE Communications Letters*, vol. 14, no. 6, pp. 551–553, 2010.

[13] E. Z. Tragos, G. Tsiropoulos, G. T. Karetsos, and S. A. Kyriazakos, "Admission control for QoS support in heterogeneous 4G wireless networks," *IEEE Network*, vol. 22, no. 3, pp. 30–37, 2008.

[14] B. Chen, *Optimization Theory and Algorithms*, Tsinghua University Press, Beijing, China, 2nd edition, 2005.

[15] J. Stoer, R. Bulirsch, R. Bartels, W. Gautshi, and C. Witzgall, *Introduction to Numerical Analysis*, Springer, New York, NY, USA, 3rd edition, 2002.