International Journal of Distributed Sensor Networks

Selected Papers from "The 6th China Conference of Wireless Sensor Networks"

Guest Editors: Fu Xiao, Ruchuan Wang, Limin Sun, Qun Li, Hongwei Zhang, and Liusheng Huang

Selected Papers from "The 6th China Conference of Wireless Sensor Networks"

International Journal of Distributed Sensor Networks

Selected Papers from "The 6th China Conference of Wireless Sensor Networks"

Guest Editors: Fu Xiao, Ruchuan Wang, Limin Sun, Qun Li, Hongwei Zhang, and Liusheng Huang

Copyright © 2013 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

Prabir Barooah, USA
Richard R. Brooks, USA
Periklis Chatzimisios, Greece
W.-Y. Chung, Republic of Korea
George P. Efthymoglou, Greece
Frank Ehlers, Italy
Yunghsiang S. Han, Taiwan
Tian He, USA
Baoqi Huang, China
Chin-Tser Huang, USA
S. S. Iyengar, USA
Rajgopal Kannan, USA
Miguel A. Labrador, USA
Joo-Ho Lee, Japan
Minglu Li, China
Shijian Li, China
Shuai Li, USA
Jing Liang, China
Weifa Liang, Australia
Wen-Hwa Liao, Taiwan
Alvin S. Lim, USA
Zhong Liu, China
Donggang Liu, USA
Yonghe Liu, USA
Seng Loke, Australia
Jun Luo, Singapore
J. R. Martinez-de Dios, Spain
Shabbir N. Merchant, India
Aleksandar Milenkovic, USA
Eduardo Freire Nakamura, Brazil
Peter Csaba Ölveczky, Norway
Marimuthu Palaniswami, Australia
Shashi Phoha, USA
Hairong Qi, USA
Joel Rodrigues, Portugal
Jorge Sa Silva, Portugal
Sartaj K. Sahni, USA
Weihua Sheng, USA
Zhi Wang, China
Sheng Wang, China
Andreas Willig, New Zealand
Qishi Wu, USA
Qin Xin, Norway
Jianliang Xu, Hong Kong
Yuan Xue, USA
Fan Ye, USA
Ning Yu, China
Tianle Zhang, China
Yanmin Zhu, China

Contents

Selected Papers from "The 6th China Conference of Wireless Sensor Networks", Fu Xiao, Ruchuan Wang, Limin Sun, Qun Li, Hongwei Zhang, and Liusheng Huang
Volume 2013, Article ID 746171, 3 pages

Stair Scheduling for Data Collection in Wireless Sensor Networks, Jinbiao Chen, Yongcai Wang, Yuexuan Wang, and Changjian Hu
Volume 2013, Article ID 124786, 11 pages

L-Shaped-Sensor-Array-Based Localization and Tracking Method for 3D Maneuvering Target, Xing Zhang and Xue Wang
Volume 2013, Article ID 741284, 8 pages

An Adaptive Opportunistic Network Coding Mechanism in Wireless Multimedia Sensor Networks, Hang Shen, Guangwei Bai, Lu Zhao, and Zhenmin Tang
Volume 2012, Article ID 565604, 13 pages

Interference-Aware Fault-Tolerant Energy Spanner in Wireless Ad Hoc Networks, Dong Wang, Wang-Chen Long, and Xiao-Hong Li
Volume 2012, Article ID 235374, 8 pages

Probability Model Based Coverage-Enhancing Algorithm for WSNs of Nodes' Adjustable Movement Pattern, Junjie Huang, Haiping Huang, Peng Sun, Ruchuan Wang, and Xun Wei
Volume 2012, Article ID 570643, 10 pages

MPD-Model: A Distributed Multipreference-Driven Data Fusion Model and Its Application in a WSNs-Based Healthcare Monitoring System, Jibing Gong, Li Cui, Kejiang Xiao, and Rui Wang
Volume 2012, Article ID 602358, 13 pages

Optimal Convergecast Scheduling Limits for Clustered Industrial Wireless Sensor Networks, Xiaoling Zhang, Wei Liang, Haibin Yu, and Xisheng Feng
Volume 2012, Article ID 891321, 12 pages

Data Gathering in Opportunistic Wireless Sensor Networks, Yongxuan Lai and Ziyu Lin
Volume 2012, Article ID 230198, 13 pages

An Open Conformance Test System towards the Standardization of Wireless Sensor Networks, Zhonghua Zhao, Wei Huangfu, Linmin Sun, Zhiqiang Shi, and Wei Gan
Volume 2012, Article ID 931869, 15 pages

A Camera Nodes Correlation Model Based on 3D Sensing in Wireless Multimedia Sensor Networks, Chong Han, Lijuan Sun, Fu Xiao, Jian Guo, and Ruchuan Wang
Volume 2012, Article ID 154602, 9 pages

A Stochastic *k*-Coverage Scheduling Algorithm in Wireless Sensor Networks, Jiguo Yu, Shaohua Ren, Shengli Wan, Dongxiao Yu, and Guanghui Wang
Volume 2012, Article ID 746501, 11 pages

Truthful Relay Assignment for Cooperative Communication in Wireless Networks with Selfish Source-Destination Pairs, Gang Liu, Liusheng Huang, Yu-e Sun, Hongli Xu, He Huang, and Xueyong Xu
Volume 2012, Article ID 345316, 14 pages

Novel Side Information Generation Algorithm of Multiview Distributed Video Coding for Multimedia Sensor Networks, Fu Xiao, Jinkai Liu, Jian Guo, and Linfeng Liu
Volume 2012, Article ID 582403, 7 pages

A Virtual-Ring-Based Data Storage and Retrieval Scheme in Wireless Sensor Networks, Xingpo Ma, Jianliang Gao, Weiping Wang, and Jianxin Wang
Volume 2012, Article ID 763015, 10 pages

Target Q-Coverage Problem with Bounded Service Delay in Directional Sensor Networks, Deying Li, Hui Liu, Xianling Lu, Wenping Chen, and Hongwei Du
Volume 2012, Article ID 386093, 10 pages

Two-Layer Storage Scheme and Repair Method of Failure Data in Wireless Sensor Networks, Yulong Shen, Xiaowei Dang, Min Shu, Ning Xi, and Jianfeng Ma
Volume 2012, Article ID 868943, 9 pages

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 746171, 3 pages
http://dx.doi.org/10.1155/2013/746171

Editorial

Selected Papers from "The 6th China Conference of Wireless Sensor Networks"

Fu Xiao,[1] Ruchuan Wang,[1] Limin Sun,[2] Qun Li,[3] Hongwei Zhang,[4] and Liusheng Huang[5]

[1] *School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2] *Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*
[3] *Department of Computer Science, College of William and Mary, Williamsburg, VA 23187, USA*
[4] *Department of Computer Science, Wayne State University, Detroit, MI 48202, USA*
[5] *School of Computer Science and Technology, University of Science and Technology of China, Hefei 230022, China*

Correspondence should be addressed to Ruchuan Wang; wangrc@njupt.edu.cn

Received 30 January 2013; Accepted 30 January 2013

Copyright © 2013 Fu Xiao 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.

Integrating the capabilities of distributed sensing, computing, and communication, wireless sensor networks (WSNs) have been transforming the way we interact with the physical and human worlds. While traditional WSNs (e.g., those for smart vineyard) have gone from research to real-world deployments, WSNs continue evolving along multiple dimensions such as content, mobility, space, and architecture. For instance, wireless multimedia sensor networks, mobile wireless sensor networks, wireless underwater sensor networks, wireless underground sensor networks, and wireless sensor-actuator networks are being explored to extend WSNs for supporting rich-media sensing, mobile sensing, underwater/underground sensing, and closed-loop sensing and control. These new WSNs will significantly expand the reach of WSNs and have far-reaching impact on WSN applications. In the meantime, they also introduce great challenges to the theory and practice of wireless sensor networking, together with The 6th China Conference of Wireless Sensor Networks (CWSN2012) sponsored by China Computer Federation, cosponsored by China Computer Federation Technical Committee on Sensor Network, and organized by Nanjing University of Posts and Telecommunication. CWSN2012 serves as a forum for sensor network researchers, developers, and users to communicate the productions and experience of sensor network research and applications. Authors of selected English papers in CWSN2012 were invited to submit their extended versions of their papers to this special issue. Also, some research articles were submitted from the manuscript tracking system as regular submissions to this special issue. This special issue aims at providing researchers and practitioners from the academia, industry, and government an opportunity to reflect upon these new developments of WSNs, including their architectures, protocols, novel characteristics, challenges, and real-world deployments. We have accepted a few papers that address the state of the art and the future directions of these research and application areas for wireless sensor networks in this issue.

The paper "*Two-layer storage scheme and repair method of failure data in wireless sensor networks*" describes a two-layer distributed data storage scheme in wireless sensor networks. The proposed storage scheme and repair method has lower repair communication overhead, which can reduce the repair communication overhead to times and is suitable for resource-constrained distributed wireless sensor networks.

The paper "*Target Q-coverage problem with bounded service delay in directional sensor networks*" addresses the target Q-coverage (TQC) problem to prolong the network lifetime with bounded service delay constraint in directional sensor networks. A protocol to find a collection of coverage sets that satisfy the coverage quality requirement and the bounded service delay constraint is proposed, and simulation results show that performance of this protocol is close to the upper bound of the optimal solution.

The paper "*A virtual-ring-based data storage and retrieval scheme in wireless sensor networks*" proposes a virtual-ring-based data storage and retrieval scheme, which is called VRS,

to take the frequencies of event and query into consideration for wireless sensor networks. In VRS, the whole sensor network field is divided into some virtual rings. According to the frequencies, one of the virtual rings is selected as the rendezvous ring, which plays the role of a bridge between the information consumers and the information producers.

The paper "*Novel side information generation algorithm of multiview distributed video coding for multimedia sensor networks*" proposes a multiview distributed video coding algorithm by gaining the intense motion regions of the Wyner-Ziv frame according to the criteria of ROI, and for nonintense motion regions, motion compensation interpolation (MCI) is utilized to generate side information. Finally, side information based on fusion of temporal and spatial side information will be gained.

The paper "*A stochastic k-coverage scheduling algorithm in wireless sensor networks*" focuses on the K-coverage scheduling problem to guarantee coverage sensing and network connectivity. Deterministic and stochastic sensing models of the sensors were both considered and the authors adapted the results of deterministic sensing model to solve the sensor scheduling problem under the stochastic sensing model. Theoretical analysis and simulation results show that this algorithm can reduce the number of active nodes and extend the network lifetime significantly.

The paper "*Truthful relay assignment for cooperative communication in wireless networks with selfish source-destination pairs*" aims at propose a relay assignment protocol (RA-VCG) for cooperative communication to maximize the total social value while guaranteeing truthfulness in an auction-theoretic sense by charging each pair an extra payment. Simulation results demonstrate the effectiveness of this protocol.

The paper "*A camera nodes correlation model based on 3D sensing in wireless multimedia sensor networks*" studies the visual information retrieved from adjacent camera nodes that usually exhibits high levels of correlation for wireless multimedia sensor networks. Firstly, a correlation model is proposed by measuring the intersection area of multiple camera nodes' field of views. Then, cluster structures are established to cooperate on image processing and transmission tasks. A set of experiments are performed to show the proposed network topology and image fusion, and transmission scheme released the pressure of camera node greatly and reduced the network energy consumption of communication efficiently.

The paper "*An open conformance test system towards the standardization of wireless sensor networks*" analyzes IPv6-based WSN protocols and the conformance testing techniques and methods for IPv6-based WSNs. A novel conformance test system for IPv6-based WSNs is designed and implemented, which is open, flexible, full featured, and practical. The related outcomes will promote the standardization and commercialization of WSNs.

The paper "*A data gathering in opportunistic wireless sensor networks*" studies the challenging problem to effectively collect the sensing data in opportunistic wireless sensor networks, and an efficient data gathering algorithm based on location prediction is proposed for opportunistic wireless sensor networks. Extensive experimental results show that the proposed algorithm is effective to reduce the message transmissions and improve the data collection coverage rate.

The paper "*Optimal convergecast scheduling limits for clustered industrial wireless sensor networks*" investigates the performance bounds of the convergecast scheduling. Firstly, by establish the lower bounds on the number of time slots to finish the intracluster and the intercluster convergecast transmissions. Then, the lower bounds on the number of channels based on the lower bounds on the number of timeslots and maximum available channels in a multichannel scenario is established. Lastly, extensive analysis-taking packet retransmissions is carried out to meet the reliability requirement. Experiment results validate the correctness and tightness of theory analysis is this paper.

The paper "*MPD-model: a distributed multipreference-driven data fusion model and its application in a WSNs-based healthcare monitoring system*" proposes a novel distributed multipreference-driven data fusion model (MPD-Model) for WSNs using distributed multipreference feature-level fusion algorithm, light-weight adaptive feature extraction algorithm, and SVM-based algorithm for health status detection tasks. We implement the proposed methods in a real wearable healthcare monitoring system and validate the proposed methods using real-world data sets with 2046 pulse samples. Experimental results show that the proposed MPD-Model is reasonable and effective.

The paper "*Probability model based coverage-enhancing algorithm for WSNs of nodes' adjustable movement pattern*" proposed a novel algorithm named as PRMCA (probability model based rotate or move along fixed direction coverage-enhancing algorithm), which determines whether the sensor node rotates or moves along a fixed direction according to the coverage effect. A set of simulation experiments verify the performance of the proposed algorithm.

The paper "*Interference-aware fault-tolerant energy spanner in wireless ad hoc networks*" addresses the power assignment in wireless ad hoc networks which induced communication graph that meets the following properties: (1) it is an energy-t-spanner which is energy efficient; (2) it is k-fault resistant which can withstand up to node failures; (3) the interference is minimal. Both the theoretic analysis and the simulations in the paper prove that these algorithms can induce a k-fault resistant energy spanner and furthermore the interference is minimized.

The paper "*An adaptive opportunistic network coding mechanism in wireless multimedia sensor networks*" proposed an adaptive opportunistic network coding mechanism (AONC) to improve video communication performance in wireless multimedia sensor networks. A novel asymmetric coding method to improve data exchange gain is proposed, and opportunistic forwarding strategy based on dynamic priority to achieve much higher is designed. Simulation results demonstrate that AONC can greatly enhance video transmission quality and efficiently utilize bandwidth and energy resources.

The paper "*L-shaped-sensor-arrays-based localization and tracking method for 3D maneuvering target*" analyzes localization and tracking technology for three-dimensional target in wireless sensor networks. A novel closed-loop system to

detect 3D maneuvering targets is designed, which firstly uses L-shaped sensor arrays to sample the signals of maneuvering targets, then 2D ESPRIT algorithm and maximum likelihood algorithm are introduced to achieve the positions of the spatial targets; thirdly, an autoregressive (AR) particle filter (PF) algorithm is realized to predict the locations in next moment; finally, localization process is directed by predicted positions to form a positive feedback closed-loop. Experiment results show that robustness and accuracy of localization and tracking for three-dimensional targets can be enhanced using this system.

The paper "*Stair scheduling for data collection in wireless sensor networks*" addresses the low-tier stair scheduling method problem for two-tiered sensor networks. The energy and communication reliability models of stair scheduling with respect to the size of clusters and LN parameters are analyzed and used to optimize the tier design of the network. The simulation results show that the performances of stair scheduling are much better than that of random scheduling algorithms.

Fu Xiao
Ruchuan Wang
Limin Sun
Qun Li
Hongwei Zhang
Liusheng Huang

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 124786, 11 pages
http://dx.doi.org/10.1155/2013/124786

Research Article

Stair Scheduling for Data Collection in Wireless Sensor Networks

Jinbiao Chen,[1] Yongcai Wang,[1] Yuexuan Wang,[1] and Changjian Hu[2]

[1] *Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing 100084, China*
[2] *NEC Laboratories, Beijing, China*

Correspondence should be addressed to Yongcai Wang; wangyc@tsinghua.edu.cn

Received 19 August 2012; Accepted 15 January 2013

Academic Editor: Liusheng Huang

Copyright © 2013 Jinbiao 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.

Spatially organized clusters are basic structure for large-scale wireless sensor networks. A cluster is generally composed by a large amount of energy-limited low-tier nodes (LNs), which are managed by a powerful cluster head (CH). The low-tier nodes that are close to the cluster head generally become bottlenecks in data collection applications. Energy efficient scheduling is important for the low-tier sensors to be longevous while guaranteeing reliable communication. In this paper, based on three aspects of performance considerations including network longevity, multihop communication reliability, and sensing system cost minimization, we propose a stair duty-cycle scheduling method for the low-tier sensors. It is designed to make the LNs in the same cluster sleep cooperatively for most of the time and wake up in assigned sequence for multihop communication. Stair scheduling cannot only improve the energy efficiency of the network but also guarantee high communication reliability and low transmission delay. Efficiency of the proposed stair scheduling is verified by analysis and intensive simulations. The results show that the performances of stair scheduling are much better than that of random scheduling algorithms.

1. Introduction

Large-scale sensor networks are attracting great research interests, because they are promising in various applications such as precision agriculture and environment monitoring [1]. In order to cover the broad area of interest where information should be monitored, such a sensor network often contains thousands or tens of thousands of small and energy limited sensors. If these sensors are directly managed by the base station, the network will suffer large communication overhead, energy inefficiency problems, and unreliability multihop communication problems. Clustering scheme was proposed by researchers to organize the sensors into two-tiered structure [2]. In the higher tier, some energy-rich sensors are deployed as backbones to organize the energy-limited sensors within their geographic neighborhood to form clusters. In the lower tier, the energy-limited sensors capture, encode, and transmit relevant information of the designated area to the cluster head (CH). Since the CHs are rich in resources, the system performances are mainly determined by the lower tier nodes (LNs.)

Sleeping and scheduling technique is the main solution to conserve energy of the LNs. As reported in INSIGHT [3], an LN works in "sleep" state in terms of radio OFF, sensor OFF, and with HPL management can save energy up to two magnitudes than the basic "listen" state. This suggests putting LNs into sleep state for most of the time and only wake them up in periodical short slot for data capturing and transmitting. But because the LNs do not serve in sleep state, this leads to an LN scheduling problem for both energy saving and QOS preserving.

Many single-hop LN scheduling schemes were proposed. In single-hop cluster, every LN transmits data directly to the CH. The LNs need not forward messages from other nodes, so they can turn to the sleep state independently to save energy [3, 4]. In [4], high density LNs are deployed in each cluster, and a linear-distance-based scheduling has been used to define the sleep schedule of the LNs. In [5], small portion of LNs are scheduled to be activated among redundant deployed SNs for coverage preserving. In [6], LNs are scheduled based on analytical hierarchy process which considers residual energy, sensing coverage overlapping, and

so forth. However, single-hop cluster is limited in scale. Many expensive CHs will be needed to cover a broad land.

Multi-hop cluster in which LNs transmit data along a multi-hop route towards the CH is proposed by the researchers of [7, 8], which has better scalability and is more cost-efficient. But the LN scheduling problem becomes difficult in a multi-hop scenario. The LNs can no longer sleep independently, because if a relay node turns to sleep, the multi-hop route which it serves will be shut down. To support multi-hop communication, joint LN scheduling becomes necessary. In [9], distributed data gathering scheduling in multihop sensor networks was proposed by using greedy algorithm to extract a rooted spanning tree. Their work focuses on routing. It did not address jointly scheduling problem for reliable communication. In [8], hops-based sleep scheduling algorithm was proposed to assign different active probabilities to different hop LNs to balance the energy consumption. In [10], localized probabilistic routing algorithm was proposed for optimizing network lifetime. However, in their studies, the communication was assumed to happen round by round. How the LNs are jointly scheduled for multi-hop communication was not explicitly discussed.

We propose a stair scheduling method in this paper. Its basic idea is to schedule the LNs in one cluster to sleep and work cooperatively to conserve energy as well as to support multi-hop communication. With this purpose, in stair scheduling, we act as follows.

(1) We propose "Stair Scheduling" to control the working slots of the LNs based on the level-based routing tree [8]. The cluster is divided into levels, and a child node in the ith level always transmits data to a parent in the $(i-1)$th level. The child always activates and sleeps one slot earlier than its parent. Each slot contains m slices, and an LN randomly selects one slice to transmit data in its active transmission slot for collision avoidance.

(2) hop-by-hop time synchronization is proposed to avoid time drifting and make the "Stair Scheduling" work in a fully distributed manner.

(3) Average function hop-by-hop data aggregation is proposed to further enhance the energy saving as well as to balance the energy consumption.

We further analyze the energy and communication reliability models of the joint scheduled cluster with respect to the cluster size and LN parameters. The results are applied to optimize the design of the network to choose a suitable number of CHs for system cost minimization. Simulation results verify the energy efficiency and communication reliability of stair scheduling and further show that it is more efficient and applicable than traditional single-hop random scheduling (SRS) and multi-hop random scheduling (MRS).

The rest of this paper is organized as follows. In Section 2, we describe the system model and formulate the problem. In Section 3, we develop the stair scheduling algorithm. Section 4 presents the energy and reliability models of the joint scheduled cluster and the network optimization results using stair scheduling. Section 5 summarizes the simulation results and the paper is concluded in Section 6, with remarks and future work discussions.

2. System Model and Problem Formulation

2.1. Assumption and Network Model. The basic assumptions and network model are outlined as follows.

(i) We consider two-tiered sensor network, which contains low-tier energy-limited nodes (LNs) and high-tier cluster heads (CHs). The LNs are not redundant. They are deployed in an economic way for exact land coverage. Each node is responsible for the monitoring of its own vicinity. They capture and transmit data periodically. The sampling frequency is denoted as U.

(ii) The CHs do not sense data but receive and aggregate raw data from LNs and report the result to the base station. The CH has enough energy and is never considered to be a bottleneck.

(iii) The cluster area is in circular shape with radius R and centered by the CH. LNs are evenly distributed in the cluster area. All the LNs have the same sensing range r_s, maximum transmission range r_c, and the initial energy E, and we assume $r_c = 2r_s$.

(iv) Each LN generates l-bit data in a period and transmits the data towards the CH using a level-base routing tree [8, 11]. Symmetry link is assumed in data forwarding that if A can hear B, then B can hear A.

2.2. Problem Formulation. We consider large-scale sensor monitoring system, such as habitat or agriculture monitoring. In these systems, application requires that each LN reports its vicinity's information U times per hour. This prohibits an LN from sleeping for an arbitrary long time. Every LN must be active periodically. Since only CH reports results to the base station, the sampling frequency of LNs should be the same with the reporting frequency of the CH. Therefore, U is denoted as the *sampling frequency* of the cluster, and the *sampling period* is $T = 1/U$.

For an LN, each sampling period can be further divided into *task slots*. In each task slot, the LN can choose to sleep, sense, process data, receive message, or transmit message. We omit sensing and data processing time. Therefore, in a task slot, the LN either sleeps, receives, nor transmits message. We suppose that each slot is equal in length and denote the slot length by T_m. Each sample period is therefore divided into $M = T/T_m$ task slots. The joint scheduling problem becomes slot state assignment problem with joint consideration of the system performance. Formally, in a sample period, the ith LN has slots $\mathbb{S}^i = \{S_1, S_2, \ldots, S_M\}$, and the task of scheduling is to determine the state of each slot.

If a cluster has N LNs, each sampling period contains M task slots, each slot can be assigned v states, and the solution space of the joint scheduling problem will be v^{MN}. We can see that it encounters a combination explosion problem with respect to v, M, and N. In addition, because an LN is commonly not aware of other LNs' status, the joint scheduling problem is challenging.

But a heuristic is that, for energy saving and in order to support multi-hop communication, the LNs in one cluster should sleep for most of the time and should have some overlaps in their active time slots to exchange message. In view of this, we propose a multi-hop joint scheduling (stair scheduling) algorithm. The details will be given in the next section.

3. Multihop Joint Scheduling

3.1. Overview of Stair Scheduling. The basic idea of stair scheduling is to schedule the LNs in one cluster to sleep and work cooperatively to conserve energy as well as to support multi-hop communication. The design of joint scheduling basically contains three schemes as follows.

(1) "Stair scheduling" is proposed to assign every LN three continuous active slots in a sample period: "R-Slot" to listen to child nodes, "T-Slot" to sense and forward data, and "Syn-Slot" to synchronize time. A child node will be always activated one slot earlier than its parent, so that its forwarding can be heard by its parent. Therefore, most energy is conserved by sleeping and reliable multi-hop communication is supported. For collision avoidance, every task slot is further designed to contain m slices. The length of each slice is equal to the atomic data transmission time, which is denoted by t. So the slot length $T_m = mt$. In "T-slot" anLN randomly selects one slice to transmit data for collision avoidance.

(2) A hop-by-hop time synchronization scheme is proposed to avoid sensor time drifting and make stair scheduling work in fully distributed manner. Due to "Stair scheduling," when a parent LN forwarding its message in its "T-Slot," for the symmetric link, its children will hear this broadcasting in their "Syn-Slot." The children synchronize time with the parent. When the hop-by-hop synchronization reachs the CH, the whole link is synchronized and "Stair scheduling" can work distributedly.

(3) Average-function-based hop-by-hop data aggregation is proposed to further enhance energy saving and to balance the energy consumption.

The detailed design of the proposed stair scheduling is as follows.

3.2. Stair Scheduling. "Stair scheduling" is to assign active task slots to LNs based on the hop counts of the LNs. It is initialized by the CH and is continually maintained by the LNs themselves during the data collection process.

3.2.1. Premise for Stair Scheduling. The level-based energy-balance routing (LEB) tree [8] and time synchronization among LNs and CH should be achieved at the cluster formation phase. The level-based energy-balance routing tree [8] is different from the shortest-path tree. Sensors are divided into levels based on the average hop progress, which can be easily obtained by flooding during the initialization phase [2, 12]. In LEB, for energy balance, a flag is assigned to each LNs, which is initially set to 0 and is changed to 1 once it acts as a relay for any other sensors. So when the node in the ith level finds its closest neighbor in the $(i-1)$th level and has flag $= 1$, it will try to select the second closest neighbor in the $(i-1)$th level as the relay. If it happens that all the neighbors in the $(i-1)$th level have flag $= 1$, the sensor in the ith level will randomly choose a neighbor in the $(i-1)$th level as the relay node.

Figure 1: Level-based energy-balance multihop communication in a cluster.

After initialization of LEB, the CH is aware of the levels of all the LNs and all the links in its cluster. Figure 1 shows a graph illustration of the CH's knowledge after LEB construction in a cluster. The square in the center is the CH and the surrounding points are LNs. The dashed circle is the levels.

3.2.2. Stair Scheduling Scheme. If the largest level is n, and the sample period of each LN is divided into M task slots, the "stair scheduling" works as follows.

(1) For an ith level of the LN, it will sleep from its 1st to $(n-i-1)$th slots and will activate at the $(n-i)$th slot. The active status will last three slots, and then the LN turns to sleep, till the end of the sample period. Among the three active slots, we have the following.

(i) The first slot is used to receive messages from the child LN in the $(i+1)$th level. We call it "R-Slot."

(ii) The second slot is used to sense local area and forward message to a parent node in the $(i-1)$th level. We call it "T-Slot."

(iii) The third active slot is used to synchronize time between the parent node and this node. Since the link is symmetry, when the parent node relays the message towards CH, the broadcasting will be overheard by this child. The overheard message is processed to synchronize time between the parent and this node. We call this slot "Syn-Slot."

For communication collision avoidance, each slot is designed to contain m slices. In "T-Slot," the LN randomly selects one slice for data transmission. Carrier sense and retransmission scheme is not used in stair scheduling.

(2) A special case will appear for the level n LNs. Since they do not relay other LNs' message, they only assign their first two slots to be active.

(i) The first slot is a "T-Slot" to sense local area and transmit message.

(ii) The second slot is used to synchronize time with a parent node in the $(n-1)$th level, which is a "Syn-Slot."

Figure 2 shows a graph illustration of the task slot assignments for LNs in the different levels. We can see two features of "stair scheduling" as follows.

(1) The active task slots of LNs form a stair shape with respect to the hop count; that is, a child LN will be activated one slot earlier than its parent.

(2) For LNs in two neighboring levels, they have two overlapped active slots. The "T-Slot" of the child overlaps with the "R-Slot" of the parent. This establishes the link for data transmission. The "Syn-Slot" of the child overlaps with the "T-Slot" of the parent. This establishes the hop-by-hop time synchronization.

3.3. Time Synchronization. Time synchronization plays an important role in stair scheduling. It is carried out during two phases: initialization phase and run-time hop-by-hop time synchronization.

In the initialization phase, time synchronization is carried out by Flooding Time Synchronization Protocol (FTSP) [13]. This is done simultaneously with the setting up of the LEB routing tree. Initially, all the LNs are active and their timers are not synchronized. The CH periodically broadcasts m CH messages for both time synchronization and LEB routing tree. The message generation time $t_{0,j}$ is broadcasted in each message, where j is the message index. When an LN receives the CH message, it uses the MAC-layer time-stamping to measure the packet forwarding delay. With the recorded delays of the previous hops, the LNs in the ith hop adjust their clock to

$$t_{i,j} = t_{0,j} + \sum_{k=1}^{i-1} s_{k,j}, \tag{1}$$

where $s_{k,j}$ is the delay at the kth hop. Since m CH messages are broadcasted periodically each LN will obtain a sequence of synchronized time $\{t_{i,1}, \ldots, t_{i,m}\}$. Linear regression is used to achieve high accuracy of the clock synchronization [13].

In the running phase, time synchronization is done hop-by-hop to avoid clock drifting. In "T-slot," LN broadcasts sensing data together with its local time. This broadcasting will be heard by its child nodes due to the symmetry link. These child nodes use the MAC-layer time stamping to record the processing delay and adjust their clock to the clock of their parents. With this hop-by-hop scheme, children synchronize clock with their parents, and the process repeats until the CH is reached as the last parent. Hence the clocks of the whole cluster can be continuously synchronized. And based on this, all LNs can maintain the "stair scheduling" in a fully distributed manner.

Figure 2: Joint active task slot scheduling with the "stair scheduling" method.

3.4. Hop-by-Hop Data Aggregation. In addition to stair scheduling, we adopt hop-by-hop data aggregation to enhance energy saving. For data collection with reverse shortest-path tree, the ideal data aggregation will be chain-based hop-by-hop data aggregation as discussed in PEGASIS [14]. However, such chain-based data aggregation poses high requirement to the scheduling of node transmission. Any child node should transmit earlier than its parent sensor; otherwise, redundant packet will be generated and additional energy will be consumed [15]. However, the authors of [14] have not discussed the details of the transmission scheduling.

The "stair scheduling" of stair scheduling provides fine solution to such chain-based hop-by-hop data aggregation, because in stair scheduling any child sensor will transmit data one slot earlier than its parent sensor. To implement hop-by-hop data aggregation, for an LN node, in R-Slot, it listens to messages from its child nodes. The data is extracted from the messages and is stored. In T-Slot, the LN senses its vicinity. The local readings will be merged with the stored data, and the results will be forwarded towards CH. We propose *average function* in data aggregation. If the size of the received data is l_r and the size of the local sensed data is l_s, using *average function*, the size of the merged data to be forwarded will be $l_f = \max\{l_r, l_s\}$. The energy performance of stair scheduling using average aggregation functions will be analyzed and evaluated in Sections 4 and 5.

4. Performance Analysis of Stair Scheduling for Network Optimization

In this section, we analyze the multi-hop reliability and energy consumption model of stair scheduling with respect to the cluster parameters. The model will be applied to optimize the tier-structure design of the sensor network.

4.1. Basic Analysis. We consider large-scale sensor systems, where sensors are not redundantly deployed for the cost consideration. Each sensor is responsible for the monitoring of its own vicinity. The sampling frequency U is commonly very low, for example, two samples per hour, and so forth.

For the monitoring completeness, there must be enough LNs deployed in the field to provide full land coverage. Based on the result of 1-coverage [16] analysis that calculates the probability of any point covered by at least one LN, the density of LN deployment can be derived as

$$\lambda = -\frac{\log\left(1 - P_{\text{1-cover}}\right)}{\pi r_s^2}, \tag{2}$$

where $P_{\text{1-cover}}$ is the desired 1-coverage probability. Since CHs do not sense data, whatever the cluster size chosen, for coverage preserving, the required LN density is λ. When CHs are deployed, the LNs are organized into clusters, and the number of LNs in each cluster can be calculated as $N = \lambda\pi R^2$. The following performance analysis and simulations are all carried out in such a coverage preserving scenario.

A simple but effective energy consumption model for sensor operations is assumed in this paper [1, 8, 17]. To transmit l-bits data over distance d, the sender will expend energy as

$$E_{TX}(l, d) = l * E_{\text{elec}} + l * \epsilon_{\text{amp}} * d^{\gamma}, \tag{3}$$

where E_{elec} is the unit and ϵ_{amp} is the amplifier energy. γ is the path loss factor, usually $2 < \gamma < 4$, and here we assume $\gamma = 2$. The energy expended by a receiver depends only on the length of the data and can be expressed as

$$E_{RX}(l) = l * E_{\text{elec}}. \tag{4}$$

Therefore, the key issue of energy consumption model of stair scheduling counts the transmission and reception times of a sensor in a sample period. In [8], the average hop progress of the sensor network r_h was derived as

$$r_h = \sqrt{3}\lambda \int_0^{r_c} x^2 e^{(\pi/3)\lambda(x^2 - r_c^2)} dx, \tag{5}$$

where r_c is the sensor's maximum transmission range and λ is the density of the sensors. The result divides the circular shape cluster into $n = R/r_c$ levels and verifies that sensors in the ith level have the most probability to reach CH by i hops. With this result, due to the uniform sensor distribution, the average number of sensor nodes in each level is

$$N_i = \pi\lambda(2i - 1) r_h^2, \tag{6}$$

where $i = 1, \ldots, n$. We assume that cluster radius R can be divided exactly by r_h.

According to the uniform distribution, an outer level will have more sensors than an inner level. So the average number of children for a sensor in level i is

$$C_i = \frac{N_{i+1}}{N_i}. \tag{7}$$

Figure 3: The interference region.

4.2. Multihop Communication Reliability. With "stair scheduling," for an arbitrary LN, if its "T-slot" comes, it randomly selects a slice within the m slices of the "T-slot" to transmit data. If two LNs share the same parent and select the same slice to transmit data, their messages will collide, and the transmissions will fail. Although the link can be enhanced by retransmission schemes, and so forth [18], we focus more on the performance of the basic stair scheduling and leave the enhanced methods to future work.

We suppose that an LN node C in the ith level tries to transmit. For the "stair scheduling" of stair scheduling, only the LNs in the same level share the same parent of C may collide C's transmission. Recall that the parent node found children nodes by broadcasting with radius r_c in the initialization phase. The "coparent circle" of C is therefore a circle centered at its parent node, and with radius r_c. We suppose that C's parent is node P, which locates in the $(i-1)$th level. Figure 3 shows the scenario. In the figure, the region of the ith level is ring centered at CH and with inner radius $(i-1)r_h$ and outer radius ir_h. The "interference region" is the region indicated by ABCD, which is the overlapped area of the "coparent circle" and the ith level ring.

4.2.1. Lower Bound of Transmission Reliability. Because even two LNs are within the "interference region," they may have different parents and their transmissions may not collide. So based on the following assumptions, we will get a lower bound of the transmission reliability. The assumption is as follows: "if two LNs are within interference region of each other and transmit data in the same time slice, their transmissions will collide." The lower bound will help us to understand the worst case of the transmission reliability of stair scheduling to direct cluster design. We can see that the area of the "interference region" is a function of i and h, where h is the distance from P to the inner border of the ith level ring. We denote the area by $S_c(i, h)$. Since LNs are uniformly deployed, the probability that there are k LNs within S_c is $p(k, S_c) = ((\lambda S_c)^k/k!)e^{-\lambda S_c}$. If all the other $k-1$ LNs select

different slices from C's transmission slice, C's transmission will success. This probability is

$$p(i,h) = \sum_{k=0}^{\infty}\left(1-\frac{1}{m}\right)^{k-1}\frac{(\lambda S_c(i,h))^k}{k!}e^{-\lambda S_c(i,h)} = \frac{m}{m-1}e^{-\lambda S_c(i,h)/m}, \quad (8)$$

where λ is node density and m is the number of slices. We can see that $p(i,h)$ is a monotone decreasing function of $S_c(i,h)$. If we can find an upper bound for $S_c(i,h)$, we will get the lower bound of $p(i,h)$.

We firstly consider the case $r_c < (i-1)r_h$, where the "co-parent" radius is smaller than the inner radius of the ring. This is corresponding to the case when C locates in some outer levels in which $i > 1 + r_c/r_h$. Since $r_c < 2r_h$ [8], this is corresponding to the case when $i \geq 3$.

It is easy to give an upper bound to $S_c(i,h)$ for this case, because the parent node P should locate in the $(i-1)$th level. As shown in Figure 4, when it locates on the inner border of the ith level, the "co-parent circle" and the ith level ring will form the maximum overlapped area, which is indicated by EFGH. It is easy to verify that the area of EFGH is an upper bound of $S_c(i,h)$. We denote this upper bound interference area as $S_{\text{up}}(i)$ as follows:

$$\begin{aligned} S_{\text{up}}(i) &= \frac{2r_c^2}{\pi(\beta-\theta)} + \frac{2\alpha(ir_h)^2}{\pi} \\ &\quad - r_c(i-1)r_h\sin\beta - \frac{2\gamma((i-1)r_h)^2}{\pi} \\ &\quad + r_c((i-1)r_h)\sin\theta \\ &\geq S_c(i,h), \quad \forall h \in [0, r_h], \end{aligned} \quad (9)$$

where

$$\begin{aligned} \alpha &= \arccos\frac{(ir_h)^2 - r_c^2 + ((i-1)r_h)^2}{2i(i-1)r_h^2}, \\ \beta &= \arccos\frac{((i-1)r_h)^2 + r_c^2 - (ir_h)^2}{2(i-1)r_h r_c}, \\ \theta &= \arccos\frac{((i-1)r_h)^2 + r_c^2 - (ir_h)^2}{2(i-1)r_h r_c}, \\ \gamma &= \pi - 2\theta. \end{aligned} \quad (10)$$

Now we consider the cases when $i = 1$ or $i = 2$. When $i = 1$, the LNs transmit directly to CH. The maximum "interference area" will be $S_{\text{up}}(i) = \pi r_h^2$. When $i = 2$, the upper bound of the "interference area" will be $S_{\text{up}}(i) = \pi r_c^2 - r_h^2$, as illustrated in Figure 5.

Therefore, we got the lower bound of the expected communication reliability of the ith hop as

$$\underline{P(i)} = \frac{m}{m-1}e^{-\lambda S_{\text{up}}(i)/m} \leq P(i). \quad (11)$$

Figure 4: Upper bound when $i \geq 3$.

Figure 5: Upper bound of $S_c(i,h)$ when $i = 2$.

Therefore, the lower bound of the multi-hop communication reliability when a message is forwarded from the level $i+j$ to the level i is

$$\underline{P(i+j,i)} = \prod_{k=i+1}^{i+j}\underline{P(i+k)}. \quad (12)$$

With the parameter settings in Table 1, where $P_{\text{1-cover}} = 0.99$, $n = 12$ and $r_c = 200$, the lower-bound of the hop-by-hop communication reliability and the lower bound of the multi-hop communication reliability are shown in Figure 6. We can see basically the following

(i) The lower bound varies with i. The lower bound increases as i decreases from 12 to 3. This is because the interference region becomes smaller as the level decreases. $\underline{P(i)}$ is small in the first and second levels, because, in such cases, the LNs are close. The number of LNs in the interference region becomes larger.

(ii) The multi-hop communication reliability decreases with the increasing of the forwarding hops. This gives us hints that in designing cluster, we should limit the number of forwarding hops.

4.3. Energy Consumption Model of Stair Scheduling. After analyzing the communication reliability in each hop, we derive the energy model for LNs in different hops. We assume that when two messages collides the energy for message transmission and reception will be consumed.

For an LN in the outmost nth level, it transmits its sensed data to its parent in its "T-Slot" and receives a time

Figure 6: The lower bound of $P(i)$ and the multihop reliability derived using the lower bound of $P(i)$.

Figure 7: Energy consumption in one sample period where LNs locate in different levels.

synchronization message from this parent in the "Syn-Slot." Its energy consumption in one sample period is

$$E(e_n) = l * E_{\text{elec}} + l * \epsilon_{\text{amp}} * (r_h)^{\gamma} + l * E_{\text{elec}}. \quad (13)$$

For an LN in the ith level, it receives data in "R-Slot" from its children. In "T-Slot," it senses local area and transmits the aggregated data to its parent. Since the received message is l-bits, and the local generated data is l-bits, the aggregated data with also be l-bits due to the average aggregation function. Then, in its "R-Slot," it receives a time synchronization message from this parent. Its energy consumption in one sample period is

$$E(e_i) = \left(2 + \frac{N_{i+1}}{N_i}\right) * l * E_{\text{elec}} + l * \epsilon_{\text{amp}} * (r_h)^{\gamma}. \quad (14)$$

The comparison of the energy consumption for LNs in different levels is shown in Figure 7, with parameter settings in Table 1. We can see the following.

Table 1: Setting of parameters.

Parameter	Value
E_{elec}	10^{-10}
l	100 (bits)
r_s	100 (m)
$P_{\text{1-cover}}$	0.999
ϵ_{amp}	10^{-10}
γ	2
r_c	200 (m)
n	10

(1) LN in the first level has the highest energy consumption rate, and LN in the outmost level has the lowest energy consumption rate. This is due to that the first level has the fewest number of sensors, so they averagely have the heaviest burden to relay data. But the outmost LNs need not to relay data at all.

(2) The difference among energy consumption rates is very small. The difference between energy consumption of the first level LNs and the tenth level LNs is only $(2.4763 - 2.476)/2.476 = 0.012\%$. This balance owes to the "stair scheduling" and "average aggregation function" of stair scheduling. By "stair scheduling," LNs in different levels activate for almost the same amount of slots. By average aggregation, different levels of LNs receive and forward nearly similar amount of data.

With the energy consumption model of stair scheduling, we can further analyze the energy consumption of the cluster and apply the analysis result to optimize the design of the tier structure.

4.4. Optimize the Design of Tier Structure. We can use the derived performances of stair scheduling to optimize the design of cluster. This is a typical multiple objectives optimization problem.

The controllable variable is the number of levels: n, which indeed determines the size of cluster. If a cluster has n levels, the size of one cluster is $\pi * (nr_h)^2$.

The constraint is that the cluster area should cover the full sensing field. For using the minimal number of clusters to cover the sensing field, the cell CH distribution [19] will be optimal. Suppose the area of sensing field is S, and the required number of CHs is

$$n_{\text{CH}} = \frac{2 * \pi}{3\sqrt{3}}. \quad (15)$$

The energy metric is to minimize the energy consumption of the first level, since it is the bottleneck of the lifetime of the cluster. So $m_1 = E(e_1)$.

The communication reliability metric is to maximize the reliability to forward a message from level n to level 1 as follows:

$$m_2 = \underline{P(n,1)} = \prod_{k=1}^{n-1} \underline{P(k)}. \quad (16)$$

The cost metric is to minimize the deployment cost of the sensors. Since the number of LNs is fixed by coverage preserving requirement. The cost is determined by the number of CHs, that is to minimize $m_3 = n_{\mathrm{CH}} * \text{cost}$.

By assigning weights to different metrics, we can arrive at an optimized design of n by performance trading off as

$$n = \arg\min_n \left(w_1 m_1 - w_2 m_2 + w_3 m_3\right), \quad (17)$$

where w_1, w_2, and w_3 are user assigned positive weights. Discussion of the cluster optimization will be out of the scope of this paper. We leave it to future work. We simply choose $n = 10$ in the following evaluations.

5. Performance Evaluation

5.1. Simulation Settings. We build a discrete event simulator using MATLAB 7.0 to evaluate the performances of stair scheduling. We compare the performances of the proposed stair scheduling with traditional single-hop randomized sleeping scheduling (SRS) [3] and multi-hop random sleeping scheduling (MRS). The results shown in Figures 9–13 are the means of 100 independent runs. In each simulation run, we generate a certain number of LN nodes and randomly place them in a rectangle sensing area with size 5000 $*$ 5000 (m^2). The number of nodes is determined by the node density, which is calculated by (2) by varying the 1-coverage probability from 0.9 to 0.99. Each LN captures data and reports data to CH. The sensing radius of LN is $r_s = 100$ meters and the communication radius $r_c = 200$ meters. The sample period of LN is half an hour $T = 1800{,}000$ ms and the time slice to finish a transmission event is fixed at $t = 9$ ms. The length of the task slot is $T_m = mt$, where m varies in the range of (10, 100).

SRS uses single-hop cluster in which every LN directly transmits data to CH. The number of clusters is decided based on the constraint of cluster coverage [19]. In each cluster, the LNs are independently scheduled, with active probability $p_{\mathrm{RSS}} = T_m/T$. Transmission will succeed if two transmissions do not collide.

Stair scheduling and MRS uses the same size of clusters and they both use LEB routing and divide the cluster into the same n levels. In MRS, LNs in each level are scheduled independently to capture, transmit, and forward data. When an LN in the ith level is activated, it broadcasts message towards level $i-1$. If the message reaches its parent LN in the level $i-1$, the transmission will succeed; otherwise, the transmission in this hop will fail. For the fairness of SRS, stair scheduling, and MRS, we set the average active duration in one period to be the same $p_{\mathrm{SRS}} = p_{\mathrm{MRS}} = T_m/T$. LNs in SRS, MRS, and stair scheduling share the same parameters and the same energy coefficients.

Figure 8: System cost as a function of 1-coverage probability for SRS, stair scheduling, and MRS.

Figure 9: Number of alive LN sensors versus system running time.

5.2. Simulation Results

5.2.1. System Cost. We focus on the energy and communication reliability performances with the joint consideration of the system cost.

The system cost means the total cost of deployed LNs and CHs. We suppose that the cost of one CH node is 50 times of the cost of an LN. In SRS, the radius of the cluster is the same with the communication radius of LN. In MRS and stair scheduling, the radius of the cluster is fixed at 10 times of the average hop progress.

Figure 8 shows the influence of node density on the system cost performance. When the required 1-coverage

Figure 10: Alive LNs in each level versus system running time in MRS.

Figure 11: Alive LNs in each level versus system running time in stair scheduling.

probability varies from 0.9 to 0.99, the corresponding node density varies from $0.73 * 10^{-4}$ to $0.1466 * 10^{-3}$. The corresponding system cost to cover the sensing field using SRS, stair scheduling, and MRS is plotted in Figure 8. We set the cost of one LN to one unit. We can see that the system cost of SRS is nearly seven times larger than the cost of stair scheduling and MRS. The cost of stair scheduling and MRS is the same since they use the same cluster structures.

5.2.2. Energy Performance. Figure 9 shows the number of alive LNs in the network as a function of system running time. The results show the following.

(i) Basically, SRS has better LN lifetime performance. This is not a surprise because in SRS, each LN only senses and transmits local data. They do not relay other sensors message.

(ii) The difference between SRS and stair scheduling and MRS is not very large. This owes to the average function hop-by-hop data aggregation method that we used in stair scheduling and MRS. With it, LNs in stair scheduling and MRS transfer a similar amount of data in each hop.

Figures 10 and 11 further show the number of alive LNs in each level of the cluster for MRS and stair scheduling. We can see the following

(1) Basically, the energy is almost evenly consumed. Most LNs in stair scheduling and MRS can work for more than 2000 periods. In both MRS and stair scheduling, the LNs in the outmost level live the longest time and the LNs in the first level consume energy more quickly. This coincides with our analysis that LNs in the inner levels will have higher burden with message forwarding.

(2) The curves of stair scheduling are much steeper than the curves of MRS. This means that LNs in different levels consume energy more evenly in stair scheduling than that in MRS. It will help LNs work together for longer time without performance degradation. The reason is due to the active slot control of the "stair scheduling."

5.2.3. Communication Reliability. For communication reliability evaluation, to concentrate on the performance of transmission scheduling, we do not consider the path loss coefficient. In SRS, communication will fail only when more than one LN start data transmission to the same CH in the same time slot. In MRS, communication may fail due to two cases. (1) When a child node transmits data, its parent node is sleeping and is not aware of the transmission. (2) Multiple children transmit data to the same parent in the same time slot. In stair scheduling, communication will fail when more than one LN start data transmission to the same CH in the same time slot. Figure 12 compares the communication reliability of SRS, stair scheduling, and MRS in different levels. Cluster using SRS only has one level and each LN directly transmits to CH. We can see that the communication reliability is very high. Clusters using stair scheduling and MRS have multiple levels. The communication reliability of MRS is very low, because the child and parent are not jointly scheduled. The probability that a child and its parent are active together is very small. The reliability of stair scheduling is much better than that of MRS, because children and parents are jointly scheduled. When a child transmits data, his parent is active and is waiting for its transmission.

Figure 13 compares the multi-hop communication reliability. The communication reliability of stair scheduling is

Figure 12: Communication reliability in each level.

Figure 13: Multi-hop communication reliability versus number of hops.

still larger than 0.9 after nine hops. But the communication reliability of MRS is lower than 0.001 after only two hops. The results further confirm that joint scheduling is necessary in multi-hop clusters and stair scheduling provides satisfactory communication reliability.

5.2.4. Performance Summary. With the above simulation results, we can see that stair scheduling is more applicable for large-scale sensor networks than SRS and MRS. Although SRS has good energy and reliability performances, its main drawback is the high cost for deployment of large amount of CHs. The main drawback of MRS is the low reliability in multi-hop communication. The proposed stair scheduling on one hand has good energy performance and on the other hand can provide satisfactory multi-hop communication reliability.

6. Conclusion

In this paper, we have proposed a stair scheduling method for data collection in tiered large-scale sensor network. The hop-by-hop time synchronization and data aggregation are proposed to make sensors in one cluster sleep and work cooperatively to conserve energy as well as to support multi-hop communication.

Particularly, in "Stair Scheduling," Child node in the ith level always activates and sleeps one slot earlier than its parent, so that LNs in each level can sleep for most of the time and can forward data reliably. Hop-by-hop time synchronization is proposed to avoid time drifting and make the "Stair Scheduling" work in a fully distributed manner. Simulation results have verified that the proposed stair scheduling provides efficient energy and communication reliability performances, which makes it more applicable than MRS and SRS. In future work, we will study the link reliability by introducing enhanced mechanisms of communication collision avoidance and the cluster optimization methods.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grants 61202360, 61073174, and 61033001.

References

[1] Y. Wang and Y. Wang, *Distributed Storage and Parallel Processing in Large-Scale Wireless Sensor Networks*, vol. 20 of *High Performance Computing: From Grids and Clouds to Exascale*, IOS Press, 2011.

[2] M. Medidi and Y. Zhou, "Extending lifetime with differential duty cycles in wireless sensor networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 1033–1037, November 2007.

[3] M. Demirbas, K. Y. Chow, and C. S. Wan, "INSIGHT: internet-sensor integration for habitat monitoring," in *Proceedings of the International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '06)*, pp. 553–558, IEEE Computer Society, June 2006.

[4] J. Deng, Y. S. Han, W. B. Heinzelman, and P. K. Varshney, "Scheduling sleeping nodes in high density cluster-based sensor networks," *Mobile Networks and Applications*, vol. 10, no. 6, pp. 825–835, 2005.

[5] D. Tian and N. D. Georganas, "A coverage-preserving node scheduling scheme for large wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pp. 32–41, ACM, September 2002.

[6] X. Wu, J. Cho, B. dAuriol, and S. Lee, "Sleep nodes scheduling in Cluster-Based heterogeneous sensor networks using AHP," in *Embedded Software and Systems*, Y. Lee, H. Kim, J. Kim, Y. Park, L. Yang, and S. Kim, Eds., vol. 4523 of *Lecture Notes in Computer Science*, pp. 437–444, Springer, Berlin, Germany, 2007.

[7] J. Paek, B. Greenstein, O. Gnawali et al., "The tenet architecture for tiered sensor networks," *ACM Transactions on Sensor Networks*, vol. 6, no. 4, article 34, 2010.

[8] Y. Wang, Y. Wang, H. Tan, and F. C. M. Lau, "Maximizing network lifetime online by localized probabilistic load balancing," in *Proceedings of the 10th International Conference on Ad-hoc, Mobile, and Wireless Networks (ADHOC-NOW '11)*, pp. 332–345, Springer, 2011.

[9] S. Bhattacharjee and N. Das, "Distributed data gathering scheduling in multihop wireless sensor networks for improved lifetime," in *Proceedings of the International Conference on Computing: Theory and Applications (ICCTA '07)*, pp. 46–50, IEEE Computer Society, March 2007.

[10] Y. Wang, D. Wang, W. Fu, and D. P. Agrawal, "Hops-based sleep scheduling algorithm for enhancing lifetime of wireless sensor networks," in *Proceedings of the IEEE International Conference on Mobile Ad Hoc and Sensor Sysetems (MASS '06)*, pp. 709–714, October 2006.

[11] W. Pak, J. G. Choi, and S. Bahk, "Tier based anycast to achieve maximum lifetime by duty cycle control in wireless sensor networks," in *Proceedings of the International Wireless Communications and Mobile Computing Conference (IWCMC '08)*, pp. 123–128, August 2008.

[12] I. Caragiannis, C. Kaklamanis, and P. Kanellopoulos, "New results for Energy-Efficient broadcasting in wireless networks," in *Proceedings of the 13th International Symposium on Algorithms and Computation (ISAAC '02)*, pp. 332–343, Springer, 2002.

[13] M. Maroti, B. Kusy, G. Simon, and A. Ledeczi, "Robust multihop time synchronization in sensor networks," in *Proceedings of the International Conference on Wireless Networks (ICWN '04)*, pp. 454–460, June 2004.

[14] S. Lindsey and C. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proceedings of the IEEE Aerospace Conference Proceedings*, vol. 3, pp. 3-1125–3-1130, 2002.

[15] S. Commuri and V. Tadigotla, "Dynamic data aggregation in wireless sensor networks," in *Proceedings of the IEEE 22nd International Symposium on Intelligent Control (ISIC '07)*, pp. 1–6, October 2007.

[16] C. Wu and Y. Chung, "Heterogeneous wireless sensor network deployment and topology control based on irregular sensor model," in *Proceedings of the 2nd International Conference on Advances in Grid and Pervasive Computing (GPC '07)*, pp. 78–88, Springer, 2007.

[17] L. Zhao, X. Hong, and Q. Liang, "Energy-efficient self-organization for wireless sensor networks: a fully distributed approach," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '04)*, pp. 2728–2732, December 2004.

[18] J. Yuan, Z. Li, W. Yu, and B. Li, "A cross-layer optimization framework for multicast in multi-hop wireless networks," in *Proceedings of the 1st International Conference on Wireless Internet (WICON '05)*, pp. 47–54, July 2005.

[19] Y. Wang, D. Han, Q. Zhao, X. Guan, and D. Zheng, "Clusters partition and sensors configuration for target tracking in wireless sensor networks," in *Proceedings of the 1st International Conference on Embedded Software and Systems (ICESS '05)*, pp. 333–338, Springer, 2005.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 741284, 8 pages
http://dx.doi.org/10.1155/2013/741284

Research Article

L-Shaped-Sensor-Array-Based Localization and Tracking Method for 3D Maneuvering Target

Xing Zhang and Xue Wang

State Key Laboratory of Precision Measurement Technology and Instrument, Department of Precision Instruments, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Xue Wang; wangxue@mail.tsinghua.edu.cn

Received 7 September 2012; Accepted 19 December 2012

Academic Editor: Ruchuan Wang

Copyright © 2013 X. Zhang and X. 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.

The localization and tracking technology for a three-dimensional target, which is a kernel problem in the military area, has received more and more attention. This paper proposes a closed-loop system to detect 3D maneuvering targets, including data acquisition, the direction of arrival (DOA) estimation, the triangle localization, and a trajectory prediction. This system firstly uses several L-shaped sensor arrays to sample the signals of maneuvering targets. Then the 2D ESPRIT algorithm and a maximum likelihood algorithm are introduced to achieve the positions of the spatial targets. Thirdly an autoregressive (AR) particle filter (PF) algorithm is realized to predict the locations in the next moment. Finally the localization process is directed by using the predicted positions to form a positive feedback closed loop. Experiment results show that this system can enhance the robustness and accuracy of the localization and tracking for three-dimensional maneuvering targets.

1. Introduction

Localization and tracking play a critical role in military applications and such as radar warning, battlefield monitoring. The kernel problem of each application is the ability of accurate localization. The most popular location tool is the Global Positioning System (GPS) [1]. Sometimes the location accuracy of the GPS is less than 0.1 m, but this sensor is usually used for self-localization. The multimedia sensor is another big type of location tools, such as CCD sensor [2, 3], stereo camera [4], and infrared sensor. However, multimedia sensor needs an extremely high image resolution, a fast network flow rate, and distinguished feature which limits its extensive use. RFID [5], wireless sensor networks [6], hybrid sensor networks [7] are generally used for the short-range target localization because their wireless transmission is easily influenced by the circumstance and the distance.

The reasons why the sensor array [8, 9, 10, 11, 12] is chosen in this paper to localize targets are as follows: sensor array is able to localize multiple targets simultaneously; sensor array can certainly eliminate the same noise and interference which can increase the signal noise ratio of the received data; and the localization and tracking system using sensor array is little impacted by the environment which can increase the accuracy reliability and robustness. Sensor array is widely used for localization. For example, Tidd used an 8-element uniform circular sensor array to precisely localize the far-field target by multiple signals classification algorithm [9]. Fallon and Godsill [10] and Talantzis [11] use a microphone array to locate the acoustic source. Lin compared the performance of DOA estimation with different types of sensor array, such as cross array, circular arrays, and L-shaped array. The Clay Metro lower boundary of uniform L-shaped array was almost lowest except circular array. The direction-finding resolution and orientation consistency of the L-shaped array were much better than circular array [12].

After well localization, tracking process is always carried out to enhance the performance of the localization and tracking system. Tidd use the Kalman filter to deal with the DOA and speed of the target to offer strategic, fast, and low-cost advantages [9]. Fallon and Godsill designed a track before detect framework to obtain the probability distribution of the particle filter which could guide the resample process to create better particles in a comparative

Figure 1: A uniform L-shaped sensor array receiving far field narrowband signal.

computational load [10]. This framework could increase the stability of the tracking system. Talantzis used the particle filter to predict the target trajectory [11]. The predicted position was used to estimate the accuracy of the localization by the microphone array and then replaced the inaccurate localization, sometimes as the true localization value when targets could not be detected.

This paper proposes a closed-loop localization and tracking system for three-dimensional targets. Firstly, choose several uniform L-shaped arrays to sample the spatial target signals and estimate the DOA of the 3D target with a two-dimensional estimating signal parameter via a rotational invariance techniques (2D ESPRIT) algorithm. Then the positions of the targets are calculated by a novel maximum likelihood (ML) algorithm according to the estimated DOA and the location of L-shaped sensor arrays. Thirdly the autoregressive particle filter algorithm is proposed to predict the target trajectory the in next time. Finally, the system uses the predicted position to guide the localization process. This system combines the localization and tracking algorithm together to form a positive feedback closed loop. This loop can increase the stability, robustness, and accuracy of the localization and tracking system.

The outline of this communication is as follows. In Section 2, the signal reception model based on the sensor array is developed. The proposed closed-loop localization and tracking system is described in Section 3. In Section 4, a series of experiment results are made to validate that the regenerative feedback system can work well to enhance the performance of the localization and tracking system. Finally the conclusion and perspectives are presented in Section 5.

2. Signal-Reception Model of Sensor Array

We consider a uniform L-shaped sensor array to localize and track the 3D maneuvering target. Make the intersected sensor as the origin and the arms of L type array as the axis to build a virtual coordinate system (shown in Figure 1). Each arm of the array have M elements. The sensors in the array are of the same nature. The distance between adjacent elements in x-axis is dx and in y-axis is dy.

L-shaped sensor array is used for collecting the narrow band signals emitted from the far-field target located at (x, y, z). The wavelength of signal is λ. (ϕ, θ) is the DOA of the spatial target. The elevation angle $\phi \in [0, 90]^\circ$. Azimuth angle $\theta \in (-180, 180]^\circ$.

The array manifold vector, called the steering vector, of the uniform L-shaped array is given as

$$\begin{aligned} \mathbf{a}_x(\phi,\theta) &= \left[1, u(\phi,\theta), \ldots, u^{M-1}(\phi,\theta)\right]^T, \\ \mathbf{a}_y(\phi,\theta) &= \left[1, v(\phi,\theta), \ldots, v^{M-1}(\phi,\theta)\right]^T, \\ \text{with } &\begin{cases} u^k(\phi,\theta) = \exp\left(j2\pi d_x \times k \times \cos\phi \cos\dfrac{\theta}{\lambda}\right) \\ v^k(\phi,\theta) = \exp\left(j2\pi d_y \times k \times \cos\phi \sin\dfrac{\theta}{\lambda}\right), \end{cases} \end{aligned} \tag{1}$$

where $(\cdot)^T$, j, x, y denote matrix transposition, imaginary unit, x-axis, and y-axis, respectively. The snapshot signal vector $\mathbf{X}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T$, $\mathbf{Y}(t) = [y_1(t), y_2(t), \ldots, y_M(t)]^T$ on x- and y-axis sampled by the L-shaped array is written as

$$\begin{aligned} \mathbf{X}(t) &= \mathbf{A}_x\mathbf{S}(t) + \mathbf{N}_x(t), \\ \mathbf{Y}(t) &= \mathbf{A}_y\mathbf{S}(t) + \mathbf{N}_y(t), \end{aligned} \tag{2}$$

where $\mathbf{S}(t) = [\mathbf{s}_1(t), \mathbf{s}_2(t), \ldots, \mathbf{s}_n(t)]^T$ are the complex envelope vector of the incoming signals. n is the number of targets which can be detected. $\mathbf{N}_x(t)$ and $\mathbf{N}_y(t)$ are the Gaussian white noise vectors whose mean is 0 and variance is σ^2 decided by signal noise ratio (SNR). $\mathbf{A}_x$ and $\mathbf{A}_y$ are array manifold matrixes as

$$\begin{aligned} \mathbf{A}_x &= \left[\mathbf{a}_x(\phi_1,\theta_1), \ldots, \mathbf{a}_x(\phi_n,\theta_n)\right], \\ \mathbf{A}_y &= \left[\mathbf{a}_y(\phi_1,\theta_1), \ldots, \mathbf{a}_y(\phi_n,\theta_n)\right]. \end{aligned} \tag{3}$$

3. Closed-Loop Localization and Tracking System for the 3D Target

With the data received by L-shaped sensor arrays, this paper uses 2D EPSRIT and ML algorithms to get the targets' positions. Then predict the trajectories by integrating AR model into PF algorithm. Finally guide the targets' localization process with the predicted positions. The whole system forms a closed-loop feedback and improves the localization and tracking accuracy.

3.1. 2D ESPRIT Algorithm. 2D ESPRIT algorithm [11] is the expansion of ESPRIT algorithm. According to the snapshot data $\mathbf{X}(t)$, $\mathbf{Y}(t)$ received by arrays, the special covariance matrix is calculated. Then construct the signal subspace using a singular value decomposition (SVD) of the covariance matrix. With the eigenvalue derived from SVD, get $u(\phi_i, \theta_i)$ and $v(\phi_i, \theta_i)$ mentioned in (5). The elevation and azimuth angle (ϕ_i, θ_i) can be achieved by $u(\phi_i, \theta_i)$ and $v(\phi_i, \theta_i)$.

The 2D ESPRIT algorithm can be described as follows.

Step 1. Decompose $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ into 4 an same size of the submatrixes $\mathbf{X}_1(t)$, $\mathbf{X}_2(t)$, $\mathbf{Y}_1(t)$, $\mathbf{Y}_2(t)$.

$$\begin{aligned} \mathbf{X}_1(t) &= \mathbf{X}(1:M-1,:)(t), \quad \mathbf{X}_2(t) = \mathbf{X}(2:M,:)(t), \\ \mathbf{Y}_1(t) &= \mathbf{Y}(1:M-1,:)(t), \quad \mathbf{Y}_2(t) = \mathbf{Y}(2:M,:)(t). \end{aligned} \tag{4}$$

Define $\mathbf{A}_{x1} = \mathbf{A}_x(1, M-1, :)$, $\mathbf{A}_{x2} = \mathbf{A}_x(2, M, :)$, $\mathbf{A}_{y1} = \mathbf{A}_y(1, M-1, :)$, $\mathbf{A}_{y2} = \mathbf{A}_y(2, M, :)$, so

$$\begin{aligned} \mathbf{A}_{x2} &= \mathbf{A}_{x1}\mathbf{\Phi}_x, \quad \mathbf{A}_{y2} = \mathbf{A}_{y1}\mathbf{\Phi}_y, \\ \mathbf{X}_1(t) &= \mathbf{A}_{x1}\mathbf{X}(t), \quad \mathbf{X}_2(t) = \mathbf{A}_{x2}\mathbf{X}(t), \\ \mathbf{Y}_1(t) &= \mathbf{A}_{y1}\mathbf{Y}(t), \quad \mathbf{Y}_2(t) = \mathbf{A}_{y2}\mathbf{Y}(t), \end{aligned} \tag{5}$$

where $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_x$ imply the translation characterization of the sensor array. They are denoted as

$$\begin{aligned} \mathbf{\Phi}_x &= \operatorname{diag}(u(\phi_1,\theta_1), \ldots, u(\phi_n,\theta_n)), \\ \mathbf{\Phi}_y &= \operatorname{diag}(v(\phi_1,\theta_1), \ldots, v(\phi_n,\theta_n)). \end{aligned} \tag{6}$$

Step 2. Calculate the covariance matrixes of these four submatrixes

$$\begin{aligned} \mathbf{R}_{\mathbf{X}_1\mathbf{Y}_1} &= E\left[\mathbf{X}_1(t)\,\mathbf{Y}_1(t)^H\right] = \mathbf{A}_{x1}\mathbf{R}_s\mathbf{A}_{y1}^H + \mathbf{F}_{\mathbf{N}_x,\mathbf{N}_y} \\ \mathbf{R}_{\mathbf{X}_2\mathbf{Y}_1} &= E\left[\mathbf{X}_2(t)\,\mathbf{Y}_1(t)^H\right] = \mathbf{A}_{x1}\mathbf{\Phi}_x\mathbf{R}_s\mathbf{A}_{y1}^H + \mathbf{F}_{\mathbf{N}_x,\mathbf{N}_y} \\ \mathbf{R}_{\mathbf{X}_2\mathbf{Y}_2} &= E\left[\mathbf{X}_2(t)\,\mathbf{Y}_2(t)^H\right] = \mathbf{A}_{x1}\mathbf{\Phi}_x\mathbf{R}_s\mathbf{\Phi}_y^H\mathbf{A}_{y1}^H + \mathbf{F}_{\mathbf{N}_x,\mathbf{N}_y}, \end{aligned} \tag{7}$$

where $\mathbf{R}_s = E[\mathbf{S}(t)\mathbf{S}^H(t)]$ is the signal self-related matrix. $\mathbf{H}$ is the complex conjugate transpose matrix. $\mathbf{F}_{\mathbf{N}_x,\mathbf{N}_y}$ conclude the information of the noise. Usually the noise is too less than the received signals to ignore, so we have $\mathbf{F}_{\mathbf{N}_x,\mathbf{N}_y} = 0$.

Step 3. Construct a special matrix $\mathbf{C}$ in formula (8). And then adopt SVD for the matrix $\mathbf{C}$ to achieve the signal subspace $\mathbf{E}_s$

$$\begin{aligned} \mathbf{C} &= \left[\mathbf{R}_{\mathbf{X}_1\mathbf{Y}_1} \quad \mathbf{R}_{\mathbf{X}_2\mathbf{Y}_1} \quad \mathbf{R}_{\mathbf{X}_2\mathbf{Y}_2}\right], \\ [\mathbf{U},\mathbf{S},\mathbf{V}] &= \operatorname{svd}(\mathbf{C}); \end{aligned} \tag{8}$$

$$\mathbf{E}_s = \begin{bmatrix} \mathbf{E}_0 \\ \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}\mathbf{\Phi}_x \\ \mathbf{A}\mathbf{\Phi}_x\mathbf{\Phi}_y^H \end{bmatrix} \mathbf{T} = \mathbf{U}(:, 1:n). \tag{9}$$

There are two unknown quantities $\mathbf{A}$ and $\mathbf{T}$. $\mathbf{A}$ is the transfer matrix produced by the $\mathbf{A}_{x1}$ and $\mathbf{A}_{y1}$. $\mathbf{T}$ is a middle matrix which is invertible. $\mathbf{E}_0$, $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_s$ are the signal subspace.

Step 4. Let $\mathbf{\Psi}_x = \mathbf{T}^{-1}\mathbf{\Phi}_x\mathbf{T}$, $\mathbf{\Psi}_y = \mathbf{T}^{-1}\mathbf{\Phi}_y^H\mathbf{T}$, we have $\mathbf{E}_1 = \mathbf{E}_0\mathbf{\Psi}_x$, $\mathbf{E}_2 = \mathbf{E}_1\mathbf{\Psi}_y$; therefore,

$$\begin{aligned} \mathbf{\Psi}_x &= \left(\mathbf{E}_0^H\mathbf{E}_0\right)^{-1}\mathbf{E}_0^H\mathbf{E}_1, \\ \mathbf{\Psi}_y &= \left(\mathbf{E}_1^H\mathbf{E}_1\right)^{-1}\mathbf{E}_1^H\mathbf{E}_2. \end{aligned} \tag{10}$$

Step 5. Use eig-function to deal with the matrixes $\mathbf{\Psi}_x$ and $\mathbf{\Psi}_y$ and gain eigenvalues which are equal to the diagonal element of the matrix $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_y^H$

$$\begin{aligned} [\mathbf{V}_x, \mathbf{\Phi}_x] &= \operatorname{eig}(\mathbf{\Psi}_x), \\ \left[\mathbf{V}_y, \mathbf{\Phi}_y^H\right] &= \operatorname{eig}(\mathbf{\Psi}_y). \end{aligned} \tag{11}$$

Step 6. Both $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_y^H$ have the information of the DOA. Due to the respective calculation, the eigenvalues from the same targets are not pairing. It needs post-process. Define a new matrix $G = \mathbf{V}_y^H\mathbf{V}_x$. Adjust the correspondence of the $\mathbf{\Phi}_x$, $\mathbf{\Phi}_y^H$ according to the maximum value in each line in $\mathbf{G}$.

Step 7. Compute the DOA of targets

$$\begin{aligned} \theta_k &= \arctan\left(\frac{\operatorname{angel}(v(\phi_k,\theta_k))}{\operatorname{angel}(u(\phi_k,\theta_k))}\right), \\ \phi_k &= \arccos\left(\frac{\lambda\sqrt{v^2(\phi_k,\theta_k) + u^2(\phi_k,\theta_k)}}{2\pi d}\right). \end{aligned} \tag{12}$$

3.2. Localization through the Maximum Likelihood Algorithm. Using the estimated DOA and the state of the sensor arrays, a maximum likelihood algorithm is proposed to realize the triangulation location. As shown in Figure 2(a), there are three uniform L-shaped arrays which have the same character. The positions of sensor arrays are $(0, 0, 0)$, $(a, b, 0)$, and $(c, d, 0)$ and meet the conditions $a \neq c \neq 0$ and $b \neq d \neq 0$.

The problem of triangulation location can be transformed into the one of seeking the intersection S for the three known radials. It has one unique answer theoretically. But due to the error of DOA estimation, the three radials will not intersect (seen in Figure 2(b)). So the paper uses a novel ML algorithm to search for partial areas and get the nearest point far from the three radials.

Definition 1. The angle between the point and radial is the one no more than 180° which is intersected by the radial and the line of the point and radial's jumping-off point.

Theorem 2. *In 2D plane, the point nearest to the three radials lies in the intersected triangle by the radials.*

For example, the angle between the point S and the radial L_2P_3 is $\angle SL_2P_3$. The nearest point S to three radials (L_2P_3, L_1P_3 and L_3P_1) lies in the triangle $P_1P_2P_3$.

This theorem can be deduced to a 3D space: in 3D space, the point nearest to three radials lies in the intersection space of projection triangle formed by the three radials on the Oxy, Oxz, and Oyz planes.

The maximum likelihood algorithm is described as follows

Firstly, get the projection of the three radials on the Oxy, Oxz, and Oyz planes. Then obtain the intersection points of the three radials on each plane. After that achieve the value ranges on each axis. Take the Oyz plane as an example, the points of intersected triangle are $P_1(y_1, z_1)$, $P_2(y_2, z_2)$, and $P_3(y_3, z_3)$,

(a) Three L-shaped arrays locate 3D target

(b) Projection on PL (Oyz) plane

Figure 2: Localization by three uniform L-shaped sensor arrays.

and the range value of y is $Y_{Oyz} \in [\min(y_1, y_2, y_3), \max(y_1, y_2, y_3)]$. We can also get the range of z on the Oyz plane.

Secondly, intersect the value range of xyz axis on each plane to create a spatial space for the maximum likelihood points. Take y-axis for example, the intersection of y is $[\max(\min(Y_{Oxy}), \min(Y_{Oyz})), \min(\max(Y_{Oxy}), \max(Y_{Oyz}))]$. Then divide this 3D space into $N_X \times N_Y \times N_Z$ pieces by a self-adaptive interval. Suppose that the probability of the angle of point and radial accords to the normal distribution. For all the divided points, then calculate the total probability of the angle to the three radials

$$P(x, y, z) = \frac{1}{2\pi} \exp\left(-\frac{\theta_1^2}{\pi^2}\right) + \frac{1}{2\pi} \exp\left(-\frac{\theta_2^2}{\pi^2}\right) + \frac{1}{2\pi} \exp\left(-\frac{\theta_3^2}{\pi^2}\right), \tag{13}$$

where θ_i is the angle between the point (x, y, z) and the ith radial.

Finally, search for the maximum of $P(x, y, z)$ and find the target position (x, y, z).

3.3. Autoregressive Particle Filter Algorithm. After localization, this paper introduces an autoregressive particle filter (ARPF) which combines AR models [13] with the particle filter [14] to predict the state of the targets in the next moment. This algorithm conquers the difficulty of designing some unknown state equations, noise matrixes, and initial state in particle filter and also provides a method to improve the accuracy of AR model.

The p-rank AR model (AR(p) in short) used for prediction can be denoted as

$$y_t = \epsilon_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p}, \quad t = p+1, p+2, \ldots T, \tag{14}$$

where y_t is the predicted position, $y_{t-1}, \ldots, y_{t-p}$ is the historical state of the same target. $\phi_1, \phi_2, \ldots, \phi_p$ is the autoregressive model coefficient and ϵ_t is the Gaussian white noise with the mean 0 and the variance ξ^2.

The ARPF algorithm is carried out. (1) Form the initial particles. (2) Use the AR model to calculate the positions of the particles at the next moment instead of the motion equation. (3) Get the prediction by the weighted sum of each particle. (4) Update the probability of each particle and do importance resampling process to form the new particle cluster. (5) Skip to (2) to recur until target is out of the detecting area.

3.4. Localization and Tracking System. The localization and tracking system presented here contains the whole process of signal collection (model building), DOA estimation, and tracking and prediction. The pseudocode of the system is described in Algorithm 1.

This closed-loop system makes the localization and tracking algorithms described into an integral whole to form a positive feedback. The loop increases the robustness and the accuracy of the target localization and tracking.

4. Simulation

The paper chooses the American F22 battle plane as the emulation target. The parameters of F22 can be achieved in http://www.aerospaceweb.org/aircraft/fighter/f22/. Create

Table 1: DOA estimation with 2D ESPRIT algorithm.

		Array L_1	Array L_2	Array L_3
Real angle/°	F22-1	(9.25, −5.84)	(39.99, 139.28)	(15.07, −67.31)
	F22-2	(5.41, 2.46)	(12.85, 23.66)	(8.67, −19.22)
Estimated angle/°		(5.14, 2.49)	(40.02, 139.28)	(14.99, 112.65)
		(9.69, 174.12)	(12.63, 23.69)	(8.58, 160.79)

```
While (the target is in the detecting area of L-shaped sensor arrays)
  If (it's the first moment the arrays detect)
      (1) Use all the sensor arrays to sample the signal of the target.
      (2) Calculate the position of the target by maximum likelihood algorithm.
      (3) Select three sensor arrays which are nearest to the targets.
  Else
      (1) Use the selected three sensor arrays to sample the signal.
      (2) Get the location by ML algorithm.
  End
  If (the number of the measured position for the target <4)
      The value achieved by ML algorithm is regarded as the final position.
  Else
      (1) Use ARPF algorithm to predict the position of the target at the next moment.
      (2) If (the number >4)
            Use the predicted position to guide targets' localization process which is discussed in Section 4.
          End
  End
  Time++;
End
```

Algorithm 1

Figure 3: Three L-shaped arrays sensing 3D maneuvering targets.

the random trajectory of F22 by the navigation and positioning toolbox of GPSoft in MATLAB, as shown in Figure 3. There are 2 pieces of moving orbits. The time range when F22-1 is flying is [195, 399] s, and the range of F22-2 is [171, 397] s. These planes emit sine complex envelope signals omnidirectionally when it moves along the trail.

We choose three uniform L-shaped sensor arrays of the same nature to localize and track 3D targets. The fixtures of the arrays are (0,0,0) m, (60000, −12000,0) m, and (40000,24000, 0) m. Set the parameters of the sensor array mentioned in Section 2, $M = 8$, $dx = dy = 0.5$ m, $\lambda = 1$ m, and SNR = 20 dB. The number of the snapshot is 64. The frequency of the localization is 1 Hz.

The experiment adopts Windows XP SP3 as the software platform, and AMD Athlon(tm) II X2 250 3.01 GHz and 2 GB memory as the hardware platform. The programming environment is MATLAB R2010b.

4.1. DOA Estimation. At a certain time, F22-1 is located in (52266, −5342, and 8561) m, and target F22-2 is located in (96898, 4164, and 9191) m, see the rhombuses shown in Figure 3. With the signal vector $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ sampled by the sensor arrays, DOA of the two targets is estimated, using the 2D ESPRIT algorithm. The results are shown in Table 1. The first item in brackets is elevation, and the second item is azimuth.

By doing tens Monte Carlo simulation and calculate the mean of DOA, it can be found that the average error of angle estimation is less than 0.2°. Most of the time the 2D ESPRIT algorithm works well, but when the incident DOA is very close to each other, the errors is a little big. The 2D ESPRIT algorithm needs to be improved in such situation.

Besides, there are other two problems: first, the DOA estimations of the same target are not in the same line.

(a) Triangulation localization with the maximum likelihood algorithm

(b) MSE of localization with the ML algorithm

Figure 4: Performance of the maximum likelihood algorithm.

(a)

(b)

Figure 5: MSE of the ARPF algorithm.

Second, the estimated azimuth is all in $[0, 180]^{\circ}$. When the incident azimuth is minus, the result is always equal to this angle plus 180°.

So before localization by maximum likelihood algorithm, the DOA from the same target must be paired in the same group, and the azimuth must be well decided. When there is no other prior information, the positions of targets at the preview moment are used to confirm azimuth and group the DOA estimations which are out of order.

4.2. Triangulation Location. With the estimated DOA and the location of sensor arrays, the maximum likelihood algorithm is employed to calculate the positions of the targets. The localization results and errors are displayed in Figure 4.

We can find in Figure 4 that most localization mean square errors (MSEs) of F22-1 are less than 300 m, and F22-2 localization deviation is less than 1500 m by the maximum likelihood algorithm. Because of the larger distance between F22-2 and the sensor arrays, MSE of F22-2 is bigger than MSE of F22-1.

For far-field targets' localization in this application, the relative errors are less than 3% which is acceptable. So the ML method meets the localization needs. There are also some outliers that have exceptional large localization MSE. These outliers always appear when the target is far away from sensor arrays. So when there are redundant sensor arrays, select three sensor arrays which are nearest to the target for localization.

Table 2: Performance of closed positive feedback loop.

	F22-1 MSE/m	F22-2 MSE/m	Average consuming/s
Localization by 2D ESPRIT and ML algorithm	98.55	290.38	0.033
Using a closed-loop system	96.97	276.14	0.047

4.3. Trajectory Prediction. After localization by the 2D ESPRIT and ML algorithms, autoregressive particle filter is implemented to form a positive feedback closed loop which enhances the accuracy and stability of the localization. The forecast MSE is seen in Figure 5.

Compare prediction MSE by ARPF and the distance between the location at current and preview moment. It can be found that most of the time, the prediction MSE by ARPF is smaller. This illustrates that the prediction is more appropriate to direct the process mentioned above which confirms azimuth and groups the DOA estimations. There are also some moments that the prediction MSE is less than the localization MSE, so when the localization result is far away from the trajectory, in other words, when the outlier appears, we would better use the prediction as the true location instead of the measurement. For example, the location MSE of F22 by ML algorithm at 358 s is 4477 m, and the forecast MSE is 1035 m; this time uses the predicted position as the real measurement.

The performance of the closed loop is shown in Table 2.

Table 2 illustrates that 2D ESPRIT and ML algorithms are available to locate the target because the localization MSE is acceptable. The close loop system can slightly improve the accuracy of the localization. The time consumption is a little increased, but still fast enough to locate and track the target.

5. Conclusion

This paper proposes a localizing and tracking system for a three-dimensional target which combines the DOA estimation, the triangulation location, and the trajectory prediction. Firstly we establish a signal-reception model for a far-field target. Then the 2D ESPRIT algorithm is carried out to estimate the DOA of the 3D target. After that, a novel maximum likelihood algorithm is introduced to solve the triangulation location problem. Next the autoregressive particle filter is implemented to predict the target trajectory. Finally the localization process is guided according to the predictions. The experiment shows that the localization by 2D ESPRIT algorithm and the ML algorithm are useful for the 3D target. ARPF algorithm can well predict position at the next moment and guide the localization process. The localization and prediction processes form a passive feedback loop which can raise the accuracy and robustness of the localization and tracking system.

Acknowledgments

The authors would like to acknowledge the support from the National Grant Fundamental Research 973 Program of China (Grant no. 2006CB303000), the National Natural Science Foundation of China (Grants no. 60673176, 60373014, 50175056, and 60970103), and the Ph.D. Programs Foundation of the Ministry of Education of China (Grant no. 20090002110016).

References

[1] N. Abdelkrim, N. Aouf, A. Tsourdos, and B. White, "Robust nonlinear filtering for INS/GPS UAV localization," in *Proceedings of the 16th Mediterranean Conference on Control and Automation (MED '08)*, pp. 695–702, Ajaccio, France, June 2008.

[2] S. Sohn, B. Lee, J. Kim, and C. Kee, "Vision-based real-time target localization for single-antenna GPS-guided UAV," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 4, pp. 1391–1401, 2008.

[3] C. Liu, X. Wang, and Z. Xing, "Fast face detection method based on real- time prediction and learning classification," *Chinese Journal of Scientific Instrument*, vol. 33, no. 1, pp. 36–41, 2012.

[4] A. Nemra and N. Aouf, "Adaptive decentralised cooperative vision based simultaneous localization and mapping for multiple UAV," in *Proceedings of the 19th Mediterranean Conference on Control and Automation*, pp. 546–551, Corfu, Greece, 2011.

[5] A. S. Paul and E. A. Wan, "RSSI-based indoor localization and tracking using sigma-point Kalman smoothers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 860–873, 2009.

[6] S. Wang, X. Wang, Y. Wang, and X. Sun, "Distributed lightweight target tracking for wireless sensor networks," in *Proceedings of the 6th IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS '09)*, pp. 870–875, Macau, China, October 2009.

[7] A. Hurst, A. Wickenheiser, and E. Garcia, "Localization and perching maneuver tracking for a morphing UAV," in *Proceedings of the IEEE/ION Position, Location and Navigation Symposium (PLANS '08)*, pp. 1238–1245, Monterey, Calif, USA, May 2008.

[8] X. Zhang, X. Wang et al., "Sensor arrays based predicted spatial multi-signals classification method for target localization and tracking," *Chinese Journal of Scientific Instrument*, vol. 33, no. 5, pp. 970–975, 2012.

[9] W. Tidd, R. J. Weber, Y. Huang, and Y. Zhao, "A RF source localization and tracking system," in *Proceedings of the IEEE Military Communications Conference (MILCOM '10)*, pp. 858–863, San Jose, Calif, USA, November 2010.

[10] M. F. Fallon and S. Godsill, "Acoustic source localization and tracking using track before detect," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 6, pp. 1228–1242, 2010.

[11] F. Talantzis, "An acoustic source localization and tracking framework using particle filtering and information theory,"

IEEE Transactions on Audio, Speech and Language Processing, vol. 18, no. 7, pp. 1806–1817, 2010.

[12] L. M. lin, *Study on Estimation of Two Dimensional Direction of Arrival Using a DBF Receiver*, Harbin Institute of Technology, Harbin, China, 2008.

[13] W. Liu and M. Farooq, "ARMA model based scheme for maneuvering target tracking," in *Proceedings of the 37th Midwest Symposium on Circuits and Systems*, pp. 1408–1411, August 1994.

[14] R. van der Merwe and A. Doucet, "The unscented particle filter," Tech. Rep. CUED/F-INFENG/TR 380, 2000.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 565604, 13 pages
doi:10.1155/2012/565604

Research Article

An Adaptive Opportunistic Network Coding Mechanism in Wireless Multimedia Sensor Networks

Hang Shen,[1,2] Guangwei Bai,[1,2,3] Lu Zhao,[1] and Zhenmin Tang[1]

[1] *Key Laboratory of Intelligent Perception and System for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China*
[2] *State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing 210093, China*
[3] *Department of Computer Science and Technology, Nanjing University of Technology, Nanjing 210009, China*

Correspondence should be addressed to Guangwei Bai, bai@njut.edu.cn

Received 5 September 2012; Accepted 21 November 2012

Academic Editor: Ruchuan Wang

Copyright © 2012 Hang Shen 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 network coding technique is promising for improving the performance of video communication in wireless multimedia sensor networks. However, some special characteristics of existing wireless network coding mechanisms degrade the performance of video data delivery. This work begins with a thorough investigation and understanding of the performance limitations of existing wireless network coding mechanisms. On this basis, we propose an Adaptive Opportunistic Network Coding mechanism (AONC) to improve the transmission quality of video stream in wireless multimedia sensor networks. First, we propose a novel asymmetric coding method to process the video data of different lengths. The aim is to improve data exchange gain. Second, we design an opportunistic forwarding strategy based on dynamic priority to ensure that packets have a better chance to be coded and transmitted, thus achieving much higher throughput. Finally, we present a traffic-aware data scheduling algorithm, working together with the above network coding mechanism, to reduce the loss of potential coding opportunities. Our simulation results demonstrate that, compared with the existing typical network coding mechanisms, AONC can greatly enhance video transmission quality and efficiently utilize bandwidth and energy resources.

1. Introduction

As the bandwidth of wireless channels increases and low-cost hardware device such as CMOS cameras appears, wireless multimedia sensor networks (WMSNs), that is, networks of wireless embedded devices that allow retrieving video and audio streams in real time from the physical environment, have drawn tremendous attention from academia and industry [1–3]. WMSNs can enhance a lot of potential applications such as video surveillance, traffic enforcement, control systems, remote health monitoring, and industrial process control. Multimedia sensors in WMSNs have severe resource constraints including bandwidth, energy, storage, processing capability, and achievable data rate [4], while the delivery of multimedia data is resource demanding. These constraints and challenges, in combination with the transmission reliability required by video applications, make video communication in such environment a challenging proposition.

Network coding is one of the most important achievements on the theory of information processing and transmission. The concept is first proposed for achieving the min-cut capacity in the context of multicast communication [5]. Since then, there are various studies concerning theoretical limitations and practical applications of network coding. COPE [6] is the first practical network coding mechanism for supporting efficient unicast communication in wireless network. Motivated by COPE, recent research includes loss-aware coding mechanism [7, 8], opportunistic forwarding strategy [9, 10], coding-aware routing [11–13], rate-adaptive transmission scheme [14–16], QoS-aware scheduling algorithm [17–19], data broadcast protocol [20, 21], and so forth. These studies demonstrate that network coding has an excellent performance in terms of network throughput, energy efficiency, robustness, and so forth. As a result, network coding is a promising technique for supporting video streaming over WMSNs. However, real-time video

Figure 1: Packet mixing of different methods.

data transmission is fundamentally different from traditional data communication. Some special characteristics of existing network coding mechanisms degrade the performance of wireless video delivery. In recent years, the existing typical network coding mechanisms proposed in literature do not take the characteristics of video streaming into account and cannot be used directly in wireless video delivery.

This work focuses on performance improvement of video communication in WMSNs. We begin with a thorough investigation and understanding of the performance limitations of existing typical wireless network coding mechanisms. On this basis, we propose an Adaptive Opportunistic Network Coding mechanism (AONC) in WMSNs. The main mechanisms of the AONC consist of three parts. First, unlike these traditional network coding methods, AONC adopts a novel asymmetric coding method to improve data exchange gain efficiently. Second, an opportunistic forwarding strategy with dynamic priority is designed to ensure that packets have a better chance to be coded and transmitted by one forwarder, thus achieving much higher throughput. Finally, we present a traffic-aware data scheduling algorithm, working together with above network coding mechanism. The purpose is to reduce the loss of potential coding opportunities. Our simulation results show that, compared with the existing typical network coding mechanisms, AONC improves the video transmission quality greatly and achieves significant gains in terms of bandwidth and energy efficiency.

The remainder of this paper is organized as follows. Section 2 presents a thorough investigation and understanding of the performance limitations of combining network coding and video streaming. In Section 3, we propose the AONC and present the details of its implementation. Section 4 involves thorough analysis and evaluation of the AONC performance in simulation methodology. Finally, Section 5 concludes this paper and outlines some perspectives for future work.

2. Problem Description

Although these existing wireless network coding mechanisms have demonstrated its capability of improving the network throughput, they still cannot efficiently support video communication. The aim of this paper is to overcome these limitations. We now describe the limitations from the following three aspects.

(1) Traditional Symmetric Coding Method Degrades the Performance of Video Communication. To illustrate the performance limitation of traditional symmetric coding method, we consider two simple examples shown in Figure 1. As shown in Figure 1(a), the data packets of different lengths saved in the output queue are divided into three sets, and we assume that data packets in any two sets meet coding condition. We will present the model of output queue later in Section 3.1. Figure 1(b) provides an example of a simple network coding method, that is, XOR operation in 1 : 1 symmetric mode, to demonstrate the restriction of this coding method. The length of an encoded packet is approximately equal to the size of the largest packet. Due to the difference in the length of video data packets, the coding space utilization is maintained at a low level. A node needs to transmit three times. Obviously, this method is inefficient,

(a) Network coding scenario ($F1$-$F2$ meets coding condition)

(b) The amount of queued data of different flows in output queue at t_c

(c) Arriving traffic volume from different flows during $(t_c, t_c + d]$

Figure 2: Data scheduling mechanism analyses (Case 1).

and the throughput gain is hard to be improved greatly. Considering the above severe limitation, we propose a novel asymmetric coding method. Figure 1(c) shows the advantage of XOR operation in asymmetric mode. A node only needs to transmit once. The method can adapt the video data of different lengths and has great potential to improve the performance of video communication. We will present the details of its implementation later in Section 3.2.

(2) The Priority Assignment of Coded Packets May Be Unreasonable. To maximize coding opportunities for throughput gain, a mechanism is needed to assign coded transmissions higher priority in scheduling. A common method to assign priority for a coded packet is based on the number of original packets. The greater the number of original packets is, the higher the priority of a coded packet is assigned. However, due to the difference in the length of video data, this method could not accurately indicate the actual amount of original data carried in a coded packet of a certain length. In addition, for the reason that there are only a few priority levels, collisions are most likely to occur. This property enables us to consider an opportunistic forwarding strategy based on dynamic priority. We will discuss it later in Section 3.3.

(3) With Traditional Data Scheduling Mechanism, Some Potential Coding Opportunities May Be Lost. Compression algorithms such as MPEG that are widely used in low-bitrate videos employ coding prediction, motion compensation, and varied length coding technologies to achieve a high efficient compression. However, the compressed video stream is a variable bitrate (VBR) and has very complex traffic characteristics. As a general queue scheduling mechanism, FIFO does not take these factors into account, resulting in loss of potential coding opportunities. Here we, respectively, give two examples, illustrated in Figures 2 and 3, to explain the reason why some potential coding opportunities are lost.

Figure 2(a) shows a general network scenario where there are five wireless nodes. Suppose that there exist three different flows, that is, $F1$, $F2$, and $F3$, and only $F1$-$F2$ meets coding condition. There has no coding possibility at current time t_c, if the amount of queued data of different flows is shown in Figure 2(b). We assume that arriving traffic volume from different flows during the next short interval $(t_c, t_c + d]$ is shown in Figure 2(c). In this case, some potential coding opportunities may be lost if data packets of $F1$ are sent more during interval $(t_c, t_c + d]$.

Figure 3(a) shows another general network scenario where there are five wireless nodes. Suppose that there exist four different flows, that is, $F1$, $F2$, $F3$, and $F4$. Under this scenario, $F1$-$F2$ and $F3$-$F4$ meet coding condition. In accordance with the amount of queued data from different flows shown in Figure 3(b), node R has no coding possibility. We assume that arriving traffic volume of different flows of the next short interval $(t_c, t_c + d]$ is shown in Figure 3(c). In this case, some potential coding opportunities may be lost if data packets of $F1$ are sent more during interval $(t_c, t_c + d]$.

AONC is designed to seize these potential coding opportunities effectively. The design details will be present later in Section 3.4.

3. Adaptive Opportunistic Network Coding Mechanism

We first describe our network coding model and introduce the notations used in this section. On this basis, the main mechanisms of the AONC are comprised of the following

(a) Network coding scenario ($F1$-$F2$ and $F3$-$F4$ meet coding condition)

(b) The amount of queued data of different flows in output queue at t_c

(c) Arriving traffic volume of different flows during $(t_c, t_c + d]$

Figure 3: Data scheduling mechanism analyses (Case 2).

Figure 4: Flow-oriented virtual queue structure.

three parts: (1) an asymmetric coding method is used to improve data exchange gain; (2) an opportunistic forwarding strategy with dynamic priority is designed to achieve higher throughput; (3) a traffic-aware data scheduling algorithm is used to reduce the loss of coding opportunities.

3.1. Network Coding Model. We consider the application of network coding in WMSNs. We assume that intermediate nodes can perform network coding operations and combine packets from multiple incoming flows into a single outgoing packet. This packet is broadcasted to the entire neighborhood, thus reaching corresponding neighboring nodes. We assume that node network interface is set in promiscuous mode. The sensor nodes can snoop on all transmissions in their neighborhood and store the overheard data packets, whether they are intended for them or not. In this way, the sensor nodes can know the overheard and routed packets that each neighboring node possesses, such that it can make network coding operations based on this information.

As shown in Figure 4, each sensor node groups data packets in the output queue into data flows, assembling data packets of the same previous forwarder and next-hop receiver into one flow. The set of data flows is expressed as $Q = \bigcup f_m$, $m \in M = \{1, 2, \ldots, |Q|\}$, where $|Q|$ denotes the number of flows in Q. Let $f_m = \bigcup p_{mn}$, $n \in N_m = \{1, 2, \ldots, |f_m|\}$ be the mth flow in Q, where p_{mn} and $|f_m|$, respectively, represent the nth data packet of f_m and corresponding number of packets in it. Let $l_{mn} \in [1, L]$ be the length of p_{mn}, where L denotes the length of maximum transmission unit. Each sensor node monitors flow information in real time and deletes f_m when $|f_m| = 0$ in a period of time.

We define "coding flows" as flows that transmit via an intermediate node, and their packets have opportunities to be encoded. The set of groups of coding flows is defined as

Table 1: Notations used in this work.

Symbols	Definition
M	The set of signs of flows in output queue
f_m	The mth flow in output queue, $\lvert f_m \rvert$ represents the number of packet belonging to f_m, $m \in M$
N_m	The set of signs of mth flow in output queue, $m \in M$
p_{mn}	The nth packet of mth flow
l_{mn}	The length of p_{mn}
L	The length of maximum transmission unit
C	The set of groups of coding flows, $\lvert C \rvert$ represents corresponding number of groups
K	The set of signs of groups of coding flows, $\lvert K \rvert$ represents corresponding number of signs
g_k	The kth group of coding flows, $\lvert g_k \rvert$ represents the number of flows in g_k, $k \in K$
M_k	The set of signs of kth group of coding flows, $\lvert M_k \rvert$ represents corresponding number of signs, $k \in K$
M'_k	The set of signs of nonempty flows belonging to kth group of coding flows, $\lvert M'_k \rvert$ represents corresponding number of signs, $k \in K$
U	The set of flows that can not be encoded, $\lvert U \rvert$ represents corresponding number of flows
M_U	The set of signs of flows that cannot be encoded
M'_U	The set of signs of nonempty flows in M_U, $\lvert M'_U \rvert$ represents corresponding number of signs
φ	Set to be 1 if network coding conditions are met and is set to be 0 otherwise

$C = \bigcup g_k$, $k \in K = \{1, 2, \ldots, |C|\}$, where $|C|$ represents the number of groups of coding flows in C. The kth group of coding flows is defined as $g_k = \bigcup f_m$, $m \in M_k = \{1, 2, \ldots, |g_k|\}$, where $|g_k|$ represents the number of flows in g_k. Let nh_m and pf_m be next-hop receiver and previous forwarder of f_m. The set of neighboring nodes of pf_m is defined as $nb(pf_m)$. The relationship between any two flows in a group of coding flows must satisfy

$$R = \Big\{(i, j) \mid \Big(nh_i = pf_j \vee nh_i \in nb\big(pf_j\big)\Big) \wedge \Big(nh_j = pf_i \vee nh_j \in nb(pf_i)\Big)\Big\}, \quad i, j \in M_k. \tag{1}$$

The set of signs of nonempty flows in g_k is defined by $M'_k = \{m \mid m \in M_k, |f_m| > 0\}$. Let $U = \bigcup f_m$, $m \in M_U = \{1, 2, \ldots, |U|\}$ be the set of flows that cannot be encoded, where $|U|$ represents corresponding number of flows. The set of signs of nonempty flows in M_U is defined by $M'_U = \{m \mid m \in M_U, |f_m| > 0\}$.

We define parameter φ to determine whether network coding conditions are met or not. If yes, φ is set to be 1, and otherwise, φ is set to be 0

$$\varphi = \begin{cases} 1, & \left|M'_k\right| \geq 2,\ k \in K, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The summarized notations in this paper are listed in Table 1.

3.2. Asymmetric Network Coding Mechanism

3.2.1. Algorithm Design and Analysis. When a sensor node wants to send data, according to (2), if φ is set to be 1, the proposed asymmetric coding method can be used. The core idea is as follows. We, respectively, select the appropriate data packets from each flow in a predetermined group of coding flows and then splice these data packets belonging to the same flow. Finally, we combine corresponding splicing units in a single broadcast transmission. This method can greatly increase the amount of original data carried of a coded packet, thus increasing the information content per transmission. We believe that it is beneficial to provide better quality of video in network bandwidth constrained environments. The optimal selection for data packets belonging to one flow in a group of coding flows with maximize utilization can be formulated as follows:

$$\begin{aligned} &\text{maximize} \quad \sum_{n \in N_m} l_{mn} \times x_{mn}, \\ &\text{s.t.} \quad \sum_{n \in N_m} l_{mn} \times x_{mn} \leq S, \quad x_{mn} \in \{0, 1\}, \\ &\qquad n \in N_m,\ m \in M'_k,\ k \in K. \end{aligned} \tag{3}$$

Here, we refer to the space length limit as S. The random variable x_{mn} is set to be 1 if p_{mn} is selected and is set to 0 otherwise.

Let $Y_m(S)$ be the optimal length of the splicing unit according to (3). We define the set of signs of data packets in the splicing unit as follows:

$$N_{m,S} = \arg Y_m(S), \quad m \in M'_k, k \in K. \tag{4}$$

Definition 1. Space utilization is a metric which is used to measure the amount of original data carried of a coded packet of a certain length.

This metric accurately reflects the enhancement of transmission efficiency that can be achieved during the transmission period of encoded packet. The space utilization based on asymmetric coding method can be expressed as

$$\mathrm{ASU}(k, S) = \frac{\sum_{m \in M'_k} Y_m(S)}{\max_{m \in M'_k} Y_m(S)}, \quad k \in K,\ S \in [1, L]. \tag{5}$$

Traditional symmetric coding method uses FIFO queue scheduling mechanism; that is, the first data packet belonging to one flow (i.e., $p_{m1}, m \in M'_k$) in a group of coding flows will be selected for transmission. Therefore, we can express the space utilization based on traditional symmetric coding method as

$$\mathrm{SSU}(k) = \frac{\sum_{m\in M'_k} l_{m1}}{\max_{m\in M'_k} l_{m1}}, \quad k \in K,\ l_{m1} \in [1, L]. \tag{6}$$

Among all possible candidate data packets, AONC selects appropriate parameters k and S that maximizes space utilization

$$Z = \max_{k\in K} \max_{S\in[1,L]} \mathrm{ASU}(k, S). \tag{7}$$

Theorem 2. *The proposed asymmetric coding method can provide higher space utilization than symmetric coding method can.*

The proof is as follows:

$$\begin{aligned} Z &\geq \max_{S\in[1,L]} \mathrm{ASU}\,(k, S) \geq \mathrm{ASU}\left(k, \max_{m\in M'_k} l_{m1}\right) \\ &\geq \mathrm{SSU}\,(k), \quad k \in K,\ l_{m1} \in [1, L]. \end{aligned} \tag{8}$$

Definition 3. Data exchange gain indicates that the number of transmissions reduced by a successful encoded packet exchange.

Let $\mathrm{ASG}(k, S)$ and $\mathrm{SSG}(k)$, respectively, be the data exchange gain based on asymmetric and symmetric coding method

$$\mathrm{ASG}(k, S) = \sum_{m\in M'_k} |N_{m,S}|, \quad k \in K,\ S \in [1, L], \tag{9}$$

$$\mathrm{SSG}(k) = \left|M'_k\right|, \quad k \in K. \tag{10}$$

Combining (9) and (10), we have

$$\begin{aligned} \mathrm{ASG}(k, S) &\geq \left|M'_k\right| \times \min_{m\in M'_k} |N_{m,S}| \\ &\geq \mathrm{SSG}(k), \quad k \in K,\ S = L. \end{aligned} \tag{11}$$

The throughput provided by a successful coded packet exchange, which is based on asymmetric coding method, can be expressed as follows:

$$\mathrm{ATG}(k, S) = \sum_{m\in M'_k} \sum_{n\in N_{m,S}} l_{mn}, \quad k \in K,\ S \in [1, L]. \tag{12}$$

According to (8) and (11), we determine that asymmetric coding method has great potential to improve the performance of wireless video communication.

Algorithm 1 shows the pseudocode for the packet selection algorithm based on proposed asymmetric coding method.

3.2.2. Design Details. The communication modules of sensor node are based typically either on IEEE 802.11 [22] standard or on IEEE 802.15.4 [23] standard. IEEE 802.15.4 compliant radio can only support a maximum rate of 250 kbps. At such low data rates, video streaming is possible only if sufficient preprocessing steps, such as dimensional reduction and descriptive representations (color histograms, object shape) are undertaken [24]. As a result, it cannot provide acceptable real-time video transmission in WMSN. Our solution is implemented on IEEE 802.11.

AONC performs packet coding and tagging at the MAC layer. It requires a modification to the DATA and ACK frame headers of the existing 802.11 MAC standard. Figure 5 shows the header fields modified or added for AONC. If encoded, it has a list RA[CFN] about receiver addresses, where CFN denotes the number of coding flows in a coded packet. PktNum[CFN] represents the number of original packets belonging to corresponding flow. For all the encoded packets, the corresponding list PktInfo[CFN][PktNum[CFN]] is used to record packet ID and packet length. The packet ID is generated by creating a 4-byte hash value out of the address of source node and the sequence number carried by the packet, similar to COPE [6]. The packet length is used to obtain the original data packets of a splicing unit. In order to efficiently acknowledge a splicing unit, we extend the structure of original ACK frame to include a new field, that is, PktID[PktNum]. In this way, a single ACK frame can acknowledge multiple data packets of a splicing unit.

3.3. Opportunistic Forwarding Strategy Based on Dynamic Priority. The diversity among the intended forwarder or potential forwarder helpers provides the packet various options to be combined with different numbers of and flows of packets, or not coded at all. To maximize throughput provided by a coded packet, AONC adopts opportunistic forwarding strategy based on a dynamic priority assignment strategy. The priority settings should take data exchange gains and other factors into account. BEND [9] assigns coded transmissions higher priority, based on the number of the original packets, in packet scheduling. However, due to the difference in the length of video data, it cannot accurately indicate the actual amount of original data carried in a coded packet of a certain length. Considering the characteristics of video data, we propose a dynamic priority transmission strategy. When the forwarder nodes contend with each other to transmit their scheduled packets, AONC prioritizes them by assigning them dynamic back-off durations before medium access based on the metric; that is, parameter Z, defined in (7), used to indicate the degree of importance of a coded packet. The prioritization is achieved as follows:

$$\begin{aligned} \mathrm{CW}_{\max} &= \left\lceil \frac{1024}{2^{(Z-1)\times 2}} \right\rceil, \\ \mathrm{CW}_{\min} &= \left\lceil \frac{512}{2^{(Z-1)\times 2}} \right\rceil. \end{aligned} \tag{13}$$

In this way, an encoded packet with higher space utilization is transmitted with a higher priority, thus achieving enhancement of transmission efficiency. For the noncoded

Input: $\bigcup g_k, k \in K$
Output: $\bigcup p_{mn}, m \in M_i, n \in N_{m,J}$
set i, J and $\text{ASUR}_{\max}$ to be 0
for $k = 1 \cdots |C|$
    for each $m \in M'_k$
        calculate $Y_m(S)$ and record $N_{m,S}$
    end for
end for
for $k = 1 \cdots |C|$
    for $S = 1 \cdots L$
        calculate $\text{ASUR}(k, S)$
        if $(\text{ASUR}(k, S) > \text{ASUR}_{\max})$ then
            $\text{ASUR}_{\max} = \text{ASUR}(k, S)$
            $i = k$
            $J = S$
        end if
    end for
end for

Algorithm 1: Packet selection algorithm based on asymmetric coding method.

Figure 5: MAC headers of AONC.

data packets, parameter Z in (13) is set to be 1. On the one hand, it coordinates potential forwarders' accesses to make best use of the coding repository. The nodes with more efficient combination (i.e., an encoded packet with higher space utilization) can capture the media with higher possibilities. This is beneficial for improving data exchange gain and throughput gain. On the other hand, this method is necessary to effectively alleviate such contention and reduce the collision probability, thereby improving the video transmission reliability.

3.4. Traffic-Aware Data Scheduling Algorithm

3.4.1. Traffic Prediction Method. A prediction method is introduced to predict arriving traffic volume. On this basis, we design a traffic-aware data scheduling mechanism. The aim is to reduce the loss of coding opportunities.

In this work, we make use of adding-weight one-rank local-region method, based on chaos time series theory, to predict the arriving traffic volume in a short period of time. The essence of this method is to find $m+1$ nearest neighbors of the center reference point in the reconstructed phase space and carry out single-step prediction in accordance with the evolution law of these $m+1$ neighbor points. Suppose that we have a time series $\{x(k), k = 1, 2, \ldots, N\}$ of a traffic flow, and the state vector in reconstruction of phase space can be expressed as $X(t) = \{x(t),\ x(t-\tau), \ldots, x[t-(m-1)\tau]\}$. The N, m and τ, respectively, represent the number of phase points, embedded dimension number, and time interval. Then the values around $X(t)$ can be fitted by formula: $X(t+1) = a + bX(t)$.

The above prediction method can be used to predict traffic volume for each fixed interval of time. We refer to the interval of time as I. $X(t+1)$ represents traffic prediction value of the next time interval I. Let t_c be current time. The traffic volume during time interval $(t_c, t_c + \delta I]$ can be denoted as

$$F(t_c + \delta I) = \sum_{\delta \in N^+} X(t_c + \delta). \tag{14}$$

A simulation experiment was designed to verify the applicability of the predicted method. Four different YUV QCIF video clips, that is, foreman, hall, coastguard, and mother-daughter [25], are used as video sources in WMSNs. We randomly select a video forwarding node and then collect the traffic information received from a video flow in a period of time. Data in successive intervals is collected. The interval I is set to be 0.1 s. We use the above prediction method to obtain predicted traffic volume and compare the bursty characteristic of it with that of original traffic volume. Figure 6 shows the comparison of the original traffic data and the predicted data. We can observe that there is no significant difference between original data and predicted data in the first 30 points, while the difference gradually increases in subsequent 30 points. The chaotic sequence is highly sensitive to the initial value. As time goes by, the

Figure 6: Comparison of bursty nature of original data and prediction data.

prediction error is getting larger and larger. In order to obtain a relatively accurate data, we set δ to be 1 in our experiments, that is, predicted traffic volume within neighboring time interval $(t_c, t_c + I]$.

3.4.2. Algorithm Design. We define parameter ω to determine whether the traffic-aware packet scheduling mechanism can be used when a sensor node needs to send data. The mechanism can be used if ω is set to be 1 and is set to be 0, otherwise

$$\omega = \begin{cases} 1, & \sum_{k\in K} |M'_K| + |M'_U| > 0 \wedge \varphi = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

The main idea of data scheduling algorithm is as follows. Data packets belonging to U are selected for transmission if $|M'_U| > 0$. Otherwise, we need to select a appropriate flow to schedule so as to reduce the loss of coding opportunities if $\exists! m \in M_k \wedge |f_m| \geq 1$. The flow selection algorithm is as follows.

Let R_m be the amount of queued data belonging to a data flow at t_c

$$R_m = \sum_{n\in N} l_{mn}, \quad m \in M,\ n \in N_m,\ l_{mn} \leq L. \tag{16}$$

We estimate the amount of queued data belonging to f_m at $t_c + \delta I$ as follows:

$$T_m(t_c + \delta I) \approx R_m + F_m(t_c + \delta I), \quad m \in M. \tag{17}$$

Note that we set δ to be 1 in our experiments. Thus, we can obtain a relatively accurate estimation value with respect to the amount of queued data.

The mean and variance of $T_m(t_c + \delta I)$ can be expressed as follows:

$$\mu_k(t_c + \delta I) = \frac{1}{|g_k|} \sum_{m\in M_k} T_m(t_c + \delta I),$$

$$\sigma_k(t_c + \delta I) = \sqrt{\frac{1}{|g_k|} \sum_{m\in M_k} \left(T_m(t_c + \delta I) - \mu_k(t_c + \delta I)\right)^2}. \tag{18}$$

Reducing the variance is conducive to reduce the loss of coding opportunities. A natural step is to select appropriate flow that minimizes the variance. As a result, the corresponding flow to be selected can be denoted as follows:

$$k = \arg\min_{k\in K} \sigma_k(t_c + \delta I). \tag{19}$$

The pseudocode for flow selection is summarized as Algorithm 2.

4. Performance Evaluation

This section involves thorough performance analyses and evaluation of AONC in simulation methodology. We design and conduct a series of simulation experiments. Section 4.1 describes the evaluation metrics and detailed simulation parameter settings. Section 4.2 presents and analyses the simulation results.

4.1. Performance Metrics. In this work, we use network simulator NS-2 [26] to implement the proposed algorithms. We now evaluate the AONC performance for supporting

```
Input: ⋃ f_m, m ∈ M
Output: flow sign i
set i and σ_max to be 0
if (|M'_U| > 0) then
    selects m ∈ M'_U according to FIFO
    i = m
    return;
end if
for k = 1 · · · |C|
    if (|M'_k| > 0)
        calculate σ_k(t_c + δI)
        if (σ_k(t_c + δI) > σ_max) then
            σ_max = σ_k(t_c + δI)
            for m = 1 · · · |M'_k|
                if (|f_m| > 0) then
                    i = m
                end if
            end for
        end if
    end if
end for
```

Algorithm 2: Flow selection algorithm.

video communication in various scenarios, and compare it with typical network coding mechanisms, that is, COPE and BEND. Four different YUV QCIF video clips, that is, foreman, hall, coastguard, and mother-daughter [25], are used as video sources for our experiments. Each video clip is encoded as a standard MPEG-4 sequence, which consists of 300 video frames. Each video frame is fragmented into packets before transmission, and the maximum transmission packet size is 1000 bytes. The structure of the group of pictures (GOP) is IBBPBBPBB ($N = 9$, $M = 3$). The video source delivers video data to receiver at 30 fps. We fix the data rate at 2 Mbps, without any rate adaptation. Our PHY layer model adopts bit-error-rate (BER), which is set to be 2×10^{-6}, to introduce random packet loss to simulate more realistic.

Three case studies are designed and conducted for evaluating the video transmission performance. We begin with a cross topology. Each peripheral node has three neighbors (the center and two other orthogonal peripheral nodes) and the center node has four neighbors (the peripheral nodes). We place four video flows originating from each of the peripheral nodes and terminating at the opposite node. Next, we use a 3-tier scenario, where tiers 1, 2 and 3 each consist of 3 nodes, referred to as 3-3-3 topologies, respectively. We set the separation distance between tiers, so that video flows between tiers 1 and 3 must use tier 2 node(s) as forwarders.

The distance among nodes of the same tier is small. We place four video flows randomly between tiers 1 and 3, two in the forward direction and two in the reverse direction. Then we generalize to a 5×5 mesh topology to test the performance of AONC in supporting video streaming in larger networks. Under this scenario, a nonperipheral node has eight neighbors, and a corner node has three neighbors. The diameter of the network is four hops. There are two video flows (using foreman and hall as video sources) and three constant bitrate (CBR) flows in the scenarios. The packet rate of these CBR flows is set to be 10 packets per second.

4.2. Simulation Experiments and Analyses. We now analyze and evaluate the performance of proposed AONC in terms of video transmission quality, network throughput, and energy utilization efficiency.

Figures 7(a)–7(c), respectively, provide PSNR value for each video frame in different scenarios. Note that the value is the average of each video frame of different video sequences. As shown in the result, AONC provides the best video transmission quality, next is BEND, followed by COPE. The 802.11 scheme, as expected, performs poorly. When using AONC, the PSNR values fluctuate little, and the average is higher. These adaptive schemes in AONC are not only able to achieve higher network throughput, but also improve the stability of the video transmission.

Figures 8(a)–8(c) show the average PSNR value of each video clip in different scenarios. On the one hand, the average PSNR of AONC is higher than that of the other three schemes. On the other hand, when using AONC, we can see that there is a little difference between the average PSNR values of different video clips. Therefore, it can be concluded that AONC can provide higher quality of video and higher reliability of video transmission. Actually, the video quality (in PSNR) of received sequences, achieved by AONC, is close to that of the original encoded sequences (no Error).

Figure 9 illustrates the reconstructed foreman video in 3-tier scenario, demonstrating that the received video quality of our proposed AONC is better than that obtained with the other three schemes. It is well known that video transmission has a high bandwidth requirement. However, the coding

(a) Cross topology

(b) 3-tier topology

(c) 5 × 5 mesh topology

Figure 7: PSNR value for each video frame.

(a) Cross topology

(b) 3-tier topology

(c) 5 × 5 mesh topology

Figure 8: Average PSNR value for each video clip.

(a) AONC (b) BEND

(c) COPE (d) 802.11

Figure 9: Visual comparison of the reconstructed video (foreman).

method and data scheduling scheme of existing network coding mechanisms degrade the network bandwidth utilization. As shown in Figures 9(b)–9(d), a large amount of packets are lost when BEND, COPE, and 802.11 are used to transmit video data, thereby causing interruptions, freezes, or jerkiness in the received video signal.

The bursty is one of the characteristics of video streaming. With the symmetric coding method, these traditional network coding mechanisms could not efficiently absorb video traffic and thus have limitations on the enhancement of throughput when available bandwidth is limited and insufficient to meet demand. Using traffic-aware data scheduling algorithm, AONC has the capability to capture these potential coding opportunities, thus eliminating the performance limitation in BEND and COPE. Figure 10 compares the throughput achieved of different mechanisms under corresponding scenarios, where cross and 3-tier topology may provide much more coding opportunities than 5×5 mesh topology. As shown in Figure 10(a), the difference of average receiving data rate between the cross topology and the 3-tier topology is the greatest, and AONC always performs better than the other three mechanisms. The throughput gains of AONC, BEND and COPE over 802.11 are plotted in Figure 10(b). We can observe that the throughput gain of AONC is obviously larger than that of BNED, and COPE. The gain we have is in the improvement of end-to-end throughput and reduction in bandwidth consumption.

Figure 11 shows the energy utilization efficiency of different mechanisms. Here we define the energy utilization efficiency as the number of obtained decodable video frames per unit of energy consumption. In WMSNs, sensor nodes consume most energy for data transmission and reception. Obviously, the enhancement of data exchange gain can greatly reduce energy consumption for data transmission. Comparing with the other three schemes, AONC achieves significant improvement in energy utilization efficiency.

(a) Average receiving data rate

(b) Throughput gain

Figure 10: Throughput.

5. Conclusion

The limited resources identify significant challenges for the reliable video transmission over WMSNs. There is great potential for network coding to enhance the performance of video transmission and bandwidth utilization. Recent studies demonstrate that network coding is a promising technique for performance improvement of video communication over WMSNs. However, these existing wireless network coding mechanisms have three limitations: (1) traditional symmetric coding method degrades the performance of video communication; (2) the priority assignment of coded packets may be unreasonable; (3) some potential coding opportunities may be lost when using traditional data scheduling mechanism. The aim of this paper is to overcome these limitations. In particular, we propose an Adaptive Opportunistic Network Coding mechanism (AONC).

Figure 11: Energy utilization efficiency.

The main contributions in our work are as follows. First, considering the characteristics of compressed video stream, we introduce a novel asymmetric coding method to improve data exchange gain. Second, an opportunistic forwarding strategy with dynamic priority is designed to ensure that packets have a better chance to be coded and transmitted. The purpose is to achieve higher throughput. Finally, we present a traffic-aware data scheduling algorithm, working together with above network coding mechanism, to reduce the loss of potential coding opportunities. We implement the AONC in NS-2 and carry out extensive evaluation. The simulation results demonstrate that, compared with the existing typical network coding mechanisms, AONC greatly improves the video transmission quality and achieves significant gains in terms of bandwidth and energy efficiency. As a result, we determine that the proposed new techniques are feasible for video communication in WMSNs.

Our ongoing work focuses on the implementation and validation of the proposed algorithms on the prototype system. In addition, one possible future direction of this work is how to provide resiliency network coding opportunity when traffic congestion occurs.

Acknowledgments

The authors would like to thank the anonymous reviewers for the constructive and valuable comments. The authors gratefully acknowledge the support and financial assistance provided by the National Natural Science Foundation of China under Grant no. 60673185 and 61073197, the Natural Science Foundation of Jiangsu Province under Grant no. BK2010548, the Scientific and Technological Support Project (Industry) of Jiangsu Province under no. BE2011186, the State Key Laboratory Program for Novel Software Technology (Nanjing University) under Grant no. KFKT2010 B08, and the Research Innovation Program for College Graduates in Jiangsu Province under Grant no. CXLX11 0262.

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdury, "Wireless multimedia sensor networks: a survey," *IEEE Wireless Communications*, vol. 16, no. 6, pp. 32–39, 2007.

[2] H. D. Ma and D. Tao, "Multimedia sensor network and its research progresses," *Journal of Software*, vol. 17, no. 9, pp. 2013–2028, 2006.

[3] R. Dai, P. Wang, and I. F. Akyildiz, "Correlation-aware QoS routing with differential coding for wireless video sensor networks," *IEEE Transactions on Multimedia*, vol. 14, no. 5, pp. 1469–1479, 2012.

[4] T. Melodia and I. F. Akyildiz, "Cross-layer QoS-aware communication for ultra wide band wireless multimedia sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 5, pp. 653–663, 2010.

[5] 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.

[6] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: practical wireless network coding," in *Proceedings of ACM SIGCOMM*, pp. 243–254, Pisa, Italy, 2006.

[7] H. Seferoglu, A. Markopoulou, and K. K. Ramakrishnan, "I^2NC: intra- and inter-session network coding for unicast flows in wireless networks," in *Proceedings of the IEEE INFOCOM 2011*, pp. 1035–1043, Shanghai, China, April 2011.

[8] S. Rayanchu, S. Sen, J. Wu, S. Baneriee, and S. Senqupta, "Loss-aware network coding for unicast wireless sessions: design, implementation, and performance evaluation," in *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '08)*, pp. 85–96, Annapolis, Md, USA, June 2008.

[9] J. Zhang, Y. P. Chen, and I. Marsic, "MAC-layer proactive mixing for network coding in multi-hop wireless networks," *Computer Networks*, vol. 54, no. 2, pp. 196–207, 2010.

[10] Y. Yan, B. Zhang, J. Zheng, and J. Ma, "CORE: a coding-aware opportunistic routing mechanism for wireless mesh networks," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 96–103, 2010.

[11] S. Han, Z. Zhong, H. Li, G. Chen, E. Chan, and A. K. Mok, "Coding-aware multi-path routing in multi-hop wireless networks," in *Proceedings of the IEEE International Performance Computing and Communications Conference (IPCCC '08)*, pp. 93–100, Austin, Tex, USA, December 2008.

[12] J. Le, J. C. S. Lui, and D. M. Chiu, "DCAR: distributed coding-aware routing in wireless networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 4, pp. 596–608, 2010.

[13] S. Sengupta, S. Rayanchu, and S. Banerjee, "Network coding-aware routing in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1158–1170, 2010.

[14] H. Seferoglu, A. Markopoulou, and U. Kozat, "Network coding-aware rate control and scheduling in wireless networks," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '09)*, pp. 1496–1499, New York, NY, USA, July 2009.

[15] K. Chi, X. Jiang, and S. Horiguchi, "Joint design of network coding and transmission rate selection for multihop wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2435–2444, 2010.

[16] C. S. Hwang and H. Moon, "Rate adaptation for wireless network coding using partial overhearing," *IEEE Communications Letters*, vol. 13, no. 12, pp. 890–892, 2009.

[17] H. Seferoglu and A. Markopoulou, "Video-aware opportunistic network coding over wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 1–16, 2009.

[18] H. Seferoglu and A. Markopoulou, "Delay-optimized network coding for video streaming over wireless networks," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, Cape Town, South Africa, May 2010.

[19] Y. Xuan and C. T. Lea, "Network-coding multicast networks with QoS guarantees," *IEEE/ACM Transactions on Networking*, vol. 19, no. 1, pp. 265–274, 2011.

[20] S. Yang and J. Wu, "Efficient broadcasting using network coding and directional antennas in MANETs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 2, pp. 148–161, 2010.

[21] A. Asterjadhi, E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "Toward network coding-based protocols for data broadcasting in wireless Ad Hoc networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 662–673, 2010.

[22] IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 802. 11 Standard, 1997.

[23] IEEE Standard 802. 15. 4, Part 15. 4: Wireless Medium Access Control and Physical Layer Specification for Low Rate Wireless Personal Area Networks. IEEE Std. 802. 15. 4, December 2003.

[24] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "Wireless multimedia sensor networks: applications and testbeds," *Proceedings of the IEEE*, vol. 96, no. 10, pp. 1588–1605, 2008.

[25] Video Trace Library, http://trace.eas.asu.edu/yuv .

[26] "The Network Simulator-ns," http://nsnam.isi.edu/nsnam/index.php/Main_Page.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 235374, 8 pages
doi:10.1155/2012/235374

Research Article

Interference-Aware Fault-Tolerant Energy Spanner in Wireless Ad Hoc Networks

Dong Wang, Wang-Chen Long, and Xiao-Hong Li

College of Information Science and Engineering, Hunan University, Changsha 410082, China

Correspondence should be addressed to Xiao-Hong Li, lixhong@263.net

Received 20 July 2012; Accepted 21 November 2012

Academic Editor: Ruchuan Wang

Copyright © 2012 Dong 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.

Power assignment in wireless ad hoc networks is an important issue of topology control which assigns power for each wireless node so that the induced communication graph satisfies some desired properties such as the connectivity and the energy spanner. In this paper, we study the problem of power assignment in order that its induced communication graph meets the following properties: (1) it is an energy-t-spanner which is energy efficient; (2) it is k-fault resistant which can withstand up to $k-1$ node failures where $k \geqslant 1$; (3) the interference is minimal. We propose algorithms to address this problem. Both the theoretic analysis and the simulations in the paper prove that our algorithms can induce a k-fault resistant energy spanner and furthermore the interference is minimized. To the best of our knowledge, this is the first paper to study the power assignment problem simultaneously considering spanner properties, the fault tolerance, and the interference reduction.

1. Introduction

Ad hoc networks are formed by autonomous nodes communicating via radio without any additional backbone infrastructure. Ad hoc networks have received significant attention in recent years due to their potential civilian and military applications. In wireless ad hoc networks, each node has limited resources such as energy, computing power, storage capacity; there are more challenges and problems compared with traditional fixed infrastructure networks.

A fundamental problem in wireless ad hoc network is to find a power assignment so that the induced communication graph can satisfy some properties such as connectivity and energy spanner.

An energy-t-spanner G' is a subgraph of G, such that for any two nodes u and v in G', there exists a path from u to v, whose energy is at most t times the energy of a minimum-energy path from u to v in the original communication graph G. The constant t is called the power stretch factor. A small power stretch factor implies low energy spent by relay nodes in propagating a message, which is extremely useful for prolonging the lifetime of the network. Due to limited power sources, the idea of energy-t-spanner becomes an important design consideration in ad hoc networks. Much effort has been devoted to finding a power assignment that the induced graph is energy-t-spanner [1–3].

Nodes in a wireless network are typically battery powered, and it is infeasible or unable to recharge the device. Due to constrained power capacity, hostile deployment environment, and other factors, node failures are more likely to happen, which might cause network partitions and badly degrade network performance. Therefore, it is important to construct k-fault resistant topologies which can withstand up to $k-1$ node failures or we can say it is k-connected. Shpungin and Segal [4] studied how to find a power assignment that the induced communication graph is energy spanner and fault tolerant.

In wireless networks, a node is not able to receive correct data from its neighbor if any of other neighbors is transmitting at the same time. This mutual disturbance of communication is called interference. Interference which causes collisions and retransmissions has a negative impact on prolonging network lifetime in wireless networks. Reducing interference in wireless network leads to fewer collisions

and packet retransmissions, which indirectly extends the lifetime of the network and improves network performance. Reducing the interference in wireless networks is consequently considered one of the foremost goals [5]. In order to make topology withstand up to $k-1$ node failures, node's transmission power must be improved. And interference will increase dramatically if the transmission power improved improperly. However, previous work studies the fault-tolerant spanner problem but ignores interference reduction. In this paper, we will study the power assignment problem that the induced communication graph is k-fault resistant and energy-t-spanner; in addition, the interference is minimized.

The rest of the paper is organized as follows. Section 2 presents an overview of previous related work. In Section 3, we state the problem studied in this paper. The models will be presented in Section 4. Then the details of our power assignment algorithms and simulation results will be given in Section 5. Finally, we conclude this paper in Section 6.

2. Related Work

To assign power to each node of a wireless network so that the induced communication graph can satisfy some desired properties (such as connectivity, spanner) is an important research topic. Chew [6] first introduced the concept of spanners. Recently, several papers focus on dealing with the problem of power assignment simultaneously with spanning properties.

Wang and Li [1] first studied how to find a power assignment that the induced communication graph is energy spanner with objectives that minimizes the maximum node power and the total energy consumption (also referred to as the cost of the power assignment $c(p)$). They also presented two heuristics for the construction of a low cost power assignment with an energy spanner property for unit disk graphs.

Shpungin and Segal [2] studied the spanner problem from a theoretical point of view under two optimization objectives: minimizing the cost of the power assignment and keep the spanner property. They studied both energy spanner and distance spanner. For the energy spanner model, they present a basic construction of a power assignment p, so that the resulting network is an energy-2-spanner with a total energy of at most $O(\log n)$, the optimum cost power assignment $c(p^*)$ in $O(n^4 \log n)$, where n is the number of nodes in the network, p^* denote an optimal power assignment; that is, a power assignment of minimum cost, for which the induced communication graph is strongly connected.

Abu-Affash et al. [3] studied the minimal energy spanner problem, they proved that for any constant $t > 1$, one can compute in $O(n^2 \log n)$ time a planar energy-t-spanner G', such that the cost of power assignment $c(p)$ implied by G' is at most $2(1+2/(t-1)) \times c(p^*)$, where p^* is a minimum-cost power assignment.

All these papers study the spanner problem with additional objective of minimizing the total node power. In many scenarios wireless ad hoc networks are deployed in hostile environments where node failures are very likely to happen. In order to keep network connectivity in wireless networks which is a node-failure prone environment, fault tolerant is especially important. Previous works addressing fault tolerance usually construct a k-connected topology by improving node's transmission range greatly; there are some recent works addressing fault tolerance [7–9]. However, in these works addressing fault tolerance fails to consider the spanner of the networks. Shpungin and Segal [4] studied the power assignment problem which combines fault resistance with spanner property. They addressed the minimum power k-fault resistant energy spanner problem (MPkES). For $k \in \{1, 2\}$, they proposed several power assignments which obtain a good bicriteria approximation on the total cost and spanner property. For $k > 2$, they also analyzed a power assignment P_k presented in [10] that the result topology after the power assignment P_k can satisfy both k-fault resistant and energy spanner property. In order to consider the fault resistant, we only discuss the situation when $k > 2$. For $k = 2$, Shpungin and Segal [4] developed a power assignment p_2 such that the result topology is 2-strongly connected and both the energy and distance stretch factors are 1. It is obvious that all the minimum-energy paths and distance paths are preserved in the resulting topology, or we can say that the transmission range of each node is still large. For $k = 3$, a power assignment P_k is analyzed. As Figure 1 illustrated, firstly, node t determines all the neighbor nodes and let N_t be a set of k closest nodes to t. Then the transmission power of node r_t^* is assigned to be the value that node t can reach the farthest nodes in N_t. After the power assignment of r_t^*, an MST of the original topology is computed, for each edge $e = (t, s)$ of MST increase the range of the nodes in $N_t \cup N_s$ such that each node $t' \in N_t$ can reach all nodes in N_s, and vice versa. In [4] Shpungin and Segal proved that the result topology after the power assignment of P_k is k-fault resistant and an energy spanner of the original topology. In [4], Shpungin and Segal addressed the k-fault resistant energy spanner problem; however, interference is ignored. As can be seen from Figure 1, if any one node of $N_t \cup N_s$ is working, all the other nodes in $N_t \cup N_s$ will be interfered.

Rickenbach et al. [5] argued that reducing the interference in wireless networks is considered one of the foremost goals. Interference which causes collisions and retransmissions has a negative impact on prolonging network lifetime in wireless networks. An interesting combination of fault resistance, energy spanner, and interference reducing is studied in this paper.

3. Problem Statement

In this paper, we study the Interference Minimal Fault-tolerant Energy Spanner (IMFES) problem: Let $V = \{v_1, v_2, \ldots, v_n\}$ be a set of n wireless nodes distributed in wireless network. For any $t \geq 1$, $k \geq 1$, find a power assignment that its induced communication graph is k-fault resistant energy-t-spanner G' of the original communication

Figure 1: The process of P_k.

graph G. Furthermore, the interference of G' is minimal. More formally, the problem can be defined as follows:

Input: a set of n wireless nodes $V = \{v_1, v_2, \ldots, v_n\}$, and two real constants $t \geq 1$, $k \geq 1$.

Output: a k-fault resistant energy-t-spanner $G' = (V, E')$ of original communication graph G.

Object: minimize the interference in G'.

4. Models

We consider a 2D wireless network that consists of a set V of n wireless nodes distributed in a 2D plane R^2. Each node can adjust the transmission range from 0 to $T_{\max}$ and each node has a unique *id* (such as IP/MAC address). A generic model of an ad hoc network is a unit disk graph (UDG) in a 2D plane. There is an edge $e = (u, v)$ between two nodes u and v if and only if the Euclidean distance $\|u, v\|$ between u and v in R^2 is at most $T_{\max}$.

When using common path loss model, the signal strength received by a node can be described as p/d^{β}, where p is the transmission power used by the sending node, β is a real constant between 2 and 5 depending on the wireless transmission environment, and d is the distance between the pair of communication node. Consequently, the energy cost to send a message of fixed length from node u to node v is $\|u, v\|^{\beta}$. Each node can calculate the distance from its neighbor node by the received power using the signal degradation model. For simplicity, in this paper we assume that β is 2, although our results can be easily extended to any constant.

For two nodes $u, v \in V$, let $\gamma(u, v, G)$ denote the minimum energy consumption from u to v in G.

We define the notion of an energy-t-spanner of G.

Definition 1. A graph $G' = (V, E')$ is an energy-t-spanner of $G = (V, E)$, if $E' \subseteq E$, and for each pair $u, v \in V$, $\gamma(u, v, G') \leq t \times \gamma(u, v, G)$.

Here, we also give the definition of k-connected used in this paper.

Figure 2: Interference of edge (u, v).

Definition 2. A graph $G = (E, V)$ is k-connected if for any two vertices u and $v \in V$, there are k vertex disjoint paths from u to v. Or equivalently, a graph is k-vertex connected if the removal of any $k - 1$ nodes does not partition graph G.

In order to measure the interference, we define the interference model first. Generally, a node v may not correctly decode the transmission from node u if the signal to interference and noise ratio (SINR) perceived by v is below a certain threshold. While this threshold is dependent on various factors like antenna sensitivity of receivers, signal modulation techniques, and other environmental factors, it is well understood that a third node w interferes node v's signal reception from node u if w is located at a nearby position of v and transmitting simultaneously with u. The distance (region) within which a node w interferes another node is called the interference range (region) of w. For simplicity of analysis, we assume that the interference range and the transmission range are the same for any node w. However, the solution is also applicable where the assumption does not hold. All common models of interference define the notion of coverage, which is the number of nodes or edges that are affected by a transmission over a specific link in the induced communication graph. For simplicity, in this paper, the interference of a node v is defined as the number of nodes covered by v with its transmission range.

Definition 3. The interference value of a single node v is defined as

$$I(v) = |\{u \mid u \in V - \{v\},\ u \in D(v, r_v)\}|, \tag{1}$$

where $D(v, r_v)$ stands for the transmission circle with node v in its center and the transmission radius is r_v.

Definition 4. The interference of an edge $e = (u, v)$ is defined as the number of nodes which are covered by either of the nodes u and v in edge e,

$$I(e) = |\{w \mid w \in D(u, r_u) \cup w \in D(v, r_v), w \in V\}|. \tag{2}$$

Definition 5. The interference of graph $G = (V, E)$ is defined as the maximum edge interference in G. That is;

$$I(G) = \operatorname*{Max}_{e \in E} I(e). \tag{3}$$

As can be seen in Figure 2, the interference of node u is $I(u) = 3$, the interference of node v is $I(v) = 4$, and the

interference of edge $e = (u, v)$ is $I(e) = 6$, where w is an interference node to both u and v, but it will be calculated only once.

Based on these definitions, we will address the problem described in Section 3.

5. Interference Minimal Fault-Tolerant Energy Spanner

In this section, we address the Interference Minimal Fault-tolerant Energy Spanner problem and propose algorithm IMFES, which produces a k-fault resistant energy-t-spanner; furthermore, the interference is minimized.

IMFES first calculates the interference of each edge in the original communication graph according to the interference model and then sorts all edges by ascending order of weight. In order to guarantee the unique outcome of the greedy algorithms that will be proposed in the second step, we have to ensure that two edges with different end nodes have different weights;

$$\begin{aligned}
&\text{Weight } (u_1, v_1) > \text{Weight } (u_2, v_2)\\
&\quad \Longleftrightarrow I(u_1, v_1) > I(u_2, v_2)\\
&\quad \Longleftrightarrow I(u_1, v_1) = I(u_2, v_2) \&\& \max\{id(u_1), id(v_1)\} >\\
&\quad \max\{id(u_2), id(v_2)\}\\
&\quad \Longleftrightarrow I(u_1, v_1) = I(u_2, v_2) \&\& \max\{id(u_1), id(v_1)\} =\\
&\quad \max\{id(u_2), id(v_2)\} \&\& \min\{id(u_1), id(v_1)\} >\\
&\quad \min\{id(u_2), id(v_2)\}.
\end{aligned}$$

In the first step, IMFES computes an interference minimal k-connected graph based on the greedy algorithm. In the second step, IMFES would check all the rest edges which are not added to G_{IMFES} in the first step. To ensure that G_{IMFES} is an energy-t-spanner, an edge $e = (u, v)$ would be added to G_{IMFES} if $\gamma(u, v, G_{\text{IMFES}}) > t \times \gamma(u, v, G)$. The detail of IMFES is described as in Algorithm 1.

IMFES:
Input: A set V of n wireless nodes, $t \geq 1, k \geq 1$.
Output: A k-fault resistant energy-t-spanner.
First Step:
1: Compute original topology $G = (V, E)$
2: For each edge $e \in E$, compute $I(e)$
3: sort all edges in E by ascending weight.
4: $G_{\text{IMFES}} = (V, E_{\text{IMFES}} = \varnothing)$
5: For each edge $e = (u, v)$ in $E(G)$
6: if u and v is not k connected in G_{IMFES}
7: $E_{\text{IMFES}} = E_{\text{IMFES}} \cup \{e\}$
8: $E(G) = E(G) - \{e\}$
9: else if all nodes are k connected in G_{IMFES}
10: go to step two
11: end if
12: end for
Second Step:
13: For each edge $e = (u, v)$ in $E(G)$
14: if $\gamma(u, v, G_{\text{IMFES}}) > t \times \gamma(u, v, G)$
15: $E_{\text{IMFES}} = E_{\text{IMFES}} \cup \{e\}$
16: end if
17: end for
18: Return G_{IMFES}

Algorithm 1

5.1. Analyze. In this section, we will prove that the graph induced by IMFES is a k-connected energy spanner with minimal interference. In the first step, IMFES sorts all edges by ascending weight then uses greedy algorithm to construct a k-connected subgraph of original communication graph. Lines 5–12 are a generalized version of Kruskal's algorithm [11] for $k \geq 2$.

Let the path $P = \{u, w_1, w_2, \ldots, w_n, v\}$ from u to v be represented by an ordered set P of vertices on the path. Let $S_{uv}(G)$ be the sum of pairwise internally vertex disjoint paths from u to v in graph G. Thus, for any path $P_1, P_2 \in S_{uv}(G)$, we have $P_1 \cap P_2 = \{u, v\}$.

Lemma 6. *Let u_1 and u_2 be two vertices in a k-connected undirected graph G. If u_1 and u_2 are k-connected after the removal of edge (u_1, u_2), then $G - (u_1, u_2)$ is still k-connected.*

Proof. In order to prove $G - (u_1, u_2)$ is k-connected, let $G' = G - (u_1, u_2)$, and it is equivalent to prove that G' is connected after the removal of any $k - 1$ vertices in G'. Without loss of generality, assume $\{u_1, u_2\} \cap \{v_1, v_2\} = \varnothing$ (other cases can be proved using a similar approach). We now prove that v_1 is still connected to v_2 after removal of the set of any $k - 1$ vertices $W = \{w_1, w_2, \ldots, w_{k-1}\}$, where $w_i \in V(G') - \{v_1, v_2\}$. Let $\mathbf{G}''$ be the resulting graph after (u_1, u_2) and the $k - 1$ vertices in W are removed from G. Now we consider two cases.

(1) There is an edge (v_1, v_2) in G:

since all the paths in $S_{v_1 v_2}(G)$ are pairwise internally vertex disjoint paths, the removal of (u_1, u_2) and the $k - 1$ vertices does not break the path $\{v_1, v_2\}$ which contain only one edge (v_1, v_2). Thus, v_1 is still connected to v_2 after removal of the set of any $k - 1$ vertices $W = \{w_1, w_2, \ldots, w_{k\text{-}1}\}$.

(2) There is no edge from v_1 to v_2 in G:

according to the definition of $S_{v_1 v_2}(G')$, it can be got that the removal of $k - 1$ vertices at most breaks $k - 1$ paths in $S_{v_1 v_2}(G')$. If we can prove that $|S_{v_1 v_2}(G')| \geq k$, then v_1 is still connected to v_2 after removal of the set of any $k - 1$ vertices $W = \{w_1, w_2, \ldots, w_{k-1}\}$.

For the sake of contradiction, we assume that $|S_{v_1 v_2}(G')| < k$. If $|S_{v_1 v_2}(G')| < k$, this occurs only when the removal of (u_1, u_2) breaks one path $P_0 \in S_{v_1 v_2}(G)$, because G is k-connected and $S_{v_1 v_2}(G) \geq k$. Since P_0 is internally

disjoint with all paths in $S_{v_1v_2}(G\prime)$, we have $P_0 \cap W = \emptyset$. Thus, v_1 is connected to u_1 and u_2 is connected to v_2 in G''. We already know that u_1 and u_2 are k-connected after the removal of edge (u_1, u_2), so we can get that $S_{u_1u_2}(G') \geq k$. Since v_1 is connected to u_1 and u_2 is connected to v_2, we can get that $|S_{v_1v_2}(G')| \geq k$, which is a contradiction. Therefore, $|S_{v_1v_2}(G')| \geq k$.

We have proved that for any two vertices $v_1, v_2 \in G'$, v_1 is connected to v_2 after the removal of any $k-1$ vertices from $G' - \{v_1, v_2\}$. Therefore, G' is k-connected. □

Lemma 7. *Let G and G' be two undirected simple graphs such that $V(G) = V(G')$. If G is k-connected, and every edge $(u, v) \in E(G) - E(G')$ satisfies that u is k-connected to v in $G - \{(u_0, v_0) \in E(G) : \text{Weight } (u_0, v_0) \geq \text{Weight } (u, v)\}$, then G' is also k-connected.*

Proof. Let $E = E(G) - E(G') = \{(u_1, v_1), (u_2, v_2), \ldots, (u_m, v_m)\}$ be a set of edges in an descending order of interference weight, that is, Weight $(u_1, v_1) \geq$ Weight $(u_2, v_2) \geq \cdots \geq$ Weight (u_m, v_m). We define a series of graphs that is subgraphs of $G : G^0 = G$, *and* $G^i = G^{i-1} - (u_i, v_i)$, $i = 1, 2, \ldots, m$. Now we prove the following by induction.

(1) Base: $G^0 = G$ is k-connected.

(2) Induction: if G^{i-1} is k-connected, and u_i is k-connected to v_i in $G - \{(u, v) \in E(G) : \text{Weight } (u, v) \geq \text{Weight } (u_i, v_i)\}$. Since $G - \{(u, v) \in E(G) : \text{Weight } (u, v) \geq \text{Weight } (u_i, v_i)\} \subseteq G^{i-1} - (u_i, v_i)$, u_i is k-connected to v_i in $G^{i-1} - (u_i, v_i)$. Applying Lemma 6 to G^{i-1}, we can prove that G^i is still k-connected.

Now we have proved that G^m is k-connected. Since $E(G^m) \subseteq E(G')$, G' is also k-connected. □

We denote by G^1_{IMFES} the graph that produced by IMFES after the first step and by G_{IMFES} the final induced communication graph.

Lemma 8. *G^1_{IMFES} is k-connected if G is k-connected.*

Proof. Since edges are inserted into G^1_{MIFES} in an ascending order, whether u is k-connected to v at the moment before (u, v) is inserted depends only on the edges of smaller interference weight. Assume that edge (u_0, v_0) is the last edge added into G^1_{IMFES}. Therefore, every edge $(u, v) \in E(G) - E(G^1_{\text{IMFES}})$ satisfies that u is k-connected to v in $G - \{(u, v) \in E(G) : \text{Weight } (u, v) > \text{Weight } (u_0, v_0)\}$. We can prove that G'_{IMFES} preserves the k-connectivity of G by applying Lemma 7. □

Theorem 9. *G_{IMFES} is k-connected if G is k-connected.*

Proof. From Lemma 8 we can get that after the first step of IMFES, G^1_{IMFES} is k-connected if G is k-connected. While in the second step of IMFES, the only operation is adding edges into G^1_{IMFES}. It is obvious that after the second step of IMFES, G_{IMFES} is k-connected if G is k-connected. □

Lemma 10. *Let u_1 and u_2 be two vertices in an energy-t-spanner G_s of graph G. If $\gamma(u_1, u_2, G_s - (u_1, u_2)) \leq t \times \gamma(u_1, u_2, G)$ after the removal of edge (u_1, u_2), then $G_s - (u_1, u_2)$ is still an energy-t-spanner of graph G.*

Proof. Let $G'_s = G_s - (u_1, u_2)$ and LP_{uv} the minimum-energy path from u to v in G. We prove the following by contradiction. Assume that G'_s is not an energy-t-spanner of graph G. Without loss of generality (other cases can be proved using a similar approach), assume $\{u_1, u_2\} \cap \{v_1, v_2\} = \emptyset$ and $\gamma(v_1, v_2, G'_s) > t \times \gamma(v_1, v_2, G)$. This occurs only because edge (u_1, u_2) belongs to $LP_{v_1v_2}(G_s)$, the minimum-energy path from v_1 to v_2. Since after the removal of edge (u_1, u_2), it still has $\gamma(u_1, u_2, G'_s) \leq t \times \gamma(u_1, u_2, G)$, by substituting edge $e = (u_1, u_2)$ with $LP_{u_1u_2}(G_s)$, it can get $\gamma(v_1, v_2, G'_s) \leq t \times \gamma(v_1, v_2, G)$ which leads to a contradiction. □

Theorem 11. *G_{IMFES} is an energy-t-spanner of G.*

Proof. To show that G_{IMFES} meets the spanner property, it is equivalent to prove that for any pair of nodes $\{u, v\} \in V$, $\gamma(u, v, G_{\text{IMFES}}) \leq t \times \gamma(u, v, G)$.

Let $E_0 = E(G) - E_{\text{IMFES}}$. We prove the following by contradiction; assume there exists two nodes $\{u_0, v_0\} \in V$, that $\gamma(u_0, v_0, G_{\text{IMFES}}) > t \times \gamma(u_0, v_0, G)$. Let p_0 be the minimum-energy path from u_0 to v_0 in G_{IMFES}, in order to satisfy that $\gamma(u_0, v_0, G_{\text{IMFES}}) \leq t \times \gamma(u_0, v_0, G)$, at least one edge $e = (u_m, v_m)$ in E_0 has to be added into p_0, and let $p_1 = p_0 + (u_m, v_m)$. Without loss of generality, assume $\gamma(u_0, v_0, G_{\text{IMFES}} + (u_m, v_m)) \leq t \times \gamma(u_0, v_0, G)$. According to lines 14–16 of IMFES, it follows that any edge $(u_i, v_j) \in E_0$ satisfies $\gamma(u_i, v_j, G_{\text{IMFES}}) \leq t \times \gamma(u_i, v_j, G)$. By applying Lemma 10, it can lead $\gamma(u_0, v_0, G_{\text{IMFES}}) \leq t \times \gamma(u_0, v_0, G)$ which leads to a contradiction. Therefore, G_{IMFES} is an energy-t-spanner of G. □

Lemma 12. *The interference of G^1_{MIFES} is minimized.*

Proof. Let $SS_k(G)$ be the set of all$k$-connected subgraphs of G. By Lemma 8 we can get that G^1_{IMFES} is also k-connected if G is k-connected. Suppose that (u, v) is the last edge inserted into G^1_{IMFES} in the first step. We can get that Weight (u, v) is larger than any edge which has been already inserted into G^1_{IMFES} and $I(G^1_{\text{IMFES}}) = I(u, v)$. Let $G_2 = G^1_{\text{IMFES}} - (u, v)$, and we can get that $|S_{uv}(G_2)| < k$ otherwise; according to the line 6 of algorithm IMFES, edge (u, v) should not be included in G_{IMFES}.

Now consider a graph $C = (V(C), E(C))$, where $V(C) = V(G)$ and $E(C) = \{(u_0, v_0) \in E(G) : \text{Weight } (u_0, v_0) < \text{Weight } (u, v)\}$. If we can prove that C is not k-connected, we will be able to conclude that any subgraph $F \in SS_k(G)$ must have at least one edge with interference equal to or larger than $I(u, v)$, which means is $I(G^1_{\text{IMFES}})$ is minimal interference.

We prove by contradiction that C is not k-connected. Assume that C is k-connected and hence $|S_{uv}(C)| \geq k$. We have $E(C) \not\subset E(G_2)$; otherwise, $|S_{uv}(G_2)| \geq |S_{uv}(C)| \geq k$ which contradict with the definition of G_2. Therefore, $E_0 = E(C) - E(G_2) \neq \emptyset$. Since edges are inserted into G^1_{IMFES} in

an ascending order, for all $(u_1, v_1) \in E_0$ satisfies that u_1 is k-connected to v_1 in $C - \{(u_0, v_0) \in E(C) : \text{Weight } (u_0, v_0) \geq \text{Weight } (u_1, v_1)\}$. By Lemma 7, we can prove that u is still k-connected to v after the removal of all edges in E_0. This means $|S_{uv}(G_2)| \geq k$, which is a contradiction. So after step one, the interference of G^1_{IMFES} is minimized. □

Theorem 13. *The interference of G_{IMFES} is minimized.*

Proof. Lemma 12 shows that the interference of G^1_{IMFES} is minimized; in order to prove the interference of G_{IMFES} is minimized, we only consider are all the edges added in the second step of IMFES are minimal interference. Assume that edge (u, v) is the last edge inserted into G_{IMFES} in the second step. Let $G_3 = G_{\text{IMFES}} - (u, v)$; according to line 14 of IMFES, G_3 is not an energy-t-spanner.

Assume that F is k-connected energy-t-spanner, $E(G^1_{\text{IMFES}}) \subseteq E(F)$ and $I(F) < I(G_{\text{IMFES}})$; it follows that for any edge $(u_i, v_j) \in E(F), I(u_i, v_j) < I(u, v)$. Since edges are checked in an ascending order in the second step of IMFES, it follows that for any edge $(u_i, v_j) \in E(F) - E(G^1_{\text{IMFES}})$, edge (u_i, v_j) is checked before edge (u, v). Let $E_1 = E(F) - E(G_3)$; according to the line 14 of IMFES, any edge $(u_m, v_m) \in E_1$ satisfies that $\gamma(u, v, G_3) \leq t \times \gamma(u, v, G)$. By applying Lemma 10, it can be got that G_3 is an energy-t-spanner which is a contradiction.

Thus, The interference of G_{IMFES} is minimized. □

Theorem 14. *The time complexity of IMFES is $O(m(n + m))$ and can be $O(m \log m)$ when $k \leq 3$, where n is the number of vertices and m is the number of edges in the graph.*

Proof. All the edges can be sorted in $O(m \log m)$ time in line 3, where m is the number of edges in the graph. A query on whether two vertices are k-connected can be answered in $O(n + m)$ time for any fixed k by using network flow techniques [12], where n is the number of vertices in the graph. For $k \leq 3$, there also exits $O(1)$ time algorithms [13]. Therefore, the time complexity of lines 5–12 is $O(m(n + m))$ and can be improved to $O(m)$ for $k \leq 3$. A query on the minimum-energy path between two nodes can be finished in $O(n^2)$ using Dijkstra's algorithm [14]. Thus, the time complexity of lines 13–17 is $O(mn^2)$.

Therefore, the time complexity of IMFES is $O(m(n + m))$ and can be $O(mn^2)$ when $k \leq 3$. □

5.2. Simulations. In this section, we evaluate the performance of IMFES and P_k, the power assignment analyzed in [4], by conducting simulations in random networks. The reason why we compare our work with [4] is that [4] is the previous best work closely matches with our work.

We analyze the performance of IMFES with P_k using the following evaluation metrics.

Interference. We measure the level of interference of a network based on the interference graph model proposed in Section 4.

Figure 3: Average edge interference when the induced topologies are 2-connected.

Power Stretch Factor. A low value of power stretch factor implies low energy spent by relay nodes in propagating a message, or we can say that this topology is energy efficient.

We will evaluate IMFES and P_k under different fault tolerance requirements: $k = 2$ and $k = 3$. We also evaluate IMFES under different values of parameter t. The parameter t can affect the power stretch factor of the resulting topology and also have an impact on the network interference reduction.

In our experiments, we randomly generate a set V of n wireless nodes in a 1000×1000 square, only if the original communication graph is k-connected and then we run these algorithms. The maximum transmission range R is set to 250; the power constant of path loss exponent β is set to 2. We vary the number of node from 50 to 175. Each result is the average of 100 runs.

Figure 3 shows the average edge interference of the resulting 2-connected topologies induced by the power assignment P_2 and IMFES. As Figure 3 shows, the average edge interference of the topology induced by IMFES is much lower then the topology induced by P_2, especially when $t = 3$. When the number of nodes increases, the interference of the topology induced by P_2 increases dramatically, while IMFES can keep the interference at a low level no matter what the value of t is.

Figure 4 shows the average edge interference of the resulting 3-connected topologies induced by the power assignment P_k and IMFES. Figure 4 shows that IMFES outperforms P_k, which is almost the same result as Figure 3 shows. From Figures 3 and 4, we can also conclude that the larger the parameter t is, the lower the value of edge interference will be in the topology induced by IMFES, that is because more edges would be kept in the resulting topology in order to achieve a lower spanner property. More edges are kept in a

Figure 4: Average edge interference when the induced topologies are 3-connected.

Figure 5: Power stretch factor when the induced topologies are 2-connected.

topology which implies that nodes in this topology have a larger transmission range; consequently, the interference of the network increases.

Figures 5 and 6 show the power stretch factor of the resulting topologies induced by P_2/P_k and IMFES under different fault tolerance requirements: $k = 2$ and $k = 3$. The power stretch factors of all the topologies are less than 3, which means that they are energy efficient.

From Figures 3 to 4, we have got that the smaller the parameter t is, the larger the value of edge interference will be in the topology induced by IMFES. Figures 5 and 6 show that

Figure 6: Power stretch factor when the induced topologies are 3-connected.

the smaller the parameter t is, the smaller the power stretch factor will be in the topology induced by IMFES. Therefore, there is a tradeoff between energy efficient and interference reduction.

6. Conclusion

In this paper, we study how to find a power assignment that the induced communication graph is k-fault resistant energy-t-spanner, in addition that the interference is minimal in it. We propose IMFES to address this problem in two-dimensional wireless ad hoc networks. To the best of our knowledge, this is the first paper to study the problem. We also prove that the communication graph induced by IMFES is an interference minimal k-connected energy-t-spanner, and the energy stretch factor t ($t \geqslant 1$) can be a low value, which means that the induced communication graph is energy efficient.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (no. 61272061, no. 61202289, and no. 61003305), the Fundamental Research Funds for the Central Universities of China (no. 531107040263), and Hunan Natural Science Foundation of China (no. 10JJ5069).

References

[1] Y. Wang and X. Y. Li, "Minimum power assignment in wireless ad hoc networks with spanner property," *Journal of Combinatorial Optimization*, vol. 11, no. 1, pp. 99–112, 2006.

[2] H. Shpungin and M. Segal, "Near optimal multicriteria spanner constructions in wireless ad-hoc networks," in *Proceedings*

of the 28th Conference on Computer Communications (IEEE INFOCOM '09), pp. 163–171, Rio de Janeiro, Brazil, April 2009.

[3] A. K. Abu-Affash, R. Aschner, P. Carmi, and M. J. Katz, "Minimum power energy spanners in wireless ad hoc networks," in *Proceedings of the 28th Conference on Computer Communications*, pp. 1–6, San Diego, Calif, USA, March 2010.

[4] H. Shpungin and M. Segal, "Near-optimal multicriteria spanner constructions in wireless Ad Hoc networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1963–1976, 2010.

[5] P. V. Rickenbach, R. Wattenhofer, and A. Zollinger, "Algorithmic models of interference in wireless Ad Hoc and sensor networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 172–185, 2009.

[6] L. P. Chew, "There is a planar graph almost as good as the complete graph," in *Proceedings of the 2nd Symposium on Computational Geometry*, pp. 169–177, Yorktown Heights, NY, USA, 1986.

[7] R. E. N. Moraes, C. C. Ribeiro, and C. Duhamel, "Optimal solutions for fault-tolerant topology control in wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5970–5981, 2009.

[8] L. Li, L. Lian, and H. Bin, "Algorithms for k-fault tolerant power assignments in wireless sensor networks," *Science China Information Sciences*, vol. 53, no. 12, pp. 2527–2537, 2010.

[9] S. Indranil, K. S. Lokesh, K. G. Subhas, and K. P. Ranjeet, "Distributed fault-tolerant topology control in wireless multihop networks," *Wireless Networks*, vol. 16, no. 6, pp. 1511–1524, 2010.

[10] P. Carmi, M. Segal, M. J. Katz, and H. Shpungin, "Fault-tolerant power assignment and backbone in wireless networks," in *Proceedings of the 4th annual IEEE International Conference on Pervasive Computing and Communications Workshops*, pp. 80–84, Washington, DC, USA, March 2006.

[11] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48–50, 1956.

[12] S. Even and R. E. Tarjan, "Network flow and testing graph connectivity," *SIAM Journal on Computing*, vol. 4, pp. 507–518, 1975.

[13] G. Di Battista, R. Tamassia, and L. Vismara, "Output-sensitive reporting of disjoint paths," *Algorithmica*, vol. 23, no. 4, pp. 302–340, 1999.

[14] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 53, no. 8, pp. 41–47, 1959.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 570643, 10 pages
doi:10.1155/2012/570643

Research Article

Probability Model Based Coverage-Enhancing Algorithm for WSNs of Nodes' Adjustable Movement Pattern

Junjie Huang,[1] Haiping Huang,[2] Peng Sun,[2] Ruchuan Wang,[2] and Xun Wei[2]

[1] *College of Internet of Things, Nanjing University of Posts and Telecommunications, Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, and Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing 210003, China*
[2] *College of Computer, Nanjing University of Posts and Telecommunications, Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, and Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing 210003, China*

Correspondence should be addressed to Junjie Huang, 117579663@qq.com

Received 20 July 2012; Accepted 23 October 2012

Academic Editor: Fu Xiao

Copyright © 2012 Junjie 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.

Aimed at the defects of current coverage-enhancing algorithms, firstly the probability model is introduced into the coverage-enhancing algorithm of the directional sensor network based on the virtual potential field, and the probability model would have an impact on the location of the centroid. Secondly, to the boundary coverage, through the simulation of typical algorithms which consider the boundary repulsion, it finds that the coverage rates of the corners are much lower than the overall coverage. To add a vertex force on the corner is essential and to some extent would reduce the gap between them. Finally, based on the random deployment of sensor nodes, a novel algorithm is proposed which determines whether the sensor node rotates or moves along fixed direction according to the coverage effect, named as PRMCA (probability model based rotate or move along fixed direction coverage-enhancing algorithm). A set of simulation experiments verify the performance of the proposed algorithm.

1. Introduction

At present, coverage, which is regarded as monitor of service quality, has been an essential problem of the wireless sensor networks (WSNs) [1]. One of these research directions is to use less quantity of sensor nodes or shorter time to achieve the same effect under the premise of ensuring a certain quality of coverage by some algorithm strategies. As the nodes' deployment reflects the cost and performance of WSN [2, 3], the covering control algorithms can greatly reduce the number of nodes, can also prolong the active time, and naturally can reduce the energy consumption of WSNs [4]. In the WSN coverage research, these aspects need to be considered, including the nodes' deployment, the nodes' sensing and communication range, energy efficiency, the algorithm characteristics, and sensor mobility. In different application fields [5, 6], researchers have proposed a lot of coverage mechanisms.

In the traditional research of utilizing the virtual potential field to optimize the quality of WSNs, while calculating the virtual force, some algorithms simply calculate the composition of the force based on the Boolean model so that they ignore the unique attribute of the probability model. For example, Li et al. in [7] proposed the virtual force algorithm (VFA). Tao and Ma in [8] proposed the target involved virtual force algorithm (TIVFA). Tao et al. in [9] proposed the PFCEA for the direction model (potential field based coverage-enhancing algorithm, PFCEA), and so forth. The authors in [10] proposed a coverage configuration algorithm based on the probability detection model (CCAP). The authors in [11] proposed a coverage preservation protocol based on the probability detection model (CPP) that makes working nodes in sensor networks as few as possible when network coverage is guaranteed. However, this protocol configures a network using centred control algorithm which limits the network scale, and at present most of the literatures

have not introduced this probability coverage model into three-dimensional sensor networks. In fact, most practical applied wireless sensor networks are deposited in three-dimensional sensor networks so that it will be more accurate if it is simulated in a three-dimensional space [12, 13]. In [12], Bai et al. proposed and designed a series of connected coverage models in three-dimensional wireless sensor networks with low connectivity and full coverage. In [13], Nazrul Alam and Haas studied a truncated octahedron deployment strategy to monitor the network coverage situation.

In the random deployment, the distribution model is mainly based on the following three assumptions: Poisson's distribution, normal distribution, and random distribution using software simulation [14–17]. In actual deployment, especially for the harsh environment, sensor nodes are usually laid by the air shedding manner [18]. However, it is inevitable to bring about nodes that are not uniformly distributed in the region. If the sensor nodes achieve the enhancement of coverage just only by adjusting the perception angle, the result will be impacted by the nodes' initial position. The maximization of the coverage cannot be really achieved. In the meantime, the free movement of the sensor nodes is not practical due to ignoring the limitation of movement direction.

According to the methods of calculating the centroid position physically and the characteristics of the probability model, the position of the node centroid based on the probability model can be figured out. The position describes the impact of the centroid on the probability model. On the basis of the above, this paper proposes an algorithm based on coverage effect to determine the sensor to rotate or move along fixed direction which is named as probability model based rotate or move along fixed direction coverage-enhancing algorithm PAMCA. This algorithm can be proved to make up some defects of current algorithms.

2. Coverage Enhancement Issues

2.1. Problem Formulation. The probability model based on the coverage-enhancing problem of directional sensing network composed of the nodes which can rotate and move along fixed direction can be described as follow: When a certain number of sensor nodes are deployed randomly in a specific region, how to calculate the composition force of the centroid in the perception region and determine the sensor to rotate or move along fixed direction according to the coverage effect.

2.2. Analysis and Definitions on the Coverage Enhancement Issue in Directional Sensor Networks. Before research, the following assumptions should be firstly made.

(1) All nodes in WSNs are isomorphic, which means that all the nodes have the same perception radius (R_S), the same communication radius (R_C), the same perception angle (α), the same parameters associated with the physical properties, and so forth. The communication radius is more than or equal to 2 times the perception radius.

(2) The nodes work independently and have the capacity to move towards the specific target location accurately.

(3) The nodes have the sufficient energy to complete the process of movement.

(4) Each node can acquire the information of perception direction and control the direction.

(5) In a rectangular target area A, any two sensor nodes deployed randomly are not in the same location.

The relevant definitions are as follows.

Definition 1. Probability Perceptual Model : under the premise that the node i does not have any neighbour nodes, the perception probability of i to j in the monitoring region can be described as follows:

$$p_{ij} = \begin{cases} 1, & d(i,j) \le r_1, \\ e^{-\beta[d(i,j)-r_1]}, & r_1 < d(i,j) \le r_2, \\ 0, & d(i,j) > r_2, \end{cases} \tag{1}$$

where r_1 and r_2 are parameters which are adjustable and related to the sensor physical properties, e denotes the natural logarithm, and $d(i,j)$ represents the distance between node i and node j.

Definition 2. Coverage: the coverage of the WSNs W deployed in the target region A is marked as $C_r(W,A)$. The definition is as follows:

$$C_r(W,A) = \frac{\int_A v(p)dS}{V(A)}, \tag{2}$$

where $V(A)$ is the area of the target region A, for any point $p \in A$, and

$$v(p) = \begin{cases} 1, & Cd(p) \ge 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

$Cd(p)$ represents the count of the nodes which cover the point p.

Definition 3. Coverage Factor: the coverage factor μ implies the impact of the network coverage when the node moves once, which is defined as follows:

$$\mu = \begin{cases} 1, & C_r'(W,A) > C_r(W,A), \\ 0, & C_r'(W,A) \leqslant C_r(W,A), \end{cases} \tag{4}$$

where $C_r'(W,A)$ means the network coverage after the node movement.

3. Description and Analysis of the Direction Sensor Based on the Probability Model

Due to the reality, the probability model based on directional sensor has a wider area in application. Sensor energy will decay as the time goes, and the sensing radius will decrease; therefore, how to ensure stable time of the regional coverage

will be a focus of the future research, which means to add the timeline to the four-dimensional coverage. And we should also pay attention to the value of the coverage probability of the regional point when the regional coverage is studied; the greater the probability is, the longer time the points are covered.

Based on the model and abovementioned definitions, the concept of the centroid is introduced in the probability model. It is reasonable to assume that the mass density is proportional to the size of the probability, so the different probabilities have different effects on the position of the centroid. The centroid position can be calculated by the method in physics.

It is assumed that points have the same mass density if the points have the same distance from the sensor, and the density is proportional to the perception probability, the ratio is set as k (k is a constant); therefore, in a circle that the node is regarded as the center and 2α as the central angle, mass is uniformly distributed in the same arc. It means that the centroid position of the arc which has the distance r from the node is in the symmetry axis and has the distance $(r\sin\alpha)/\alpha$ from the node, so it can be deduced further that the centroid position of probability model is based on fan-shaped region, as shown in Figure 1.

It is clear that the centroid position is in the symmetry axis AB and set as x-axis. According to the centroid formula of rigid object, it can be drawn that

$$X_c = \left(\int_0^{r_1} (x\sin\alpha/\alpha)\cdot 2\alpha\cdot k\cdot x\,dx + \int_{r_1}^{r_2} (x\sin\alpha/\alpha)\cdot 2\alpha\cdot k\cdot e^{-\beta(x-r_1)}\cdot x\,dx\right) \times\left(\int_0^{r_1} 2\alpha\cdot k\cdot x\,dx + \int_{r_1}^{r_2} 2\alpha\cdot k\cdot e^{-\beta(x-r_1)}\cdot x\,dx\right)^{-1}. \tag{5}$$

Simplifying further,

$$X_c = \left(2\beta^3 r_1^3\sin\alpha - 6\beta^2\left[r_2^2 e^{\beta(r_1-r_2)} - r_1^2\right]\sin\alpha - 12\left[\beta r_2 e^{\beta(r_1-r_2)} + e^{\beta(r_1-r_2)} - \beta r_1 - 1\right]\sin\alpha\right) \times\left(3\alpha\beta^3 r_1^2 - 6\alpha\left[\beta^2 r_2 e^{\beta(r_1-r_2)} + \beta e^{\beta(r_1-r_2)} - \beta^2 r_1 - \beta\right]\right)^{-1}, \tag{6}$$

where $r_1 = AC$, $r_2 = AD$, α, β are the positive constants and related to the sensor physical properties, and the centroid position can be calculated directly in the probability model based on fan-shaped perception region. It is in the symmetry axis and has the distance X_c from the center.

Obviously, when $r_1 = r_2$, the centroid is in

$$X_c = \frac{2R\sin\alpha}{3\alpha}. \tag{7}$$

It can be concluded that the uniform sensing region is a special case of the probability perception model. When $\alpha = \pi$, the directional sensor becomes the omnidirectional perception model and the centroid is in the center of the circle.

Figure 1: Sensory domain of node i (SD_i).

4. The Study of the Visual Force

When the distance between the sensor nodes is less than $2R_S$, it is possible that the perception region is overlapped, which provides the possibility of repulsion. It is reasonable to assume that if several perception areas are overlapped, its probability should be the maximum of each probability. In addition, according to the probability perception model, it can be concluded that when the probability is less than a certain value, the effect of the node can be ignored. The threshold can be determined by the sensor parameters of r_1, r_2, and β. In this paper the possibility threshold is $e^{-\beta[r_2-r_1]}$ and the perception radius can be considered as $R_S = r_2$.

Under the impact of the repulsion, the nodes spread as far as possible in order to ensure the maximum area covered. It is well known that the condition that the distance between sensor nodes is less than $2R_S$ is a necessary condition of the existence of overlapped area instead of sufficient condition. Therefore, it is unreasonable to only rely on the distance between the nodes to judge whether the repulsion exists. Reference [19] points out that there are 11 overlapping situations for perception areas, as shown in Figure 2.

When the distance of the two nodes is less than $2R_S$, there will be no repulsion between the nodes because the overlapping regions are not existed. Therefore, it will not consume energy caused by the action of nodes; in this case the energy can be saved.

Reference [20] provides another judgment for the existence of perception overlapping area: while the perception model is simplified as a triangle, according to the condition of whether the two triangles have intersecting edge, the existence of the overlapping area can be judged. The method is called the straddle experiment. As is shown in Figure 3, there must be at least an intersecting edge for the two triangles with the overlapping regions; in Figure 3, S_1a_1 and S_2a_2 intersect, and the following conditions must be met:

$$[(a_2 - S_1)\times(a_1 - S_1)] * [(a_1 - S_1)\times(S_2 - S_1)] \geq 0, \tag{8}$$

$$[(S_1 - a_2)\times(S_2 - a_2)] * [(S_2 - a_2)\times(a_1 - a_2)] \geq 0. \tag{9}$$

Figure 2: The situations of overlap of sensory region in directional sensor model.

Figure 3: Two sensor models with overlapping sensory region.

Figure 4: The corner in the sensory region.

If the relationship between the nodes meet conditions in (8) and (9), there will exist repulsion between nodes, and it will be defined as

$$F_{\text{rep}}(i,j) = \begin{cases} \dfrac{K_{\text{rep}}}{d(i,j)^2}\left(\dfrac{i-j}{d(i,j)}\right), & SD_i \cap SD_j \neq \varnothing, \\ 0, & SD_i \cap SD_j \neq \varnothing, \end{cases} \tag{10}$$

where K_{rep} is repulsion coefficient, a positive constant. Therefore, the centroid n_i is the vector sum of the repulsion of adjacent K centroid points. The definition of the force mode $\vec{F}$ is

$$F_{\text{rep}}(i) = \begin{cases} \displaystyle\sum_{j=1, j\neq i}^{K} F_{\text{rep}}(i,j), & K \geq 1, \\ \vec{0}, & K = 0. \end{cases} \tag{11}$$

In the random deployment, it is not easy to realize that the nodes are distributed uniformly. The randomness of the nodes' position is great, so it is possible that many points are centralized. In the calculation of the virtual joint force, the boundary problem should be taken into account, which is not considered in most of the papers, or just one boundary direction, as shown in Figure 4.

In the simulation of the proposed algorithm in this paper, it takes four same squares in the corner, whose area is 1/9 of the total area but other algorithms only consider the repulsion nearest to boundary nodes. The number of adjustment is set 100.

As is shown in the simulation experiment results of the Table 1, it is not enough to only consider the repulsion within a certain distance to the node D in the corner. The attraction of the uncovered vertex B should also be taken into consideration, namely, the attraction of the vertex to the sensor.

This paper assumes that the force of the nearest boundary points and four vertexes of the rectangular frame to the sensor node exist. Obviously, in the region that needs to be covered, the force of all directions should be considered simultaneously if all the vertexes are in the communication range of the sensor node.

As is shown in Figure 4, C is the centroid point, D is a boundary point within a certain distance to sensor node, and B is a border vertex; in this case the force of D and B on C must be considered.

Table 1: Coverage and border coverage.

Condition	Coverage (%)	Border coverage (%)
$\alpha = \pi/8, N = 100$	56.9199	30.7778
$\alpha = \pi/6, N = 100$	64.6202	39.4444
$\alpha = \pi/4, N = 100$	81.4776	51.5556
$\alpha = \pi/8, N = 150$	72.9448	47.8678
$\alpha = \pi/6, N = 150$	85.1117	54.6876
$\alpha = \pi/4, N = 150$	98.4397	76.7778
$\alpha = \pi/8, N = 200$	91.0510	52.1036
$\alpha = \pi/6, N = 200$	98.4391	77.7778
$\alpha = \pi/4, N = 200$	99.3756	87.6667

The definition of the boundary force is as follow:

$$F_{\text{bor}}(i) = \begin{cases} \dfrac{k_{\text{bor}}}{d(i,q)^2}\left(\dfrac{x_i - q}{d(i,q)}\right), & d(i,q) \leqslant \sqrt{3}r_s, \\ 0, & d(i,q) > \sqrt{3}r_s, \end{cases} \tag{12}$$

where k_{bor} is ratio coefficient, a constant. $\sqrt{3}r_s$ is a threshold distance, which is mentioned above as the value of a certain distance, when the distance is below this value, the force of the boundary point and vertex starts to take effect on nodes. For the direction of the force, because the node is close to the boundary, the coverage area may be reduced due to being out of the region. Therefore, it is reasonable to assume that there exists repulsive force between the boundary points and the sensor node. However, if the vertexes are not covered, there will exist a blind zone in the region. In this case, if all the force is repulsion, the blind zone will not be erased because the sensor node cannot get close to the corner. So in this paper, a new solution will be introduced: if the vertex is not covered, the attractive force will take effects while if the vertex has been covered, the repulsive force will make the overlapping area eliminated quickly. The attractive force is set by a negative value.

To sum up, the resultant force of the sensor centroid i can be expressed as

$$F(i) = F_{\text{rep}}(i) + F_{\text{bor}}(i). \tag{13}$$

5. Descriptions of Node Movements

Now, in most papers, when considering the forces of the virtual potential field to nodes, the researchers only consider the rotation or the omnidirectional movement of the nodes, while the omnidirectional movement is unpractical and the rotation of the node costs too much energy when dispersing without a high speed. For instance, when considering the rotation merely, only when the force has a tangential component, the node can rotate along this direction. As Figure 5(a) shows, though there is an overlapping region, it has no forces in the tangential direction. Therefore, the node can't rotate; even if it rotates, it also needs to rotate for more times to make the overlapping region disappear. As Figure 5(b) shows, when considering the movement merely, it only needs to rotate for a small angle to make the overlapping region disappear. But in this condition, the movement will certainly cause the unnecessary energy cost. Hence, a method can be introduced that movement and rotation are considered synthetically to avoid consumption of the node's energy.

In the movement procedure of nodes, their moving directions are not arbitrary. Such as cameras, they can only move forward and backward by focusing or modulate the viewpoint by rotating. In conclusion, most moving directions of sensor nodes are limited. Therefore, this paper assumes that the node only rotates or moves along with the sensor's direction.

In physics, acceleration has been defined. It is a vector that measures the speed changing in unit time. It reflects the moving trend of the object. This conception can be applied in coverage researches. According to the analysis above, a criterion must be introduced to measure the moving change of objects. We can use the coverage improvement (influence factors of coverage) as a measurement and call it acceleration.

Objects' movement can be judged by the influence factors of coverage. We consider that the influence factors of coverage (acceleration) in this direction are bigger than the other direction; the node will move towards this direction. Otherwise, the node will move towards the other direction. In particular, we define that if the accelerations in two directions are equal, the sensor will rotate. It shows as Figure 1.

That, the node can only move along with the center line ($\vec{AB}$ direction or the opposite direction) or the tangential direction of the fan-shape. Therefore, we can determine the final moving direction of the node according to the influence of resultant force $\vec{F}$ and θ.

$$\vec{F} \cdot (\cos\theta, \sin\theta) \geq 0, \tag{14}$$

$$(\cos\theta, \sin\theta) \times \vec{F} \geq 0. \tag{15}$$

When (14) was founded, the angle between resultant force $\vec{F}$ and $\vec{AB}$ is less than $\pi/2$ and the node moves along with $\vec{AB}$. Otherwise, it will move towards the negative direction of $\vec{AB}$. When (15) was founded, $\vec{AB}$ rotates along with the tangential direction and the angle θ increases. On the contrary, θ decreases. More attentions need to be paid to the boundary value. When the node crosses the region boundary in its movement process, the node must be on the boundary. In the rotation process, when θ is greater than π, $2 * \pi$ needs to be subtracted from θ. Similarly, when θ is less than $-\pi$, θ needs to add $2 * \pi$.

6. Algorithm Description

A coverage-enhancing algorithm is introduced as follows according to the virtual forces to nodes. This algorithm is a distributed one, which runs simultaneously in each sensor. Taking node i for example, its formal description is shown as in Algorithm 1.

```
//initialization
(1)  t ← 0;
(2)  Set relevant parameters and coefficients;
(3)  Set the max cycle-index t_max;
While (t < t_max) do
{
(4)  Renew the neighbor nodes set of node i;
(5)  Read the current coordinate position of node i;
(6)  Read all neighbor nodes of node i;
(7)  Calculate resultant force to node i according to formulas
     (8)~(13);
(8)  Calculate the target position of node i at time t + 1
     according to formulas (14) and (15);
(9)  Calculate the coverage rate of this movement and rotation
     according to formula (2);
(10) Determine the mode of movement according to the
     coverage rate. Calculate influence factors μ. And determine
     the final moving direction of the node.
(11) If μ = 1, move node i according to the mode and direction
     in step (10), or keep the current position of node unchanged.
(12) t ← t + 1;
}
(13) End.
```

Algorithm 1

Figure 5: Two special situations in the WSNs.

Figure 6: Convergent Curve with different N.

7. Algorithm Simulations and Performance Analysis

This simulation experiment uses VS2010 to develop and realize a probability model based coverage-enhancing algorithm for WSN of nodes' adjustable movement pattern. And it uses Matlab 2009 to achieve many simulation experiments, through which the effectiveness and advantages of this algorithm have been proved. To simplify these experiments, it is assumed that all assumptions in Section 2.2 are met.

7.1. Experiment Parameters' Setting. In the simulation, it sets a 300 m $*$ 300 m sized square target region. Some nodes are distributed in this region randomly. Repulsion coefficient among nodes is defined as $k_{\text{rep}} = 100$. Gravitation coefficient among nodes can be described as $k_{\text{att}} = 0.025$. Gravitation index $\partial = 1$. Repulsion coefficient among boundaries is $k_{\text{bor}} = 1$. In the probability model, we can set the parameter $\beta = 1$.

7.2. Algorithm Convergence Analysis. To discuss the convergence of PRMCA, some experiments are carried out using four kinds of different network nodes' number. We produce ten scenarios randomly in terms with different topological

Figure 7: Coverage optimization under PRMCA.

Figure 8: Comparison with PFCCEA.

constructs and network nodes' number, meanwhile calculating the average value of the algorithm's convergence. Experiment data are showed in Table 2. Other experiment parameters are defined, respectively, as $R_1 = 20$, $R_2 = 40$ m, and $\alpha = \pi/6$.

Obviously, it can be concluded as follows: the convergence of PRMCA (regulating times) will not have an obvious change as the increase of the number of sensor nodes. The range of variation is usually between 20 to 30. Therefore, PRMCA proposed in this paper has a satisfactory convergence. It can accomplish the coverage-enhancing process in a short limited time.

Table 2: Experiment data convergence analysis.

Number of nodes/N	Original coverage rate/%	Final coverage rate/%	Cycle-index/t
50	30.4891	41.6233	22.3
100	50.3642	67.2216	24.7
150	61.6025	85.3278	28.2
200	77.5234	98.127	27.6

7.3. Case Study. In this section, PRMCA will be analyzed through cases:

$$N \geq \frac{\ln(1-P)}{\ln(S-\alpha R_s^2)-\ln S}, \tag{16}$$

where P can be defined as the coverage rate you expected and S can be described as the total region area. According to the experiment parameters, the number of sensor's nodes can be estimated if the network coverage rate reaches P as expected. Such as these cases, if $P = 70\%$, then $N = 129$; if then $P = 80\%$, then $N = 172$; if $P = 90\%$, then $N = 246$. In these simulations, $N = 100$, $N = 150$, and $N = 200$.

For the example above, when $N = 150$, the movement process can be shown in Figure 7.

In Figure 7, nodes tend to uniformly distributve gradually in the influence of the virtual force. The proposed algorithm can achieve the enhancement by the adjustment which can eliminate the overlapping and the blind regions. After 30-step adjustments, the coverage rate increases from 62.375% to 78.496% with the enhancement of 16.121%.

Figure 6 shows that due to the influence of the coverage factor, the coverage rate is constantly improving. When it reaches around 15 times, coverage rate tends to be stable,

(a) Effect of node scale α on the condition of $R_S = 40\,\text{m}$, $N = 100$

(b) Effect of node scale N on the condition of $R_S = 40\,\text{m}$, $\alpha = \pi/6$

(c) Effect of sensory distance of nodes R_S on the condition of $N = 150$, $\alpha = \pi/6$

Figure 9: Effect on different condition.

the effect on the coverage rate enhancement is not quite remarkable. The movement of the node would make energy consumption, so it will save energy without moving any longer.

7.4. Simulation Analysis. The node in each movement must choose the path containing the maximum coverage rate, because the algorithm proposed in this paper is based on the acceleration which determines the movement mode, [9] for the two-dimensional space of the wireless sensor network provides an enhancing algorithm PFCEA based on the virtual force. It introduces the coverage factor into the algorithm and names the improved algorithm as PFCCEA. By comparison, as Figure 8 shows, PRMCA converges much faster than PFCCEA. In addition, node in PFCCEA needs to be randomly and uniformly distributed, so PRMCA obtains better performance than PFCCEA. But at the same time, the initial location of the node based on PFCCEA has the greater impact.

This section is also to see the influence by changing the parameters of the proposed algorithm, which includes the number of nodes N, perception angle α, and sensing radius R_S. Performance analysis and comparison between PFCCEA and PRMCA are demonstrated by Figure 9.

From the curve of Figure 9, it can be seen that when N and R_S are fixed, the smaller the value of α, the smaller coverage rate will be. With the increasing of α, the increasing of coverage rate Δp tends to be higher, and when $\alpha = \pi/3$, network coverage increase can reach 20.7076%. After that,

Figure 10: Boundary effect on different conditions.

Δp drops. This is because with the increasing of the perceived angle, the probability that covers blind spot is reduced greatly, which will undoubtedly sacrifice the PRMCA performance.

Obviously, from Figures 9(b) and 9(c), it can be seen that the sensing radius R_S and the node number N have the similar effects on the performance of PRMCA algorithm. When the sensing angle α is certain, the value of sensing radius R_S or node number N is larger, the possibility that blind spot exists is smaller, and the performance of PRMCA becomes not obvious.

In addition, through the comparison, parameters such as nodes scale, sensing radius, and sensing angle have similar effect on the performances of PFCCEA and PRMCA. However, due to the limitations of PFCCEA that PFCCEA does not consider the boundary effects and PFCCEA requires uniform deployment of the nodes, it is bound to be limited by the initial coverage. At the same time, it will limit the improvement in the network coverage performance. By contrast, PRMCA is dominant.

7.5. Analysis on the Boundary. As shown in Figure 10, through the introduction of the force in corner and vertices, to a certain extent, the algorithm can reduce the gap between the overall coverage and boundary coverage. Owing to omitting the local effect of the nodes, the boundary coverage is lower than the overall coverage.

8. Conclusions

In this paper, the probability model based on virtual potential field is introduced into the coverage-enhancing algorithm, whose probability is reflected by the position of "centroid" and transforms the sensor network coverage enhancement problem into a problem that the center of mass gradually spreads by the virtual force, and the blind region and the overlap zone are eliminated.

It is difficult for the nodes to achieve the uniform distribution in the actual application, especially in the dangerous area. At the same time, the moving direction of the node was restricted, such as camera. In this way, the initial node will ultimately affect the entire regional coverage. The proposed PRMCA can let the nodes select a movement form according to the "acceleration," which can let the node achieve the convergence faster and save the energy of nodes.

Because the algorithm is an optimal algorithm, so the experimental data can be seen that the corner coverage must be smaller than the overall coverage. By considering the gravity in boundary vertices, it improves the corner coverage, but it is still lower than the overall coverage.

Finally, through a series of simulation experiments, it verifies the convergence and effectiveness of PRMCA, and describes several important parameters of PRMCA that influence performance. In this paper, it effectively improves the WSNs coverage rate, but when one point is covered by many sensors, the integrated calculation of this point is not considered, and it would be taken into account in the further research in the four-dimensional space.

Acknowledgments

The subject is sponsored by the National Natural Science Foundation of China (nos. 61171053, 61003039), Scientific & Technological Support Project (Industry) of Jiangsu Province (no. BE2012183), Natural Science Key Fund for Colleges and Universities in Jiangsu Province (11KJA520001, 12KJA520002), Science & Technology Innovation Fund for higher education institutions of Jiangsu Province (CXZZ11-0409), Doctoral Fund of Ministry of Education of China (20113223110002).

References

[1] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, "Coverage problems in wireless ad-hoc sensor networks," in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 3, pp. 1380–1387, IEEE Press, New York, NY, USA, April 2001.

[2] J. Huang, L. Sun, R. Wang, and H. Huang, "Improved virtual potential field algorithm based on probability model in three-dimensional directional sensor networks," *International Journal of Distributed Sensor Networks*, vol. 21, no. 6, pp. 45–53, 2012.

[3] E. Onur, C. Ersoy, and H. Deliç, "Analysis of target detection probability in randomly deployed sensor networks," *IEEE Communications Letters*, vol. 11, no. 10, pp. 778–780, 2007.

[4] Y. Wu, M. Li, Z. Cai, and E. Zhu, "A distributed algorithm to approximate node-weighted minimum α-connected (θ,k)-coverage in dense sensor networks," in *Proceedings of the International Frontiers of Algorithmic Workshop*, vol. 5059, pp. 221–232, Springer, Berlin, Germany, 2008.

[5] M. Cardei and J. Wu, "Energy-efficient coverage problems in wireless ad-hoc sensor networks," *Computer Communications*, vol. 29, no. 4, pp. 413–420, 2006.

[6] Y. Zou and K. Chakrabarty, "Sensor deployment and target localization in distributed sensor networks," *ACM Transactions on Embedded Computing Systems*, vol. 3, no. 1, pp. 61–91, 2004.

[7] S. J. Li, C. F. Xu, Z. H. Wu, and Y. H. Pan, "Optimal deployment and protection strategy in sensor network for target tracking," *Acta Electronica Sinica*, vol. 34, no. 1, pp. 71–76, 2006 (Chinese).

[8] D. Tao and H. Ma, "Coverage-enhancing algorithm for directional sensor networks," in *Proceedings of the of the 2nd International Conference on Mobile Ad-Hoc and Sensor Networks*, I. Stojmenovic and J. N. Cao, Eds., pp. 256—267, Springer, Berlin, Germany, 2006.

[9] D. Tao, H. D. Ma, and L. Liu, "A virtual potential field based coverage-enhancing algorithm for directional sensor networks," *Journal of Software*, vol. 18, no. 5, pp. 1152–1163, 2007 (Chinese).

[10] D. Zhang, M. Xu, Y. Chen, and S. Wang, "Probabilistic coverage configuration for wireless sensor networks," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '06)*, pp. 1–4, September 2006.

[11] J. P. Sheu and H. F. Lin, "Probabilistic coverage preserving protocol with energy efficiency in wireless sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 2633–2638, March 2007.

[12] X. Bai, C. Zhang, D. Xuan, and W. Jia, "Full-coverage and k-connectivity ($k = 14, 6$) three dimensional networks," in *Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM '09)*, pp. 388–396, April 2009.

[13] S. M. Nazrul Alam and Z. J. Haas, "Coverage and connectivity in three-dimensional networks," in *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MOBICOM '06)*, pp. 346–357, Los Angeles, Calif, USA, September 2006.

[14] B. Liu and D. Towsley, "A study of the coverage of large-scale sensor networks," in *Proceedings of the IEEE International Conference on Mobile Ad-Hoc and Sensor Systems*, pp. 475–483, October 2004.

[15] M. Cardei and D. Z. Du, "Improving wireless sensor network lifetime through power aware organization," *Wireless Networks*, vol. 11, no. 3, pp. 333–340, 2005.

[16] D. Brinza and A. Zelikovsky, "DEEPS: deterministic energy-efficient protocol for sensor networks," in *Proceedings of the 7th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD '06)*, pp. 261–266, June 2006.

[17] D. Tian and N. D. Georganas, "A coverage-preserving node scheduling scheme for large wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pp. 32–41, Atlanta, Ga, USA, September 2002.

[18] J. Wang, C. Niu, and R. Shen, "Priority-based target coverage in directional sensor networks using a genetic algorithm," *Computers and Mathematics with Applications*, vol. 57, no. 11-12, pp. 1915–1922, 2009.

[19] D. Pescaru, C. Istin, D. Curiac, and A. Doboli, "Energy saving strategy for video-based wireless sensor networks under field coverage preservation," in *Proceedings of the IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR '08)*, pp. 289–294, May 2008.

[20] J. J. Huang, L. J. Sun, R. C. Wang, and H. P. Huang, "Virtual potential field and covering factor based coverage-enhancing algorithm for three-dimensional wireless sensor networks," *Journal on Communications*, vol. 31, no. 9, pp. 16–21, 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 602358, 13 pages
doi:10.1155/2012/602358

Research Article

MPD-Model: A Distributed Multipreference-Driven Data Fusion Model and Its Application in a WSNs-Based Healthcare Monitoring System

Jibing Gong,[1,2,3] Li Cui,[1] Kejiang Xiao,[1,3] and Rui Wang[1]

[1] *Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China*
[2] *Department of Computer Science and Engineering, Yanshan University, Qinhuangdao 066004, China*
[3] *Graduate University of Chinese Academy of Sciences, Beijing 100190, China*

Correspondence should be addressed to Jibing Gong, gongjibing@gmail.com

Received 6 September 2012; Accepted 31 October 2012

Academic Editor: Fu Xiao

Copyright © 2012 Jibing Gong 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 first propose an MPD-Model, a novel distributed multipreference-driven data fusion model for WSNs. Here, *preferences* are looked as the core elements of collaboration mechanism in a data fusion procedure. We then present MFA, a distributed multipreference feature-level fusion algorithm based on weighted average method. Next, to implement feature extraction of wrist-pulse data, we propose FEA, a light-weight adaptive feature extraction algorithm for time series sensed data. Simultaneously, we design TFD-Pattern that is a unique human pulse pattern. Based on historical data, we propose an SVM-based algorithm for health status detection tasks. Finally, we implement the proposed methods in a real wearable healthcare monitoring system which had been previously developed in-house. We validate the proposed methods using real-world data sets with 2046 pulse samples. Experimental results show that the proposed methods outperform the baseline methods, and the proposed MPD-Model is reasonable and effective.

1. Introduction

The rapid development of Wireless Sensor Networks (WSNs) brings some new situations. On the one hand, WSNs applications rapidly expand from traditional fields, for example, Military and National Defense and Environmental Monitoring, to emerging civil fields such as intelligent transportation, healthcare monitoring, and smart home. On the other hand, more mobile devices (e.g., smartphones) or local network systems (e.g., Body Sensor Networks) are frequently treated as one single node of WSNs. Undoubtedly, the situations will lead to form a large-scale and complex sensor network. We think the network is a typical Internet of Things (IoT) [1] and it have some characteristics such as interactivity and sociality. Simultaneously, It emphasizes intelligence and sensing-actuating ability of WSNs applications. Consequently, sensed data generated under the situations have the following features [1–4]. (1) The data has polymorphism and heterogeneity. (2) The data is massive. (3) The data is real time. (4) Storage locations of the data have much diversity. (5) The data has high complexity and strong relationships.

In the new situations, users play more and more importance to intelligence of sensor network systems because of the growing actual demands. But it is very difficult to implement the intelligence due to seriously limited resources on nodes and the lack of historical sensed data. This is particularly true in the healthcare monitoring applications based on WSNs. Fortunately, system detection/identification accuracies and intelligence can be greatly improved if multisource data fusion technologies are fully leveraged. However, if so, some challenges still exist below. (1) How to select and evaluate those multisource sensing parameters to fuse? (2) How to establish novel data fusion model/mechanism to improve detection accuracy? (3) How to leverage limited resources on nodes to automatically recognize complex parameter patterns?

To address these challenges, we propose an MPD-Model, a novel-distributed multi-preference-driven data fusion model for WSNs. Based on traditional data fusion architecture, MPD-Model takes regard preference information as core elements of collaboration mechanism in a data fusion procedure. It introduces large-scale historical sensed data into machine learning for data fusion, treats the massive data as real-world test set, and thus effectively ensures enough identification accuracy and high intelligence of WSNs application systems. Here, *preference* is defined as a kind of description information being different from existing nature, accuracy, and detail level information of monitoring tasks in traditional data fusion method. It can reflect subjective wishes of individuals or a group. For example, the *preference*, Task Parameter Relevance, just is a kind of importance degree information telling that the accuracies of detection tasks depend on sensing parameters selected for fusion. It has two characteristics: uncertainty and dynamics. The former means that preference information varies with both the size and the type of WSNs applications. This well embodies individual data fusion. The latter indicates that preference information can be quantified and then evaluated by formalizing them.

We systematically address the problems of data fusion and health status detection in a healthcare monitoring system based on WSNs with the following contributions.

(i) To address the basic problem of data fusion, we propose an MPD-Model, a novel-distributed multi-preference driven data fusion model for WSNs. Based on MPD-Model, this paper conducts the individual data fusion of multisource sensing parameters in a real-world healthcare monitoring system. MPD-Model can not only solve incompleteness, inaccuracy, and uncertainty problems of preference information, but also significantly improve detection accuracy and intelligence of WSNs applications. According to MPD-Model, this paper presents several fusion algorithms to finish practical health status detection tasks.

(ii) In MPD-Model, we design and formalize four kinds of preference information (or impact factors). Through analyzing multi-preference information, we can quantify and evaluate importance degree of different multisource sensing parameters during data fusion. We propose MFA, a distributed and unified multi-preference fusion algorithm based on weighted average method for generating complex parameter patterns that will be further used for health status detection tasks.

(iii) To implement MPD-Model in a real wearable health status detection system [1], we propose EFA, a light-weight adaptive feature extraction algorithm for pulse data, to obtain physiological features of human wrist pulse waveforms. According to these features, we design TFD-Pattern, a novel time series pattern, and make the pattern as the input of MFA algorithm. Considering massive historical pulse data, this paper also presents SVM-based feature-decision-making algorithm to detect human health statuses, for example, subhealth. The algorithm is simultaneously an instance of y_i element in MPD-Model.

(iv) We validate our proposed MPD-Model and fusion algorithms using large-scale real-world healthcare data sets. Experimental results show that FEA is effective and outperforms the existing derivative-based method at an over 16% feature extraction accuracy; an average fusion ratio of 98.73% is available when running MFA. SVM-based health status detection method well support intelligence of MPD-Model via outperforming (20∼30%) AR and WPT methods. Different weighted values of multi-preference parameters perform distinct diagnosis performance. It proves that our individual fusion model based on preferences and weighted average method is reasonable and effective.

2. Multipreference-Driven Data Fusion Model for WSNs

2.1. MPD-Model. In this paper, we propose an MPD-Model, a novel-distributed multi-preference-driven data fusion model for Wireless Sensor Networks, as in Figure 1. Explanations of MPD-Model are given below. Its input may include environment monitoring information, human physiological parameters, and movements/activities information. R_i $(1 \leq i \leq n)$ indicates the output results of signals/data processing procedure P_i on the sensor node S_i. $R_i \blacklozenge R_j$ $(1 \leq i \neq j \leq n)$ states the fusion operation between any two output results. D_k $(1 \leq k \leq m)$ means one motoring target state or decision-making as the output of the model. y_i stands for any method, algorithm, or model for recognition or decision-makings.

The model is suitable for all kinds of WSNs applications for human sensing, particularly in healthcare monitoring, due to the following flexible features. (1) P_i, a signal/data processing procedure, may be any kind of processing method or technology that can run on the sensor nodes, and they should be distributed, light-weight, or interactive. (2) Fusion operation, namely, $R_i \blacklozenge R_j$, may also be any kind of calculation, algorithm, or technology. (3) The decision-making y_i may be a complicated classification method (e.g., Support Vector Machine and Time-Space Factor Graph Model), a user-defined rule, or even a simplest Threshold Value Judgement Operation such as IF statement. (4) Values of D_k depend on different WSNs applications. It can identify one monitored target state with Yes or No when $n = 2$. Otherwise, multiple states of one target can be recognized and output.

MPD-Model views intelligence and historical sensed data as core elements and key features. This is reasonable and scientific because they are the development tendency of data fusion model. There are several reasons as follows. (1) The two features are strongly required by user demands of WSNs applications in the new situations. Another point to

Figure 1: MPD-Model: a distributed multipreference driven data fusion model for Wireless Sensor Networks.

illustrate is that characteristics of sensed data generated in WSNs have dramatically varied. (2) Historical sensed data is paying more and more important role in a data fusion procedure. Thus, the role should be clearly embodied in a novel data fusion model, or fusion performance conducted by intelligent methods based on analyzing historical data will be seriously weakened. (3) With rapid expanding of WSNs applications from traditional domains to civil fields, they have dramatically increased demands on the intelligence. (4) Data processing ways depend on data types. But only data fusion model can effectively solve the requirement when considering the intelligence. This model should be driven by physical model of WSNs, and aiming at the intelligence of WSNs applications, this model should combine all kinds of data processing technologies and establish a unified architecture or a novel fusion mechanism to efficiently implement users' requirements.

2.2. Formalization of Preferences' Information/Impact Factors. Influences on detection decision-making results of sensing parameters depend on specific WSNs applications. Thus, we propose an individual data fusion model for WSNs. The model first computes fusion weights of different impact factors according to one WSN application's own features, and then combines and fuses these factors using weighted average method. Finally, more reasonable and accurate event detection results can be obtained. In the model, we innovatively design four kinds of ubiquitous preferences (impact factors or constraints) for data fusion in WSNs. They can clearly describe uniqueness or features of every WSNs application. They are formalized and described below.

2.2.1. Task-Parameter Relevance, TPR. The impact factor measures the relevance degree between sensing parameters and task types of event detection. For instance, although some parameters, for example, temperature, humidity, air composition, and so on, are often combined together to detecting fire event, temperature parameter will obviously have more influences on fire event detection than others. TPR is designed from the perspective of the event detection types and computed using the following formula:

$$t_\text{score}(P,E) = \frac{\vec{q}_P \cdot \vec{q}_E}{||\vec{q}_P||||\vec{q}_E||}, \tag{1}$$

where $\vec{q}_E$ is the vector of event types and $\vec{q}_P$ is the vector of sensing parameters for detection or monitoring task.

2.2.2. Event-Location-Collection-Position Relevance (ECR). The impact factor reflects the relevance degree between the event location and the sensing parameter collection position. As we all know, those parameters that are collected from event locations will have more influences on event detection performance. For example, in the fire detection, temperature parameters at fire location will have higher reference value than the ones at other locations. ECR is computed according to formula (2)

$$e_\text{score}(P,E) = \begin{cases} \#P's \text{ es of } E's \text{ location} & \text{if } E's \text{ location} \in La, \\ 0 & \text{if } E's \text{ location} \notin La, \end{cases} \tag{2}$$

where es indicates event and $La = \{l_1, l_2 \ldots, l_m\}$ means the set of position where sensing parameters are collected from.

2.2.3. Observer Recommendation Influence (ORI). It indicates one parameter's importance degree in the eyes of observers or domain experts during event detection. It is

well known that different experts will emphasize different sensing parameters in one same event detection task. Thus, the preference, that experts assign to one sensing parameter, will be very the important information for fusion. For example, environmental experts can obtain more accurate detection results of Algae outbreak by combining water quality, PH value, and water temperature parameter together, and ordinary observers can directly do it by using the value of Algae content information. ORI is calculated using the following formulas:

$$o_\text{score}(O,E) = \begin{cases} 1 - \text{spw}(O,E) & \text{if spw}(O,E) < 1, \\ 1 & \text{if spw}(O,E) \geq 1, \end{cases}$$
$$o_\text{score}(P,E) = \frac{\sum_{P's\ \text{observers}} o_\text{score}(O,E)}{\#P's\ \text{recommenders}}, \tag{3}$$

where O denotes an observer or a domain expert, E an event, and spw means the weight of shortest path between O and E. The term "observers" is equivalent with "recommenders."

2.2.4. Collection Cost Preference (CCP). It measures collection cost of sensing parameters selected in data fusion of WSNs applications in order to design more reasonable detection solution. For example, we can easily detect Algae outbreak event using Algae content sensors. However, for most environmental managers, it is impossible to do it by this way due to the technology difficulty or the high-cost solution. But according to suggestions of environmental experts, they can combine water quality, water temperature, water visibility, and PH value together to detect the same event. These kinds of information can be obtained using sensors or devices with high quality and low price. In this case, collection cost of sensing parameters has undoubtedly effect on the detection precision of Algae outbreak event.

Computing CCP needs to map all kinds of information from both sensing parameters and collecting sensors/devices into one same number interval so that comparisons might be made at one same order of magnitude. Therefore, Sigmoid function is used to map all inputs of real number field into the interval $[0,1]$, and then rank fraction function is applied to map ranking information into the interval $[0,1]$. They are defined as in (4), respectively,

$$\text{sigmoid} : S(t) = \frac{1}{1+e^{-t}},$$
$$\text{rank_score} : R(r,c) = 1 - \log_{10}\left(1 + 9 \times \frac{r-1}{c-1}\right), \quad c > 1, \tag{4}$$

where r indicates the rank of a sensing parameter or a device, and c indicates the number of the ranks. We first adjust input errors of devices and then carry out formatting and normalization processes. At last, combining sensing parameters' information, we compute $t_$rank—the rank of technology maturity degree of sensors/devices, $t_$count—the number of technology maturity degrees, $p_$rank—the rank of popularity degree of devices, $p_$count—the number of popularity degrees, $d_$rank—the rank of difficulty grade/degree for collecting sensing parameters, $d_$count—the number of difficulty grades, and gpa_score—the percentage transferred from the price of devices. Thus, we can compute collection cost score of a sensor/device for a parameter using formula (5), the final formula for computing CCP impact weight is given, as in (6):

$$\begin{aligned} c_\text{score}(P,I) = S(&R(I's\ t_\text{rank}, t_\text{count}) \\ &+ (I's\ p_\text{rank}, p_\text{count}) \\ &- R(P's\ d_\text{rank}, d_\text{count}) \\ &- I's\ \text{gpa_score}), \end{aligned} \tag{5}$$

$$c_\text{score}(P,E) = \frac{\sum_{P's\ \text{devices for}\ E} c_\text{score}(P,I)}{\#P's\ \text{devices for}\ E}, \tag{6}$$

where $d_\text{count} = n$, $t_\text{count} = m$, $p_\text{count} = k$, P means a kind of sensing parameter, I indicates a sensor/device, and E means a monitoring task or an event.

2.3. Historical Sensed Data in MPD-Model. Though lots of WSN-based applications had employed machine learning methods to conduct recognition in WSNs [1, 2], to the best of our knowledge, few work definitely introduce Historical Sensed Data (or Historical Data), shortly HSD, as the core component of fusion model [6]. In our opinion, HSD should be involved in data fusion model design due to the following facts.

(i) In WSNs, long-term continuous monitoring will inevitably produce massive data. The importance of these data is that it not only provides query services for users, but also lays a solid foundation of decision-making for future data mining.

(ii) In WSNs, most parameters, such as pulse, ECG, and physical image, have complex data patterns. Accordingly, diverse machine learning methods are needed to train, learn, and classify these patterns. Therefore, HSD should be undoubtedly involved in data fusion procedure of WSNs systems.

(iii) Parameter types often determine what decision-making methods should be employed to finish health status detection task. Specifically, in the case of unattended monitoring, only some simple data processing method can be employed.

In the case of human sensing, complex factors such as users' social relationships might be considered and accordingly call for more complicated data processing methods (e.g., Time-Space Factor Graph Model).

To illustrate the role of HSD in data fusion tasks, we conduct comparative diagnosis experiments based on AR [7] and WPT [8] methods to recognize the five kinds of human health statuses. SVM and user-defined rules are employed as decision-making methods. The comparison results, as in Figure 2, show that SVM-based methods with HSD often have higher recognition accuracies (30∼40%) than user-defined rules without HSD.

3. Health Status Detection

Chinese Pulse Diagnosis Theory (CPDT) [9] has been used in clinical diagnosis for thousands of years and has been proved to be valuable. According to CPDT, wrist-pulse data contains the abundant physiological and pathological parameter information of human body. Pulse information has different patterns which can identify different kinds of people with a specific health status or symptom/disease. Thus, sensed pulse data can be leveraged to effectively detect human health status.

3.1. TFD-Pattern: A Novel Time Series Pattern for Health Status Detection. In this paper, a unique kind of pulse pattern is designed based on CPDT [9] and our previous researches on pulse diagnosis [1, 5]. We first compute 90 time-domain feature variables based the five feature points (see Figure 3). For example, the feature variable T_1/T can be obtained through leveraging the T-point and the P-point. Then we compute 36 frequency-domain feature variables based on the discrete Fourier transform formula, as in (7), and the discrete power spectral density function, as in (8). Finally we obtain 6 feature variables for pressure value. Each of them is the average value of all the real-time external pulse pressures in one pulse waveform data. Actually, these above computations are just suboperations of $f(\cdot)$ in Figure 1:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)kn}, \tag{7}$$

$$\sum_{n=0}^{N-1} x^2(n) = \frac{1}{N} \sum_{k=0}^{N-1} |\mathrm{X}(k)|^2. \tag{8}$$

We merge the three types of feature variables into one 132-dimensional pulse pattern that stands for a wrist-pulse waveform and call it time-frequency-domain Pattern, shortly TFD-Pattern, as in Figure 4.

3.2. SVM-Based Feature-Decision-Making Algorithms for Health Status Detection. Since the SVM performs well on problems with low training set sizes, we conduct SVM-based identification experiments on these pulse data sample groups with different human health statuses. Based on decision function of SVM theory, as in (9), we propose a health status detection algorithm which runs on the central server:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{k} \alpha_i^* y_i K(E_i, x) + b^* \right), \tag{9}$$

where k is the number of pulse data samples; E is a set of feature vectors. K might be one of Linear, Quadratic, Polynomial, RBF, and MLP kernel functions. $\alpha_i y_i$ is weight vector and $\sum_{i=1}^{n} \alpha_i y_i = 0$.

It needs to be noted that the algorithm was designed just according to our proposed MPD-Model and is corresponding to $\alpha_i y_i$ in the model. Another characteristic of this algorithm is that it is based on our designed TFD-Pattern which is totally different from other pulse diagnosis (or health status detection) methods [7, 8].

Figure 2: Contribution and importance of Historical Sensed Data (HSD) for data fusion (user-defined rules are designed according to expert's prior knowledge. At the x-axis, **H** shows the comparison results of identifying the Hypertensive from the Normal, **S**—Subhealthy versus Normal, **Co**—Coronary Heart disease versus Normal, **Ci**—Cirrhosis disease versus Normal, and **P**—Pregnant versus Normal female.)

4. Distributed Multipreference-Driven Feature-Level Fusion Algorithms Based on Weighted Average Method

In terms of $f(\cdot) : R_1 \blacklozenge R_2 \blacklozenge \cdots \blacklozenge R_m$ in MPD-Model, we present MFA, a novel distributed and unified multi-preference fusion algorithm based on weighted average method. MFA first computes multi-preference parameter's impact weights of detecting health statuses using probabilities of four impact factors mentioned above, as in (10), and

Figure 3: The five most important physiological feature points of human pulse waveform data [5].

Figure 4: Illustration of our designed TFD-Pattern.

then fuses all physiological parameters to generate parameter patterns for further health status detection decision-making, as in (11):

$$\begin{aligned}
&\text{impact_weight}(P, D) \\
&\quad = (t_\text{score}(P, D) \times t_\text{weight} \\
&\qquad + m_\text{score}(P, D) \times m_\text{weight} \\
&\qquad + e_\text{score}(P, D) \times e_\text{weight} \\
&\qquad + c_\text{score}(P, D) \times c_\text{weight}) / \\
&\qquad (t_\text{weight} + m_\text{weight} + e_\text{weight} + c_\text{weight}),
\end{aligned} \tag{10}$$

$$\text{Fusion}(P_1, P_2, \ldots, P_m) = \bigcup_{i=1}^{m} f(P_i) \blacklozenge \text{impact_weight}(P_i, D), \tag{11}$$

where the initial values of all weights are set to 1 and $\sum_{i=1}^{m} \text{impact_weight}(P_i, D) = 1$. With receiving feedbacks from users, the model would use feedback data to optimize these weight values.

Figure 5: Wrist-pulse samples distribution in the real-world health care data sets.

5. Experimental Results

5.1. Experimental Setup. In our previous work, we had developed a wearable healthcare monitoring sensor network system [1] for health status detection. This system helps us to collect the real-world health care data set in a large scale clinic experiments which happened at the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS) in 2009 and 2010 Hi-tech Fair of China in Shenzhen City, respectively. Wrist-pulse data samples distribution is shown in Figure 5.

We validate the effectiveness of the proposed methods comparing with the following baseline methods.

5.1.1. Derivative-Based Feature Extraction Method [11]. In this method, an automated derivative-based time-domain feature extraction method on wrist pulse waveform is proposed to extract magnitude-type feature h_i ($i = 1$ to 5). Here, the set of h_i is equal to the five feature points (see Figure 3) that are employed in our work.

5.1.2. AR Pulse Diagnosis Method [7]. In this health status detection method, an autoregressive- (AR-) based method is proposed to extract the pulse signal features. The mean and variance of the prediction error are calculated and selected as features. The selected features are then taken as inputs to a support vector machine (SVM) for diseases classification.

5.1.3. WPT Pulse Diagnosis Method [8]. This health status detection made pulse waveform feature extraction based on optimal Wavelet Packet Transform (WPT). Subband energies contained in the best basis were extracted as features and SVM classifiers were trained, differentiating cholecystitis patients and nephritic syndrome patients from normal people.

We evaluate our method from the following four angels.

(1) *Precision*. We evaluate the proposed method in terms of Extraction Precision and Diagnosis Precision comparing to corresponding baseline methods.

(2) *Fusion Ratio*. It states the ratio of the fused file size versus the original file size. The former file is used to save multi-preference parameter data. Fusion ratio is computed by the formula: (1 − fused size/original size)∗100%.

(3) *CPU Time*. It indicates how long learning procedures would take for different SVM_based methods. It is employed to measure real time of these comparative methods.

(4) *Significant Difference*. It can tell users what feature variables in TFD-Pattern can much help in revealing the human symptom condition, and then, it would prove that our proposed method is effective if any valid feature variable with significant difference exists.

5.2. Feature Extraction. According to CPDT [9], we must first extract the five most important time-domain feature points (see Figure 3) before running MFA algorithm in our health status detection. Thus, we propose FEA, an adaptive feature extraction algorithm for pulse data. It can be considered as an instance of $P_i \rightarrow R_i$ $(1 \leq i \leq n)$ in MPD-Model. Specifically, we formalize the slope variability as a trend decision function, as in (12). Then, the final desirable feature points can be selected from the candidate ones according to some adaptive parameters. The most important adaptive parameter Threshold_of_PULSE is used to find the PB-point, the TD-point, and the PD-point, and it is computed by (13). Another adaptive parameter Amp_pulse indicates the mean value of all pulse cycles in one pulse waveform. It can be computed by (14):

$$f(i) = \begin{cases} 0 & \text{Other}, \\ 1 & s(i) < 0,\ s(i+1) < 0, \\ 2 & s(i) > 0,\ s(i+1) < 0, \\ -2 & s(i) < 0,\ s(i+1) > 0, \end{cases} \tag{12}$$

$$\text{Threshold_of_PULSE} = \frac{(1/Fs + m)}{\text{Amp_pulse}} \tag{13}$$

$$\text{Amp_pulse} = \frac{1}{\text{Peak_Num}} \sum_{k=1}^{\text{Peak_Num}-1} (T\text{-point}(k+1) - T\text{-point}(k)). \tag{14}$$

5.2.1. Effectiveness of Adaptive Parameters. To illustrate the effectiveness of Threshold_of_PULSE, we run FEA algorithm on normal pulse data with 50 Hz, 100 Hz, and 1000 Hz sampling frequencies, respectively. The results are shown in Figure 6. From this figure, we can see that all the beacon features points can be accurately extracted.

Figure 6: Effectiveness of the adaptive parameter Threshold_of_PULSE.

Similarly, according to our prior knowledge, we may set Amp_pulse = $T/2$ when the irregular slippery pulse is processed, and Amp_pulse = $T/3$ when the irregular normal pulse is dealt with. T indicates the length value of pulse waveform cycle. The results in Figure 7 indicate that FEA algorithm also well performs on different wrist-pulse conditions.

5.2.2. Performance Evaluation of Feature Extraction. We compare the proposed FEA algorithm with a typical derivative-based method [11] via extracting these five feature points on 354 healthy people's wrist-pulse samples. Table 1 shows comparison results of extraction accuracies of 10 random examples.

From Table 1, FEA outperforms (+16%) the derivative-based method due to the following. (1) Though the derivative-based method is also light-weight, it is vulnerable to heavy noise from human body's activities. (2) The FEA not only is light-weight and provides great support to real-time pulse signal processing tasks, but also makes full use of many adaptive parameters based on automatic computing or prior knowledge.

5.3. Fusion Effectiveness. In our experiments, different types of data at each processing stage would be saved into the respective files. Considering the existing information fusion estimator, for example, sample-based estimator using measurements [12], we calculate data fusion\reduction ratios by comparing file sizes before and after processing sensed pulse data. Reduction degree of these files size directly shows the performance of MFA algorithm. Table 2 displays MFA's data fusion ratios. In this table, the first column indicates

Figure 7: Effectiveness of prior knowledge: Amp_pulse.

Figure 8: Wrist-pulse from diverse sources has different effects on revealing health status conditions according to Traditional Chinese Medicine [10].

Table 1: Comparison results of FEA versus the derivative-based method.

Number	ID	Derivative-based method	FEA
	...	...	...
26	A19_709A	71%	86%
27	A20_809C	70%	84%
28	A22_737E	67%	90%
29	A37_719E	72%	82%
30	A66_524G	69%	89%
31	B20_735E	74%	89%
32	B9_0724C	64%	84%
33	B10_723E	70%	78%
34	B18_706K	73%	89%
35	B19_706G	70%	88%
	...	...	...
	Average	70%	86%

the number of one wrist-pulse ample, the second column stands for the sizes of these files before running MFA algorithm, the third column shows the sizes of these files after running MFA algorithm, and the final one gives data fusion ratios. The results show MFA algorithm has excellent fusion effect for wrist-pulse data.

5.4. Health Status Detection Performance. We make evaluations from the following aspects.

Table 2: Data fusion ratios of MFA algorithm.

Data file name	Original size (KB)	MFA	Ratios
....	....	..	...
NORMAL0613B_4_01	1491	14	99.06%
hBP_0612A_0_01	658	6	99.08%
hBP_AO0709B_1_01	192	3	98.44%
NORMAL0612B_0_01	271	5	98.15%
NORMAL0709A_5_02	858	8	99.07%
hBP_AO0709_1_01	119	2	98.32%
NORMAL0715A_1_01	1058	11	98.96%
....	....	..	...
Average			98.73%

Table 3: Different weights of multi-preference parameters for health status detection.

	P_1	P_2	P_3	P_4	P_5	P_6
Weights solution-I of impact factors						
H	0.5	0	0	0	0	0.5
Co	1	0	0	0	0	0
Ci	1	0	0	0	0	0
P	0	0	0.5	0	0	0.5
Weights solution-II of impact factors						
H	0	0.25	0.25	0.25	0.25	0
Co	0	0.2	0.2	0.2	0.2	0
Ci	0	0.2	0.2	0.2	0.2	0
P	0.25	0.25	0	0.25	0.25	0
Weights solution-III of impact factors						
H	0.3	0.1	0.1	0.1	0.1	0.3
Co	0.7	0.06	0.06	0.06	0.06	0.06
Ci	0.8	0.04	0.04	0.04	0.04	0.04
P	0.02	0.02	0.4	0.02	0.02	0.4

Table 4: Feature variables with significant difference.

Feature variable number	Left/right wrist	Cun, Guan, or Chi	Feature variable name
Number 22	Left	Cun	Real-time external pulse pressure (EPP)
No. 59	Left	Chi	S_5/S

5.4.1. Impact Factor Weight Validation. According to Chinese Pulse Diagnosis Theory (CPDT) [9, 10], we investigate that what wrist-pulse parameters from diverse sources (i.e., Cun, Guan and Chi on left and right human wrists [13]) have great impacts on revealing disease/symptom information. Based on this investigation (see Figure 8), we further compute different weight values that are listed in the 2–6 columns of Table 3. According to formula (1), we detect four kinds of health statuses or diseases/symptoms: hypertension (H), coronary (Co), cirrhosis (Ci), and pregnancy (P). These weights are selected and grouped into three weight solutions shown in Table 3. According to these weight solutions we conduct comparative experiments and give the results shown

Figure 9: AR-based comparison results of different weights solutions for multi-preferences parameters (characters **L**, **Q**, **P**, **R**, and **M** stand for Linear, Quadratic, Polynomial, RBF, and MLP kernel function of SVM, respectively. The title of every subfigure tells which health status had recognized).

in Figures 9 and 10. From the two figures, we can see that the greater weight values of related impact factor (P_i) become the higher health status detection precisions have. This proves that our multi-preferences-driven data fusion-based impact factor evaluation can well work.

5.4.2. Comparative Evaluation. In this paper, we compare with the baseline methods, AR [7] and WPT [8] methods, to validate the effectiveness of the proposed MPD-Model. The two methods and ours all focus on feature extraction of wrist-pulse data and are based on SVM to distinguish patients with different diseases from normal people. The difference is that distinct wrist-pulse pattern is employed to recognize. Thus, we can get solid comparison results shown in Figure 11. These comparative experiments are conducted on the real healthcare data sets mentioned above (see Figure 6).

CPU time of ours, AR, and WPT health status detection method is approximately 17.3121 s, 164.7448 s and 42.0289 s, respectively. Among all comparative experiments, our SVM-based health status detection method takes the least running time and thus it is proved to be light-weight and more suitable to fit wearable healthcare monitoring system based on WSNs. As for hypertension, coronary, cirrhosis, and pregnancy symptoms, our method outperforms (average 15∼30%) AR and WPT methods. Thus, our method is more effective to detect human health status. However, identification accuracy of our method is a little lower than the two methods when detecting subhealth status. There are two reasons to explain it: (1) low accuracies of feature extraction for pulse waveform data will cause that it is hard

Figure 10: WPT-based comparison results of different weight solutions for multipreference-parameters (explanations of this figure have been given in Figure 9).

to clearly distinguish some pulse patterns from others. But AR and WPT methods have no such problem just because algorithms of computing feature vector are totally different from ours. (2) Generally, it is very difficult to identify people with subhealth status from the normal ones since such patients often have not clear disease characteristics.

5.4.3. Effectiveness Evaluation. At last, feature variables with significant difference in the two groups, as in Table 4, are obtained by conducting T statistical analysis method.

As Figure 3 shows, feature variable S_5/S indicates the ratio of TD-point–T-point area to the whole area of one pulse cycle. It reveals clear physiological meaning. That is, patient's reflected wave of blood flow in one heartbeat cycle are stronger than the healthy. This is may be one reason that blood pressure values of a hypertension people are always high. Another feature variable, real-time External Pulse Pressure (EPP), indicates that real-time blood pressure values recorded by pulse-sensor nodes have distinct effects on the patients. It is more reasonable than S_5/S to explain such the pathological phenomenon that patient's blood pressure is higher than healthy people's.

6. Related Work

WSN-based healthcare solutions become a hot research topic. It has been well developed through the research work in activity recognition [14], physiological data gathering [15], pattern recognition [16], and so forth. Though lots of existing work collect blood pulse data for a healthcare-monitoring purpose, few work study systematically how to fuse parameters from diverse sources based on impact factors.

Figure 11: Comparison with AR and WPT methods (Part (a), (b), (c) and (d) show SVM-based diagnosis experiments with Linear, Quadratic, Polynomial and RBF kernel function, respectively. The y-axis indicates Precision (%) of health status or disease/symptom identification, and the explanations of the x-axis has already been given in Figure 2.).

The five most important feature points (see Figure 3) considered as the basis of the objectification of pulse study by most researchers [17]. Currently, most extraction methods of time domain features are derivative based [11]. This kind of method is easy to be implemented on sensor nodes, but is just suitable to regular pulse waveform with high Signal-to-Noise Ratio (SNR). To address the problem of limited resources, feature extraction in our work focuses on sensed pulse data with low SNR. The proposed algorithm is light-weight and adaptive. Much work focus on pulse detection and classification. For example, [18] proposed a Lempel-Ziv complexity analysis-based approach to classify seven pulse patterns that exhibit different rhythms. Besides, Chinese Pulse Diagnosis Theory (CPDT) is also objectified in order to pulse acquisition, feature extraction of pulse, and pulse diagnosis [17]. As a powerful method for pattern classification with low training set sizes, SVM is often leveraged to diagnose pulse signals/patterns that are computerized using Modified Auto-Regressive Models [7] or spatial and spectrum features [19].

Though there already exist some data fusion methods, for example, Crash Fault Correction [20], Privacy Protection [21], position-based Aggregator node election in WSNs [22, 23], and environmental Monitoring [24], they are not suitable for Wearable Sensor Networks due to complicated sensed data patterns and high intelligence requirements [6]. Reference [25] also focus on multisource fusion model, but employing Chi-square distribution with n degrees of freedom method instead of weighted average method. Reference [12] discusses information fusion estimation problems and provides useful estimation methods for our proposed data fusion model. A continuous classification of multisource sensed data in [26, 27] inspired us to introduce historical data into designing multisource-driven data fusion model.

7. Conclusion

In this paper, we try to systematically investigate the problems of data fusion and health status detection and propose MPD-Model, a novel-distributed multi-preference-driven data fusion model for WSNs. Based on MPD-Model, we present MFA algorithm (an instance of $f(\cdot)$ in MPD-Model), EFA algorithm (P_i), and SVM-based health status detection method (y_i). We also design TFD-Pattern, a novel time serials pattern according to CPDT. It is a kind of human pulse pattern as well as is both EFA's output and MFA's input. Experimental results show that they all outperform baseline methods with the feature extraction accuracy of +86% and the system diagnosis accuracy of +75%, and our proposed MPD-Model is reasonable and effective.

Acknowledgments

This paper is supported in part by the National Basic Research Program of China (973 Program) under Grant no. 2011CB302803, the "Strategic Priority Research Program" of the Chinese Academy of Sciences under Grant no. XDA060307000, the National Natural Science Foundation of China under Grant no. 61003292, and the Key Project of Natural Science Foundation of Jiangsu Province under Grant no. BK201100.

References

[1] J. J. Zhang, R. Wang, S. L. Lu et al., "EASiCare: design and implementation of a portable chinese pulse-wave retrieval system," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, (SenSys '11)*, pp. 149–161, Seattle, Wash, USA, 2011.

[2] M. Keally, G. Zhou, G. L. Xing et al., "PBN: towards practical activity recognition using smartphone-based body sensor networks," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, (SenSys '11)*, pp. 246–259, Seattle, Wash, USA, 2011.

[3] T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, and J. Song, "Demo: 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. 359–360, Seattle, WA, USA, 2011.

[4] E. Ertin, N. Stohs, S. Kumar et al., "AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, (SenSys '11)*, pp. 274–287, Seattle, Wash, USA, 2011.

[5] J. B. Gong, S. L. Lu, R. Wang, and L. Cui, "PDhms: pulse diagnosis via wearable healthcare sensor network," in *Proceedings of the IEEE International Conference on Communications, (ICC '11)*, Kyoto, Japan, 2011.

[6] E. F. Nakamura, A. A. F. Loureiro, and A. C. Frery, "Information fusion for wireless sensor networks: methods, models, and classifications," *ACM Computing Surveys*, vol. 39, no. 3, Article ID 1267073, pp. A9/1–A9/55, 2007.

[7] Y. Chen, L. Zhang, D. Zhang, and D. Y. Zhang, "Computerized wrist pulse signal diagnosis using modified auto-regressive models," *Journal of Medical Systems*, vol. 35, no. 3, pp. 321–328, 2011.

[8] Q. L. Guo, K. Q. Wang, D. Y. Zhang, and N. M. Li, "A wavelet packet based pulse waveform analysis for cholecystitis and nephrotic syndrome diagnosis," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, (ICWAPR '08)*, pp. 513–517, August 2008.

[9] L. Xu, D. Zhang, and W. Kong, "Wavelet-based Cascaded Adaptive Filter for Removing Baseline Drift in Pulse Waveforms," *IEEE Transaction on Biomedical Engineering*, vol. 52, no. 11, pp. 1973–1975, 2005.

[10] Z. F. Fei, *Contemporary Sphygmology in Traditional Chinese Medicine*, People's Medical Publishing House, Beijing, China, 2003.

[11] C. Xia, Y. Li, J. Yan et al., "Wrist pulse waveform feature extraction and dimension reduction with feature variability analysis," in *Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering, (ICBBE '08)*, pp. 2048–2051, May 2008.

[12] R. N. Madan and N. S. V. Rao, "A perspective on information fusion problems," *International Journal of Distributed Sensor Networks*, vol. 5, no. 1, 4 pages, 2009.

[13] Y. Chen, L. Zhang, D. Zhang, and D. Y. Zhang, "Wrist pulse signal diagnosis using modified Gaussian models and Fuzzy C-Means classification," *Medical Engineering and Physics*, vol. 31, no. 10, pp. 1283–1289, 2009.

[14] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, "epSICAR: an emerging patterns based approach to sequential, interleaved and concurrent activity recognition," in *Proceedings of the 7th Annual IEEE International Conference on Pervasive Computing and Communications, (PerCom '09)*, pp. 1–9, Galveston, Tex, USA, March 2009.

[15] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies, (INFOCOM '04)*, pp. 2571–2582, March 2004.

[16] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[17] L. Xu, K. Wang, D. Zhang et al., "Objectifying researches on traditionalchinese pulse diagnosis," *Informatica Medica Slovenica*, pp. 56–63, 2003.

[18] L. Xu, D. Zhang, K. Wang, and L. Wang, "Arrhythmic pulses detection using Lempel-Ziv complexity analysis," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 018268, 12 pages, 2006.

[19] D. Y. Zhang, W. Zuo, D. Zhang et al., "Wrist blood flow signal-based computerized pulse diagnosis using spatial and spectrum features," *Journal of Biomedical Science and Engineering*, vol. 4, no. 3, pp. 361–366, 2010.

[20] B. Balasubramanian and V. K. Garg, "Fused data structures for handling multiple faults in distributed systems," in *Proceedings of the International Conference on Distributed Computing Systems, (ICDCS '11)*, pp. 677–688, 2011.

[21] W. B. He, X. Liu, H. Nguyen, and K. Nahrstedt, "A cluster-based protocol to enforce integrity and preserve privacy in data aggregation," in *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops, (ICDCS '09)*, pp. 14–19, 2009.

[22] L. Buttyán and P. Schaffer, "Position-based aggregator node election in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2010, Article ID 679205, 15 pages, 2010.

[23] A. Mpitziopoulos, D. Gavalas, C. Konstantopoulos, and G. Pantziou, "CBID: a scalable method for distributed data aggregation in WSNs," *International Journal of Distributed Sensor Networks*, vol. 2010, Article ID 206517, 13 pages, 2010.

[24] S. Dauwe, T. Van Renterghem, D. Botteldooren, and B. Dhoedt, "Multiagent-based data fusion in environmental monitoring networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 324935, 15 pages, 2012.

[25] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1367–1377, 1999.

[26] G. Xing, J. Wang, K. Shen, Q. Huang, X. Jia, and H. C. So, "Mobility-assisted spatiotemporal detection in wireless sensor networks," in *Proceedings of the 28th International Conference on Distributed Computing Systems, (ICDCS '08)*, pp. 103–110, July 2008.

[27] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3858–3866, 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 891321, 12 pages
doi:10.1155/2012/891321

Research Article

Optimal Convergecast Scheduling Limits for Clustered Industrial Wireless Sensor Networks

Xiaoling Zhang, Wei Liang, Haibin Yu, and Xisheng Feng

Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

Correspondence should be addressed to Haibin Yu, yhb@sia.cn

Received 27 July 2012; Revised 12 October 2012; Accepted 14 October 2012

Academic Editor: Hongwei Zhang

Copyright © 2012 Xiaoling 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.

Increased mobility coupled with a possible reduction of cabling costs and deployment time makes wireless communication an attractive alternative for the industrial process monitoring and control. The major obstacles toward the utilization of wireless industrial networks are predominantly the timeliness and reliability requirements. In this paper, orienting to clustered industrial wireless sensor networks, we analyze the performance bounds of the convergecast scheduling, which is a typical many-to-one communication in wireless industrial networks. The analysis aims to the cluster-line and cluster-tree topologies. Each kind is future divided into three scenarios according to an application parameter, named data update rate. Firstly, we establish the lower bounds on the number of timeslots to finish the intracluster and the intercluster convergecast transmissions. Secondly, we establish the lower bounds on the number of channels based on the lower bounds on the number of timeslots and maximum available channels in a multichannel scenario. Lastly, we carry out the extensive analysis-taking packet retransmissions into consideration so as to meet the reliability requirement. Experiment results validate the correctness and tightness of our theory analysis.

1. Introduction

With the success of wireless technologies in consumer electronics, standard wireless technologies are envisioned for the deployment in industrial environments to improve the functionality and the efficiency, which boost the formation of the industrial wireless sensor networks (IWSNs) [1]. Recently, some international organizations are actively promoting the standardized process of IWSNs and have achieved several productions, such as WirelessHART [2] and Wireless network for industrial automation-process automation (WIA-PA) [3]. Convergecast is a typical many-to-one communication pattern in IWSNs. In convergecast, many or all nodes in the network send data to a sink node during a relatively short time period [4]. In these applications all packets generated in the network have to reach a sink node either for record or for computationally intensive analysis.

A guarantee on packet delivery and a bound on convergecast latency are highly desirable in mission critical applications. Guarantee on packet delivery ensures availability of accurate information about the sensing field. An optimal bound on convergecast latency leads to timely detection of the events. Furthermore, a known time bound on convergecast latency can help the sink node to schedule convergecast requests and related computations [4].

As concluded in [4], collisions present a major challenge in convergecast when contention-based MAC protocol like CSMA are employed. Collisions will result in loss of packets. The recovery methods, such as retransmissions method, will increase the latency. Therefore, the convergecast latency incurred by radio interference is far from the optimal [4]. Contention-free MAC protocols like TDMA can be used to eliminate collisions and obtain a bound on the time required to complete convergecast. Hence, we are interested in determining a TDMA schedule such that the entire convergecast can be completed in minimal number of timeslots and in minimal number of channels.

Choi et al. [5] proved that convergecast scheduling problem is NP-hard. Florens et al. [6] consider that the convergecast is the reverse operation of the broadcast scheduling. If node i sends packets to node $i+1$ by using timeslot t during broadcast scheduling, then node $i+1$ send packets to node i by using timeslot $T-i+1$. T is the minimal schedule length and is calculated according to the number of packets that

Figure 1: Superframe structure.

sink node sends to node i. Gandham et al. [4] proved that once convergecast scheduling requires at most $3N$ timesltos, where N represents the number of nodes in the network. They also proved that at lease $3N - 3$ timeslots are required to complete convergecast in a linear network with N nodes; at least $\max(3n_k - 3, N)$ timeslots are required to complete convergecast in a multiline network; at least $\max(3n_k - 1, N)$ timeslots are required to complete convergecast in a tree network. Yu et al. [7] consider a network as a disc and proved that the upper bound on the number of timeslots is $24D + 6\Delta + 16$, where D is the network diameter and Δ is the maximum node degree. Soldti [8, 9] consider the jointly time- and channel-optimal convergecast scheduling, and proved the lower bound on the number of timeslots and channels for convergecast transmission in mesh network. Shang et al. [10] considered how to collect data from sensors deployed in the Euclidean plane in a time-efficient way. Zhang et al. [11] focused on bursty convergecast and proposed a windowless block acknowledgement scheme to achieve reliable and real-time error control and contention control. Liu et al. [12] estimated the probabilistic path delay bounds in an agile manner and designed a delay-sensitive routing protocol.

In summary, there are three problems in previous researches. First, the network topology is ad hoc or mesh and the network deployment is plane or quadrate, which are not suitable for industrial applications. Second, the cyclic data feature is not taken into consideration. Third, previous research is based on one channel, which are inherently inefficient. To solve these three problems, a new result specific to clustered IWSNs, where in particular, cyclic data collection over the network is assumed. In addition, we consider data update rates (DURs) under various assumptions and derive corresponding time and channel performance. This work extends the results in [8], where only mesh topology is assumed and the DURs are not taken into consideration. It should be noted that most of industrial networks use hierarchical topology and faced to different applications with different data features. Therefore, when requirements are more stringent, the results in [8] may no longer be practical. In particular, this paper presents the following contributions.

(i) For a cluster-line IWSN, we analyze and prove the performance limits for completing the convergecast transmission. We establish the lower bounds on the number of timeslots and channels based on three scenarios: all the data update rates (DURs) in a cluster are same, all the DURs in a cluster are different, and the DURs in a cluster are partly same.

(ii) For a cluster-tree IWSN, the same problem as the first item is studied.

The rest of this paper is organized as follows. First, the system model and problem formulation of convergecast scheduling are described in Section 2. The analysis of the time- and channel-optimal convergecast scheduling for networks with cluster-line and cluster-tree topologies are studied in Sections 3 and 4, respectively. This paper is concluded in Section 5.

2. Problem Formulation

Time is equally divided into timeslots. A collection of timeslots repeating constitute one superframe. Our superframe structure extends the IEEE STD 802.15.4-2006 superframe [13], and divides the inactive period into intracluster communication period, intercluster communication period, and sleep period. The extended superframe structure is shown in Figure 1. The superframe length of the gateway device or a routing device can be configured and the basic superframe length $\overline{S_f}$ is also configurable.

One field device initially generates at most one packet destined to the gateway device every DUR time during each convergecast. Packets are routed along $T = (V, E_r)$, where $E_r \subseteq E$ is the set of routing paths. For every device i, $I(i)$ and $O(i)$, respectively, denote the set of links that arrive at or leave device i; $p_t(i)$ denotes the capability of buffer after timeslot t, where we defined that $p_0(i) = 1$. Let g_{ij} be the number of packets of a cluster with R_{ij} being the cluster head in one superframe cycle. g_{ij} is decided by the DURs of all field devices in a cluster. Let L_s be the superframe length (in timeslots). Let x_{ijt} be a binary variable indicating whether timeslot t has been allocated to link (i, j). If timeslot t has been allocated to link (i, j), $x_{ijt} = 1$; otherwise, $x_{ijt} = 0$.

The time- and channel-optimal scheduling to complete convergecast with the objective to minimize both the number of timeslots and the number of channels can be informally formulated as [8].

Problem 1. Time-optimal convergecast scheduling:

$$\text{minimize } L_s, \tag{1}$$

s.t.

$$\sum_{(i,j)\in I(i)} x_{ijt} + \sum_{(i,j)\in O(i)} x_{ijt} \le 1, \quad \forall i \in V,\ \forall t \in [0, L_s], \tag{2}$$

$$p_{L_s}(\text{GW}) = \sum_{i,j} g_{ij}, \tag{3}$$

$$p_t(i) = p_{t-1}(i) + \sum_{(i,j)\in I(i)} x_{ijt} - \sum_{(i,j)\in O(i)} x_{ijt}, \quad \forall i \in V,\ \forall t \in [0, L_s], \tag{4}$$

$$x_{ijt} \in \{0, 1\}, \quad \forall (i, j) \in E_r. \tag{5}$$

Objective (1) guarantees the minimization of the convergecast scheduling length. Constraint (2) is the half-duplex requirement, which restricts devices not to transmit and receive simultaneously. Constraint (3) guarantees that all packets generated by field devices have been sent to the gateway device. Constraint (4) is the memory status. And constraint (5) indicates the link status.

Problem 2. Time- and channel-optimal convergecast scheduling:

$$\text{minimize } C_s = \max_{t\in[0,L_s]} \sum_{(i,j)\in E_r} x_{ijt}, \tag{6}$$

s.t.

$$L_s \text{ calculated by } (1) \sim (5), \tag{7}$$

where C_s is the number of channels used for transmission.

3. Time- and Channel-Optimal Convergecast Scheduling for Cluster-Line Topology

Cluster-line topology that is shown in Figure 2 is instrumental for more general routings, and is the preferred topology in certain applications such as pipeline monitoring and unmanned offshore gas production [8]. Without loss of generality, the GateWay device (GW) is placed at the right end of the line, and N routing devices $R_{11} \sim R_{1N}$ are placed from right to left. For routing device R_{1i}, $i \in [1, N]$, it acts as a cluster head and manages several field devices. f_{1i} denotes the number of field devices in the cluster with R_{1i} being cluster head; and F_{1ij} denotes the jth field device, $j \in [0, f_{1i}]$; C_{1ij} denotes the DUR of F_{1ij}. The following content studies three scenarios: all the DURs in a cluster are same, all the DURs in a cluster are different, and the DURs in a cluster are partly same, which are briefly denoted as ADS, ADD, and PDS.

3.1. Scenario I: ADS. Each field device in this cluster initially generates one packet at the begin of each superframe cycle. Considering the single-packet memory, the generated data must be sent out before the end of each superframe. Otherwise, the old data will be overlapped by newly generated data.

Figure 2: Cluster-line network topology.

3.1.1. Performance Limits on Intracluster Convergecast Scheduling. A routing devices manages one intracluster convergecast scheduling, and assigns timeslots and channels for communications between it and field devices. Because of the half-duplex radio transceiver, the length of intracluster convergecast scheduling should not be less than the number of field devices. For network scalability, we assume that the intracluster communication periods of all clusters are equal, which is defined as $\max_{i\in[1,N]} f_{1i}$. Each field device is assigned one timeslot during the intracluster communication period.

The intracluster convergecast scheduling is simple and intuitive. In one cluster, different field devices using different timeslots, while using one channel can eliminate collisions completely. We only need to consider the interference among clusters. Neighboring clusters using different channels can avoid collisions, which will be an extension in our future work.

3.1.2. Performance Limits on Intercluster Convergecast Scheduling. GW manages the intercluster convergecast scheduling, and allocates timeslots and channels for communications among routing devices and GW. The routing device nearest GW such as R_{11} in Figure 2 is the "bottleneck" of all communications, so that the lower bound on the number of timeslots for intercluster convergecast scheduling is determined by R_{11}.

Theorem 1. *If all the data update rates of field devices in a cluster are same, that is $C_{1i1} = C_{1i2} = \cdots = C_{1if_{1i}}$ and $i = 1, 2 \ldots N$, the lower bound on the number of timeslots used for the intercluster convergecast in a cluster-line routing network is $L_s = 2\sum_{i=2}^{N} f_{1i} + f_{11}$.*

Proof. At the beginning of each intercluster communication period, if $C_{1i1} = C_{1i2} = \cdots = C_{1if_{1i}}$ $(i = 1, 2 \ldots N)$, the number of packets in R_{1i} satisfies $g_{1i} = f_{1i}$. In Figure 2, R_{11} is responsible for receiving and forwarding packets from $R_{12} \sim R_{1N}$ to GW; R_{11} is also responsible for forwarding g_{11} packets in its cluster to GW. Due to the half-duplex constraint, R_{11} needs $2\sum_{i=2}^{N} g_{1i} + g_{11}$ timeslots. Reminded that $g_{1i} = f_{1i}$, then R_{11} needs $2\sum_{i=2}^{N} f_{1i} + f_{11}$ timeslots to forward all packets to GW. Therefore, the lower bound on the number of timeslots used for the intercluster convergecast in a cluster-line network is $L_s = 2\sum_{i=2}^{N} f_{1i} + f_{11}$. □

Theorem 2. *Given any scheduling algorithm S that can complete intercluster convergecast in $L_s = 2\sum_{i=2}^{N} f_{1i} + f_{11}$ timeslots*

in a cluster-line routing network, the lower bound on the number of channels used in S is

$$\left\lceil \frac{2\sum_{i=2}^{N} f_{1i} + f_{11} + 1}{2} - \sqrt{\left(\frac{2\sum_{i=2}^{N} f_{1i} + f_{11} + 1}{2}\right)^2 - \sum_{i=1}^{N} i f_{1i}} \right\rceil. \tag{8}$$

Proof. Let C_s be the number of channels that can be used in a network, and $\mathrm{PT}_{\max}(t)$ denotes the number of parallel transmissions scheduled in timeslot t ($t \in [1, L_s]$). In order to complete convergecast in L_s timeslots, the following conditions should be satisfied.

(i) When $t = L_s$, there should be only one packet in the network and R_{11} forwards it to GW. That is, only one device could be scheduled to transmitted if $t = L_s$.

(ii) When $t = L_s - 1$, there exist two situations. First, R_{12} sends its last packet to R_{11} in the timeslot t; and R_{11} forwards this packet to GW in timeslot L_s. Second, there exist only two packets in R_{11}; and R_{11} sends these two packets to GW, respectively, in timelsot $L_s - 1$ and L_s. In all, only one device could be scheduled to transmitted if $t = L_s - 1$.

(iii) Deduced as so forth, the maximum number of devices allowed to transmit in timeslot t is $\lceil (L_s - t + 1)/2 \rceil$.

Since $\lceil (L_s - t + 1)/2 \rceil \geq C_s$, when $t \in [1, L_s - 2(C_s - 1)]$ and the upper bound on the number of channels is C_s, we can obtain the following conclusion as in [8]:

$$\mathrm{PT}\max(t) = \begin{cases} C_s, & t \in [1, L_s - 2(C_s - 1)] \\ \left\lceil \dfrac{L_s - t + 1}{2} \right\rceil, & t \in [L_s - 2(C_s - 1) + 1, L_s]. \end{cases} \tag{9}$$

By (9), the maximum number of transmissions that can be scheduled in L_s using C_s channels is

$$\begin{aligned} \sum_{t=1}^{L_s} \mathrm{PT}\max(t) &= \sum_{t=1}^{L_s - 2(C_s - 1)} C_s + \sum_{t=L_s - 2(C_s-1)+1}^{L_s} \left\lceil \frac{L_s - t + 1}{2} \right\rceil \\ &= -C_{s^2} + C_s(L_s + 1). \end{aligned} \tag{10}$$

For intercluster convergecast, only routing devices and GW along the line routing transmit packets. Thus, the number of transmissions for completing convergecast is at least $\sum_{i=1}^{N} i f_{1i}$. Combined with (10), the number of transmissions satisfy the following inequality:

$$-C_{s^2} + C_s(L_s + 1) \geq \sum_{i=1}^{N} i f_{1i}. \tag{11}$$

Replacing L_s with $2\sum_{i=2}^{N} f_{1i} + f_{11}$, the lower bound on the number of channels used to complete convergecast in $2\sum_{i=2}^{N} f_{1i} + f_{11}$ timeslots is

$$\left\lceil \frac{2\sum_{i=2}^{N} f_{1i} + f_{11} + 1}{2} - \sqrt{\left(\frac{2\sum_{i=2}^{N} f_{1i} + f_{11} + 1}{2}\right)^2 - \sum_{i=1}^{N} i f_{1i}} \right\rceil. \tag{12}$$

□

From Theorems 1 and 2, we can discover that the results deduced in [8, 9] is special cases of our work when all the data update rates of field devices in a cluster are same.

3.2. Scenario II: ADD. If all the DURs of field devices in a cluster with R_{1i} ($i \in [1, N]$) being cluster head are different, is satisfied the following.

$$C_{1ij} \neq C_{1ik}, \quad \forall j, k \in [1, f_{1i}],\ j \neq k,\ i \in [1, N]. \tag{13}$$

3.2.1. Performance Limits on Intracluster Convergecast Scheduling. Field devices with different DURs can realize the timeslot multiplex. For example, if the superframe cycle is 1 and the DURs of two field device R_{11i} and R_{11j} ($i \neq j$) are 2 and 4, respectively, then the timeslot of R_{11j} can be occupied by R_{11i} when R_{11j} has no data to be transmitted.

Theorem 3. *Given a cluster with more than one field device, if all the data update rates of field devices are different, that is $C_{1ij} \neq C_{1ik}$, for all j, $k \in [1, f_{1i}]$, $j \neq k$ & $i \in [1, N]$, the lower bound on the number of timeslots used for the intracluster convergecast in a cluster-line routing network is 2.*

Proof. Obviously, for a cluster with only one field device, the intracluster convergecast can be completed by using only one timeslot. We consider the cluster with more than one field device below. For a cluster with R_{1i} ($i \in [1, N]$) being cluster head, the number of field devices is f_{1i} ($f_{1i} \geq 2$) and DUR of the jth field device is C_{1ij} ($j \in [1, f_{1i}]$). We assume that the DUR of a field device is 2^n, where n is a natural number. Without loss of generality, we suppose that $C_{1i1} < C_{1i2} \ldots < C_{1if_{1i}} = 2^{\Theta}$ (Θ is a natural number).

Denoted that the superframe length of R_{1i} is $\overline{S_f} \times C_{1i1}$. $\overline{S_f}$ denotes the basis superframe length. F_{1ij} that has the minimal DUR needs one timeslot per superframe cycle to transmit its packet. Other field devices cannot occupy its timeslot and only one timeslot can not finish the intracluster convergecast.

During $C_{1if_{1i}}/C_{1i1}$ superframe cycles, the first timeslot in every intracluster communication period is allocated to F_{1i1} and is denoted as $TC_{1i1} + k_1(C_{1i1} \times \overline{S_f})$, where TC_{1i1} is the index of the timeslot during the first superframe cycle and $k_1 = 0, 1, \ldots (C_{1if_{1i}}/C_{1i1} - 1)$.

For F_{1i2}, the available timeslot is the second timeslot in each intracluster communication period. We choose one timeslot numbered as 2 in the intracluster communication period from the first C_{1i2}/C_{1i1} superframe cycles and denoted as TC_{1i2}, which is to guarantee the packet in F_{1i2} can be sent out before DUR. Since F_{1i2} updates its data every $\overline{S_f} \times (C_{1i2}/C_{1i1})$ timeslots, F_{1i2} will occupied the timeslot indexed by $TC_{1i2} + k_2(C_{1i2} \times \overline{S_f}/C_{1i1})$ during $C_{1if_{1i}}/C_{1i1}$ superframe cycles, where $k_2 = 0, 1, \ldots (C_{1if_{1i}}/C_{1i2} - 1)$. During $C_{1if_{1i}}/C_{1i1}$ superframe cycles, the occupied number of timeslots number as 2 in the intracluster communication period is $C_{1if_{1i}}/(C_{1i1} \times C_{1i2})$ and the remainder number of the second timeslots in the intracluster communication period is $(C_{1if_{1i}}/C_{1i2}) \times (C_{1i2}/C_{1i1} - 1)$.

The timeslot assignment for the rest field devices can be done in the same manner as F_{1i2}. In all, the available timeslot

for F_{1ij} ($j \in [1, f_{1i}]$) is the second free timeslot in the intracluster communication period during the first C_{1ij}/C_{1i1} superframe cycles and occupied number is $C_{1if_{1i}}/C_{1ij}$ during $C_{1if_{1i}}/C_{1i1}$ superframe cycles.

For the last field device $F_{1Nf_{1N}}$, the number of idle timeslot numbered as 2 in the intracluster communication period during $C_{1if_{1i}}/C_{1i1}$ superframe cycles is:

$$\begin{aligned}&\frac{C_{1if_{1i}}}{C_{1i1}} - \frac{C_{1if_{1i}}}{C_{1i2}} - \frac{C_{1if_{1i}}}{C_{1i3}} - \cdots C_{1if_{1i}}/C_{1i(f_{1i}-1)} \\ &\quad \geq 2^{\Theta} - 2^{\Theta}/2 - 2^{\Theta}/4 - \cdots 2^{\Theta}/2^{(\Theta-1)} = 2,\end{aligned} \tag{14}$$

which illustrates that the number of free timeslot number as 2 in the intracluster communication period during $C_{1if_{1i}}/C_{1i1}$ superframe cycles is enough for all field devices completing their intracluster convergecast. □

3.2.2. Performance Limits on Intercluster Convergecast Scheduling. The following Corollaries 4 and 5 can be deduced from Theorems 1 and 2.

Corollary 4. *If all the data update rates of field devices in a cluster are different, that is $C_{1ij} \neq C_{1ik}$, for all $j, k \in [1, f_{1i}]$, $j \neq k$ & $i \in [1, N]$, the lower bound on the number of timeslots used for the intercluster convergecast in a cluster-line routing network is $4N - 2$, where N is the number of routing devices in a network.*

Proof. Let g_{1i} be the number of packets in R_{1i} ($i \in [1, N]$) at the beginning of each superframe cycle. The proving approach is similar to Theorem 1. The only difference is that the lower bound of g_{1i} is 2 because of Theorem 3. R_{11} needs $2\sum_{i=2}^{N} g_{1i}$ timeslots for forwarding packets from $R_{12} \sim R_{1N}$ and needs g_{11} timeslots for forwarding own packets to GW. In all, the lower bound on the number of timeslots used for intercluster convergecast scheduling is the sum of these two parts, which can be formulated as $L_s = 2\sum_{i=2}^{N} g_{1i} + g_{11} = 2 \times 2(N-1) + 2 = 4N - 2$. □

Corollary 5. *Given any scheduling algorithm S that can complete intercluster convergecast in $4N - 2$ timeslots in a cluster-line network, the lower bound on the number of channels used in S is $\lceil ((4N-1)/2) + \sqrt{(12N^2 - 12N + 1)/4} \rceil$, where N is the number of routing devices in a network.*

Proof. The proving approach is similar to Theorem 2. The only difference is to replace $L_s = 2\sum_{i=2}^{N} f_{1i} + f_{11}$ in Theorem 2 with $L_s = 4N - 2$. Then, the lower bound on the number of channels satisfy the following inequality:

$$C_s \geq \left\lceil \frac{(4N-2)+1}{2} - \sqrt{\left(\frac{(4N-2)+1}{2}\right)^2 - \sum_{i=1}^{N}(2i)} \right\rceil. \tag{15}$$

□

3.3. Scenario III: PDS. If the DURs of field devices in a cluster with R_{1i} ($i \in [1, N]$) being cluster head are partly same, the following is satisfied:

$$C_{1ij} = C_{1ik}, \quad \exists j, k \in [1, f_{1i}],\ j \neq k,\ i \in [1, N]. \tag{16}$$

3.3.1. Performance Limits on Intracluster Convergecast Scheduling. In this scenario, field devices with different DURs can also realize the timeslot multiplex.

Theorem 6. *Given a cluster with more than one field device and the data update rates of field devices in a cluster with R_{1i} being cluster head are partly same, that is $C_{1ij} = C_{1ik}$, $\exists j, k \in [1, f_{1i}]$, $j \neq k$, $i \in [1, N]$, the lower bound on the number of timeslots used for the intracluster convergecast in a cluster-line network is $\sum_{j=1}^{Q_{1i}} y_{1ij}$. N_{1ij} is the number of field devices with DUR being C_{1ij} and satisfies $\sum_{j=1}^{Q_{1i}} N_{1ij} = f_{1i}$ ($Q_{1i} \leq f_{1i}$); Q_{1i} is the number of DUR kinds in this cluster; y_{1ij} satisfies:*

$$\begin{aligned} y_{1i1} &= \left\lceil \frac{N_{1i1} \times C_{1i1}}{C_{1i1}} \right\rceil = \lceil N_{1i1} \rceil; \\ y_{1i2} &= \left\lceil \frac{N_{1i2} \times C_{1i1}}{C_{1i2}} \right\rceil; \\ y_{1ik} &= \begin{cases} 0, & \dfrac{C_{1i1}\left(N_{1i(k-1)}C_{1ik} + N_{1ik}C_{1i(k-1)}\right)}{C_{1i(k-1)}C_{1ik}} < y_{1i(k-1)} \\ \left\lceil \dfrac{N_{1ik} - \left(y_{1i(k-1)}C_{1i(k-1)}/C_{1i1} - N_{1i(k-1)}\right) \times \left(C_{1ik}/C_{1i(k-1)}\right)}{C_{1ik}/C_{1i1}} \right\rceil, & \text{else,} \end{cases} \end{aligned} \tag{17}$$

where $k = 3, 4, \ldots Q_{1i}$.

Proof. The basic idea of intracluster scheduling is to search free timeslots without increasing the number of total occupied timeslots during each superframe cycle. The scheduling cycle is designated as $C_{1iQ_{1i}}/C_{1i1}$ superframe cycles. We assume that $C_{1i1} \leq C_{1i2} \leq \cdots \leq C_{1iQ_{1i}}$.

The number of field devices that have the minimum DURs being C_{1i1} is N_{1i1}. These field devices can not occupy each other's timeslots. Therefore, the number of timeslots needed by field devices with DURs being C_{1i1} is $y_{1i1} = \lceil N_{1i1} \times C_{1i1}/C_{1i1} \rceil = \lceil N_{1i1} \rceil$.

The field devices with DURs being C_{1i2} need to be allocated the free second timeslot in the intracluster communication period during the first C_{1i2}/C_{1i1} superframe cycles. If these timeslots have already been occupied, the total number of timeslots should be increased. And the increased number of timeslots is $y_{1i2} = \lceil N_{1i2} \times C_{1i1}/C_{1i2} \rceil$.

For N_{1i3} field devices with DUR being C_{1i3}, the available timeslots can be searched from the free second timeslot in the intracluster communication period during the first C_{1i3}/C_{1i1} superframe cycles, which is numbered from 2 to $y_{1i1}+y_{1i2}$. If these timeslots are not enough for the N_{1i3} field devices, the number of timeslots should be increased. The number of free timeslots indexed from 2 to $y_{1i1}+y_{1i2}$ is $(\lceil N_{1i2}\times C_{1i1}/C_{1i2}\rceil\times(C_{1i2}/C_{1i1})-N_{1i2})\times(C_{1i3}/C_{1i2})$. The increased number of timeslots is determined by

$$y_{1i3}=\begin{cases}0, & \dfrac{C_{1i1}}{C_{1i2}C_{1i3}}(N_{1i2}C_{1i3}+N_{1i3}C_{1i2})<\left\lceil\dfrac{N_{1i2}\times C_{1i1}}{C_{1i2}}\right\rceil\\ \left\lceil\dfrac{N_{1i3}-(\lceil N_{1i2}\times C_{1i1}/C_{1i2}\rceil\times(C_{1i2}/C_{1i1})-N_{1i2})\times(C_{1i3}/C_{1i2})}{C_{1i3}/C_{1i1}}\right\rceil, & \text{else.}\end{cases}\tag{18}$$

Generally speaking, for N_{1ik} field devices with DUR being C_{1ik} $(k=3,4,\ldots Q_{1i})$, the available timeslots are from the free second timeslot to $\sum_{j=1}^{k-1}y_{1ij}$ in the intracluster communication period during the first C_{1ik}/C_{1i1} superframe cycles. If these timeslots are not enough for these N_{1ik} field devices, the number of timeslots should be increased. The number of free timeslots indexed from 2 to $\sum_{j=1}^{k-1}y_{1ij}$ is $(y_{1i(k-1)}\times(C_{1i(k-1)}/C_{1i1})-N_{1i(k-1)})\times(C_{1ik}/C_{1i(k-1)})$. The increased number of timeslots is determined by

$$y_{1ik}=\begin{cases}0, & \dfrac{C_{1i1}}{C_{1i(k-1)}C_{1ik}}(N_{1i(k-1)}C_{1ik}+N_{1ik}C_{1i(k-1)})<y_{1i(k-1)}\\ \left\lceil\dfrac{[N_{1ik}-(y_{1i(k-1)}\times(C_{1i(k-1)}/C_{1i1})-N_{1i(k-1)})\times(C_{1ik}/C_{1i(k-1)})]\times C_{1i1}}{C_{1ik}}\right\rceil, & \text{else.}\end{cases}\tag{19}$$

In all, the total number of timeslots used for intracluster convergecast is the sum of the increased number of timeslots, which is $\sum_{j=1}^{Q_{1i}}y_{1ij}$. □

3.3.2. Performance Limits on Intercluster Convergecast Scheduling. The following Corollaries 7 and 8 can be deduced from Theorems 1 and 2.

Corollary 7. *If the data update rates of field devices in a cluster are partly same, that is $C_{1ij}=C_{1ik}$, $\exists j,k\in[1,f_{1i}]$, $j\neq k$, $i\in[1,N]$, the lower bound on the number of timeslots used for the intercluster convergecast in a cluster-line routing network is $L_s=2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}$, where N is the number of routing devices in a network and Q_{1i} is the number of DUR kinds in a cluster with R_{1i} being the cluster head.*

Proof. Let g_{1i} be the number of packets in routing device R_{1i} $(i\in[1,N])$ at the beginning of each superframe cycle. The proving approach is similar to Theorem 1. The only difference is that the lower bound of g_{1i} is $\sum_{j=1}^{Q_{1i}}y_{1ij}$ according to Theorem 6.

R_{11} needs $2\sum_{i=2}^{N}g_{1i}$ timeslots for forwarding packets from $R_{12}\sim R_{1N}$ and needs g_{11} timeslots for forwarding own packets to GW. In all, the lower bound on the number of timeslots used for intercluster convergecast scheduling is the sum of these two parts, which can be formulated as $L_s=2\sum_{i=2}^{N}g_{1i}+g_{11}=2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}$. □

Corollary 8. *Given any scheduling algorithm S that can complete intercluster convergecast in $2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}$ timeslots in a cluster-line routing network, the lower bound on the number of channels used in S is*

$$\left\lceil\frac{2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}+1}{2}-\sqrt{\left(\frac{2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}+1}{2}\right)^2-\sum_{i=1}^{N}\left[i\sum_{j=1}^{Q_{1i}}y_{1ij}\right]}\right\rceil,\tag{20}$$

where N is the number of routing devices in a network and Q_{1i} is the number of DUR kinds in a cluster with R_{1i} being the cluster head.

Proof. The proof is similar to Theorem 2. The only difference is to replace $L_s=2\sum_{i=2}^{N}f_{1i}+f_{11}$ in Theorem 2 with $L_s=2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}$. Then, the lower bound on the number of channels used for intercluster convergecast scheduling is

$$\left\lceil\frac{2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}+1}{2}-\sqrt{\left(\frac{2\sum_{i=2}^{N}\sum_{j=1}^{Q_{1i}}y_{1ij}+\sum_{j=1}^{Q_{11}}y_{11j}+1}{2}\right)^2-\sum_{i=1}^{N}\left[i\sum_{j=1}^{Q_{1i}}y_{1ij}\right]}\right\rceil.\tag{21}$$

□

Figure 3: Cluster-tree network topology.

4. Time- and Channel-Optimal Convergecast for Cluster-Tree Topology

Figure 3 is a cluster-tree network. There exist N routing devices in the network, which manage several field devices and constitute clusters. Let D denote the depth of the routing tree. The GW is located at level 0 and has m subtrees $T_1, T_2, \ldots T_m$ rooted at routing devices $R_{11}, R_{21}, \ldots, R_{m1}$, respectively. The number of routing devices in subtrees $T_1, T_2, \ldots, T_m$, respectively, are $n_1, n_2, \ldots n_m$. Without loss of generality, it is assumed that $n_1 \geq n_2 \geq \cdots \geq n_m$. Therefore, T_1 is the largest subtree. Let R_{ij} denote the jth routing device in the ith subtree and f_{ij} be the number of field devices in the cluster with R_{ij} being the cluster head ($i \in [1, m]$, $j \in [1, n_i]$). Next, we will investigate the convergecast scheduling problems in cluster-tree networks from three scenarios: ADS, ADD, and PDS.

4.1. Scenario I: ADS. For a cluster-tree network, if all the DURs in a cluster are same, the number of timeslots for intracluster convergecast with R_{ij} being cluster head is at least equal to the number of field devices in this cluster, that is, the number of timeslots for intracluster convergecast is f_{ij} ($i \in [1, m]$, $j \in [1, n_i]$). For network scalability, we specify that the number of timeslots for intracluster convergecast is $\max_{i\in[1,m], j\in[1,n_i]} f_{ij}$.

The lower bound on the number of timeslots used for intercluster convergecast scheduling is decided by both the largest subtree and the number of packets in network. For ADS, the number of packets in routing devices are equal to the number of timeslots needed for intracluster convergecast. That is to say, the number of packets in routing device R_{ij} is f_{ij} ($j \in [1, n_i]$, $i \in [1, m]$).

Theorem 9. *If all the data update rates of field devices in a cluster are same, the lower bound on the number of timeslots required to complete the intercluster convergecast in a cluster-tree network and N routing devices is $L_s = \max\{2\sum_{j=2}^{n_1} f_{1i} + f_{11}, \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}\}$, where f_{ij} is the number of field devices in the cluster with R_{ij} being cluster head ($i \in [1, m]$, $j \in [1, n_i]$).*

Proof. For R_{i1} ($i \in [1, m]$), all packets in subtree T_i must go through R_{i1} to GW. Hence, R_{i1} needs at least $2\sum_{j=2}^{n_i} f_{ij} + f_{i1}$ timeslots to forward all packets in T_i. Since the largest subtree is T_1, at least $L_s = 2\sum_{j=2}^{n_i} f_{1j} + f_{11}$ timeslots are needed to complete the convergecast. Meanwhile, since the GW can receive only one packet per timeslot, at least $\sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}$ timeslots are needed to complete the intercluster convergecast in the network having $\sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}$ packets. In all, the lower bound on the number of timeslots needed to complete the intercluster convergecast is $L_s = \max\{2\sum_{j=2}^{n_1} f_{1j} + f_{11}, \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}\}$. □

Theorem 9 shows that the structure of the routing tree plays a fundamental role in minimizing the convergecast time and dominates the minimum network latency. If $2\sum_{j=2}^{n_1} f_{1i} + f_{11} \leq \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}$ (i.e., $\sum_{j=2}^{n_1} f_{1j} \leq \sum_{i=2}^{m}\sum_{j=1}^{n_i} f_{ij}$), then the lower bound on the number of timeslots to complete the intercluster convergecast is $L_s = \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}$, which is equal to the number of packets in the whole network. Otherwise, the lower bound on the number of timeslots to complete the intercluster convergecast is $L_s = 2\sum_{j=2}^{n_1} f_{1i} + f_{11}$, which is equal to the number of packets in the largest subtree T_1. Finding a minimum spanning tree subject to cardinality constraints on the number of routing devices in any subtree, the so called capacitated minimum spanning tree problem, is known to be NP-hard [8], and the effective heuristics algorithm should be carefully designed.

Theorem 10. *Given any scheduling algorithm S that can complete intercluster convergecast in $L_s = \max\{2\sum_{j=2}^{n_1} f_{1j} + f_{11}, \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}\}$ timeslots in a cluster-tree network, the lower bound on the number of channels used in S is*

$$\left\lceil \left(L_s + \frac{1}{2}\right) - \sqrt{\left(L_s + \frac{1}{2}\right)^2 - 2\sum_{d=1}^{D}(d \times n(d))} \right\rceil, \qquad (22)$$

where $n(d)$ denotes the total number of packets in routing devices at level d.

Proof. Let C_s be the number of channels that can be used in a network, and $\mathrm{PT}_{\max}(t)$ denotes the number of parallel transmissions scheduled in timeslot t ($t \in [1, L_s]$).

To complete convergecast in L_s timeslots, the following conditions should be satisfied [8].

(i) For $t = L_s$, there should be only one packet in the network and R_{11} is required to send it to GW. That is, only one device could be scheduled to be transmitted if $t = L_s$.

(ii) For $t = L_s - 1$, two devices could be scheduled to be transmitted in the same timeslot.

(iii) Deduced as so forth, the maximum number of devices allowed to transmit in timeslot t is $L_s - t + 1$.

Since $\lceil (L_s - t + 1)/2 \rceil \geq C_s$ when $t \in [1, L_s - 2(C_s - 1)]$ and the upper bound on the number of channels is C_s, we can obtain the following conclusion [8]:

$$\mathrm{PT}\max(t) = \begin{cases} C_s, & t \in [1, L_s - C_s] \\ L_s - t + 1, & t \in [L_s - C_s + 1, L_s]. \end{cases} \tag{23}$$

By (23), the maximum number of transmissions that can be scheduled in L_s using C_s channels is

$$\begin{aligned} &\sum_{t=1}^{L_s} \mathrm{PT}\max(t) \\ &\quad = \sum_{t=1}^{L_s - C_s} C_s + \sum_{t=L_s - C_s + 1}^{L_s} L_s - t + 1 \\ &\quad = (L_s - C_s) \times C_s + C_s \times (C_s + 1) - \frac{C_s \times (C_s + 1)}{2} \\ &\quad = -\frac{1}{2} C_{s^2} + C_s \left(L_s + \frac{1}{2} \right). \end{aligned} \tag{24}$$

For the intercluster convergecast, only routing devices and GW along the routing tree transmit packets. Thus the number of transmissions for completing convergecast is at least $\sum_{d=1}^{D} dn(d)$. Combined with (24), the number of transmissions satisfy the following inequality:

$$-\frac{1}{2} C_{s^2} + \left(L_s + \frac{1}{2} \right) C_s \geq \sum_{d=1}^{D} dn(d). \tag{25}$$

The inequality (25) can rewritten as

$$C_s \geq \left\lceil \left(L_s + \frac{1}{2} \right) - \sqrt{\left(L_s + \frac{1}{2} \right)^2 - 2 \sum_{d=1}^{D} (d \times n(d))} \right\rceil, \tag{26}$$

where $L_s = \max\{2 \sum_{j=2}^{n_1} f_{1j} + f_{11}, \sum_{i=1}^{m} \sum_{j=1}^{n_i} f_{ij}\}$. □

Figure 4: Experiment platform.

4.2. Scenario II: ADD. In Scenario II, the DURs of field devices in the cluster with R_{ij} ($i \in [1, m]$, $j \in [1, n_i]$) being cluster head satisfy the following relationship:

$$C_{ijk} \neq C_{ijl}, \quad \forall k, l \in \left[1, f_{ij}\right], \ k \neq l. \tag{27}$$

Theorem 11. *Given a cluster with more than one field device, if all the data update rates of field devices in a cluster are different, that is $C_{ijk} \neq C_{ijl}$, for all $k, l \in [1, f_{ij}]$, $k \neq l$ & $i \in [1, m]$, $j \in [1, n_i]$, the lower bound on the number of timeslots used for the intracluster convergecast in a cluster-tree network is 2.*

Proof. Similar to Theorem 3. □

For ADD, the number of packets in a routing device is equal to the number of timeslots for intracluster convergecast. According to Theorem 11, at the beginning of the intercluster convergecast, the number of packets in a routing device is 2.

Corollary 12. *If all the data update rates of field devices in a cluster are different, that is $C_{ijk} \neq C_{ijl}$, for all $k, l \in [1, f_{ij}]$, $k \neq l$ & $i \in [1, m]$, $j \in [1, n_i]$, the lower bound on the number of timeslots used for the intercluster convergecast in a cluster-tree network is $L_s = \max\{4n_1 - 2, \sum_{i=1}^{m} n_i\}$, where n_i is the number of routing devices in subtree T_i ($i \in [1, m]$).*

Proof. Let g_{ij} be the number of packets in R_{ij} during one superframe cycle. Reminded that the number of packets in a routing device is 2 at the beginning of the intercluster communication period, we can conclude that $g_{ij} \geq 2$ ($i \in [1, m]$, $j \in [1, n_i]$). Similar to the proof of Theorem 9, the lower bound on the number of timeslots used for the intercluster convergecast in a network with cluster-tree routing is $L_s = \max\{4n_1 - 2, \sum_{i=1}^{m} n_i\}$. □

Corollary 13. *Given any scheduling algorithm S in a network with cluster-tree routing that can complete intercluster convergecast in $L_s = \max\{4n_1 - 2, \sum_{i=1}^{m} n_i\}$ timeslots, the lower bound on the number of channels used in S is*

$$\left\lceil \left(L_s + \frac{1}{2} \right) - \sqrt{\left(L_s + \frac{1}{2} \right)^2 - 2 \sum_{d=1}^{D} (d \times n(d))} \right\rceil, \tag{28}$$

where $n(d)$ denotes the total number of packets in routing devices at level d.

Figure 5: Network topology and monitoring data.

Proof. The proof of Corollary 13 is similar to Theorem 10. The only difference is that $L_s = \max\{4n_1 - 2, \sum_{i=1}^{m} n_i\}$ according to Corollary 13. □

4.3. Scenario III: PDS. In Scenario III, the DURs of field devices in the cluster with R_{ij} $(i \in [1, m],\ j \in [1, n_i])$ being cluster head satisfy the following relationship:

$$C_{ijk} = C_{ijl}, \quad \exists k, l \in \left[1, f_{ij}\right],\ k \neq l. \tag{29}$$

Corollary 14. *Given a cluster with R_{ij} $(i \in [1, m], j \in [1, n_i])$ being cluster head, the data update rates of field devices in a cluster are partly same, that is $C_{ijk} = C_{ijl}$, $\exists k, l \in [1, f_{ij}]$, $k \neq l$. Supposing that the number of field devices with DUR being C_{ijk} is N_{ijk} and $\sum_{j=1}^{Q_{ij}} N_{ijk} = f_{ij}$ $(Q_{ij} \leq f_{ij})$, the lower bound on the number of timeslots used for the intracluster convergecast in a cluster-tree network is $\sum_{j=1}^{Q_{ij}} y_{ijk}$, where $i \in [1, m]$, $j \in [1, n_i]$, $k \in [1, Q_{ij}]$. Q_{ij} is the number of DUR kinds in the cluster with R_{ij} being cluster head, and y_{ijk} is defined as:*

$$\begin{gathered} y_{ij1} = \left\lceil \frac{N_{ij1} \times C_{ij1}}{C_{ij1}} \right\rceil = \left\lceil N_{ij1} \right\rceil; \\ y_{ij2} = \left\lceil \frac{N_{ij2} \times C_{ij1}}{C_{ij2}} \right\rceil, \\ y_{ijk} = \begin{cases} 0, & \dfrac{C_{ij1}\left(N_{ij(k-1)}C_{ijk} + N_{ijk}C_{ij(k-1)}\right)}{C_{ij(k-1)}C_{ijk}} < y_{ij(k-1)} \\ \left\lceil \dfrac{N_{ijk} - \left(y_{ij(k-1)} \times \left(C_{ij(k-1)}/C_{ij1}\right) - N_{ij(k-1)}\right) \times \left(C_{ijk}/C_{ij(k-1)}\right)}{C_{ijk}/C_{ij1}} \right\rceil, & \text{else,} \end{cases} \end{gathered} \tag{30}$$

where $k = 3, 4, \ldots Q_{ij}$.

Proof. similar to Theorem 6. □

The performance limits on the intercluster convergecast scheduling are determined by the performance limits on the intracluster convergecast scheduling, which have been established in Corollary 14.

Corollary 15. *If the data update rates of field devices in a cluster are partly same, that is $C_{ijk} = C_{ijl}$, $\exists k, l \in [1, Q_{ij}]$, $k \neq l$, the lower bound on the number of timeslots used for the intercluster convergecast in a cluster-tree network is $L_s = \max\{2\sum_{j=2}^{n_1}\sum_{k=1}^{Q_{1j}} y_{1jk} + \sum_{k=1}^{Q_{11}} y_{11k}, \sum_{j=1}^{n_i}\sum_{k=1}^{Q_{ij}} y_{ijk}\sum_{i=1}^{m}\}$, where n_i is the number of routing devices in subtree T_i $(i \in [1, m]$, $j \in [1, n_i]$, $k \in [1, Q_{ij}])$.*

Proof. Let g_{ij} be the number of packets in routing device R_{ij} during one superframe cycle. Reminded that the number of packets in the routing device is $\sum_{j=1}^{Q_{ij}} y_{ijk}$ at the beginning

of the intercluster convergecast, we can conclude that $g_{ij} \geq \sum_{j=1}^{Q_{ij}} y_{ijk}$ $(i \in [1,m],\ j \in [1,n_i],\ k \in [1,Q_{ij}])$. Similar to the proof of Theorem 9, the lower bound on the number of timeslots used for the intercluster convergecast in a network with cluster-tree routing is $L_s = \max\{2\sum_{j=2}^{n_1}\sum_{k=1}^{Q_{1j}} y_{1jk} + \sum_{k=1}^{Q_{11}} y_{11k}, \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{Q_{ij}} y_{ijk}\}$. □

Corollary 16. *Given any scheduling algorithm S that can complete intercluster convergecast in* $L_s = \max\{2\sum_{j=2}^{n_1}\sum_{k=1}^{Q_{1j}} y_{1jk} + \sum_{k=1}^{Q_{11}} y_{11k}, \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{Q_{ij}} y_{ijk}\}$ *timeslots in a cluster-tree network, the lower bound on the number of channels used in S is*

$$\left\lceil \left(L_s + \frac{1}{2}\right) - \sqrt{\left(L_s + \frac{1}{2}\right)^2 - 2\sum_{d=1}^{D}(d \times n(d))} \right\rceil, \qquad (31)$$

where $n(d)$ denotes the total number of packets in routing devices at level d.

Proof. The proof of Corollary 16 is similar to Theorem 10. The only difference is that $L_s = \max\{2\sum_{j=2}^{n_1}\sum_{k=1}^{Q_{1j}} y_{1jk} + \sum_{k=1}^{Q_{11}} y_{11k}, \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{Q_{ij}} y_{ijk}\}$ according to Corollary 16. □

5. Extensive Analysis

The theoretical analysis of the lower bounds on the number of timeslots and the number of channels gives deep insights into the network planning and reliable scheduling algorithm designing. In order to increase the reliability for IWSNs, a retransmission scheme is introduced.

We consider the general cluster-tree network. According to the theoretical analysis in Section 4, the lower bound on the number of timeslots for intracluster convergecast in Scenario I is $\max_{i\in[1,m],j\in[1,n_i]} f_{ij}$ and the lower bound on the number of timeslots for intercluster convergecast is $\max\{2\sum_{j=2}^{n_1} f_{1i} + f_{11}, \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}\}$. If we consider x retransmissions for each packet (default value of *mac Max Frame Retries* in IEEE STD 802.15.4-2006 [10]), the lower bounds on the number of timeslots for intracluster and intercluster convergecast in Scenario I are $x+1$ times of $\max_{i\in[1,m],j\in[1,n_i]} f_{ij}$ and $\max\{2\sum_{j=2}^{n_1} f_{1i} + f_{11}, \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}\}$, respectively. Meanwhile, the number of timeslots during the beacon, CAP and CFP periods are 16 according to the IEEE STD 802.15.4-2006 [11]. In all, the superframe length can be calculated as $(16 + (x+1) \times \max_{i\in[1,m],j\in[1,n_i]} f_{ij} + (x+1) \times \max\{2\sum_{j=2}^{n_1} f_{1i} + f_{11}, \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}\}) \times duration\ of\ timeslot$. For a network having 1000 field devices, that is $\sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij} = 1000$, and the duration of the timeslot being 10 ms, if $x = 3$ and $2\sum_{j=2}^{n_1} f_{1j} + f_{11} < \sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}$, the maximum superframe length is 80.16 s, which is allowable in most industrial process applications. In Scenario II and Scenario III, the superframe lengths are less than 80.16 ms because of smaller number of timeslots.

The analysis of superframe length indicates that the network with hierarchical topology can accommodate more devices and can guarantee better timeliness. For applications required large network and tight timeliness, if we assign field devices with different DURs in same cluster, the requirements can be well satisfied.

Figure 6: Convergecast time by experiment in network with cluster-line topology.

6. Experiment

We validate the tightness from a real implementation on the SIA2420 modules [14]. The SIA 2420 module is completely compliant to the IEEE STD 802.15.4-2006 physical layer and built in multiple specifications, such as WIA-PA, WirelessHART and Zigbee. It is a high reliable, low energy-consumption and long-distance wireless product of original equipment manufacturer (OEM), which is controlled by a 16-bit microcontroller and is installed by a MultiMedia Communication eXchange (MMCX) antenna connector. The transmission radius of SIA 2420 is 25 meters. The user interface is built on the Microsoft Visual C++ 6.0. Figures 4 and 5 exhibits the experiment platform, network topology and user operation interface, respectively.

Based on our experiment system, we investigate the bounds of the number of timeslots and channels with cluster-line and cluster-tree routing topologies, respectively. We use three parameters to represent the network with cluster-tree topology, which are M, D, O. M denotes the number of children of GW; D denotes the depth of the tree; and O denotes the number of field devices in the cluster. Each simulation setting is represented as $\{M, D, O\}$. Different combination of these three parameters will generate different topology. Figure 6 compares the experiment convergecast time with the optimal results for three scenarios in the network with the cluster-line routing topology. The experiment results

Figure 7: Convergecast time by experiment in network with cluster-tree topology.

show that there exist the optimal strategies for convergecast in the network with cluster-line topology, the conclusion of which is same as the proof results. Figure 7 compares the experiment convergecast time with the optimal results for three scenarios in the network with the cluster-tree routing topology. The experiment results of the network with cluster-tree routing topology show that our two-stage scheduling algorithms are near-optimal and the maximum degree of approximation is 99.9%.

7. Conclusion

In this paper, orienting to clustered networks including cluster-line and cluster-tree, we analyze the performance limits of the convergecast scheduling. Our analysis is divided according to the network topology (cluster-line or cluster-tree) and the application scenarios (ADS, ADD, and PDS). We establish the lower bounds on the number of timeslots to finish the intracluster and the intercluster convergecast transmissions. In addition, we establish the lower bounds on the number of channels based on the lower bounds on the number of timeslots and available channels in multichannel scenario. The analysis results are future extended to take packet retransmissions into consideration so as to meet the reliability requirement.

Working with this framework has opened up several possibilities of future research. Work currently in progress is under assumption of reliable link-level transmissions. In the future, an extension to the framework is planned to analyze the performance of the networks in presence of intermittent connectivity and channel errors. The channel-constrained convergecast scheduling when the number of available channels is less than the lower bound required for optimal convergecast is a great challenge for research. As comparison with the convergecast transmission in this paper, another interesting issue to be investigated in the future work is the dissemination communications from the gateway device to other devices in the network.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Contract 61202418 and 61174026, the National High Technology Research and Development Program of China under Contract 2011AA040101, the Special Program for Key Basic Research Founded by MOST under Contract 2010CB334705, and the Foundation of Chinese Academy of Sciences under Contract (KGCX2-EW-104).

References

[1] A. Willig, "Recent and emerging topics in wireless industrial communications: a selection," *IEEE Transactions on Industrial Informatics*, vol. 4, no. 2, pp. 102–122, 2008.

[2] HART Communication Foundation, "Wire- lessHARTTM technical data sheet," May 2009.

[3] Industrial communication network—Field-bus specifications—WIA-PA communication network and communication profile, http://www.iec.ch/.

[4] S. Gandham, Y. Zhang, and Q. Huang, "Distributed minimal time convergecast scheduling in wireless sensor networks," in *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS '06)*, pp. 462–466, Washington, DC, USA, 2006.

[5] H. Choi, J. Wang, and E. A. Hughes, "Scheduling on sensor hybrid network," in *Proceedings of the 14th International Conference Computer Communications and Networks (ICCCN '05)*, pp. 503–508, San Diego, Calif, USA, October 2005.

[6] C. Florens, M. Franceschetti, and R. J. McEliece, "Lower bounds on data collection time in sensory networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1110–1120, 2004.

[7] B. Yu, J. Li, and Y. Li, "Distributed data aggregation scheduling in wireless sensor networks," in *Proceedings of the 28th IEEE Conference on Computer Communications (INFOCOM '09)*, pp. 2159–2167, Orlando, Fla, USA, April 2009.

[8] P. Soldti, *On cross-layer design and resource scheduling in wireless networks [Ph.D. thesis]*, KTH, School of Electrical Engineering, Stockholm, Sweden, 2009.

[9] P. Soldati, H. Zhang, and M. Johansson, "Deadline-constrained transmission scheduling and data evacuation in wirelesshart networks," in *Proceedings of the European Control Conference (ECC '09)*, Budapest, Hungary, September 2009.

[10] W. Shang, P. Wan, and X. Hu, "Approximation algorithm for minimal convergecast time problem in wireless sensor networks," *Wireless Networks*, vol. 16, no. 5, pp. 1345–1353, 2010.

[11] H. Zhang, A. Arora, Y. R. Choi, and M. G. Gouda, "Reliable bursty convergecast in wireless sensor networks," *Computer Communications*, vol. 30, no. 13, pp. 2560–2576, 2007.

[12] X. Liu, H. Zhang, Q. Xiang, X. Che, and X. Ju, "Taming uncertainties in real-time routing for wireless networked sensing and control," in *Proceedings of the 13th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '12)*, Hilton Head, SC, USA, 2012.

[13] IEEE SA Standards Board, "Part 15.4: wireless medium access control (MAC) and physical layer (PHY) specification for low-rate wireless personal area networks (LR- WPANs)," Nmuber 802.15.4, 2006.

[14] SIA 2420, http://www.industrialwireless.cn/03.asp?pd=dl&nc-las-sid=628&anclassid=14.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 230198, 13 pages
doi:10.1155/2012/230198

Research Article

Data Gathering in Opportunistic Wireless Sensor Networks

Yongxuan Lai[1] and Ziyu Lin[2]

[1] *Department of Software Engineering, Xiamen University, Xiamen 361005, Fujian, China*
[2] *Department of Computer Science, Xiamen University, Xiamen 361005, Fujian, China*

Correspondence should be addressed to Ziyu Lin, ziyulin@xmu.edu.cn

Received 23 July 2012; Revised 24 September 2012; Accepted 2 October 2012

Academic Editor: Ruchuan Wang

Copyright © 2012 Y. Lai and Z. Lin. 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 wireless sensor networks and opportunistic networks have nowadays presented a trend of technology convergence. On one hand, the nodes periodically sense the environment and continuously generate sensing data; on the other hand, the movements and sparse deployment of nodes usually lead to intermitted connected links and create some form of opportunistic communications. So it is a challenging problem to effectively collect the sensing data in opportunistic wireless sensor networks. In this paper, we propose an efficient data gathering algorithm based on location prediction in opportunistic wireless sensor networks. The algorithm first collects the network metadata such as history of node encounters and contact durations; then it creates a node contact graph, based on which predictive optimal data gathering locations are dynamically calculated and updated. Finally, the sink is controlled to move to these locations to collect sensing data, avoiding lots of unnecessary data exchanges and message transmissions. Extensive experimental results show that the proposed algorithm is effective to reduce the message transmissions and improve the data collection coverage rate.

1. Introduction

The wireless sensor networks (WSNs) and the opportunistic networks (or mobile delay tolerant networks) have nowadays presented a trend of convergence [1, 2]. On one hand, the sensor network has emerged some "opportunistic" characteristics in its nature: (1) often deployed in harsh environments, the signal is susceptible to external interference and leads to loss of messages; (2) the node is resource-constrained; many applications would take initiative to turn off the wireless radio devices based on energy considerations, resulting a disconnected network; (3) in mobile sensor networks, the movement of nodes also leads to opportunistic communications; (4) in sparse sensor networks, a mobile sink node is used for message and data collection; the communication is also opportunistic [3, 4]. On the other hand, similar to the "data-centric" sensor network, the opportunistic network is also a "data-related" network. End-to-end paths are not available in the opportunistic network, and it usually adopts a "store-carry-forward" mechanism to forward messages [1, 5]. The content of packets being forwarded plays an important role in the routing scheme. So the opportunistic wireless sensor network (OWSN) is a kind of opportunistic network that consists of nodes with sensing capabilities and actively or passively adopts the form of opportunistic transmissions. Research on OWSNs is a result of technology integration between wireless sensor networks and opportunistic networks.

Data gathering is a key application in the "data-centric" sensor networks. For example, in agricultural applications the deployed sensor network would collect the sensed data such as temperature and humidity for analysis [6]; in habit monitoring applications, data such as occurrence, skin color, and locations of the animals are to be collected [7]; recently Ayaki et al. [8] proposed a data gathering scheme in urban streets using mobile phones as the relayed nodes. However, the "opportunistic" characteristic makes the data gathering a challenging problem in OWSNs. Firstly, static communication links (or paths) between the source node and the sink are not available in the opportunistic network. Nodes adopt a "store-carry-forward" mechanism for message transmissions, and most of the existing data gathering algorithms based on a stable link assumption in WSNs cannot work properly. Secondly, most of the research in opportunistic network focuses on the message diffusion rather than particularly on the issue of data gathering. The goal of message

diffusion algorithm is populating the messages to their varied destinations efficiently; yet for the data gathering algorithm, there is usually only one target (the sink node) for the large amount of sensed data that are injected into the network from different sources. So existing routing schemes cannot work properly for the data gathering applications in OWSNs.

In this paper we study the problem of data gathering in the OWSNs and propose an efficient data gathering algorithm called PDA (predictive dynamic data-gathering algorithm). The idea behind PDA is that nodes in OWSNs, for example, mobile phones, do not usually move randomly, but would roughly move according to some relatively stable patterns. If the moving pattern is known to the sink, the sink would then be programmed to move to the optimized locations to collect the sensing data, and lots of message exchanges and transmissions would be avoided to enhance the performance. More concretely, our algorithm firstly collects the network metadata such as the history of node encounters, contact durations, and so forth for modeling. Then it generates a node contact graph, based on which the predictive optimal locations are calculated. Finally, the sink is controlled to move to these locations to collect the sensing data and avoid lots of transmissions. Compared with other data gathering algorithms, PDA has the following features and advantages.

(1) Data transmissions are based on the pattern of node movements. The algorithm collects the metadata for modeling to predict the optimized data gathering points, and it selectively exchanges data among the nodes, which cuts down the transmissions of data gathering.

(2) The algorithm utilizes abundant resources of the sink node. The sink is programmed to travel among several optimized data gathering points to collect the sensing data, which avoids unnecessary data transmissions among ordinary nodes.

(3) The algorithm attaches great importance on the success of "last hop" transmissions to the sink. Nodes with large volume of sensing data from other nodes stored on their storage would have larger probability to encounter with the sink and have enough connection length to route their data to the sink.

Extensive experimental results show that the proposed algorithm can reduce about 40∼60 percent of message transmissions and improve data collection coverage rate about 8∼12 percent compared with other epidemic and probabilistic data gathering algorithms.

The rest of the paper is structured as follows. Section 2 surveys some existing research related to this paper. Section 3 gives some assumptions about the network model and defines the cost model of data gathering. Section 4 describes the detailed mechanism of the proposed scheme PDA, including the initial run, calculation of data gathering points, data collection, and update of data gathering points. Finally, Section 5 describes the experimental setup and performance evaluation, and Section 6 concludes the paper.

2. Related Work

There exists some research on the data gathering applications in traditional MANET (mobile ad hoc network) and sensor networks, yet they assume the nodes are fixed and have stable links for communication [9, 10]. For the opportunistic network, most of the existing research focuses on the forwarding and routing of messages; there is no much work on the research of data gathering [5, 6, 11]. So in this section we survey some related work from areas of data forwarding and data gathering in the opportunistic networks and sensor networks.

2.1. Data Forwarding. Opportunistic networks adopt a "Store-Carry-Forwarding" strategy for message transmission [5, 11]. Messages are temporally stored on the nodes, and if there exists chance of communication at some proper time, the messages are forwarded to other nodes and finally routed to their target nodes. So utilization of the chance of opportunistic channel, constraint of energy and storage space, and the rate of successful message transmissions are all key factors for the message routing and forwarding algorithms [12].

Epidemic routing [13] uses a flooding-like mechanism for message routing. Encountered nodes would fully take advantage of the communication chance for message exchange to increase the rate of successful message transmission and decrease the message delay. To cut down message transmissions and handle the network congestion, controlled-flooding algorithm [14] is proposed, which selectively forwards the message based on the forwarding probability, as well as time-to-live (TTL) or kill time. In the prioritized epidemic routing PREP [15], messages are forwarded or deleted based on the current cost to destination, current cost from source, expiry time, and generation time, and messages closer to the destinations would have more copies to improve the rate of successful transmission.

Besides the flooding-like scheme, another type of message forwarding scheme is to utilize context information and knowledge of the network to optimize the message transmissions. ZebraNet project [7] uses a mechanism based on the history of node movements. Each node maintains a probability to the sink, and the node with higher probability would send its messages to the node with lower probability when two nodes encounter. Similar to [7], PROPHET [16] computes a forwarding probability based on the historical record of its observed contacts, and messages are routed to its neighbor only if it has higher probability than the neighbor. The CAR (context-aware routing) scheme [17] takes the contextual factors such as residual energy, change rate of topology, and moving speed as the input and uses Kalman filter to calculate the probability to the destination node. Messages are sent to the node with the highest probability. Most of the schemes mentioned above assume a random mobility model where the chance for communication is usually by accident and uncertain.

More recently, studies have focused on mobile social networks (special type of DTNs consisting of human-carried devices) and analyzed the social network properties of these networks to assist the design of routing algorithms, where

data forwarding metric is centrality-based. bubble rap [18] uses betweenness as the centrality metric which measures the social importance of a node facilitating the communication among other nodes. In *friendship-based routing* [19], friendships in terms of their behaviors are defined between nodes (i.e., people) to facilitate message forwarding. In addition [20] studies the transient characteristics of node contact patterns. It formulates the transient social contact patterns of mobile nodes as a Gaussian function, based on which it develops data forwarding metrics to analytically predict the contact capability of mobile nodes with better accuracy. In addition [21] proposes a community-based message forwarding scheme CMTS. It organizes the network into different communities based on the contact frequencies among them. The number of message copies is determined by the communities, and messages are transmitted to target communities mainly relying on active nodes that have larger social degrees.

Yet schemes mentioned above are general message transmission schemes, which are not optimized for the data gathering application. The algorithm proposed in this paper would learn the moving pattern of nodes and actuate the sink node properly to improve the overall performance.

2.2. Data Gathering. Nodes in sensor networks continuously sense the environment and generate large volume of data. Traditional algorithms such as TAG [10], PULL/PUSH [9] all depend on infrastructures such as query trees or clustering to collect data. Yet in OWSNs these infrastructures are expensive to maintain as the full point-to-point links are usually not available, so *infrastructure-free* strategies are adopted in the data gathering algorithms.

Reference [6] introduces a real deployment of opportunistic network in agricultural applications which uses the PBR (potential-based routing) algorithm for data gathering. By setting higher potential to remote sensors and lower potential to data server, sensor data is autonomously gathered at lower potential nodes. In data mules system [22], the mule nodes collect the sensing data and route them to the access point through one or multiple hops of transmissions. In the *message ferrying* mechanism [23], ferry node loads data from the source nodes and then forwards them to the middle nodes while moving along its path. The path could be predefined or resolved from the requests of the ordinary nodes. Reference [24] uses a predictable mobility model to collect sensing data, where nodes would learn the connection time with the mobile nodes (e.g., buses) and wake up at that time to facilitate data communication. Our algorithm also learns the mobility model to improve the performance, but we assume all nodes are mobile nodes and besides the connection time, our scheme introduces other factors such as location, contact duration, and so forth for optimization.

Reference [4] uses a mobile sink node to collect static sensor readings by controlling the movement of the mobile sink. It models the data collection problem as sensor selection problem and analyzes the design of a feasible movement route for mobile sink to collect data and minimize the velocity requirements. The schemes use mobile sink node to assist in data gathering, but it assumes the ordinary nodes are static. In OWSNs, all the nodes are mobile and they adopt a style of opportunistic communication, so we need new and energy efficient algorithms that are suitable for the network to collect the sensing data.

3. Network Model and Cost Model

In this section, we define the network and analyze the cost model for data gathering within the network.

3.1. Network Model. We assume the network has the following characteristics (as in Figure 1):

(1) there are one sink node and N sensing nodes, every node has a unique id s_i;

(2) nodes periodically sense the environment and generate a tuple. These tuples are to be gathered at the sink; all copies of the same message have the same message id;

(3) the sensed nodes have constraint on energy and storage space, yet the sink has no limit on these aspects;

(4) nodes have the knowledge of their locations, and they could record their locations at proper time;

(5) the movement displays some kind of patterns in cycles; and the sink is programmed and actuated by applications.

The network uses an opportunistic way for message transmission. When two nodes encounter, they establish a temporary communication link for message exchange. Here we define some notions of network connection.

Definition 1 (chance for communication). Given $L_{i,j}(t)$ is a period of time when node s_i and s_j are within each other's radio range in cycle t, if the length of $L_{i,j}(t)$ is larger than a predetermined threshold value ($L_{i,j}(t) \geq \phi$), then there exists a *chance for communication* between the two nodes in cycle t, denoted as $s_i - s_j(\phi, t)$.

For the sake of convenience, we use $L_{i,j}$ to denote the connection length within the current cycle, and we use $s_i - s_j$ to denote there exists a chance for communication between node s_i and s_j in the current cycle.

Definition 2 (path). For node s_i and s_j, if there exists a collection of nodes $\{m_1, m_2, \ldots, m_{k-1}\}$ such that there are communication chances $s_i - m_1, m_1 - m_2, \ldots, m_{k-1} - s_j$, then there exists a path from s_i to s_j, denoted as path_{ij}. In addition $|\text{path}_{ij}| = k$ indicates the length of the path.

3.2. Cost Model of Data Gathering. As the network consists of N sensing nodes and one sink node, the collection of all encounters could be denoted as $\Omega = \{(t_k, s_i, s_j) \mid t_k \in \{1, \ldots, \text{total}\}, s_i, s_j \in \{s_1, \ldots, s_n, \text{sink}\}\}$, where (t_k, s_i, s_j) denotes that node s_i and s_j have a communication chance at time t_k, and $\{1, \ldots, \text{total}\}$ is the time span under observation. Based on Ω, the data exchange collection could be defined as

(a) Probability-based (b) PDA

Figure 1: Illustration of data gathering strategies in OWSN.

$\Phi = \{(t_k, d_{ij}) \mid t_k \in \{1, \ldots, \text{total}\}, (t_k, s_i, s_j) \in \Omega\}$, where (t_k, d_{ij}) denotes that node s_i sends data d_{ij} to s_j at time t_k. According to these two collections, the cost of data gathering for the network is

$$\text{Cost} = \sum \left| d_{ij} \right| * T_{ij}, \quad \left(t, d_{ij}\right) \in \Phi, \tag{1}$$

where $|d_{ij}|$ denotes that the size of packets of data d_{ij}, T_{ij} is the cost of transmitting unit data from node s_i to s_j. The data gathered at the sink is

$$D = \sum_i |d_{is}|, \quad (t, d_{is}) \in \Phi, \tag{2}$$

where s denotes the sink. If the total data s_i generates d_i, then the data coverage of the network is

$$\text{Cove} = \frac{D}{\sum_{i=0}^{n} |d_i|}. \tag{3}$$

Since the sink could be programmed and actuated by applications, the time and place that the sink encounters with other nodes belong to the calculation of the data gathering problem. So the goal of the data gathering algorithm is to calculate the encounter collection Ω and data exchange collection Φ to minimize the total message transmissions while the data coverage Cove is above certain threshold ρ. At the same time, the node s_i and s_j should have enough bandwidth to transmit the exchanged data d_{ij}. Therefore, the formalization of data collection could be defined as

$$\begin{aligned} \text{minimize}: \ \text{Cost s.t.} \quad & \text{Cove} \geq \rho, \\ & d_{ij} \leq \text{BW}\left(s_i, s_j\right)^* L_{i,j}(t_k), \\ & \left(t_k, d_{ij}\right) \in \Phi, \qquad \left(t_k, s_i, s_j\right) \in \Omega, \end{aligned} \tag{4}$$

where $\text{BW}(s_i, s_j)$ is the bandwidth of the link between node s_i and s_j. From (4) we could see that when the network topology is fixed, the optimized data collection could be transformed into the *minimum-cost flow problem* [25] if a virtual source node is added to the network and connects to every sensor node in the network (see The Appendix). But in opportunistic networks, no centralized node is available for computing, and the network topology changes dynamically due to the moving nodes. When a node moves, the current optimal solution for data gathering becomes invalid. So the resolution of (4) is meaningful only in cases when nodes move according to some kind of patterns, for example, office workers traveling among home, bus stations, and offices at roughly the same time every day.

4. Data Gathering Algorithm Based on Location Prediction

4.1. Algorithm Overview. This section describes details of the proposed data algorithm PDA (predictive dynamic data-gathering algorithm) in OWSNs. As an overview, Figure 1 compares the PDA strategy with a probability-based data gathering scheme in OWSNs. The circles denote the nodes and the dashed lines denote the paths along which the nodes move, and Table 1 presents the time, nodes, and duration of the connections. In Figure 1(a), the sink s encounters node d, and d encounters nodes b and c, so the data forward probability is $d > b = c > a$. According to the probability-based data forwarding algorithms (e.g., PROPHET [16]), at T1 node a would forward its data to b; at T2 nodes b and c would exchange their data; at T3 and T4 nodes c and d would route their data to d; at T5 nodes a, b, and c would exchange their data; finally at T6, after gathering data from other nodes, node d encounters the sink, yet it has only 0.5 s to route all the data to the sink. Part of the data cannot be gathered successfully by the sink, causing data loss in the application. Instead, if the nodes move according to some relatively stable patterns, after some necessary data exchanges, the sink would learn the pattern of node movements and move to the data gathering point to collect the data at T5, which cuts down several hops of transmissions and has enough time to collect the sensing data (Figure 1(b)). Moreover, as nodes a, b, and c have stable connection with the sink at T5 (10 s), PDA suppresses data transmissions between nodes a and b at T1 and between b and c at T2, which further cuts down the data transmissions.

There are 4 phases in the PDA algorithm: (1) the sink collects metadata such as historical node movements, the contact durations, and so forth for modeling; (2) the sink builds an encounter graph to dynamically compute the optimal locations of data gathering points; (3) the sink is controlled to move to data gathering points to collect the sensing data, saving lots of transmissions and improving the data collection coverage; (4) gathering points are updated if necessary.

4.2. Initial Run. At the very beginning, PDA adopts the epidemic routing strategy [13] for message transmission and data gathering. The sink moves according to the way-point mobility model for information exchange in the sensing field. Each node, for example, s_i, would send its sensing data and moving history to the sink when it encounters the sink. The data of moving history is represented and stored in *MTable* (*movement table*), which includes the timestamp, location,

Table 1: Encountered nodes and their contact duration.

Time	Nodes and flow of data	Duration (s)
T1	$a \to b$	1
T2	$b \to c$	3
T3	$c \to d$	4
T4	$b \to d$	3
T5	$a \to b$, $a \to c$	10
T6	$d \to s$	0.5

node id, contact duration, and so forth. Table 2(a) shows an example of *MTable*.

The length of initial run is 1 sliding window. Within the time span *MTables* are disseminated from the ordinary nodes to the sink and also among the ordinary nodes. To avoid the over-flooding problem, PDA uses filtered and compressed version of *MTables*. For example, Table 2(b) shows a compressed version of the *MTable* in Table 2(a), where the compressed *MTable* only includes tuples whose contact duration is greater than 1.0s. Controlled-flooding strategies are also used to cut down transmissions in this phase, which would be further discussed at Section 4.4.

4.3. Calculation of Data Gathering Points

4.3.1. Constructing the Contact Graph. The sink would merge *MTables* from ordinary nodes and use them as the input for constructing the contact graph. Firstly, for every node s_i in the merged *MTable*, PDA calculates its encounter probability M_{ij} and total connection length TL_{ij} with node s_j:

$$M_{ij} = \sum_{K=1}^{T-1} f(k),$$
$$\mathrm{TL}_{ij} = \frac{1}{T-1} \sum_{K=1}^{T-1} L(i,j)(k), \quad T > 2, \tag{5}$$

where T is the current epoch, and $\{1, 2, \ldots, k, \ldots, T-1\}$ consists of a sliding window with $(T-1)$ epochs. The encounter probability and total connectivity length are all calculated within this window. $f(k)$ is a user-defined weighting function, when s_i encounters s_j at kth epoch and contact each other (denoted as $s_i - s_j$), $f(k) = 1/2^{T-k}$; otherwise when they do not encounter each other, $f(k) = 0$. $L_{i,j}(k)$ denotes that the connection length of node s_i and s_j at kth epoch; when s_i and s_j do not encounter, $L_{i,j}(k) = 0$. As illustrated in the definition of $f(k)$, PDA assumes that later encounter has greater impact on the weight. For example, when $T = 5$, if node s_i encounters s_j at the 2nd, 4th epochs, then the encounter probability $M_{ij} = f(2) + f(4) = 1/2^{5-2} + 1/2^{5-4} = 0.625$.

If encounter probability M_{ij} meets some threshold δ ($M_{ij} \geq \delta$), then an edge that has s_i and s_j as its vertexes is added into the *contact graph* (CG). Figure 2(a) is an example graph that is transformed from Figure 1. PDA assumes that the nodes are densely deployed so that most of the nodes are included into the connected contact graph. The movements of the isolated nodes not in the graph usually lack patterns, so they have little impact on calculating the data gathering points. It is worth mentioning that Figure 2 only shows the encounter probability M_{ij} on the edges, yet other information such as timestamp, contact duration, and locations is also included in the edges.

Table 2: Examples of *MTable*.

(a) *MTable* of node s_1

Time	Location	Node	Duration ($\times 10^{-1}$ s)
5	(30, 30)	s_2	5
12	(30, 30)	s_3	13
24	(50, 40)	s_4	20
26	(50, 40)	s_5	22
52	(70, 80)	s_2	10
55	(80, 80)	s_6	2
62	(80, 90)	s_7	6

(b) Compressed *MTable* of node s_1, threshold ≥ 10

Time	Location	Node	Duration ($\times 10^{-1}$ s)
12	(30, 30)	s_3	13
24	(50, 40)	s_4	20
26	(50, 40)	s_5	20
52	(70, 80)	s_2	10

Figure 2: Contact graph G and possible positions of the sink.

4.3.2. Calculation of Data Gathering Points. The contact graph is actually a sketched description of the moving pattern of nodes, and it is used to calculate the data gathering points. Then the sink would move to these predefined gathering points to collect the sensing data. Figures 2(b) and 2(c) illustrate the possible positions of the sink when gathering the data.

We denote the position of the sink as s and denote the encounter probability of node s_i and sink as

$$M_{is} = \min\left(1, \frac{L_i}{\mathrm{Th}}\right), \tag{6}$$

where L_i is the connection length between node s_i and sink; Th is the minimal time interval for a connection. Also we denote the contact graph as G' when s is added to G, and we can construct a spanning tree routed at s from G'. Data is then transmitted from the internal nodes to the root of the tree. Suppose the cost of transmitting unit data between any two nodes is $T_{ij} = 1$, then the data gathering cost Cost and expected amount of gathered data D defined in (1) and (2) in Section 3.2 could be transformed to

$$\begin{aligned} \mathrm{Cost}(s,G) &= \sum_a |\mathrm{path}(a,s)|, \quad a \in \mathrm{Tree}, \\ D(s,G)_e &= \sum_a |d_a| * \Pi M_{ij}, \quad i-j \in \mathrm{path}_{as},\ a \in \mathrm{Tree}, \end{aligned} \tag{7}$$

where path_{as} is the path from node a to the sink $s$in the spanning tree $|\mathrm{path}_{as}|$ is the hops $i-j$ is the edge within the path; d_a is the data generated by a; ΠM_{ij} $(i-j \in \mathrm{path}_{as})$ is the probability of data gathered from node a to the sink. The optimal position s^* for sink in the graph G' meets the following conditions:

$$\mathrm{Cost}(s^*,G) = \mathrm{Minimal}(\mathrm{Cost}(s,G)), \tag{8}$$

$$D(s^*,G)_e \geq \rho^* \sum_a |d_a|, \tag{9}$$

where ρ is the predefined threshold of the data coverage. When nodes move randomly, it is undesired to calculate the optimal data gathering positions; when nodes move according to some kind of pattern, then the edges in the contact graph represents larger chance of encounters and longer connections among the nodes. So the optimal data gathering points could be calculated according to (9).

A direct method for calculating s^* in (9) consists of two steps: (1) construct an extended contact graph G' for each stable edge by adding the sink linking to the vertexes into the graph G; (2) for each G', generate the spanning tree who has the minimal transmission cost and also satisfies the minimal flow constraints. For example, Figures 3(a) and 3(b) are two generated spanning trees that have the minimal $\mathrm{Cost}(s,G)$. Yet combined with Figure 1(b), node a and b have larger connection length with the sink, so Figure 3(b) is a preferred position for the scheme. Also if more than 2 nodes are connected during the same time period, then the sink could be added to the graph to calculate the spanning tree, which is illustrated in Figure 3(c).

When the data gathering point is decided, suppose the set of vertexes connected with the sink is P, then the nodes represented by P could all connect with sink during some time period. The optimal location of the sink is defined as the geometrical center, which could be calculated from the information such as connection timestamp and location contained at edges of the contact graph.

$$s \cdot \mathrm{Interval} = \cap L_{i,j}(t), \quad \forall s_i, s_j \in P, \tag{10}$$

$$s \cdot x = \frac{1}{|P|} \sum a \cdot x, \qquad s \cdot y = \frac{1}{|P|} \sum a \cdot y, \quad a \in P, \tag{11}$$

Figure 3: Possible spanning trees when the sink is added to the graph.

where $(a \cdot x, a \cdot y)$ is the location of node a; $|P|$ is the number of nodes P contains. Note that the optimal position of the sink s^* could be a set of positions that the sink could travel along to enhance the efficiency of data gathering, and we will discuss it later in Section 5.3.

4.4. Data Gathering. When the gathering position is decided, a set of nodes P_u could also be defined, and PDA calculates the *data forwarding probability* (DFP) for the nodes in the set. If $u = 0$, P_0 is the sink; if $u = i\,(i \geq 1)$, P_i is the sink and its i-hop neighbor in the contact graph. The DFP of the sink is denoted as $F_s = 1.0$; if node a links to node s directly or indirectly, then the DFP of node a is

$$F_a = \Pi M_{ij}, \quad i-j \in \mathrm{path}_{as}, \tag{12}$$

where edge $i-j$ belongs to the path from a to sink. The data forward probability of the nodes within P_u is then flooded to the network, and nodes that do not belong to P_u could deduce their probability according to the *MTable's* when they receive the DFPs from other nodes. The DFP of a node $k \notin P_u$ is initially set to be 0; if node k ever contacts some node $a \in P_u$, then the DFP of node k is

$$F_k = \max(F_a * M_{ak}), \quad a \in P_u,\ k \notin P_u,\ s_a - s_k, \tag{13}$$

where M_{ak} is the encounter probability between node k and a. When two nodes move close enough to establish a connection, if their difference on DFP is larger than some predefined threshold, then the data is transmitted from the node with smaller DFP to the node with larger DFP. As shown in Figure 3(c), P_1 is the set of nodes $\{s,a,b,c\}$, and their DFP is 1.0; so in Figure 1, node a, b and node b, c suppress their data transmissions at T1 and T2 as they have the same DFP; then at T5, the nodes move to the radio range of the sink, so they send their data to the sink and finish the data gathering at this epoch.

P_u is diffused within the network, and its size could be adjusted according to the applications. When two nodes encounter, they first exchange P_u and then calculate their DFPs to decide whether a further data transmission is needed. According to (13) and (14), the optimal data gathering point s^* guarantees that the nodes within P_1 have larger DFP. These nodes could be viewed as agents of the sink, and they expect to send their data to the sink through 1 hop of transmissions.

4.5. Update of Data Gathering Points. Although we assume the pattern of node movements are relatively stable, as time

goes by the gathering points may be obsolete. In more detail, there are two cases when PDA needs to adjust the data gathering points.

(1) The sink encounters most of the nodes in P_u, but the data coverage is relatively small. In this case, the sink would recalculate the data gathering point according to the algorithm described in Section 4.3 as nodes also record their *MTable's* and route them to the sink periodically and hence the sink has all the metadata for the calculation.

(2) The sink only encounters a small part of the nodes in P_u, and the data coverage is small. For this case, the sink would initiate an *initial run* as described in Section 4.1 to collect the metadata and sensing data using the epidemic routing strategy to update the data gathering points.

(3) It is worth mentioning that the *MTable's* are not affected by the DFP's of the nodes; they are flooded within the network. To cut down the cost of metadata transmissions, a TTL (time-to-live) segment is inserted into the message. When a piece of metadata is copied to a node, TTL increases by 1; the copy is stopped when TTL increases up to the predefined threshold. Moreover, a probability is used for the metadata exchange. When node s_i meets s_j, the nodes would exchange the metadata with probability $\gamma < 1$, which disperses the data exchange within the time dimension. So according to the strategies of controlled flooding, sink could collect enough metadata and avoid the problem of overflooding.

5. Experimental Study

We implement PDA in C# and compare it with other data gathering schemes. The experimental result shows that PDA reduces about 40∼60 percent of message transmissions and improves data collection coverage rate about 8∼12 percent, compared with other epidemic and probabilistic data gathering algorithms.

5.1. Environment Setup. To simulate the movement pattern of nodes, we adopt the *community model* described in [16]. In the model, the network is divided into $K * K$ grids (also called community, illustrated as Figure 4). Each community has an *interest index* $c_i \in (0, 1)$. If the index is greater than a threshold, the community is called the *hot community*, and the set of hot communities in the network is denoted as $\mathbb{C}$. Similar to [8] nodes move along paths, yet in different ways: (1) node s_i could move along *Max_Path* paths at most; (2) if p_i is a path of node s_i, then p_i consists of *Max_Com* communities at most, and the start and the end points are the *Home Community* of the node; (3) along the path there are at least $m$phot communities belonging to $\mathbb{C}_i$, where $\mathbb{C}_i \subseteq \mathbb{C}$ is the set of hot communities of s_i. When a node moves along the path and travels on community C_i, it would stop for a period with probability of $P_{\text{stop}} = ps + (1 - ps) * c_i$ and would choose to move to the next community along the path with probability of $1 - P_{\text{stop}}$, where $ps$is the predefined value; c_i is the interest index of the target community. $(1 - ps) * c_i$ represents the probability that a node would stop because of the distraction within the community that it currently visits. The simulation runs periodically, a node has several predefined paths, and the path is randomly chosen for the node. Meanwhile, nodes may not travel along their path with probability η, called the disturbance factor. The simulation consists of 10 rounds, and the length of each round is 100 s. Each node generates 0.2 packets per second, where the size of the packet is 512 B. So the size of sensing data that each node generates and is to be collected by the sink is 25 k (100 s $*$ 0.2 Packet/s $*$ 512 B/Packet) each round. Also, we assume that all *MTable's* could be wrapped into 1 packet.

Figure 4: Snapshot of the simulation field ($N = 50$).

Three other data gathering algorithms are compared with our algorithm: (1) Random: the sink moves according to the way-point mobility model and collects data from the nodes it encounters. There is no data exchange between ordinary nodes; (2) Epidemic [13]: the sink and ordinary nodes take advantage of all chances of communications and data are exchanged among any nodes if possible and finally routed to the sink; (3) PROPHET [16]: data are exchanged according to the data forwarding probability which is maintained by the nodes using the movement history. As the cost of communication dominates the depletion of the limited battery energy in sensor nodes, we present only the total communication cost (number of packets) incurred by various algorithms. We assume ideal links when two nodes meet and establish a connection. Table 3 lists the default parameters in the simulations.

5.2. Performance Comparison and Analysis

5.2.1. Impact of Network Size and Radio Range. From Figures 5 and 6 we would see that the size of the network has great impact on the message transmissions, yet has little impact on the data coverage. Random has the smallest number of

Table 3: Parameters of the simulation.

Parameter	Value	Description
N	50	Number of nodes
W	120 m	Width and length of the field (120 m $*$ 120 m)
K	8	Divided into $K*K$ communities
Sim_T	2000 s	Length of simulation
E	100 s	Length of each round
O_T	5	Number of rounds in the sliding window
Th	1 s	Minimal length of time for a valid connection
R	6 m	Radio range for a node
η	0.1	Disturbance factor when a node travels along its path
st	10 s	Length when a node stops
Min, max	6, 10 m/s	Min/max speed of nodes
$\mathbb{C}$	9, 15, 20, 17, 5, 48, 56	Id set of hot communities
mp	1	Minimal number of hot communities in the paths
Max_Path	2	Maximal number of paths a node has
Max_Com	10	Maximal number of communities a path may travel
Max_Copy	5	Maximal number of versions a packet may have in Epidemic routing
Max_Buffer	500	Size of cache buffer (packet)
δ	0.6	Threshold of encounter probability when constructing the graph

Figure 5: Network size versus messages.

Figure 6: Network size versus data coverage.

transmissions (about 5% of that in Epidemic); Epidemic and PROPHET have the largest number of transmissions, and the performance of PDA is in the middle, which cuts down about 38% of transmissions. But from the view of data coverage, Random has a coverage of 80%, and PDA is about 5~10% higher than Epidemic (about 89%) and PROPHET (about 83%). This lies on two aspects: (1) PDA takes advantage of the pattern of node movements, and controls the sink to collect the data at optimal data gathering points; (2) PDA requires less storage space than other algorithms (discussed in Section 5.2.2). When $N = 50$, the sink in Random moves $2.8*10^3$ m due to its random movement; yet in PDA, the path is only about $0.4*10^3$ m long because the sink only moves among the data gathering points, which partly cuts down the energy consumption.

Figures 7 and 8 shows the impact of radio range. From the figure, we could see that the data transmissions and coverage grow with the radio range. Random has the best performance on transmissions, and PDA has the best performance on data coverage. Yet when radio range grows up to 16 m, the data coverage and transmissions increase to their maximal. Note that when radio range is large (e.g., 24 m), the number of message transmissions in Random is about $1.8*10^4$ and the data coverage is about 93%, which is a good performance for data gathering. Yet considering the increase of radio range is constrained by the hardware and would greatly increase the cost of per-unit data transmissions, to improve the performance by increasing the radio range is actually impractical in real deployments.

5.2.2. Impact of Buffer Size. As the algorithms adopt a "Store-Carry-Forward" scheme for message transmissions, the size of buffer would have a great impact on the performance. In Figures 9 and 10, every node would generate 0.2 packets per second, so each node would have 400 (2000 $*$ 0.2) packets as its own sensing data. As the figures show, the number of transmissions and data coverage are small when the buffer size is small. Random requires the least buffer as each node only stores its own sensing data in the buffer. Compared with Random, Epidemic and PROPHET have poorer performance when the buffer size is less than 150 packets. This is because

Figure 7: Radio range versus messages.

Figure 8: Radio range versus data coverage.

Figure 9: Buffer size versus messages.

Figure 10: Buffer size versus data coverage.

Epidemic is "cache-hungry", as it would exchange messages whenever nodes encounter other nodes. So many packets would be stored and cached in the buffer. When the buffer is full, a cache replacement is needed for the algorithm. Here we use a FIFO (first-in-first-out) strategy to replace the old packets for all the algorithms. When the buffer size is larger, more packets are stored and exchanged between the nodes, and the data transmissions and data coverage also increase. Yet PDA uses a selective strategy for the data exchange, so it requires relatively less buffer compared with Epidemic, which in return improves its performance.

5.2.3. Impact of Moving Speed. The moving speed of nodes reflects the activeness of the network. When nodes move faster, more nodes would meet and make data exchanges. Figures 11 and 12 show that the data transmissions and data coverage increase with the average speed of nodes. When nodes move slowly (e.g., 2 m/s), all algorithms incur small transmissions and small data coverage as there are fewer opportunistic connections. Compared with Epidemic and PROPHET, PDA cuts down about 62% and 47% of message transmissions and increases about 5∼8% of the data coverage.

5.2.4. Comparison of Delay. Figure 13 shows that besides Random, the data coverage increases quickly with time and goes to a relatively stable point when the time is greater than 600 s. This is because all the algorithms are based on the epidemic copy strategy, and for PDA and PROPHET there is an initial phase when messages are copied among the nodes, which cuts down the delay for data gathering. But in Random, data would not be exchanged between ordinary nodes, so it has larger delay on the data gathering process.

Figure 11: Average speed versus messages.

Figure 12: Average speed versus data coverage.

Figure 13: Change of data coverage with time.

Figure 14: Impact of optimal position set.

Random has a delay of about 500 s for its data coverage to climb up to 60% compared with other algorithms.

5.3. Impact of Other Factors

5.3.1. Impact of Optimal Position Set. The optimal position s^* could be a point or a set of points. If s^* is a set, then the sink would move among the positions within s^* to collect the sensing data. s^* could be computed as the procedure described in Section 4.3.2 except that there are m ($m = |s^*| > 1$) edges inserted into the graph G to compute the spanning tree with minimal Cost(s, G). Then the sink would visit each positions in s^* accordingly, which is a *feasible route design problem* [4]. As illustrated in Figure 14, when s^* increases, the transmissions goes down, and the data coverage increases when the size of s^* is less than 4 and then decreases when the size grows larger. This is because the sink moves among the positions in s^* for data gathering, and fewer hops are needed to route data from the source to the sink. When the size of s^* is as large as 50, most of the nodes assume that they could communicate with the sink. So they have similar data forward probability and hence suppress the data transmissions among the nodes. Yet the sink could not appear in all the positions at the arranged time, and not every node could communicate with the sink, so the data coverage goes down accordingly. When the size of s^* is 3∼4, PDA has a performance with fewer transmissions and higher data coverage (about 94%).

5.3.2. Impact of Metadata. PDA collects metadata based on *MTables* within every observed period (or sliding window). The default size of window is 5 epochs, and each epoch is 100 s. As illustrated in Figure 15, the cost of metadata collection takes only a little portion of the whole transmissions for data gathering. When node connection threshold Th is 0.5 s, PDA has the largest cost of metadata collection, and the transmissions are about 7% of the whole transmissions;

Figure 15: Impact of node connection threshold.

Figure 16: Impact of hot communities.

when Th > 1.0 s, the cost of metadata collection is less than 5%. This is because PDA adopts a controlled-flooding strategy with the time-to-live TTL = 5 and the metadata exchange probability $\gamma = 0.5$. Meanwhile, when node connection threshold increases, the size of *MTable* goes down; this cuts down the transmissions for metadata. Yet larger Th would also affect the precision of metadata and hence affect the prediction of data gathering points. As illustrated in Figure 15, 1.0∼1.5 s is a suitable range for the connection threshold, which decreases the cost of metadata collection and does not harm the overall data coverage.

5.3.3. Impact of Hot Communities. Figure 16 illustrates the impact of hot communities. As the number of hot communities increases, the data transmissions are relatively the same, yet the data coverage goes down from 94% to 78%. In PDA, hot communities of each node are randomly selected from the hot communities of the network, and each path of the node would contain at least one hot community of the node's hot communities. When there are fewer hot communities in the network, nodes have fewer overlapped hot communities, and the paths of nodes intersect with each other with higher probability, and there are more chances for nodes to meet each other and exchange their sensing data when traveling along their paths. So the pattern of node movements is formed here, and data is routed to the sink at predefined data gathering points and finally increases the data coverage. This also reflects that PDA is more efficient and suitable for cases when nodes move according to the patterns.

In order to study the unbalance of message transmissions in the network, we also present the maximal, minimal, and average number of message transmissions in PDA. As illustrated in Figure 17, the average number of sent messages is about 8000 for each node, and the minimal number is around 3000 messages. However, the maximal number of transmitted messages is about 30,000 when there are only 5 hot communities, and it goes down when there are more hot communities. When there are fewer hot communities

Figure 17: Unbalance of message transmissions.

(e.g., 5), the nodes are more likely to encounter each other at the communities. These nodes are the active nodes and they would exchange data from each other, making the maximal number of message transmissions larger. Here we assume that the nodes have enough energy to route their own or relayed messages. Yet it is possible that some nodes could not route their messages when they are out of energy. The unbalance of energy consumption is a common problem for the sensor network, and we leave this part of discussion as our future work.

6. Conclusion

As the integration of opportunistic networks and wireless sensor networks, data gathering is becoming an important issue in OWSNs. In this paper we have proposed an efficient data gathering algorithm that takes advantage of the pattern

of node movements. It collects the network metadata to create a contact graph, based on which the optimal data gathering positions are calculated and the sink is controlled to move to these positions to collect the sensing data, avoiding lots of unnecessary transmissions. Extensive experimental results show that the proposed algorithm can reduce about 40∼60 percent of message transmissions and improve data collection coverage rate about 8∼12 percent, compared with other epidemic and probabilistic data gathering algorithms.

For the future work, we are going to implement and deploy PDA on real test bed to validate the performance of the algorithm. We are also planing to investigate the impact of cache strategies and data redundancy of nodes to further improve the performance of the data gathering algorithms in OWSNs.

Appendix

For the static sensor network, the network topology is fixed. Each node, for example, a_i, generates some amount of data d_i $(i = 1, 2, \ldots, n)$, which are exchanged among the nodes and finally routed to the sink node s. The data gathering problem hence is to minimize the total cost of sending the data set $D = \{d_1, d_2, \ldots, d_n\}$ to the sink node s under the bandwidth constraint. The constraint could be stated as follows: any link in the network should have enough bandwidth to transmit data d_j when the link is connected.

A virtual node vs is then added into the network. vs connects to all the other nodes, and the dotted lines are the virtual edges (as in Figure 18). The sensor network then becomes a flow network, denoted as $G = (V, E)$ with nodes $a_i \in V$, virtual source vs $\in V$, and sink $s \in V$, where edge $(u, v) \in E$ has the capacity $c(u, v) > 0$, flow $f(u, v) \geq 0$, and unit cost $a(u, v) \geq 0$; the cost of sending the flow is $f(u, v) * a(u, v)$. The *data gathering problem* on a static sensor network now becomes a *minimal-cost flow problem (MCF problem)*, which is to minimize the total cost of sending a flow of data $\sum d_i$ from vs to the sink s:

$$C = \sum_{(u,e)\in E} a(u, v) * f(u, v) \tag{A.1}$$

with constraints:

capacity constraints:

$$f(u, v) \leq c(u, v), \tag{A.2}$$

skew symmetry:

$$f(u, v) = -f(v, u), \tag{A.3}$$

flow conservation:

$$\sum_{w\in V} f(u, w) = 0, \quad \forall u \neq s, t, \tag{A.4}$$

required flow:

$$f(\text{vs}, a_i) = d_i, \qquad \sum_{w\in V} f(w, s) = \sum d_i. \tag{A.5}$$

Figure 18: Data are flowed from the virtual source node vs through the sensor network and gathered at the sink node s.

The capacity constraint relates to the bandwidth constraint, and the flows from vs to all other nodes are fixed as $f(\text{vs}, a_i) = d_i$. When the cost of sending data from virtual node vs to all the other nodes a_i is subtracted from the total cost C in *MCF*, the result is the minimal cost of data gathering in the fixed network.

Acknowledgments

This work is supported by the Natural Science Foundation of China (no. 61202012), Natural Science Foundation of Fujian (no. 2011J05156), Fundamental Research Funds for the Central Universities (nos. 2012121030, 2011121049), and Open Project Foundation of the Key Laboratory of Data Engineering and Knowledge Engineering, Ministry of Education (no. KF2011002).

References

[1] L. J. Chen, C. H. Yu, C. L. Tseng, H. H. Chu, and C. F. Chou, "A content-centric framework for effective data dissemination in opportunistic networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 761–772, 2008.

[2] Y. P. Xiong, L. M. Sun, J. W. Niu, and Y. Liu, "Opportunistic networks," *Journal of Software*, vol. 20, no. 1, pp. 124–137, 2009 (Chinese).

[3] X. Wu and G. Chen, "Dual-Sink: Using mobile and static sinks for lifetime improvement in wireless sensor networks," in *Proceedings of the 16th International Conference on Computer Communications and Networks (ICCCN '07)*, pp. 1297–1302, August 2007.

[4] X. Xu, J. Luo, and Q. Zhang, "Delay tolerant event collection in sensor networks with mobile sink," in *Proceedings of the 29th Conference on Information Communications (INFOCOM '10)*, pp. 2471–2479, March 2010.

[5] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 27–34, ACM, August 2003.

[6] H. Ochiai, H. Ishizuka, Y. Kawakami, and H. Esaki, "A dtn-based sensor data gathering for agricultural applications," *IEEE Sensors Journal*, vol. 11, no. 11, pp. 2861–2868, 2011.

[7] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet," *ACM SIGOPS Operating Systems Review*, vol. 36, no. 5, pp. 96–107, 2002.

[8] R. Ayaki, H. Shimada, and K. Sato, "A proposal of sensor data collection system using mobile relay nodes," *Wireless Sensor Network*, vol. 4, no. 1, pp. 1–7, 2012.

[9] X. Liu, Q. Huang, and Y. Zhang, "Combs, needles, haystacks: balancing push and pull for discovery in large-scale sensor networks," in *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, pp. 122–133, ACM, November 2004.

[10] S. Madden, M. Franklin, J. Hellerstein, and W. Hong, "TAG: a tiny aggregation service for Ad-Hoc sensor networks," in *Proceedings of the ACM Symposium on Operating System Design and Implementation (OSDI '02)*, 2002.

[11] S. Burleigh and K. Scott, "Bundle protocol specification," IETF Request for Comments RFC, vol. 5050, 2007.

[12] L. Zhang, X. W. Zhou, J. P. Wang, Y. Deng, and Q. W. Wu, "Routing protocols for delay and disruption tolerant networks," *Journal of Software*, vol. 21, no. 10, pp. 2554–2572, 2010 (Chinese).

[13] A. Vahdat and D. Becker, "Epidemic routing for partially connected Ad hoc networks," Tech. Rep. CS-2000-06, Duke University, 2000.

[14] K. A. Harras, K. C. Almeroth, and E. M. Belding-Royer, "Delay tolerant mobile networks (DTMNs): controlled flooding in sparse mobile networks," in *Proceedings of the 4th International IFIP-TC6 Networking Conference: Networking Technologies, Services, and Protocols, Performance of Computer and Communication Networks, Mobile and Wireless Communications Systems (NETWORKING '05)*, pp. 1180–1192, May 2005.

[15] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, "Prioritized epidemic routing for opportunistic networks," in *Proceedings of the 1st International MobiSys Workshop on Mobile Opportunistic Networking*, pp. 62–66, ACM, June 2007.

[16] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *Lecture Notes in Computer Science*, vol. 3126, pp. 239–254, 2004.

[17] M. Musolesi, S. Hailes, and C. Mascolo, "Adaptive routing for intermittently connected mobile ad hoc networks," in *Proceedings of the 6th International Symposium on a World of Wireless Mobile and Multimedia Networks*, pp. 183–189, IEEE, 2005.

[18] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE rap: social-based forwarding in delay tolerant networks," in *Proceedings of the 9th ACM International Symposium on Mobile ad Hoc Networking and Computing*, pp. 241–250, ACM, May 2008.

[19] E. Bulut and B. Szymanski, "Exploiting friendship relations for efficient routing in mobile social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 12, pp. 2254–2265, 2012.

[20] W. Gao, G. Cao, T. La Porta, and J. Han, "On exploiting transient social contact patterns for data forwarding in delay tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 151–165, 2013.

[21] J. Niu, X. Zhou, Y. Liu, L. Sun, and J. Ma, "A message transmission scheme for community-based opportunistic network," *Journal of Computer Research and Development*, vol. 46, no. 12, pp. 2068–2075, 2009 (Chinese).

[22] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: modeling and analysis of a three-tier architecture for sparse sensor networks," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 215–233, 2003.

[23] W. Zhao, M. Ammar, and E. Zegura, "Controlling the mobility of multiple data transport ferries in a delay-tolerant network," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 1407–1418, IEEE, March 2005.

[24] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Using predictable observer mobility for power efficient design of sensor networks," in *Proceedings of the 2nd International Conference on Information Processing in Sensor Networks*, pp. 129–145, Springer, 2003.

[25] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows: Theory, Algorithms, and Applications*, 1993.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 931869, 15 pages
doi:10.1155/2012/931869

Research Article

An Open Conformance Test System towards the Standardization of Wireless Sensor Networks

Zhonghua Zhao,[1,2,3] Wei Huangfu,[4] Linmin Sun,[1,3] Zhiqiang Shi,[1,3] and Wei Gan[1,2,3]

[1] *Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*
[2] *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*
[3] *State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*
[4] *School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China*

Correspondence should be addressed to Zhonghua Zhao, zhaozhonghua@is.iscas.ac.cn

Received 15 August 2012; Accepted 14 October 2012

Academic Editor: LiuSheng Huang

Copyright © 2012 Zhonghua 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.

The conformance testing on the wireless sensor networks (WSNs) is fundamental for the functional large-scale deployment and interconnection with the global Internet. The WSN protocols based on the Internet Protocol version 6 (IPv6) are the major trends of the future applications, which include IEEE 802.15.4, 6LoWPAN, and RPL. These protocols are not only diverse for various applications, but also volatile in the future. Moreover, sensor nodes are resource constrained, application related and lack the standardized test interface. Hence the corresponding conformance testing is seriously insufficient and needs urgently research. In this paper, the IPv6-based WSN protocols are analyzed and the conformance testing techniques and methods for IPv6-based WSNs are investigated towards the standardization of WSNs. A novel conformance test system for IPv6-based WSNs is designed and implemented, which is open, flexible, full featured, and practical. The conformance test system is suitable for the protocol evolution and the various hardware interfaces for the sensor nodes. The related outcomes will promote the standardization and commercialization of WSNs.

1. Introduction

Wireless sensor networks (WSNs) are the emerging research areas in the current international academia and industry, which can be widely used in many fields, such as intelligent transportation, environmental monitoring, and industrial automation. WSNs have received widespread attention, with a great prospect for economic and social value.

It is fundamental to consider the interconnection between WSNs and Internet for the large-scale application of WSNs. As the core of next-generation Internet protocols, IPv6 has many advantages which include the richness in address resource, the autoconfiguration of address, supporting real-time business, higher security, and better mobility. With the combination of IPv6 and WSNs, IPv6-based WSNs can meet the demands of the current WSNs in the address, security, mobility, and the integration with the existing networks, which will greatly promote the development of WSNs and has become a hot research field of WSNs. Some lightweight IPv6 protocol standards and draft for WSNs have successively been developed in industry, such as IEEE 802.15.4 in MAC (Media Access Control) layer, 6LoWPANs (IPv6 over Low-power Wireless Personal Area Networks) in the adaptation layer, and routing protocol RPL (IPv6 Routing Protocol for Low-power and Lossy Networks) in the network layer. The IETF is accelerating the standardization process of the above-mentioned protocols.

IEEE 802.15.4 is one of most widely used protocol standards in the field of WSNs, which describes the technical specifications suitable for the low-power devices and the short-range wireless communication. The 6LoWPAN group was established by the IETF (the Internet Engineering Task Force) in November 2004, which is dedicated to the IPv6 protocol over IEEE 802.15.4 based networks and has submitted several copies of RFCs (Request for Comments) and the drafts. 6LoWPAN adds an adaptation layer between the IPv6

network layer and the IEEE 802.15.4 MAC layer, which will realize ultimately the seamless connectivity between the IPv6 technology and WSNs. The IETF is also promoting the standardization process of RPL routing protocol on the top of 6LoWPAN adaptation layer.

In the development of WSNs and their protocol stacks, it must be considered whether the protocol implementations in different sensor devices are consistent to the protocol specifications, that is the protocol conformance testing. Conformance testing is verification that an implementation meets the formal requirements of the referenced standards and, more precisely, that it meets the conformance clauses contained in the standards ISO/IEC 9646. In order to ensure the interconnection and interoperability between protocol implementations, the protocol conformance testing must be conducted. Therefore, with the development of WSNs, the key technologies of the protocol conformance testing must be broken to assure the coincidence between protocol implementation and the protocol specifications.

Specifically, the conformance testing on protocols for IPv6-based WSNs is to execute a group of targeted conformance test cases of IPv6-based sensor network protocols (e.g., 6LoWPAN), observe the output behaviors of SUT (System Under Test), analyze the test results, and determine whether the functionality or performance of SUT to meet the specifications of IPv6-based WSN protocols.

At present, there is a serious shortage of research on protocol conformance testing and the equipment development for IPv6-based WSNs. Although some related work exists, the standardized conformance testing specifications on protocols for IPv6-based WSNs have not still been established; the conformance test suites and the engineering conformance test system have not been designed and developed so far. So it is very meaningful and important to carry out the related research and development of the protocol conformance testing in order to promote the commercialization of WSNs. However, there are many difficulties and challenges to face, as follows.

Firstly, at present IEEE 802.15.4 contains five standards, 6LoWPAN contains two RFCs and five drafts, and the RPL contains five RFC and nine working group drafts. The preparation of the comprehensive conformance test specifications and the development of the conformance test suites for IPv6-based WSN protocols are tedious and the workload is very large. Therefore, it is the key issue how to combine the technology of the automation and the manual to reduce labor costs, improve the preparation speed of test specifications, and to choose a reasonable test description language for the efficient development of the conformance test suites for IPv6-based WSN protocols.

Secondly, most of IPv6-based WSN protocols are in the draft stage and the content of the protocols will continue changing and form a new draft version, which requires that the conformance test system is open and has the good adaptability and scalability. It is best that the test system does not depend on the specific protocol test suites, and the ETS (Executable Test Suite) in the test system can be added and updated dynamically.

Thirdly, the application of WSNs is extensive and diverse; there are many node manufacturers which generate node devices with different appearances and communication interfaces. The typical communication interfaces of nodes include the air interface, serial interface, and Ethernet interface. The air interface is used for wireless communication between nodes and is relatively uniform, but the latter two interfaces, which are used for the edge nodes to communicate with the existing wired network, are usually different in appearance. Faced with the difference of the WSN hardware and software, the uncertainty of the evolution of the tested IPv6-based WSN protocol standards, the diverse needs of users, and the test system need designing with the extensible hardware and software architecture to support multiple interfaces.

Fourthly, the practical test system requires the hardware to be compact and appropriate. In the specific conformance testing of the target nodes, combining with different test requirements and the system hardware volume, the test system needs to virtualize different wireless environment, to simulate multiple nodes to virtualize larger-scale networks with different network topologies, and to virtualize various events between nodes to simulate different application scenes.

In this paper, to address the above challenges, the conformance testing techniques and methods are studied, the conformance test suites are stipulated and developed, and the open conformance test system is designed and implemented towards the standardization of IPv6-based WSNs. This paper is organized as follows. Section 2 reviews related work. Section 3 describes the design and generation of protocol conformance test suites for IPv6-based wireless sensor networks. Section 4 describes the design and implementation of conformance test system for IPv6-based wireless sensor networks. Conclusions are finally drawn in Section 5.

2. Related Work

As the basic and important aspect of the protocol testing, the protocol conformance testing is the primary concern of all the protocol developers and there has been a lot of related work in different communication network areas.

In the 1990s, ISO (International Organization for Standardization) specially formulated a set of international standards ISO/IEC 9646, which provide the theoretical framework and test methods for protocol conformance testing, the design steps and description methods for developing the test suites, and the guidance on the implementation of the test system.

In the conformance testing of Internet, the conformance test suite for the RIPng protocol was developed in [1] using TTCN-3. The mobile IPv6 protocols specification, validation, and testing were dealt with using a formal approach [2]. In China, Tsinghua University has caught out the conformance testing of TCP/IP protocol from the 1990s, gotten a number of research results, and developed different versions of the protocol conformance test systems including the PITS [3] and PITSv3 [4].

In the protocol conformance testing of mobile communication network, [5, 6] developed the TTCN-3-based conformance testing tools for OMAMMS (Multimedia Messaging Service) and Push-to-talk service over Cellular in 3G or 4G networks, respectively. [7] studied the conformance testing methodology of MBBMS (Mobile Broadcast Business Management System) and designed and implemented an MBBMS protocol conformance testing tool based on TTCN-3. DTmobile developed the TD-SCDMA Terminal protocol conformance tester ECT6210. Equipped with the latest instrument hardware platform of DTmobile, ECT6210 can provide customers with a complete TD-SCDMA terminal protocol conformance testing solution.

In the wireless sensor networks, many methods are usually used to verify the effectiveness and feasibility of the proposed protocols and algorithms, which include the basic theoretical analysis, the computer simulation methods [8, 9], the small-scale testbeds for actual system test [10, 11], and the sniffer tools for real-time monitoring WSNs [12, 13]. However, the related work in conformance testing on protocols for WSNs (especially IPv6-based WSNs) is rare.

ZigBee is a low-power personal area network protocol based on the IEEE 802.15.4 standard, which is characterized by short-distance, low-complexity, self-organizing, low-power, and low-data rate. In order to ensure the reliability of ZigBee products and the stability of wireless networking, the ZigBee Alliance announced the ZigBee certified program in 2007 and is specifically responsible for the implementation and management of the certification testing. At present, the ZigBee Alliance has sanctioned certification testing by three qualified international test service providers, including TRaC Global, National Technical Systems, Inc., and TUV Rheinland Group. The official ZigBee test specification for the certification of ZigBee products uses the special test suite to test and certify whether ZigBee platform equipments meet the requirements of ZigBee specification. The ZigBee Alliance offers two levels of standard compliance which include ZigBee compliant platforms (ZCPs) and ZigBee certified products. ZCP testing includes application layer, network layer, IEEE 802.15.4 Physical, and MAC layer, as well as security testing required in the test specification. In addition to the above test content, ZigBee Certified Products testing also includes the application profile testing.

RealProct [14] is presented to use a few low-cost real sensor nodes to mimic large-scale real WSNs and then perform protocol conformance testing for WSNs. A part of the test suite is developed by the handwritten method to test the protocol implementation of the μIP TCP/IP stack in Contiki-2.4. However, RealProct mainly emphasizes the software bug detection and there are some cases which cannot be virtualized; for example, it cannot virtualize the wireless environment when interference caused by simultaneous transmissions happens. The fundamental components of conformance testing on protocols for IPv6-based WSNs are simply described in [15, 16]. Conformance testing methodology and system design of the 6LoWPAN protocol is studied in [17], but the standardized conformance testing specification on protocols for IPv6-based WSNs has not still been established; the conformance test suite and the engineering conformance test system have not been designed and developed. The industrial wireless standard ISA100.1la is studied in [18], which may provide references for conformance testing on protocols for IPv6-based WSNs.

3. Design and Generation of Protocol Conformance Test Suites for IPv6-Based WSNs

In the conformance testing on protocols for IPv6-based WSNs, the tested protocols include IEEE 802.15.4 in MAC layer, 6LoWPAN in the adaptation layer, and routing protocol RPL in the network layer, which are characterized by a wide variety of species and quite different contents. For the protocol conformance testing, the protocol conformance test specification is the basis and key, and it is used for providing the abstract test suite (ATS) and regulating the test methods. However, there have not been still the corresponding conformance test specifications for IPv6-based WSNs. Therefore, in accordance with the characteristics of WSNs and the specific content of IPv6-based WSNs protocol specifications, the specific protocol conformance test suites for IPv6-based WSNs are described and the automatic generation technology of the abstract test suite based on the formal methods is studied in this paper.

In the programming of ATS, it is important to choose a generic programming language for the testing requirements [4]. The programming language directly determines the description methods of the test suite, whose expression ability determines the quality of the written test suite and affects the ease of use and efficiency. In addition, the programming language also directly determines the test structure and execution mechanism of test systems and affects the characteristics of test system. According to the testing requirements of the IPv6-based WSNs protocols, we choose TTCN-3 [19] as the programming language.

TTCN is developed and maintained by ETSI (European Telecommunications Standards Institute) as the internationally standardized language for defining test suite. The latest version, TTCN-3, is a modern and flexible language and can be used to describe the black-box testing of a variety of interactive systems, whose typical application fields include protocol testing, system testing, interaction testing, service testing, and module testing. Compared with other programming languages (such as C/C++, JAVA, Tcl/tk, and Perl), TTCN-3 is the programming language dedicated in protocol testing. With the characteristics of the platform independence and the special testing capabilities, TTCN-3 has been used for developing the conformance test suites for many protocols, such as GSM, 3G, and Bluetooth protocols.

3.1. Design of Protocol Conformance Test Suites in the Conformance Test Specifications. In order to integrate IPv6 and WSNs, the IETF has set up three working groups to study the key technologies and carry out the standardization of the low-power IPv6 network. The 6LoWPAN Working Group mainly discusses how to adapt IPv6 protocol to IEEE 802.15.4

based networks. The ROLL (Routing Over Low-power and Lossy Networks) Working Group focuses on the routing protocols in the low-power network and develops the routing requirements in every scene. The CoRE (Constrained Restful Environment) Working Group is providing a framework for resource-oriented applications intended to run on constrained IP networks.

At present, 6LoWPAN group has completed two RFC manuscripts: "IPv6 over Low-power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals" (RFC4919, Informational) and "Transmission of IPv6 Packets over IEEE 802.15.4 Networks" (RFC4944, Proposed Standard). ROLL Working Group has developed the routing requirements for four scenarios, including "Home Automation Routing Requirements in Low-power and Lossy Networks" (RFC5826), "Industrial Routing Requirements in Low-power and Lossy Networks" (RFC5673), "Routing Requirements for Urban Low-power and Lossy Networks" (RFC5548), and "Building Automation Routing Requirements in Low-power and Lossy Networks" (draft). On the basis of the routing requirements and the quantitative indicators in the link selection, ROLL Working Group has developed the RPL.

By studying the above protocol specifications, we can understand their basic functions and the relationship between the various functions and know the essential and optional features. Thus the test cases with different testing purposes are generated.

In order to try to cover all the technical requirements of the protocol specifications, it can be implemented by both of the manual writing and the automatic generation for identifying and describing the test suites in the test specifications as follows.

(1) Manual Writing

(i) Defined by the functions, the protocol specifications contain the following.

Firstly, according to different functions of the protocols, the technical requirements are divided to determine the large function sets. Secondly, the large function sets are refined to form the individual test cases in the light of different equipment, different implementations, and the specific function process.

(ii) Defined by the transfer of the state machine or the protocol process, we have the following.

Firstly, according to the description of the protocols, the complete state machine and/or the protocol process is established. Secondly, the transfer between any two states or any step in the protocol process could be taken as an independent test case.

(iii) Defined by the data format (frame format) stipulated in the protocols

Rather than providing the function, status, and the protocol process, some protocol specifications define the data, signals, and frame format in the data exchange. At this time, the test cases could be established according to different fields or different requirements of the data signal and the frame format.

(2) Automatic Generation. The abstract test suites may be generated automatically by the formal methods and then the test specifications are formed, as shown in Figure 1.

Firstly, the IPv6-based WSN protocol specifications in the natural language are described by the formal models, such as FSM (Finite-State Machine) model, EFSM (Extended Finite-State Machine) model, SDL (Specification and Description Language), and UML (Unified Modeling Language).

Secondly, the abstract test sequences are generated with the test generation technology, according to the protocol formal models. In the generation process, the coverage of the generated test sequences is ensured by the test coverage technology. For example, the coverage criteria (such as state coverage, change coverage, and path coverage) are used for the FSM model; the coverage criteria (e.g., all-du-path coverage) are used for the EFSM model.

Finally, the abstract test sequences are used as the basis and reference to develop the conformance testing specifications of IPv6-based WSNs protocols and can also verify the test specification by the manual writing.

In describing the test suites in the test specifications, the various parameters and primitives in each test case need defining, as well as the correct response results in the protocol exchange. As a part of the test specification, the correct judgment is not only given when the test cases succeed in passing the test, but also the judgment of failure is given against the wrong response. In the actual testing process, a lot of test cases begin with a particular state, which requires that in the beginning of each test case, the initial states of the test system and the devices under test are given to ensure that the test cases will start from the correct state machine.

According to the above generation methodology of the test specifications and test suites of IPv6-based WSN protocols, the main content of the test suites of IEEE 802.15.4 in MAC layer, 6LoWPAN in the adaptation layer, and routing protocol RPL in the network layer is as follows

(1) The Conformance Test Suite of IEEE 802.15.4 Protocol in MAC Layer. IEEE 802.15.4 is a standard which specifies the physical layer and MAC layer for low-rate wireless personal area networks and is maintained by the IEEE 802.15 Working Group. The specification in MAC layer belongs to the data link layer and is one of the goals of protocol conformance testing for IPv6-based WSNs in this paper.

Based on the function requirements of the MAC section in IEEE 802.15.4, the test suite is expected to include the following: channel scanning, building networking, beacon synchronization, association and disassociation, channel access, indirect data transmission, direct data transmission, super-frame management, GTS management, CSL mechanism, RIT mechanism, subbeacon network, and security mechanisms, as shown in Figure 2.

Figure 1: The automatic generation of the abstract test suites.

Figure 2: The test suite of IEEE 802.15.4 in MAC layer.

For example, in the test cases of "association and disassociation", the "association" function allows the device to request the association through the PAN (Personal Area Network) coordinator and join in the WPAN (Wireless Personal Area Network). The association includes four primitive interfaces: the association request, the association response, the association instruction, and the association confirmation. Two devices are used for the test of the associated function in the test suite. If the PAN coordinator wants an associated device to leave the PAN, or an associated device wants to leave the PAN, the "disassociation" function will be implemented by the upper of MLME (MAC Sublayer Management Entity). The disassociation consists of three primitive interfaces: the disassociation request, the disassociation indicates, and the disassociation confirmation. Two devices are also used for the test of the disassociation function in the test suite. The test cases are divided into two kinds of the beacon-enabled network and the nonbeacon-enabled network, which are, respectively, divided to the positive disassociation and the passive disassociation. The specific test cases include IUT (Implementation Under Test) as the device sends an association request (in the beacon-enabled network or the nonbeacon-enabled network); IUT as the coordinator receives the association request (in the beacon-enabled network or the nonbeacon-enabled network); IUT as the device takes the initiative to issue the disassociation in the beacon-enabled network; IUT as a coordinator in the beacon-enabled network, the device takes the initiative to issue the disassociation; IUT as a coordinator in the beacon-enabled network, the device is asked with the disassociation; IUT as the device in the beacon network, the coordinator asks the device with the disassociation; IUT as a coordinator in the nonbeacon-enabled network, the device is asked with the disassociation; IUT as the device in the nonbeacon-network, the coordinator asks the device with the disassociation.

(2) The Conformance Test Suite of 6LoWPAN Protocol in the Adaptation Layer. The minimum packet length IPv6 requires and supports are much greater than the maximum frame length of IEEE 802.15.4 protocol. In order to implement the transmission of IPv6 data packets over the data link layer, the construction of an IPv6 link-local address, and the autoconfiguration of the stateless address, the adaptation layer must be defined. The 6LoWPAN has defined encapsulation and header compression mechanisms that allow IPv6 packets to be sent to and received from IEEE 802.15.4 based networks.

According to the function modules of the 6LoWPAN protocol specifications, the test suite is expected to include the following: the examination of packet format, fragmentation and reorganization, IPv6 address configuration, IPv6 address resolution, packet compression and decompression, neighbor discovery, and security mechanism, as shown in Figure 3.

For example, in the test cases of "adaptation layer and its frame format", 6LoWPAN defines the package head, of which the first byte indicates the distribution type. The head expressed includes IPv6 header, IPv6 compression head, mesh head, and slice head, with enough room for the expansion in future. When multiple header types appear in the same packets, they are arranged with the order of the mesh header, broadcast header, and slice header. The testing of the packet format needs to verify whether the packet format is inline with one defined in the protocol specification. The specific test cases include encapsulating the IPv6 header format and processing; encapsulating the IPv6 compression head format and processing; encapsulating the mesh head format and processing; encapsulating the broadcast head format and processing; encapsulating the slice head format and processing; the order of the encapsulated head and processing; processing of the abnormal distribution value.

(3) The Conformance Test Suite of RPL Protocol in the Network Layer. RPL is the IPv6 routing protocols based on distance vector in connection with the characteristics of the LLN network. DAG (Directed Acyclic Graph) is constructed through nodes exchanging the distance vectors and consists of one or more goal-oriented DODAG (Destination-Oriented DAG). DODAG distinguishes between the upward routes and the downward routes based on the different routing direction to or from the root node. RPL supports three different communications means, including the mobile peer-to-peer (MP2P) communication from the network node to the central control

Figure 3: The test suite of 6LoWPAN in the adaptation layer.

Figure 4: The test suite of RPL in the network layer.

device, the point-to-multipoint (P2MP) communication from the central control device to multiple nodes, and the point-to-point (P2P) communication between nodes, to effectively avoid the problem of routing loops.

Based on the function modules of the RPL protocol specifications, the test suite is expected to include the following: control message format, construction and repair of topology, the operation of sequence counter, uplink routing, downlink routing, loop detection and avoidance, and security testing, as shown in Figure 4.

For example, in the test cases of "uplink routing", the root node broadcasts the DIO packet to the node around. After a node receives multiple DIO packets, it selects another node as its parent node and joins the DODAG topology; thus an upstream routing pointing to the root node is built, according to the requirement of the target function. The DODAG node can communicate with other DODAG nodes through the root node. The new node can send the DIO information request to the neighbor nodes through the DIS information and select the parent node to join in the network. The specific test cases include the basic rules of the DIO information; uplink route discovery and maintenance in the same DODAG; uplink route discovery and maintenance in the different DODAG; the DIO information trigger and transmission; DODAG selecting; the local DODAG operation; data path validation; leaf node operation; management in the node level.

3.2. The Generation Methods and Test Structures of Protocol Conformance Test Suites. The development of the test suite is to program ATS described in the natural language into ETS with TTCN-3. The corresponding steps include the following. Firstly the test structure and test configuration are determined according to the established test specifications and the protocol specifications. Secondly, the test data in the test specifications are refined and described with TTCN-3. Thirdly, each step of the test cases in the natural language is mapped to the corresponding TTCN-3 statement. Finally, the test judgment is given in the TTCN-3 code, according to the judgment basis of each test case with succeeding in passing the testing. The final TTCN-3 codes are formed through the above steps. The detailed process is as follows.

(a) Single test component (b) Double test components

(c) Three test components.

Figure 5: Several test structures and test configuration.

Before developing the TTCN-3 code, the test structure and test configuration are determined by the test specification. The test structure and test configuration of the single test component or multiple test components is used with the characteristics of the protocols under test.

For example, the Reduced Function Device (RFD) in IEEE 802.15.4 acts as the host node in the 6LoWPAN network, which does not relay packets and interacts only with the router in its own 6LoWPAN network. When the RFD device is tested, the test structure of single test component is used for testing, shown in Figure 5(a). The test system connects with the DUT (device under test) through a main test component (MTC) and simulates the router nodes to interact with and test the device under test.

The Full Function Device (FFD) in IEEE 802.15.4 acts as a router in the 6LoWPAN network, which relays packets to perform the routing function and may interact with the host

nodes or router nodes in its own 6LoWPAN network. When the FFD device is tested, the test structure of multiple test components is used for testing, shown in Figures 5(b) and 5(c), which are, respectively, the test structures of the double test components and the three test components. The test system (logically) connects with the device under test through multiple simulated nodes (test components) and simulates the neighbor nodes to interact with and test the device under test.

After fixing the test structure, the appropriate test configuration will be defined in the description of the TTCN-3 test cases. For the test structure of multiple test components, each PTC (Parallel Test Component) is created and implemented by using "create" and "start" statements, and the test ports are mapped and connected by using "map" and "connect" statements.

Then, the test data in the test specification are refined to determine the protocol packet format and content of the TTCN-3 test data. The structure type of the corresponding protocol packet is given with the definition of type in TTCN-3. The specific protocol packet data are defined with the "template" statement.

Each test step in the test specifications is mapped to the corresponding TTCN-3 statements. For example, when a test message is sent or received on the test interface, the "send" or "receive" statements can be adopted, respectively. If the test configuration of multiple test components is adopted, the function needs defining in the TTCN-3 code to realize the test behavior of PTC. If necessary, the synchronized messages between each test component also need adding.

Finally, according to the judgment of test specifications, the "setverdict" statement is adopted in TTCN-3 code to set the test decisions, including a variety of pass and fail situations.

In addition to the above manual methods for developing the test suite, the automatic generation technology may be adopted to improve the efficiency of generating ETS described with TTCN-3, on the basis of the foregoing formal methods of generating the test specifications.

At present, some companies (such as Conformiq, Elvior, and All4tec) have provided the commercial tools for automatically generating the TTCN-3 test cases based on the formal models and they have had some successful applications, such as the All4tec's MaTeLo generates TTCN-3 test cases based on the Markov chains, but the application of this model is relatively rare. The Conformiq's Conformiq Designer and the Elvior's TestCast Generator adopt more popular UML (Unified Modeling Language) state diagrams with the text programming language describing the model, and the TTCN-3 test suite automatically generated can be applied to the test execution with less modification, which have some successful applications in communications, medical equipment, and even the aerospace industry. The working principle of these autogeneration tools is basically the same and the process is as follows. Firstly, the protocol specification is modeled with the formal methods. Secondly, the model is inputted into the autogeneration tools and ATSs are automatically generated. Finally, with the required coverage conditions specified, ATSs are automatically transformed into ETS which is described with TTCN-3 in the background.

The automatic generation of test suite can effectively reduce the development time and the errors introduced by the manual programming. However, the methods of automatically generating test cases also have their limitations. Firstly, due to the abstract model, the generated test cases are difficult to run directly and need some adjustments by hand. Secondly, the parts requiring the strengthen testing are difficult to describe in the models. Finally, the generated test suite may not be ideal in aspects of naming, organization, and management.

In order to realize the optimal generation, the combination of the automatic generation and the manual programming is used, as shown in Figure 6. The first is to generate the test suite with the automatic generation tool. And then the test suite is complemented and modified with the manual programming based on the test specification. For example, the test cases, which require the strengthen testing and failed in the automatic generation, are added. Meanwhile, the existing test specification is verified by the automatically generated test suite. The test cases which are not covered in the test specification are supplemented and the irrational norms or the errors introduced artificially are corrected.

In addition, because most of IPv6-based WSN protocols are in the draft stage and the content of the protocols will continue changing and form a new draft version, the automatic generation method will greatly reduce the overhead of the ETS development. When the content of the protocols varies, the formal models of protocols will only be adjusted on the basis of the changes and then inputted into the automatic generation tools to regenerate the corresponding test suite, instead of programming all the test cases by the manual method. Therefore, it is the method of combining the automatic generation and the manual programming that is more suitable for the evolving IPv6-based WSN protocols.

3.3. A Test Case Is Taken as an Example. The building process of DODAG (Destination-Oriented Directed Acyclic Graph) is taken as an example to explain how to use TTCN-3 to write the test cases. The RPL's high-level goal is to provide efficient routing paths for three traffic patterns (multipoint to point, point to multipoint, and point to point) in low-power and lossy networks. Specifically, once an RPL node obtains a proper global IPv6 address, it tries to join a DODAG by exchanging ICMPv6-based DODAG Information Solicitation (DIS) or DODAG Information Object (DIO) messages [20].

As shown in Figure 7, we assume that the network has five nodes: root, nodes A, B, and C, and E. In the building process of DODAG, the root node advertises DIO message as a parent node for other nodes in its neighborhood. If the nodes A, B, and C receive the DIO message, they will process the DIO, and then join in the DODAG and set the root node as their parent node. The nodes A, B, and C compute their own "rank" value within the DODAG from the root. It is easy to calculate rank(A) = 1, rank(B) = 3, and rank(C) = 2. In Figure 7(b), after the node C selects a parent, it will propagate

Figure 6: The combination of the automatic generation and the manual programming for generating TTCN-3 test cases.

Figure 7: The construction process of DODAG.

its own DIO further down the network to form its sub-DODAG. When the root node receives the DIO message from C, it will ignore the DIO from the downstream node. When node B receives the DIO message from C, it will set C as its another parent node and compute rank(B) which is still 3. When node E receives the DIO message from C, it will set C as its parent node and compute rank(E) which is equal to 3. In Figure 7(c), node A also propagates its own DIO further down the network to form its sub-DODAG. The root node will ignore the DIO from the downstream node A. When node B receives the DIO message from A, it will compute rank(B) which is equal to 2 and then no longer set the root node and node C as the parent node.

The above protocol interaction is present in a graph, as shown in Figure 8(a). After the root node finishes sending the DIO messages, the upward routing of node B points to the root node; therefore the data from B are directly forwarded to the root node. After node A finishes sending the DIO messages, the upward routing of node B becomes root A; therefore the data form B are again directly forwarded to root A.

Therefore, through the analysis of the protocol interaction, if the IUT is the node B, the part of the red box is the behavior of the test system, as shown in Figure 8(b). P_root, P_A, and P_C are the node interfaces the test system tries to simulate, and the communication between the three interfaces does not need to be reflected in the testing process. The test case is written with the TTCN-3 language as follows (Algorithm 1) (with the clock behavior and the Alt optional steps omitted).

4. Design and Implementation of Conformance Test System for IPv6-Based WSNs

When designing the protocol conformance test system for IPv6-based WSNs, a number of technical factors must be taken into account, such as the versatility, extensibility, and standardization.

The first is the versatility of the test system. IPv6-based WSN protocols under test include IEEE 802.15.4 in MAC layer, 6LoWPAN in the adaptation layer, and routing protocol RPL in the network layer. The types of the protocols are various and their content is very different, but it is inefficient and uneconomical to, respectively, develop the appropriate test system for each protocol. The conformance testing

Figure 8: The protocol interaction between nodes.

of IPv6-based WSNs needs a universal solution necessary. In addition, the test system needs to simulate and adapt different scenarios, such as different topologies and events.

The second is the extensibility of the test system. In future, the conformance test system may add new test cases and new air interfaces and implement the automated interconnect testing between multiple test systems, which requires that the conformance test system should be scalable, such as the upgraded test cases, the modular system components and the networked interconnection of the systems.

The last is the standardization of the test system. The marketing promotion of the conformance test system needs the certification license of the authoritative department; so the system itself must comply with relevant standards of conformance testing and the related equipment norms in industry.

4.1. Design of Conformance Test System for IPv6-Based Wireless Sensor Networks. Based on the above consideration, a networked protocol conformance test system for IPv6-based WSNs is designed in this paper. The system logically consists of two component units, which are the remote control unit in the user side and the test unit in the equipment side. The composition of the conformance test system and the connection with the device under test is shown in Figure 9.

The remote control unit is implemented on a general computer, which remotely controls the test system through the common LXI (the LAN eXtensions for Instrumentation) bus interface. The remote control unit can provide the friendly remote user interface for conformance testing, a convenient and practical test message window, the flow control window for test execution, the window for analysis of test results, smart decisions, and report formation.

The test unit is the main unit for providing the conformance test capability, which includes system control unit, system interface, and TTCN-3 execution environment. The system control unit provides the management capabilities for the component units of the test system and receives the protocol information from the LXI bus. The system interface provides users with the use of the interface window of the system, realizing configuration, management, and operation. TTCN-3 execution environment is the main functional module of performing the conformance testing, whose design refers to the framework of the TTCN-3 runtime environment to support the running of ETS. The PA (platform adapter) provides the clock and other underlying support for the TTCN-3 runtime system. The SA (SUT adapter) and the corresponding test components provide the adapter between the test system and the system under test. SA includes the physical layer interface components of 780 MHz and 2.4 GHz and other testability interface components, as well as a set of benchmark protocol stacks. TTCN-3 core runtime system can perform the calls of the benchmark protocol stacks to form the peer-level protocol testing capability.

The system control unit is the main control part of the component modules in the test system and implements the collaborative work between the components, whose main functions include receiving the user input request from the system interface, adjusting and controlling the basic status and working modes of the system, receiving the protocol information from the remote LXI bus, executing the corresponding test operation, submitting the test results to the user, completing the LXI bus synchronization to ensure the timing of the TTCN-3 operation, automatically detecting and configuring the system components, and providing the correct power supply, the clock, and resources configurations for the hardware components. In addition, the system control unit also provides the remotely upgrading capability for the system and allows the user to update the test suites to ensure the extensibility of instrument.

The system interface is the interactive window between the test system and the users. The user-entered information and the information outputted by the test system will be shown in the system interface. In the beginning of the test execution, the users can select the tested protocol, configure the test parameters, select the test execution mode and the ETS, and start the testing process by the user interface. In the testing process, the system interface can display the test log in real time. In the end of the testing process, the system interface can generate the test result and reproduce the execution process of test cases to facilitate the analysis of the test result.

TTCN-3 running environment (including the ETS of the relevant protocols) is the core part of the software in the test system, which provides the ETS and the supporting environment, including TM (test management), T3RRTS

```
(1) P_root.send(DIO_root);
(2)     P_C.send(Data);
(3)         P_root.receive(Data);
(4)             P_C.send(DIO_C);
(5)                 P_A.send(DIO_A);
(6)                     P_C.send(Data);
(7)                         P_A.receive(Data){
(8)                             setverdict(pass);
(9)                         }
```

Algorithm 1: The building process of DODAG.

Figure 9: The protocol conformance test system for IPv6-based WSNs.

(TTCN-3 runtime system), CD (coding and decoding), PA (platforms adapter), and SA (SUT adapter).

TM conducts the comprehensive management of the test system. Most of IPv6-based WSN protocols leave a lot of options, which may or may not be implemented in an implementation. The process of conformance testing requires a combination of the specific implementation of protocols. PICS (the Protocol Implementation Conformance Statement) and PIXIT (Protocol Implementation eXtra Information for Testing) are provided by the supplier of an implementation and jointly maintained by the test administrator, which describes protocol requirements and the supplier's support for these requirements, as well as the additional information regarding parameters and values or their ranges. According to PICS and PIXIT provided by the supplier, the test administrator selects the test cases and deploys the test parameters in the system interface to perform the initialization of test system.

After the initialization of test system, TM will maintain the list of test cases selected by the user in the testing process, transfer the parameters of the external modules set by the user to the ETS, and start to run the selected test cases. In the testing process, TM will fulfill the information exchange between the system interface and the ETS. After the end of the test execution, TM is responsible for generating the test logs and displaying them in the system interface. In addition, TM has the function of operating the database which may store/read the finishing test information (such as the configuration and the test results) into/from the database.

T3RRTS provides the support for the execution of the TTCN-3 ETS, including defining and implementing the TTCN-3 test data types that the ETS needs; during the execution of ETS, providing TCI (test control interface) to complete the interaction between ETS and other modules (e.g., TM, CD, etc.); providing the TRI(test run interface) to complete the interaction between the ETS, the clock, and SA. In short, the execution of ETS must be supported by T3RRTS.

In order to realize the seamless transition of the test data between the TTCN-3 format and the actual transmitted bit strings, the CD module is designed in the test system. TTCN-3 test data can be converted into a bit string suitable for

the transmission to the SUT and the data received by SA are also converted into the corresponding TTCN-3 format data with CD.

The PA module provides the ST (System Timer) module for the TTCN-3 running environment, which can provide the unified clock and the associated operations for each test component, including the starting and stopping of the clock, reading, and querying the clock state.

SA mainly provides the communication between the TTCN-3 running environment and the underlying hardware platform of the test system and can complete the sending and receiving of the related protocol test data by controlling the underlying test components. In the test system, different SA modules need realizing the different tested protocols.

In the scheme, the TI (test interface) component, which is the core of hardware modules in the test system, is connected to SA and provides the test system with the access to the various testability interface components of the DUT (device under test). TI is closely related with the interface types of typical sensor nodes. The sensor nodes generally have the air interfaces (the wireless communication interfaces). Some nodes at the edge of WSNs act as the role of routing forwarding and usually also have an Ethernet interface or serial interface based on the PPP (Point to Point Protocol). In addition, in order to initialize the testing environment and set the nodes into a known test status, the power and reset signal interfaces of nodes need to been controlled. Some typical network interfaces of nodes in IPv6-based WSNs are shown in Table 1.

For the frequency bands of 780 MHz and 2.4 GHz, there are the mature and certified integrated circuits in the market. Therefore the test system can be designed based on the above-integrated circuits. In order to support the protocols in MAC layer, the collisions of multiple RF (Radio Frequency) packets and power superposition need simulating. So the multichannel RF output needs supporting in the same TI. The RF test interface component is designed as shown in Figure 10. The designed scheme can support full-duplex testing of RF receiving and sending. At the same time, the RF can simulate to produce the superimposed waveform of two RF signals to ensure the testing capability in MAC layer. The sending time in the air of IEEE 802.15.4 protocol packets is about hundreds of microseconds. In order to simulate a variety of situations, the timing control logic module in the scheme will be generated by the FPGA timing to ensure providing microsecond-level synchronization timing control for different RF modules. Two TI components will be, respectively, implemented for 780 MHz and 2.4 GHz bands. With multiple RF test interface components, the test system can realize different wireless environment.

Serial test interface and Ethernet test interface are the more mature technology. However, in the conformance testing, some test cases need to precisely understand the delay between the events in the protocol interaction, which requires implementing the precise timing control of the protocol packets, sending a packet at the particular time, and obtaining the accurate timestamp of the received packet. Therefore, the timing control logic needs designing with FPGA to realize the precise timing control and the timing analysis for the serial transceiver control and Ethernet transceiver control. The serial test interface and Ethernet test interface component are designed as shown in Figure 11.

Table 1: The typical testability interfaces.

Type	Interface name
Air interface	IEEE 802.15.4-compliant 2.4 GHz RF interface
	IEEE 802.15.4-compliant 780 MHz RF interface
Wired interface	Serial interface based on the PPP protocol
	Ethernet interface
Others	Power interface
	Reset interface

The analog switch and the sequential logic are, respectively, used for the power interface and the resetting interface to control the power supply and reset the node status. The power interface and the resetting interface are designed as shown in Figure 12.

Based on the RF test interface supporting 780 MHz and 2.4 GHz bands in physical layer, the test system provides a set of benchmark protocol stacks for conformance testing, which include IEEE 802.15.4 in MAC layer, 6LoWPAN in the adaptation layer, and routing protocol RPL in the network layer. Different levels of the protocol stacks can be accessed by the SA, so as to enable data communication between the peer levels, to facilitate carrying out the conformance testing layer by layer, and to simplify the complexity of programming with TTCN-3.

4.2. Conformance Testing Methods for IPv6-Based Wireless Sensor Networks. Combined with the specific IPv6-based WSN protocols, different test methods are used for different protocol levels and different roles of nodes; so multiple test methods are designed in the test system in this paper. Taking 6LoWPAN as an example, the remote test method of the single test interface and the TTM (Transverse Test Method) of multiports are, respectively, used for testing the nodes within the network and the edge nodes (namely, the boundary router node in RPL protocol).

For the internal nodes of WSNs, the air interface usually is used for communicating with other nodes. When testing such nodes, the test structure of the remote test method and the single test node is used, as shown in Figure 13. The test system connects with DUT through the air test interface components.

If the edge nodes undertake the Internetwork forwarding of packets, they usually have the wired interfaces in addition to the air interfaces to realize data forwarding between the air interfaces and the wired interfaces. When testing such nodes, the test structure of TTM and multiple test nodes is used, as shown in Figure 14. TI components (air test interface and wired test interface) in the test system are connected with the DUT. With packets flowing through the DUT, the protocol conformance testing of the DUT is executed.

4.3. Implementation of Conformance Test System for IPv6-Based Wireless Sensor Networks. In order to verify and evaluate the feasibility of the proposed design of conformance test

Figure 10: RF test interface component for 780 MHz and 2.4 GHz.

Figure 11: Wired test interface components.

system, we have built an initial prototype of conformance test system for IPv6-based WSNs on the basis of HINT test platform [21]. HINT is a high-accuracy nonintrusive networked testing platform for WSNs, which was designed and developed by our own laboratory. With the help of HINT, users can gather information of internal signals on the sensor nodes, measure the energy consumption of sensor nodes, parse the interconnect signal for the radio packets, obtain precise timestamps of events for fine-grained phases, evaluate the network performance parameters, program and debug the sensor nodes remotely, and store the runtime data to trace files. Most of these features are disturb free and transparent to the applications.

Based on the design of the conformance test system, the main test unit of the conformance test system we have implemented is shown in Figure 15. In the prototype of conformance test system, the tested nodes adopt the SensHoc nodes, which are exploited by our own laboratory. SensHoc node is compatible to TelosB produced by Crossbow Technology. The actual picture and block diagram of SensHoc are shown in Figure 16. The micro-controller of SensHoc is TI MSP430, and the RF transceiver is CC2420. SensHoc node supports the TinyOS operating system, which is the open source operating system developed by UC Berkeley and designed specifically for wireless sensor networks. We use BLIP and TinyRPL in TinyOS 2.1, the TinyOS implementation of 6LoWPAN, and RPL protocols, which are also tested protocol implementation.

In order for the preparation and implementation of TTCN-3 test cases, the TTworkbench [22] developed by Testing Technologies GmbH is used and its operation interface is shown in Figure 17. The JAVA virtual machine is adopted as a TTCN-3 execution environment to execute the ETS. With the support of the JAVA virtual machine, the executable TTCN-3 test cases can run on different hardware and software platforms.

In the software design of the prototype system, the one-stop processing mode with data acquisition, analysis, and result processing is adopted to realize test cases as the center of the software architecture. In the management of test case information, we have designed information structure, including the basic information of the test cases (e.g., number and name), the test information (e.g., test conditions, test steps), and the configuration information (e.g., equipment configuration information). In the management of test equipment and the conjunction of test cases, we have designed the information constitution of test equipments, the conjunction modes, and management behavior patterns between test cases and test equipment. In the management of test data, the usage patterns, the flow, and the format conversion requirements of test data in the different test cases are designed, based on the collection, processing, and analysis of test data. In the processing of the test analysis results, the kernel module for matching and analysis is designed; the test data are analyzed by the automatic means of filtering, comparing, matching, and finding to obtain the test results.

Figure 12: The power interface and the resetting interface.

Figure 13: The remote test method with single test interface.

Figure 14: Transverse test method with multiple test nodes.

The test results and test logs are generated and outputted based on the test analysis results. The correct judgment is not only given when the test cases succeed in passing the test, but also the judgment of failure is given against the wrong response.

At present, the prototype of conformance test system has implemented the decomposition and reconstruction of the 6LoWPAN protocol packets, the conformance testing for the 6LoWPAN protocol header compression function, and so on, based on the nonintrusive signal capturing and automatic matching and analyzing technology. The actual operation screenshot of prototype system in conformance testing for the 6LoWPAN protocol is shown in Figure 18. At a later time, the prototype of conformance test system will be upgraded and expanded to the full realization of the protocol conformance testing for IPv6-based Wireless Sensor Networks.

4.4. Characteristics of Conformance Test System. Based on IPv6-based WSN protocol specifications, the conformance test system for IPv6-based WSN protocols proposed in this paper focuses on the versatility, extensibility, and standardization of the system. The test suites are designed according to the specifications and the standardized development program is adopted to ensure the extensibility of the test system. Thus the normative, practical, open, and configurable test

Figure 15: The main test unit of the conformance test system.

Figure 16: SensHoc sensor node.

Figure 17: The operation interface of TTworkbench.

Figure 18: The actual operation screenshot of test system.

system is established. The conformance test system for IPv6-based WSN protocols in this paper has the following characteristics.

(1) Opening and Flexible Extensibility. TTCN-3 is used in this paper for the transform from ATS to ETS. The scheme not only efficiently supports the development of ETS, but also facilitates the user to add dynamically the test suite. TTCN-3 also does not make the test system depend on the specific protocol test suite; thus it provides the test system with the good extensibility and adaptability.

In addition, in the hardware design, the component design and the slot structure are used for providing the unified extensible hardware architecture to support the additional test interfaces in future.

(2) Rich Configurability. Based on the universal conformance testing scheme with TTCN-3, the test system can flexibly configure different test suites and patch up the new test cases. Following the LXI design, the test system can realize the seamless interconnection with different system and upper software. With the TI components, the test system can support multiple testability interfaces and test configuration. The testing content and testing reports can be flexibly set by TM. The test system has multiplex test interface components and can realize different wireless environment and meet different test requirements. The above means provide the test system with rich configurability, which makes the test system adjust to the supplier's different support for protocol requirements in the specific implementation, combined with PICS and PIXIT.

(3) High Standardization. The scheme of the test system for IPv6-based WSN protocols in this paper is in strict accordance with ISO/IEC 9646. In addition, in the technology choices, the test system conforms with the related instrument design standards and the protocol test specifications in order to ensure that standardization and comparability of the test system.

(4) Engineering Practicality. In the hardware design, the test system designed in this paper emphasizes the practicality of system and the combination of theoretical study and practical application, stipulating the engineering design, such as mechanicalness, electricity, ventilation, power supply, and electromagnetic compatibility. In the software design, the convenient and practical test message windows provide the users with the accurate record of the message flow and the intelligent judgment of the test results. A variety of system configuration and the easy-to-read judgment solutions are offered to meet the testing requirements of different users. The above means provide the test system with engineering practicality.

5. Conclusion

For the large-scale application of WSNs, the interconnection between WSNs and the existing Internet based on IP protocols must be considered. The WSN protocols based on IPv6, which integrate IPv6 protocol into WSNs, are the major trends of the future applications. With the development of WSN devices and their protocol stacks, as well as the network integration of WSNs and Internet, it must be considered whether the protocol implementations in different sensor network devices are or not consistent to the referenced standards, which is the protocol conformance testing. As an important aspect of the protocol testing, the conformance testing is mainly used for testing and verifying the conformity between the protocol implementations and the standards. At present, there is a serious shortage of research on protocol conformance testing and the equipment development for IPv6-based WSNs. In this paper, faced with difficulties and challenges in the protocol conformance testing for IPv6-based WSNs, the conformance testing techniques and methods for IPv6-based WSN protocols (including IEEE 802.15.4 in MAC layer, 6LoWPAN in the adaptation layer, and routing protocol RPL in the network layer) are studied, the protocol conformance test suites are stipulated and developed, and the open conformance test system is designed and implemented towards the standardization of IPv6-based

WSNs. At present, the prototype system has been developed and the development of the complete test system is in progress. We believe that the related outcomes can fully validate the IPv6-based WSN nodes and promote the process of large-scale commercialization of WSNs.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (Grant no. 60903211), the State Key Development Program for Basic Research of China (Grant no. 2011CB302902), and the National Science and Technology Major Project (Grant no. 2011ZX03005-002).

References

[1] A. Floch, F. Roudaut, A. Sabiguero, and C. Viho, "Some lessons from an experiment using TTCN-3 for the RIPng testing," in *Proceedings of the 17th IFIP TC6/WG 6.1 International Conference on Testing of Communicating Systems (TestCom'05)*, pp. 318–332, June 2005.

[2] F. N. Noudem and C. Viho, "Modeling, verifying and testing the mobility management in the mobile IPv6 protocol," in *Proceedings of the 8th International Conference on Telecommunications (ConTEL'05)*, pp. 619–626, June 2005.

[3] J. P. Wu, X. H. Chen, and R. B. Hao, "Protocol integrated test system based on formal technology PITS," *Journal of Tsinghua University (Science and Technology)*, vol. 38, no. S1, pp. 26–29, 1998.

[4] X. Yin, Z. L. Wang, C. M. Jing, and X. G. Shi, "A TTCN-3-based protocol testing system and its extension," *Science in China Series F*, vol. 51, no. 11, pp. 1703–1722, 2008.

[5] Y. B. Lin, C. F. Liang, K. H. Chen, and H. Y. Liao, "NTP-SIOT: a test tool for advanced mobile services," *IEEE Network*, vol. 21, no. 1, pp. 21–26, 2007.

[6] Y. B. Lin, C. C. Wang, C. H. Lu, and M. R. Hsu, "NTP-PoCT: a conformance test tool for push-to-talk over cellular network," *Wireless Communications and Mobile Computing*, vol. 8, no. 5, pp. 673–686, 2008.

[7] Z. L. Wang, X. Yin, Y. Xiang et al., "TTCN-3 based conformance testing of mobile broadcast business management system in 3G networks," in *Proceedings of the 21st IFIP International Conference on Testing of Communicating Systems and the 9th International Workshop on Formal Approaches to Testing of Software (Testcom/FATES'09)*, pp. 163–178, November 2009.

[8] P. Levis, N. Lee, M. Welsh, and D. Culler, "TOSSIM: accurate and scalable simulation of entire TinyOS applications," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys'03)*, pp. 126–137, November 2003.

[9] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (Simutools'08)*, pp. 1–10, March 2008.

[10] The Network Testbed, MoteWorks, http://www.xbow.com.

[11] G. Werner-Allen, P. Swieskowski, and M. Welsh, "MoteLab: a wireless sensor network testbed," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN'05)*, pp. 483–488, April 2005.

[12] X. Kuang and J. Shen, "SNDS: a distributed monitoring and protocol analysis system for wireless sensor network," in *Proceedings of the 2nd International Conference on Networks Security, Wireless Communications and Trusted Computing (NSWCTC'10)*, pp. 422–425, April 2010.

[13] A. Koubaa, S. Chaudhry, O. Gaddour et al., "Z-monitor: monitoring and analyzing IEEE 802.15.4-based wireless sensor networks," in *Proceedings of the 6th IEEE LCN Workshop on Network Measurements (WNM'11)*, pp. 939–947, October 2011.

[14] J. J. Xiong, C. H. Edith, Y. F. Zhou, and R. L. Michael, "RealProct: reliable protocol conformance testing with real nodes for wireless sensor networks," in *Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom'11)*, pp. 572–581, November 2011.

[15] J. F. Ma and H. Tian, "The design of the conformance test system for 6LoWPAN protocol stack," *Journal of Telecommunications Network Technology*, vol. 57, no. 9, pp. 25–27, 2010.

[16] F. Song, "IPv6 wireless sensor network protocol test system," *Journal of Telecommunications Network Technology*, vol. 59, no. 1, pp. 13–18, 2012.

[17] X. H. Chen, *The Design and Implementation of a Conformance Testing System for 6LoWPAN Protocol Stack*, East China Normal University, Shanghai, China, 2006.

[18] P. Wang, P. P. Wang, and Y. Li, "Research on the conformance test method of isal00. 1la protocol," *Chinese Journal of Scientific Instrument*, vol. 31, no. 8, pp. 312–315, 2010.

[19] J. Grabowski, D. Hogrefe, G. Réthy, I. Schieferdecker, A. Wiles, and C. Willcock, "An introduction to the testing and test control notation (TTCN-3)," *Computer Networks*, vol. 42, no. 3, pp. 375–403, 2003.

[20] J. G. Ko, D. H. Stephen, and O. Gnawali, "Evaluating the performance of RPL and 6LoWPAN in tinyos," in *Proceedings of the 10th International Conference on Information Processing in Sensor Networks (IPSN'11)*, pp. 14–19, April 2011.

[21] W. Huangfu, L. M. Sun, and J. C. Liu, "A high-accuracy non-intrusive networking testbed for wireless sensor networks," *Eurasip Journal on Wireless Communications and Networking*, vol. 2010, Article ID 642531, 2010.

[22] The Test Development and Execution Environment, TTworkbench, http://www.testingtech.com.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 154602, 9 pages
doi:10.1155/2012/154602

Research Article

A Camera Nodes Correlation Model Based on 3D Sensing in Wireless Multimedia Sensor Networks

Chong Han,[1] Lijuan Sun,[1,2,3] Fu Xiao,[1,2,3,4] Jian Guo,[1,2,3] and Ruchuan Wang[1,2,3,4]

[1] *College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2] *Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China*
[3] *Key Lab of Broadband Wireless Communication and Sensor Network Technology of Ministry of Education, Nanjing University of Posts and Telecommunications, Jiangsu Province, Nanjing 210003, China*
[4] *Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou 215006, China*

Correspondence should be addressed to Lijuan Sun, sunlj@njupt.edu.cn

Received 20 July 2012; Revised 21 September 2012; Accepted 26 September 2012

Academic Editor: LiuSheng Huang

Copyright © 2012 Chong Han 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 multimedia sensor networks, multiple camera sensor nodes generally are used for gaining enhanced observations of a certain area of interest. This brings on the visual information retrieved from adjacent camera nodes usually exhibits high levels of correlation. In this paper, first, based on the analysis of 3D directional sensing model of camera sensor nodes, a correlation model is proposed by measuring the intersection area of multiple camera nodes' field of views. In this model, there is a asymmetrical relationship of the correlation between two camera nodes. Then, to farthest eliminate the data redundancy and use the node collaboration characteristic of wireless (multimedia) sensor networks, two kinds of cluster structure, camera sensor nodes cluster, and common sensor nodes cluster are established to cooperate on image processing and transmission tasks. A set of experiments are performed to investigate the proposed correlation coefficient. Further simulations based on a sample of monitoring a crossing by three correlative camera nodes show that the proposed network topology and image fusion and transmission scheme released the pressure of camera node greatly and reduce the network energy consumption of communication of the whole network efficiently.

1. Introduction

The integration of low-power wireless networking technologies with inexpensive CMOS cameras and microphones is enabling the development of distributed networked systems referred to as wireless multimedia sensor networks (WMSNs) [1, 2]. The characteristics of WMSNs diverge considerably from traditional network paradigms such as the Internet and even from "scalar" sensor networks. The applications of WMSNs, such as a surveillance sensor network, environmental and industrial monitoring, intelligent traffic congestion control, health-care, and other multimedia digital entrainments, or green city applications, require the sensor network paradigm to be rethought in view of the need for mechanisms to deliver multimedia content with a predetermined level of quality of service (QoS).

Different from conventional sensor networks, WMSNs are characterized by high data rate and directional sensing range on account of multimedia nodes' field of views (FoVs). In a densely deployed WMSN, there exists correlation among the visual information observed by cameras with overlapped FoVs. These differences are calling for new approaches for sensor networking and in particular in network data processing for reduction of data redundancy. So in this paper, aiming at the issue of correlation relationship of multiple cameras, based on the analysis of 3D directional sensing model of camera sensor nodes, we present the correlation model among the camera sensor nodes and propose two kinds of cluster structure, camera sensor nodes cluster and common sensor nodes cluster to cooperate to accomplish image processing and transmission tasks in WMSNs.

The remainder of this paper is organized as follows. Section 2 discusses some previous works on correlation degree of camera nodes which motivated our work. Section 3 presents correlation model for multiple cameras, and the correlation calculate method of multiple camera nodes

based on the property of convex is proposed. Section 4 describes correlation-based hierarchical network structure and image processing and transmission scheme. The analysis of the correlation model and performance of the proposed framework are examined in Section 5. Finally conclusions and future work are derived in Section 6.

2. Related Works

There are lots of researches on correlation model of camera sensor nodes based on 2D camera sensing model. In [3], different from conventional sensing models where omnisensing area centers on the sensor node, a directional sensing model of camera sensor nodes is first employed. Meanwhile, the correlation degree of two cameras was defined as the portion of overlapped sensing areas to the entire area of the camera's FoVs. Then two sensors cooperate with each other, image processing scheme based on correlation is proposed. But this scheme is only valid when the sensing directions of the two sensors do not differ very much.

In [4], a spatial correlation model for visual information in WMSNs is proposed. Firstly, a spatial correlation function is derived to describe the correlation characteristics of visual information observed by cameras with overlapped FoVs. Then, by using this spatial correlation model, the correlation characteristics of visual information are obtained at low communication and computation costs. The shortcoming of this spatial correlation model is it just considers the angle between two cameras. But the offset angle of camera, which expresses the area of camera's FoVs, is not investigated, especially, when the offset angle of two cameras is unequivalent. In [5], based on the spatial correlation model put forward in [4], an information-theoretic data compression framework is proposed with the objective to maximize the overall compression of the visual information retrieved by a WMSN.

Papers [6, 7] show how multimedia nodes of a randomly deployed WMSN are categorized in clusters considering the FoVs as the criterion of clustering. If the FoVs of two nodes have a wide common area, sensors are grouped in a cluster since they obtain a similar vision of the monitored area. The established clusters in [6] are disjoint and nonoverlapping while in [7] they overlap each other with common nodes.

Considering the above mentioned, the correlation coefficient, or correlation degree of camera nodes based on overlapped FoVs is an important parameter to effectively divide the monitoring task between two sensors, for example, to calculate the joint entropy of multiple cameras, or to be the criterion of camera node clustering or hibernate redundant nodes, and so forth. And all the above correlation models are based on the cameras' FoVs model which is sector shaped. However, from the human eye's point of view, this simplified 2D FoVs model is reasonable, but it is not very suitable for CCD cameras in monitoring scenario in WMSNs. In practice, the imaging surface of a camera is generally a rectangle, but the shape of the human retina is much closer to a spherical surface. Objects within some range of distances (called depth of field or depth of focus) are in acceptable focus. The field of view of a camera is the portion of scene space that actually projects onto the retina of the camera. It is not defined by the focal length alone but also depends on the effective area of the retina, for example, the area of film that can be exposed in a photographic camera, or the area of the CCD sensor in a digital camera [8]. So to investigate the 3D perception model of camera node is very necessary.

In [9], since the 2D-based schemes cannot be easily extended to address the coverage issues in 3D directional sensor networks, a 3D directional sensor coverage-control model with tunable orientations is proposed to specify the actual target-detecting scene. Then, address the issue of coverage enhancement by sensors with tunable 3D orientations.

In [10], a 3D wireless multimedia sensor nodes coverage perception model with tunable tilt angle and deviation angle is designed. Through decomposing the 3D space problem to the two aspects of horizontal plane and vertical plane, which reduce the 3D problem to 2D ground. And tilt angle and deviate angle can be adjusted to change nodes' sensing direction, also particle swarm optimization is used to eliminate sensation overlapped area and blind spots, thus ultimate coverage performance of the wireless sensor network can be enhanced. In our prior work [11], based on the 3D wireless multimedia sensor nodes coverage perception model proposed in [10], a FoVs correlation model to describe the correlation characteristics for the images observed by multiple cameras with overlapping FoVs is proposed. Two algorithms, a grid-based and a relative position-based algorithm are devised to calculate the correlation coefficient of two cameras. Eight kinds of relative positions of two cameras's FoVs are shown. Then, a relative position-based approach is designed to calculate the correlation coefficient. The shortcoming of this paper is to use the proposed relative position-based approach, and all areas of cameras's FoVs must be equivalent. This is infeasible in practice.

In this paper, based on the analysis of the 3D directional sensor coverage-control model in [9], we study the correlation characteristics of camera nodes and utilize this correlation to cluster the camera nodes in WMSNs to achieve network energy conservation and prolong network lifetime.

3. Correlation Model for Multiple Cameras

3.1. 3D Sensing Model. In [9], the camera's 3D sensing model is located at a fixed 3D point and sensing direction 3D rotatable PTZ (Pan Tilt Zoom) camera. Figure 1 illustrates that 3D sensing model. This model is denoted by a 5-tuple $(P, \vec{D}, A, \alpha, \beta)$. Where P is the location (x, y, z) of the camera node in 3D space, $\vec{D}$ is the sensing orientation of the camera node, and $\vec{D} = (dx, dy, dz)$ is of unit length, where dx, dy, and dz are the components along x-axis, y-axis, and z-axis, respectively. $\vec{D}$ controls the orientation of the sensor node and can be seen as an angle for the origin of axes. To the original $\vec{D}$ in the actual scene, it is just to control the initial orientation of the camera, and it could be represented by the angle θ. γ is the value of the tilt angle γ; α and β are the horizontal and vertical offset angles of the field of view around $\vec{D}$.

Figure 1: The 3D directional sensing model.

Figure 2: The projection of two parallel lines in image.

Through analyzing this camera node 3D sensing model, we can find that the FoVs of wireless camera sensor nodes is trapezium-shaped. Meanwhile, when $\vec{D}$ is represented by the angle θ, based on our priori work [10], the four vertexes in this trapezium can be calculated by:

$$\begin{aligned} D_1 &: (x + d_1 \times \cos(\theta + \alpha), y + d_1 \times \sin(\theta + \alpha)), \\ D_2 &: (x + d_1 \times \cos(\theta - \alpha), y + d_1 \times \sin(\theta - \alpha)), \\ D_3 &: (x + d_2 \times \cos(\theta - \alpha), y + d_2 \times \sin(\theta - \alpha)), \\ D_4 &: (x + d_2 \times \cos(\theta + \alpha), y + d_2 \times \sin(\theta + \alpha)), \end{aligned} \tag{1}$$

where d_1 and d_2 are calculated as

$$\begin{aligned} d_1 &= z \times \tan(\gamma - \beta), \\ d_2 &= z \times \tan(\gamma + \beta). \end{aligned} \tag{2}$$

In practice, when two parallel lines in the ground plane are taken picture by a camera, the projections of these two parallel lines lying in the image plane will converge on a point in infinity far away. Figure 2 simply illustrates the projection of two parallel lines in the image plane. This is why based on camera 3D sensing model, camera's FoVs is trapezium-shaped.

3.2. Correlation Coefficient. In WMSNs, multiple camera sensors are deployed to provide multiple views, multiple resolutions, and enhanced observations of the environment, meanwhile, they are always deployed in a field of interest, so the cameras' FoVs may overlap with each other. A camera can only observe the objects within its FoV. The observed images from cameras with overlapped FoVs are correlated with each other. For two arbitrary camera sensors C_i and C_j with FoVs F_i and F_j, suppose at a same time, their observed images are X_i and X_j, respectively. X_i and X_j are correlated when F_i and F_j overlap with each other. The definition of correlation coefficient between two cameras is given as below.

Definition 1 (correlation coefficient). Given the sensing areas of camera nodes C_i and C_j with FoVs F_i and F_j as Area(F_i) and Area(F_i), the correlation degree of X_j to X_i denoted by $\rho_{i,j}$, and the correlation degree of X_i to X_j denoted by ρ_{ji}, respectively are defined as

$$\begin{aligned} \rho_{ij} &= \frac{\text{Area}(F_i) \cap \text{Area}\left(F_j\right)}{\text{Area}(F_i)}, \\ \rho_{ji} &= \frac{\text{Area}(F_i) \cap \text{Area}\left(F_j\right)}{\text{Area}\left(F_j\right)}. \end{aligned} \tag{3}$$

Different from the correlation degree model of two cameras proposed in [3, 4, 6, 11]. In this paper, in most cases, ρ_{ij} is unequal to ρ_{ji} in (3). This is because when a WMSN is initially deployed, in order to gain better surveillance quality of service, lots of coverage-enhancing methods, for example, simulated annealing algorithm [9], particle swarm optimization [10], are used to change sensing orientation and adjust the tilt angle of camera. After network optimization, the area of each camera node's FoVs will be not always equivalent.

From Section 3.1, we know that a 3D space sensor node has trapezium-shaped FoVs in the ground plane. And those four vertexes in trapezium can be calculated by (1). Through this processing, the problem of the overlapping FoVs of two cameras has been formulated as determining the intersection polygon of two polygons in plane geometry. Once the area of the FoV of camera and intersection area of the FoVs of two cameras are calculated, we can get the correlation coefficient ρ of camera C_i and camera C_j in (3). To calculate the overlapping area of two arbitrary camera sensors, the simple and intuitive method is to divide the overall network area into small grids, then check for each grid if its center whether in the FoV of a camera [5] or assuming the sensing area of sensor is composed of discrete points, then examine every point therein and determine if it also falls in others' sensing areas [3]. The time complexity of these two methods is decided by the size of the grid or the number of discrete points. The better precision these two methods pursue, the higher complexity they increase. To remedy these deficiencies, in our priori work [11], a relative position-based approach was proposed to calculate the correlation of

```
(1) Get the FoVs of two cameras C_i, C_j are F_i, F_j, respectively. V_i and V_j are the sets
    of the four vertices of F_i and F_j, that is, V_i = {v_im | m = 1,2,3,4}, V_j = {v_jm | m = 1,2,3,4}.
    E_i and E_j are the sets of the four edges of F_i and F_j, that is, E_i = {e_in | n = 1,2,3,4},
    E_j = {e_jn | n = 1,2,3,4};
(2) Initialize the vertex set V = φ, the point of intersection of two edges v_e = φ;
(3) for m = 1 to 4 do
(4)     if v_im in F_j then
(5)         V = V ∪ v_im;
(6)     end
(7)     if v_jm in F_i then
(8)         V = V ∪ v_jm;
(9)     end
(10)    for n = 1 to 4 do
(11)        v_e = e_im ∩ e_jn;
(12)        if v_e ≠ φ then
(13)            V = V ∪ v_e;
(14)        end
(15)    end
(16) end
(17) V = Sort(V);
(18) The polygon P constructed by vertex set V is the intersection polygon of F_i and F_j;
(19) return ρ_{i,j} = (Area(F_i) ∩ Area(F_j))/Area(F_i), ρ_{j,i} = (Area(F_i) ∩ Area(F_j))/Area(F_j).
```

Algorithm 1: Correlation coefficient calculation.

multiple camera nodes, but the shortcoming of this approach is all area of cameras's FoVs must be equivalent. In this paper, we extend this approach to various sizes of the FoVs area of camera nodes in the following.

3.3. Correlation Coefficient Algorithm. In a WMSN, when some cameras are deployed and would not move anymore, these cameras are located in the variance height in 3D space, they maybe have different tilt angles γ, horizontal offset angles α, vertical angles β, and orientation of the camera angles θ. The trapezium area size of the camera's FoV is determined by the angle γ, α, β, and the height z of camera nodes. The angle θ just controls the main sensing orientation of the camera. But because every camera's FoV is project in the ground plane trapezium-shaped. As is known, trapezium is a kind of convex polygon, and the intersection of two arbitrary convex polygons still is convex polygon. So we can use the convex polygons knowledge of Computer Geometry to calculate the intersection area of two trapeziums.

Using the property of convex polygon, according to the position to calculate intersection of two trapeziums is feasible. The thought of this correlation coefficient calculate algorithm is as follows: first, judge the four vertices of a trapezium A whether located in another trapezium B. If it is true, add the vertex(s) that locate in B to a vertex set V, vice versa. Then calculate the point of intersection of the four edges in trapezium A with the four edges in trapezium B, respectively. Add the point of intersection to V. Finally, sort this vertex set V by clockwise or counterclockwise order. Now, the polygon constructed by arranged vertex set V is the intersection polygon of trapezium A and trapezium B. Algorithm 1 shows the pseudo code of correlation coefficient algorithm. From the details of this algorithm, we can obtain it has a processing time complexity of $O(1)$ per camera node.

4. Correlation-Based Clustering

The images observed by cameras with overlapped FoVs are correlated, thus leading to substantial redundancy in the network traffic. To remove such data redundancy, camera nodes can perform intercamera coding with each other by allowing a camera node to encode its image based on the reference of the image from another camera node, so that the coding rate could be reduced. The clustering strategy has been proved to be an effective way to improve network scalability and energy efficiency for sensor networks [12]. This strategy uses the hierarchical concept where the entire network is divided into regions. In many existing algorithms, the metrics for clustering are distance between nodes or node residual energy [13]. In this paper, we aim to construct clusters based on the correlation of cameras so as to minimize the redundancy of network traffic. We divide the entire network into different clustered regions. Each region corresponds to a cluster, in which a group of camera sensors collaboratively perform data processing.

4.1. Network Structure. In WMSNs, the multitier network architecture is recommended [1]. For example, many WMSN testbeds, such as the SensEye [14], adopt multitiered network topology, and it has been shown to improve energy efficiency. In this paper, we further investigate the collaborative visual information compression and transmission scheme in heterogeneous WMSNs. Two types of sensor nodes exist in our network: camera sensor nodes and common sensor nodes.

Figure 3: An example of the proposed network clustering architecture.

Camera nodes cluster based on the correlation, and common sensor nodes cluster around each camera node based on shorter distance. So there are two kinds of cluster structure in the network. The network model is illustrated in Figure 3. To construct this multitiered network topology, the following assumptions are made in first.

(1) The overall monitoring area is divided into some key regions, and more than one camera nodes are deployed to monitor a field of interest. The spatially proximal cameras could have highly overlapped FoVs, namely, some camera nodes would have correlation.

(2) There are lots of common sensor nodes randomly deployed around every camera sensor node. And the density of the common sensor nodes is sufficient for guaranteeing that every camera node in the wireless communication link connectivity within the adjacent region to the common sensor nodes is not empty.

(3) All the nodes in the network have a unique ID and the network is time synchronous.

4.2. Clustering Establishment. There are two manners to construct the hierarchical clusters, centralized and distributed. The thinking of the centralized manner is very intuitive. In centralized manner, there are always a central controller. It is sink node in WSNs or WMSNs in general. Sink node provides the full information of the network topology along with the detailed settings (e.g., location, various sensing parameters) for each camera. Sink node based on these information constructs the cluster and broadcasts the result to inform the role of each node.

However, in a large-scale distributed network like WMSN, the centralized operations have limited flexibility and scalability. Moreover, the energy constraint of sensor nodes prohibits network-wide information exchange. So, in this section, we will propose a distributed method that only needs local information exchange to accomplish nodes clustering.

Through network structure proposed in Section 4.1, we know that there are two kinds of clusters structure in the network. One kind is camera nodes cluster. The other kind is common nodes cluster. In the process of establishing the network architecture, we construct the camera node cluster firstly, then build every camera node's common nodes cluster.

In construction of the cluster of camera nodes, one camera node just needs to exchange its sensing parameters, for example, location, tilt angle γ, horizontal offset angles α, vertical angles β, and orientation angle θ to its neighbor camera nodes, which are located in its 1-hop communication range. It is a common assumption that the communication range is at least twice of the sensing range [15, 16], which is d_2 in (2). The details of the camera nodes clustering algorithm are presented in Algorithm 2.

Because one camera node just needs to exchange information with its relevant neighbor camera nodes, for each camera node to construct the cluster, it just needs to calculate the correlation coefficient between it and its neighbor camera nodes instead of all the camera nodes in the whole network. This manner can guarantee that the computational intensity would not increase considerably as the number of cameras goes up.

Once the cluster structure of camera nodes has been established, the next step is to construct the common sensor nodes adjacency cluster around each camera node. In this paper, we call these clusters neighbor clusters. We establish the neighbor cluster of each camera as follows.

(1) Taken a camera node as the center, a certain number of common nodes within this camera node's connective region compose a cluster.

(2) In this cluster, the node with sufficient energy and had best link quality of communicating with station is chosen as cluster head node.

(3) The cluster head broadcasts its ID to the surrounding cluster nodes.

(4) The surrounding nodes in the cluster inform their ID to the cluster head node.

(5) The cluster head node keeps the list of all cluster nodes' ID.

4.3. Network Transmission Scheme. When two kinds of hierarchical clustering structure have been set up already, we discuss the whole network transmission framework in this section.

(1) Based on the network structure proposed in Section 4.1, more than one camera nodes are deployed to monitor a field of interest. The spatially proximal cameras could have highly overlapped FoVs, which is measured by the correlation coefficient. So when the camera nodes cluster established, the common camera nodes send the captured pictures to the camera cluster head node. The camera cluster head node receives and fuses these pictures.

```
For a group of camera nodes C, C = {C_1, C_2, ..., C_N}, ordered by its (x, y) coordinates;
Get the FoV of each camera node F_i, F = {F_1, F_2, ..., F_N};
Initialize i = 1, correlation degree threshold ε, cluster X_i = φ;
while length (C) != 0 do
    X_i = X_i ∪ C_i;
    Check the camera nodes which distance with C_i is less than (d_i2 + d_j2);
    Add these camera nodes to a set Ct;
    for j = 1 to length (Ct) do
        Calculate ρ(C_i, Ct_j) by Algorithm 1;
        if ρ(C_i, Ct_j) ≤ ρ(Ct_j, C_i) && ρ(C_i, Ct_j) ≥ ε || ρ(C_i, Ct_j) ≥ ρ(Ct_j, C_i) && ρ(Ct_j, C_i) ≥ ε then
            if Ct_j ∈ X_other then
                X_other = X_other ∪ Ct_j;
            else
                X_i = X_i ∪ Ct_j;
                C = C − Ct_j;
            end
        end
    end
    if Area(F_k) has the bigest area in X_i then
        C_k becomes the cluster head in X_i;
    end
    i++;
end
return X_i = {X_1, X_2, ..., X_m}.
```

Algorithm 2: Distributed camera nodes clustering.

(2) For energy conservation, the camera cluster head node has no responsibility for compressing and sending the fused pictures to the sink or station. The camera cluster head node just sends the fused picture to the cluster head node of its neighbor cluster.

(3) The common cluster head node receives the picture from the camera cluster head node, then cooperatively compresses the picture with the member of common nodes cluster.

(4) The common cluster head node sends the compressive picture to the sink or station.

(5) In the camera nodes cluster, when the residual energy is less than a threshold, the cluster head of camera cluster will turn to the next camera node, which has sufficient energy and bigger area.

5. Performance Evaluation

In this section, we design some simulations to evaluate the effectiveness of correlation coefficient and the performance of the image transmission scheme we proposed.

5.1. Analysis of the Correlation Coefficient. To investigate the effect of camera's four parameters to correlation coefficient ρ. First, when the locations, α, β, θ of multiple cameras are different and fixed, we increase the value of tilt angel γ gradually to explore the change of ρ.

Figure 4 illustrates that the comparison of correlation coefficients among three cameras, denoted by i, j, k. We can find that the the correlation coefficient of a pair of camera is

Figure 4: The effects of increasing tilt angle γ on ρ.

an asymmetrical relationship. This is because the area of the corresponding two cameras' FoVs is inequal. Meanwhile, the correlation coefficient ρ of a pair of cameras is independent of other pair of cameras, that is, ρ_{ij} (or ρ_{ji}), ρ_{ik} (or ρ_{ki}), and ρ_{jk} (or ρ_{kj}) are irrelevant with each other, but they all have the same variation tendency. As shown in Figure 4, in the beginning, the correlation coefficient ρ increases as the tilt angles γ of these three camera i, camera j, and camera k increase, but when ρ increases to a peak value, subsequently,

Figure 5: The effects of different clustering threshold on average cluster-size.

it would decrease. This is because correlation coefficient ρ denotes the intersection of two cameras' FoVs, and the tilt angel γ increased means that the FoV of camera increasing. Because the locations of these cameras are fixed, to increase the FoV, the overlap will became bigger. So in the beginning, the correlation coefficient ρ increases. But when the FoV of camera becomes very large, the portion of overlapped FoVs in camera's FoV is decreased, so all the curves in Figure 4 are descended eventually.

Like γ, the parameters α, β, and the height z of camera nodes have similarity property to cameras' FoVs, so these parameters have the same effect to ρ, although their influence is kind of moderate.

In Section 4.2, we use the correlation degree threshold ε as a criterion for camera nodes clustering. Next, we discuss the effects of different clustering thresholds on average cluster-size in different numbers of camera nodes in the network. In a square region $S = 100 \times 100\,\mathrm{m}^2$, we randomly deploy 50, 100, 150 camera nodes, respectively. Based on the uniform distribution model, the camera parameters are setting as follows. The horizontal offset angle is $\alpha \in [30^\circ \quad 40^\circ]$. The vertical offset angle is $\beta \in [25^\circ \quad 30^\circ]$. The tilt angle is $\gamma \in [40^\circ \quad 50^\circ]$. The sensing orientation $\vec{D}$, namely, θ, is taken from -180° to 180°, and the 3D space height z of camera nodes is taken from 5 to 8. After repeating the experiment for lots of times to take the average value, Figure 5 illustrates that the variety curves of the average cluster-size followed with the different correlation coefficient thresholds ε in three different node densities.

As shown in Figure 5, the higher correlation degree thresholds obviously restrict node memberships and decrease the number of camera nodes per cluster, while lower thresholds increase the area covered by a given cluster at the cost of complexity of coordination. It should be noted that the situation that signal camera node as a cluster by itself is quite common, even in a very densely deployed network in low correlation degree threshold. This is why in Section 4.1, we assume that every camera node must have a common node cluster around.

5.2. Performance of Network Scheme. We set up a scene to evaluate the network transmission scheme proposed in Section 4.1 as follows. Three cameras are deployed to surveillance an area of interest, for example, a crossing, lots of animals would often pass by. The FoVs of these three cameras are overlapped and by using the Algorithm 2 proposed in Section 4.2. These cameras are in a cluster, which cluster head is camera 2. These three cameras periodicity take photos from their FoVs. At a time, three photos are captured by these cameras. The captured photos are shown in Figure 6. Once the photos have been taken, based on our proposed network scheme, camera 1 and camera 3 would send the pictures to the cluster head node, camera 2, respectively. The cluster head node camera 2 will use stitching technique [17] to roughly fuse the photos, which are received from camera 1, camera 3 and captured by itself. The fused photo by camera 2 is shown in Figure 7. When the photo has fused, the camera 2 will send this photo to the cluster head of its neighbor cluster. This neighbor cluster takes the responsibility to compress and send compressive photo to the sink or station.

Based on this scene, we evaluate the performance of the proposed network transmission scheme in Section 4.3 with comparing without clustering and fusing scheme by energy consumption for communication to transmit these three photos. In Figure 6, the size of photo captured by each camera is $384 \times 512 \times 8$ bit. and in Figure 7, the size of fused photo by camera 2 is $652 \times 397 \times 8$ bit. Without considering the neighbor cluster would compress the photos and reduce the data. According to the energy model for communications in [18], in two schemes, the comparison of whole energy consumption for communication in the network is shown in Figure 8.

From Figure 8, we can see that the whole energy consumption for communication to transmit these three photos by our proposed scheme is obvious less than without clustering scheme. This is very significative in the energy-constrained WMSNs.

Camera node is the key node in the network image transmission in WMSNs. We analyze the energy consumption for communication of the camera node by the following three schemes.

(1) Scheme (A): the camera node directly sends captured photos to the sink or station.

(2) Scheme (B): the camera node sends the captured photos to the neighbor cluster head node, and the neighbor cluster takes the responsibility to compress and send compressive photo to the sink or station.

(3) Scheme (C): the camera node uses the proposed scheme in this paper, to fuse the captured photos and send the fused photos to the neighbor cluster head node.

(a) (b) (c)

Figure 6: Photos captured by distributed three cameras. (a) Camera 1. (b) Camera 2. (c) Camera 3.

Figure 7: Photo fused by the camera cluster head node.

Figure 8: Energy consumption for communication in the whole network.

The comparison of energy consumption for communication of camera node in these three schemes is shown in Figure 9. Relative to Scheme (A), in Scheme (B) and Scheme (C), the camera node does not need to send data

Figure 9: Energy consumption for communication of the camera node.

to the sink or station and abundant energy is saved. In the situation of Scheme (B) and (C), the energy consumption for communication of camera node only is the communication with the neighbor cluster head node. It is unrelated to the distance between camera node and sink or station. This is why the curves for Scheme (B) and Scheme (C) are almost constant in Figure 9. Meanwhile, in Scheme (C), the camera node needs to send the fused photos to the neighbor cluster head node, so its energy consumption for communication is more than camera node in Scheme (B). Although energy consumption of few camera nodes is increased in Scheme (C), the energy consumption in the whole network is greatly reduced. From this point, Scheme (C), the proposed scheme in this paper, is more suitable for image compression and transmission in WMSNs.

6. Conclusions and Future Work

In this paper, the problem of the correlation between camera nodes is investigated. Based on the 3D sensing model of the camera sensor nodes, this problem is reduced to calculate the relative position between camera nodes, and a low complexity and effective distribute method is proposed. Then, using the proposed correlation coefficient, an image fusion and transmission scheme is designed through building network hierarchy structure. A set of experiments are performed to discuss the parameters in 3D camera nodes sensing model influence on the correlation among camera nodes and analyze the perform of the devised network scheme. Simulations demonstrate that the proposed network topology and image fusion and transmission scheme reduce the network energy consumption of communication of the whole network efficiently and save the energy of camera sensor nodes.

Different from the conventional scalar WSNs, the energy consumption of data processing is very important to camera and common sensor nodes in WMSNs. So in our future work, we would consider talking about the energy consumption of data processing to design optimum image processing method to accord with the characteristic of WMSNs.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant no. 61003236 and 61171053; the Doctoral Fund of Ministry of Education of China under Grant no. 20113223110002 and 20103223120007; the Natural Science Foundation of Jiangsu under Grant no. BK2011755; the Natural Science Major Program for Colleges and Universities in Jiangsu Province under Grant no. 11KJA520001; the Science & Technology Innovation Fund for Higher Education Institutions of Jiangsu Province under Grant no. CXZZ12_0480; the Priority Academic Program Development of Jiangsu Higher Education Institutions (Information and Communication).

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.

[2] B. Bhanu, C. V. Ravishankar, A. K. Roy-Chowdhury, H. Aghajan, and D. Terzopoulos, *Distributed Video Sensor Networks*, Springer, London, UK, 2011.

[3] H. Ma and Y. Liu, "Correlation based video processing in video sensor networks," in *Proceedings of the International Conference on Wireless Networks, Communications and Mobile Computing (WiCOM '05)*, vol. 2, pp. 987–992, Wuhan, China, June 2005.

[4] R. Dai and I. F. Akyildiz, "A spatial correlation model for visual information in wireless multimedia sensor networks," *IEEE Transactions on Multimedia*, vol. 11, no. 6, pp. 1148–1159, 2009.

[5] P. Wang, R. Dai, and I. F. Akyildiz, "Collaborative data compression using clustered source coding for wireless multimedia sensor networks," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 388–401, 2011.

[6] M. Alaei and J. M. Barcelo-Ordinas, "Node clustering based on overlapping FoVs for wireless multimedia sensor networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '10)*, Sydney, Australia, April 2010.

[7] M. Alaei and J. M. Barcelo-Ordinas, "MCM: Multi-Cluster-Membership approach for FoV-based cluster formation in Wireless Multimedia Sensor Networks," in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference (IWCMC '10)*, pp. 1161–1165, Caen, France, July 2010.

[8] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2nd edition, 2011.

[9] H. Ma, X. Zhang, and A. Ming, "A coverage-enhancing method for 3D directional sensor networks," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 2791–2795, Rio de Janerio, Brazil, April 2009.

[10] F. Xiao, L. Sun, R. Wang, and J. Weng, "Coverage-enhancing algorithm for wireless multi-media sensor networks based on threedimensional perception," *Acta Electronic Sinica*, vol. 40, no. 1, pp. 11–18, 2012.

[11] C. Han, L. Sun, F. Xiao, J. Guo, and R. Wang, "FoVs correlation model for wireless multimedia sensor networks based on threedimensional perception," *Advances in Information Sciences and Service Sciences*, vol. 4, no. 3, pp. 27–34, 2012.

[12] 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 '33)*, pp. 3005–3014, Maui, Hawai, USA, January 2000.

[13] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826–2841, 2007.

[14] P. Kulkarni, D. Ganesan, P. Shenoy, and Q. Lu, "SensEye: a multi-tier camera sensor network," in *Proceedings of the 13th annual ACM International Conference on Multimedia*, pp. 229–238, Hilton, Singapore, 2005.

[15] 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, Los Angeles, Calif, USA, November 2003.

[16] C. Liu and G. Cao, "Distributed critical location coverage in wireless sensor networks with lifetime constraint," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM '12)*, pp. 1314–1322, Orlande, Fla, USA, March 2012.

[17] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[18] W. B. Heinzelman, *Application-specific protocol architectures for wireless networks [Ph.D. thesis]*, Massachusetts Institute of Technology, 2000.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 746501, 11 pages
doi:10.1155/2012/746501

Research Article

A Stochastic k-Coverage Scheduling Algorithm in Wireless Sensor Networks

Jiguo Yu,[1,2] Shaohua Ren,[1] Shengli Wan,[1] Dongxiao Yu,[3] and Guanghui Wang[4]

[1] *School of Computer Science, Qufu Normal University, Rizhao 276826, China*
[2] *Key Laboratory for Intelligent Control Technique of Shandong Province, Qufu Normal University, Rizhao 276826, China*
[3] *Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong*
[4] *School of Mathematics, Shandong University, Shandong, Jinan 250100, China*

Correspondence should be addressed to Jiguo Yu, jiguoyu@sina.com

Received 20 July 2012; Accepted 19 September 2012

Academic Editor: Limin Sun

Copyright © 2012 Jiguo 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.

Coverage is one of the key issues to achieve energy efficiency of a wireless sensor network. Sensor scheduling is one of the most important methods to solve coverage problems. It can ensure the coverage degree of a region and prolong the network lifetime. In this paper, we focus on the k-coverage scheduling problem to guarantee k-coverage sensing and network connectivity. We consider both deterministic and stochastic sensing models of the sensors and adapt the results of deterministic sensing model to solve the sensor scheduling problem under the stochastic sensing model. We use regular pentagons to divide the sensing disks to solve the k-coverage problem. Each sensor node runs a stochastic k-coverage-preserving scheduling algorithm to determine its work modes, and redundant nodes can enter into sleep mode, while active nodes ensure the coverage of the network. Theoretical analysis and simulation results show that our algorithm can reduce the number of active nodes and extend the network lifetime significantly while maintaining a given coverage degree.

1. Introduction

Wireless sensor networks (WSNs) are composed of a large number of sensor nodes, which are densely deployed in a given region. All nodes collaborate to execute sensing and monitor tasks and to send sensed data to sinks. It has so far been employed in military activities, target acquisition, environmental activities, and civil engineering. On the one hand, each sensor is equipped with a limited power source, and it is impossible to replenish power resources in most applications. On the other hand, many applications require a durable lifetime. Thus, a major constraint for WSNs to be widely used is network lifetime.

Since wireless sensor networks are characterized by high density and limited energy. It is not necessary to have all sensor nodes operate in active mode simultaneously. Sensor scheduling, the most effective method to solve coverage problems, makes redundant nodes into sleep mode, in which energy consumption is lower, while active nodes meet specialized requirements. It can decrease the number of active nodes, thus avoiding the channel collision, reducing the network energy consumption, and prolonging the network lifetime substantially. However, most of the existing results on k-coverage are based on the deterministic sensing model, where a point in a region is guaranteed to be covered by k sensors, that is, the point is within the sensing ranges of those k sensors.

In this paper, we consider the k-coverage sensor scheduling problem. A more realistic sensing model, called stochastic sensing model, was considered. Under the stochastic sensing model, a point is covered by a sensor with some probability. Firstly, we solve the sensor scheduling problem of k-coverage under the deterministic sensing model. Then, we utilize the results to solve the problem under the stochastic sensing model. Finally, we present a stochastic k-coverage-preserving sensor scheduling algorithm to achieve lifetime extension. We use regular pentagon instead of Reuleaux Triangle to achieve k-coverage and give the minimum number of sensors to fully k-cover a region while ensuring the connectivity. Theoretical analysis and comprehensive simulation show that our approach is effective.

The rest of the paper is organized as follows. Section 2 discusses the related works of this paper. Section 3 introduces the network models and definitions. Section 4 gives a detailed description of our algorithm. Section 5 presents simulation results of our algorithm. Conclusion is given in Section 6.

2. Related Work

In [1, 2], Zorbas et al. divided sensor nodes into different cover sets, where the sensors in each set are capable of monitoring all targets in the region. By activating one cover set at a time, the sensors network lifetime can be extended. Erlebach et al. also used sensors set partition to solve events detection problem, and proposed an approximation algorithm to maximize the lifetime of WSNs [3]. In [4], Ammari used Reuleaux Triangle to solve the sensor scheduling problem for k-coverage under the deterministic sensing model, where each node runs a k-coverage candidacy algorithm and gives a distributed sleep-wakeup scheduling protocol for stochastic k-coverage (SCPk) of a region. In many applications, critical areas and common areas must be distinguished adequately, and it is more practical and useful to monitor critical areas rather than common areas, or the available budget cannot provide enough sensors to fully cover the sensor field. In [5], Ammari and Das improved the former method to decrease the bound of sensor spatial density and proved that when the ratio of the radius R of the communication range to the radius r of the sensing range of sensors be at least equal to one, the generated network is connected. In [6], Ke et al. proposed an approximation algorithm for critical-square-grid coverage. In [7], Xing et al. presented a novel protocol that can dynamically configure a network to achieve guaranteed degrees of coverage and connectivity, integrated CCP with SPAN [8] to provide both coverage and connectivity guarantees, proposed a probabilistic coverage model, and extended CCP to provide probabilistic coverage guarantees.

In [9], n nodes were placed over a unit area in an unreliable wireless sensor grid network, Shakkottai et al. proposed an algorithm that can meet coverage requirements and connectivity. Chakrabarty et al. presented a novel grid coverage strategy for effective surveillance and target location in distributed sensor networks in [10]. They gave a coding-theoretic bounds on the number of sensors and presented methods for determining their placement in the interesting region. And the result showed that grid-based sensors placement for single target provides asymptotically complete location of multiple targets in the grid. The exposure in WSNs, which is related to the quality of coverage provided by WSNs, was studied in [11] based on a general sensing model, where the sensing signal of a sensor at an arbitrary point was measured by a function that is inversely proportional to the distance between the sensor and the point. Based on the concept of connected dominating set (CDS), a distributed approach for active sensors selection in [12] was proposed to fully cover a region. This approach is based on a probabilistic sensing model, where the probability of the existence of a target is defined by an exponential function that represents the confidence level of the received sensing signal. In [13], Yang et al. selected the minimum number of sensor nodes to fulfill k-coverage and proposed a centralized solution based on integer linear programming and a distributed algorithm called PKA, which uses Pruning methods to reduce the number of active nodes.

A joint scheduling scheme based on a randomized algorithm for providing statistical sensing coverage and guaranteed network connectivity was presented in [14]. This scheme works without the availability of per-node location information, to schedule sensor nodes for energy saving and meet constraints, both sensing area coverage and network connectivity without accurate location information. Mao et al. proposed an energy efficient connected coverage protocol (EECCP) based on the analytical results and the existing algorithms, which constructs the connected dominating set in [15]. Extensive simulation studies show that the protocol can effectively maintain both high quality sensing coverage and connectivity for a long time.

Many applications do not require complete coverage all the time. For such applications, one effective method to save energy and prolong network lifetime is to partially cover the area. In [16], Zorbas et al. analyzed the problem of connected partial target coverage, where cover sets are allowed to monitor a subset of the targets at any point in time, while connectivity with the base station is retained. The simulation results show that monitoring 90% of the targets may yield twice the network lifetime provided by a full coverage approach. In [17], Li et al. proposed a framework that can transform almost any existing complete coverage algorithm to a partial coverage one with any coverage ratio by running a complete coverage algorithm to find full coverage sets with virtual radii and converting the coverage sets to partial coverage sets via adjusting sensing radii. The framework can preserve the characteristics of the original algorithms and the conversion process has low time complexity.

In [18–20], coverage problems in heterogeneous WSNs were considered. In [18], Jin et al. proposed a location-independent, energy-efficient routing algorithm EECCR, which simultaneously preserves the network m-coverage ratio and the sensor n-connectivity probability. In [19], Khedr and Osamy introduced two new efficient distributed algorithms for finding the minimum number of connected perimeter sensor nodes that are sufficient to cover the perimeter of queried region, and the algorithm can preserve full perimeter coverage. In [20], the relay nodes were deployed to provide fault tolerance with higher network connectivity in heterogeneous wireless sensor networks, where sensor nodes possess different transmission radii.

In [21], El Salti and Nasser introduced a new approach for obtaining a fully covered network such that every single point in a region is fully covered by at least one sensor node. This approach is referred to as the chipset coverage model and algorithm.

In this paper, we use regular pentagon to solve the k-coverage problem under deterministic and stochastic sensing models. We propose a stochastic k-coverage sensor scheduling algorithm, theoretical analysis and simulation results show the correctness and superiority of our algorithm.

3. Network Model and Definition

3.1. Network Model. In this section, we present some reasonable assumptions and introduce both deterministic sensing model and stochastic sensing model. All sensors and the sink are static and aware of their locations via a localization technique [22].

All sensors are randomly and nonuniformly deployed in a planar region. The sensors are homogeneous and have determined sensing and communication ranges of radii r and R, respectively. The location of a sensor s in the region is denoted by ξ_s.

Deterministic sensing model is also known as perfect sensing model, where a point ξ in a region is covered by a sensor s based on the Euclidean distance $\delta(\xi, s)$ between ξ and s. This model considers the sensing range of a sensor as a disk, and hence sensing ability of all sensors is determined and has no uncertainty.

Stochastic sensing model considers the signal attenuation and the presence of noise. It is necessary to consider a more realistic sensing model by defining the coverage $\mathrm{Cov}(\xi, s_i)$ of a point ξ by s_i using some probability function. Here, in our stochastic sensing model, the coverage $\mathrm{Cov}(\xi, s_i)$ is defined as the probability of detection $p(\xi, s_i)$ of an event at ξ by s_i as follows.

$$p(\xi, s_i) = \begin{cases} e^{-\beta\delta(\xi,s_i)^{\alpha}}, & \delta(\xi, s_i) \le r \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where β represents the physical characteristic of sensing units of sensors and $2 \le \alpha \le 4$ is the path-loss exponent. The uncertainty sensing capability of a sonar sensor is modeled by a Gaussian probability density function $p(r \mid z, \theta) = (1/2\pi\sigma_r\sigma_\theta)\exp[-(1/2)((r-z)^2/\sigma_r^2+\theta^2/\sigma_\theta^2)]$ [4]. Where r and θ represent sensing range and angel, respectively.

3.2. Definitions

Definition 1 (k-coverage). Under the deterministic sensing model, a point ξ in a region is said to be k-covered if it belongs to the intersection of the sensing disks of at least k sensors. Under the stochastic sensing model, a point ξ in a region is said to be probabilistically k-covered if the detection probability of an event occurring at ξ by at least k sensors is at least equal to some threshold probability of p_{th} $(0 < p_{\mathrm{th}} < 1)$. For both sensing models, a region is said to be k-covered if every point ξ in the region is k-covered.

Definition 2 (reuleaux triangle). Construct an equilateral triangle. On each vertex, center a compass and draw the minor arc between the other two vertices. The perimeter will be three nonconcentric; arcs, this is a Reuleaux Triangle. The minimum overlap area of two adjacent slices forms a lens.

As shown in Figure 1, the gray area is a Reuleaux Triangle and $becf$ is a lens.

Definition 3 (central area). Construct a regular pentagon, on each vertex, center a compass, and draw the arc between the other two adjacent vertices.

The five nonconcentric arcs form the central area of the regular pentagon, as shown in Figure 2.

Definition 4 (width of regular pentagon). The width of regular pentagon is the maximum of the distances between any pair of parallel lines that bound it.

Definition 5 (zero direction). We take the position of a sensor s as the datum point when dividing the sensing disk and then give another point to determine a line segment. The direction of the line segment is zero direction. We use zero direction to determine the position of regular pentagon vertices.

As shown in Figure 3, take position o as the datum point, if the zero direction is (o, p_1), the angle of $\angle p_4 o p_1 = 90^\circ$. Similarly, when we take the other points p_2, p_3, or p_4, the angles between zero direction and x axis equal 60°, 30°, and 0°, respectively.

4. A Stochastic k-Coverage Sensor Scheduling Algorithm

In this section, we analyze the k-coverage problem in WSNs using the deterministic sensing model and give the minimum required number of nodes to achieve k-coverage. Then, we utilize this result to the stochastic sensing model, get the stochastic sensing range r_s of the sensors that is necessary to probabilistically k-cover a region, and present the stochastic k-coverage-preserving sensor scheduling algorithm

4.1. The Deterministic Sensing Model. In order to characterize k-coverage, we need to compute the maximum size of an area, which is k-covered with exactly k sensors. To this end, we introduce a fundamental result of convex set theory, known as Helly's theorem [4], which characterizes the intersection of convex sets. We exploit this theorem in a regular pentagon to characterize k-coverage of sensing disks. Then, we compute the minimum number of sensors required for a k-coverage.

Theorem 6 (see Helly [4]). *Let E be a family of convex sets in R^n such that any m members $(m \ge n+1)$ of E have a nonempty intersection. Then, the intersection of all members of E is nonempty.*

The following Lemma is an instance of Helly's Theorem in two-dimensional space that characterizes the intersection of k sensing disks.

Lemma 7. *Let $D_1, D_2, \ldots, D_k$ $(k \ge 3)$ be disks in the plane R^2. If the intersection of any three disks of the k sensing disks is not empty, then the intersection of the k sensing disks is not empty.*

By Lemma 7, the following Lemma 8 gives a sufficient condition for k-coverage of a region.

Lemma 8. *A regular pentagon is k-covered if the central area of the regular pentagon contains at least k active sensors $(k \ge 3)$.*

Figure 1: A Reuleaux Triangle and One lens.

Figure 2: Central area of a regular pentagon.

Figure 3: The zero direction.

Proof. According to Definition 3, the maximum distance among those k sensor nodes and the boundary of the regular pentagon is r, as shown in Figure 2. For the central area lies in the inside of the regular pentagon, its width is smaller than r. Thus, the intersection of any three of those k sensing disks is not empty. We can know that the regular pentagon is k-covered by Lemma 7. □

We can get the following observation. When we use regular pentagons to cover a sensing disk, if any central area of regular pentagons contains at least k active sensors, the sensing disk is k-covered. If every sensing disk in the region is k-covered, then the whole region is k-covered.

Theorem 9. *Let r be the radius of sensing disks of sensors and $k \geq 3$. The minimum number of sensors $n(r,k)$ to k-cover a region with area F is given by*

$$n(r,k) = \left\lceil \frac{4Fk}{\sqrt{25+10\sqrt{5}}r^2} \right\rceil. \tag{2}$$

Proof. We use regular pentagons to cover a sensing disk, the area formula of the regular pentagon is $S = (\sqrt{25+10\sqrt{5}}/4)r^2$, we can get that the minimum area those k nodes can cover is S. Thus, the minimum number of sensors to k-cover a region of area F is $n(r,k) = \lceil F/S \rceil = \lceil 4Fk/\sqrt{25+10\sqrt{5}}r^2 \rceil$. □

Theorem 10. *A k-covered WSN is connected if $R \geq \sqrt{3}r$.*

Proof. As shown in Figure 4, the distance between point a and b is the maximum distance of sensor nodes in two adjacent regular pentagons. Thus, $\delta(a,b) = 2r\cos 30^\circ = \sqrt{3}r$. Therefore, as long as the communication radius is not less than $\sqrt{3}r$, the sensor nodes in the two regular pentagons can communicate with each other, and the network is connected. □

4.2. The Stochastic Sensing Model. In this section, we utilize the results of Section 4.1 to characterize probabilistic k-coverage. Theorem 11 computes the minimum k-coverage probability $p_{k,\min}$.

Theorem 11. *Let r be the radius of the determined sensing range of the sensors and $k \geq 3$. The minimum k-coverage probability such that each point in a region is probabilistically k-covered by at least k sensors under the stochastic sensing model is given by*

$$p_{k,\min} = 1 - \left(1 - e^{-\beta r^{\alpha}}\right)^k. \tag{3}$$

Proof. We should deploy k sensors in the central area to k-cover a regular pentagon. The sensing ability of sensors

Figure 4: Communication radius and sensing radius.

attenuates with the increase of the distance. The farthest point away from the sensors is the five vertices of the regular pentagon (see Figure 2). Thus, the vertices have the smallest probability to be k-covered by the sensors in central area. When all sensors are deployed on the arc of the curved pentagon, the distance between sensors and corresponding vertex equals r. Thus, the minimum k-coverage probability for the least k-covered point ξ by k sensors under the stochastic sensing model is given by $p_{k,\min} = 1 - \prod_{i=1}^{k}(1 - p(\xi, s_i)) = 1 - (1 - e^{-\beta r^{\alpha}})^{k}$. □

The stochastic k-coverage problem is to select a minimum subset $S_{\min} \subseteq S$ of sensors such that each point in a region is k-covered by at least k sensors and the minimum k-coverage probability of each point is at least equal to some given threshold probability p_{th}, where $0 < p_{\text{th}} < 1$. This helps us to compute the stochastic sensing range r_s, which provides probabilistic k-coverage of a region with a probability no less than p_{th}. The following lemma computes the value of r_s.

Lemma 12. *Let $k \geq 3$ and $2 \leq \alpha \leq 4$. The stochastic sensing range r_s of the sensors that is necessary to probabilistically k-cover a region with a minimum number of sensors and with a probability no lower than p_{th} is given by*

$$r_s = \left(-\frac{1}{\beta}\ln\left(1 - (1 - p_{th})^{1/k}\right)\right)^{1/\alpha}, \tag{4}$$

where β represents the physical characteristic of sensing units of sensors.

Proof. The proof is simple. Let $p_{k,\min} \geq p_{\text{th}}$, we can easily get $r_s \leq (-(1/\beta)\ln(1 - (1 - p_{\text{th}})^{1/k}))^{1/\alpha}$. □

The following theorem states a sufficient condition for probabilistic k-coverage based on our stochastic sensing model, p_{th} and k.

Theorem 13. *Let $k \geq 3$. We use regular pentagons to divide all the sensing disks in the region. The length of sides of regular pentagon is r_s. If any central area of regular pentagons contains at least k active sensors, then the region is stochastic k-covered with a probability no lower than $p_{th}(0 < p_{th} < 1)$.*

The following Lemma states a sufficient condition for connectivity among sensors under our stochastic sensing model as follows.

Lemma 14. *Let $k \geq 3$. The sensors that are selected to k-cover a region with a probability no less than p_{th} under the stochastic sensing model are connected if the radius of their communication range is at least equal to $\sqrt{3}r_s$, where r_s is the stochastic sensing range of the sensors.*

4.3. ISCPk-An Improved Stochastic k-Coverage-Preserving Sensor Scheduling Algorithm. In this section, we present the stochastic k-coverage-preserving sensor scheduling algorithm. It is a distributed sleep-wakeup sensor scheduling algorithm. One can gain stochastic k-coverage and connectivity by executing this algorithm. Suppose N sensors are randomly and nonuniformly deployed in a planar region. The sensor network has the following properties.

(1) Stochastic sensing model considers the signal attenuation and noise.

(2) Sensors cannot move after they are deployed, and battery energy cannot be recharged.

(3) Each node has a globally unique id.

(4) R and r represent communication and sensing ranges, respectively, ($R \geq \sqrt{3}r_s$).

(5) In sleep mode, sensor node will turn off its sensing and communication functions.

The algorithm can be described as follows.

Step 1. Compute stochastic sensing range r_s and the list of the neighbor-sensor of each node.

According to α, β, k, and p_{th}, calculate r_s. Each node maintains a neighbor sensor list, a work set count flag WSC, and a redundancy flag RF. Initially, RF = 0 and WSC = 1.

Step 2 (select redundant nodes). (1) Provide a zero direction, each node divides its sensing disk into three 108° and one 36° sector regions. Take the two radii of each sector region with 108° as two sides of a regular pentagon we can get three regular pentagons. The fourth regular pentagon is as shown in Figure 5, the dashed line one. We use these four regular pentagons to cover a sensing disk.

(2) If every central area of four regular pentagons contains at least k active sensors with the same value of flag WSC equalling to the central sensor's, then the sensing disk is k-covered.

(3) Decide whether the sensing disk is k-covered according to (2), then determine the node is redundant or not, select out the redundant nodes.

Step 3 (sensor scheduling). (1) Each node with RF = 0 executes Step 2 to determine itself redundant or not by ascending order according to id. Redundant nodes change

Figure 5: Regular pentagons division of a sensing disk.

Figure 6: Initial node distribution.

their RF to 1 and send an activate message to the nodes in the central areas; these activated nodes change their RF to 0, nonredundant nodes keep RF values 0.

(2) If there are still other sensor nodes in the sensing range of redundant nodes except those nodes in four central areas, then redundant nodes send a message to those nodes, change their RF to 2. Thus, they do not need to run Step 2, since they have already been k-covered by the activated nodes.

(3) If a redundant node receives activate message from other nodes, it indicates that the node activation enabling the sensing disk of one sensor is k-covered, then change RF from 1 to 0.

(4) Record the number of redundant nodes as m. Assume the area of the graph in Figure 5 equals A. The algorithm iterate execute Steps 2 and 3 until $m \geq 3\lceil F/A \rceil$, where F is the area of the region.

Figure 7: Redundant and active nodes.

Step 4 (enter into work mode). At the beginning, every node determines their mode according to RF, nodes with RF equal to 1 and 2 sleep and increase WSC by one, while the other nodes work. Those sleeping nodes will be awake after T_{sleep}. And these nodes will modify their RF to 0 and execute Step 3. Here we assume that T_{sleep} is smaller than the period of one sensor working. Since T_{sleep} is smaller, when the active nodes run out of energy, a number of these once sleep nodes have already turned into work mode. Thus, the network will not be affected by nodes failure.

5. Theoretical Analysis

In this section, we evaluate and analyze the performance of the proposed algorithm.

Theorem 15. *The ISCPk algorithm is capable of generating at least one solution, if the solution exists.*

Sketch of Proof. Note that sufficient number of nodes were deployed to ensure k-coverage of the region. During the execution of the algorithm, we take $m \geq 3\lceil F/A \rceil$ as the terminate condition, where $\lceil F/A \rceil$ indicates the optimum result of the algorithm. Three times of the optimum result can ensure full coverage of the region. For the deployed nodes are sufficient, thus, we can get a reasonable solution to k-cover the region. The algorithm is proved to be correct.

Theorem 16. *For ISCPk, the message complexity is $O(N)$ and time complexity is $O(N^2)$, respectively. Where N is the number of sensors.*

Sketch of Proof. Assume that we deployed N sensors in the region, let $\triangle$ be the maximum node degree, that is, the maximum size of neighbor sensors list. Assuming no packet loss, every node broadcasts a message to obtain its

Figure 8: The regular pentagons cover a sensing disk.

Figure 9: The regular pentagons cover the region.

neighbor sensors list in Step 1. If we can get l redundant nodes in Step 2, every redundant node sends an active or redundant message to the neighbor nodes in the sensing range of redundant nodes. The network sends at most $N + l\triangle$ messages in each iteration. Since $l\triangle$ is much smaller than N, thus, the message complexity of ISCPk is $O(N)$. Provided that each operation of the algorithm takes one unit of time. The calculation of r_s and redundant nodes selection takes $O(N)$ and $O(N^2)$ time, respectively. The other operations of the algorithm takes at most $O(N)$ time. Therefore, the time complexity for the proposed distributed algorithm is $O(N^2)$.

Our algorithm and the algorithm in [4] are both to k-cover a sensing disk before k-cover a region firstly, then to k-cover the whole region. The difference between the two algorithms is the way to divide the sensing disk. In [4], a sensor randomly decomposes its sensing disk into six Reuleaux Triangles with width r_s. If the lens of each of the three pairs of adjacent slices contains at least k active sensing

Figure 10: Active nodes 2-cover to the region.

Table 1: Simulation parameters.

Parameter	Value
Region area 1 (deterministic model)	400 m × 400 m
Region area 2 (stochastic model)	60 J
Transmission energy consumption	0.06 J/s
Receive energy consumption	0.012 J/s
Idle energy consumption	0.012 J/s
Sleep energy consumption	0

neighbors, then the sensing disk is k-covered. According to the algorithm in [4], it needs $3k$ nodes to k-cover a sensing disk, while our algorithm needs $4k$ nodes to k-cover a sensing disk. It seems that our algorithm require more nodes to k-cover a region. However, that is not the case.

We consider the node utilization from the coverage area of every k nodes, in [4]; $3k$ nodes can k-cover an area of $S_1 = \pi r^2$. $4k$ nodes can k-cover an area of $S_2 = (\sqrt{25 + 10\sqrt{5}} - 4\sin 18^\circ \cos 54^\circ)r^2$ in our algorithm; it is equal to 4 times the area of a pentagon minus the area of overlap, as shown in Figure 5. We can get that $S_1/3 < S_2/4$.

By the analysis above, we can see that the number of sensors in our algorithm k-covering a single-sensing disk is greater than the method in [4], but the average cover area for each k node is larger than [4].

6. Simulation Results

In this section, we present the simulation results of our algorithm. Simulation parameters are shown in Table 1, the algorithms in this paper and [4] are represented by ISCPk and SCPk, respectively.

6.1. Sample Topology. 900 nodes are randomly and nonuniformly deployed in a planar region of 400 m × 400 m. We

Figure 11: The relation between the degree of angle and the selected nodes.

Figure 12: Sensor spatial density versus k, p_{th} and α.

consider deterministic sensing model, the sensing radius is 70 m, use the deployed nodes to fulfill 2-coverage of the region. We take the angle equals 30°. The initial topology is as shown in Figure 6.

Figure 7 shows the distribution of redundant and active nodes after the execution of our algorithm, where red and blue nodes indicate redundant and active nodes, respectively. The green nodes in Figure 7 are the sensors in the sensing range of redundant nodes. Since the selected active nodes can fully cover the region, the green and black nodes are redundant nodes too, even though they cannot meet the redundant definition which we define. Thus, the number of redundant and active nodes is 134 and 766, respectively.

The regular pentagon division of a redundant node is shown in Figure 8; the red small circle represents the near central area of a regular pentagon.

Figure 9 shows the regular pentagon division of 17 redundant nodes in the region. The green circles represent the sensing range of the redundant nodes (red nodes), red small circles indicate the near central area of regular pentagons, and blue nodes denote the selected active nodes.

Figure 10 gives the 2-coverage of the region, blue nodes are the selected active nodes, and we draw the sensing ranges of the active nodes. We can see from Figure 10 that active nodes can 2-cover the whole region.

During the simulation process, we find that the zero direction has great influence on our experimental results.

Figure 13: Achieved degree of coverage k versus the number of sensors.

Figure 14: The number of active sensors increases with the physical characteristic β of the sensors.

Figure 11 shows the relation between the zero direction and the selected nodes. The abscissa indicates the angle of zero direction. As shown in Figure 11, the selection of zero direction has a great influence on the simulation results. Thus, we change the zero direction in the next iteration to reduce the influence of zero direction to the experimental results.

6.2. Simulation Results on Various Parameters. Figure 12 shows the sensor spatial density versus k, p_{th}, and α. Sensor spatial density equals the ratio of $n(r,k)$ and region area F, that is, $n(r,k)/F = \lceil 4k/\sqrt{25+10\sqrt{5}r_s^2}\rceil$, where $r_s = (-(1/\beta)\ln(1-(1-p_{th})^{1/k}))^{1/\alpha}$ and $\beta = 0.02$. Theoretical analysis shows that when k, p_{th}, and α increase, the number of nodes which meet threshold p_{th} increases too, so we need more nodes to achieve k-coverage. As shown in Figure 12, we find a good matching between the simulation and theoretical analysis.

The achieved degree of coverage k versus the total number of deployed sensors is shown in Figure 13.

We can get the relation between k and $n(r_s,k)$ through Theorem 9. As shown in Figure 13, the number of nodes increases as k, but our algorithm needs smaller number of nodes than that of [4], when they have the same coverage degree k.

Figure 14 shows that the number of active sensors required to provide 3-coverage increases with the physical

Figure 15: The impact of p_{th} on the network lifetime to provide 3-coverage.

characteristic β of the sensors. The number of active sensors increases with β, but our algorithm needs smaller number of active nodes to achieve 3-cover than the algorithm in [4].

Figure 15 shows the impact of p_{th} on the network lifetime to fulfill 3-coverage. As mentioned earlier, higher values of p_{th} require larger number of active sensors, and hence the network consume much energy. As shown in Figure 15, we deploy 30000 nodes in the region, the total initial energy is 1.8×10^6 J. As time progresses, the residual energy of the network gradually reduces, but our algorithm can save network energy consumption effectively.

7. Conclusion

In this paper, a distributed approach for sensor scheduling in stochastic k-covered WSNs is proposed. First, we characterize k-coverage in WSNs and provide a sufficient condition for k-coverage with a minimum number of sensors. Then, we present our stochastic k-coverage-preserving node scheduling algorithm. Theoretical analysis and simulation results indicate the correctness and superiority of the algorithm. The algorithm can k-cover a larger area than that of [4] while using the same number of nodes, and it can reduce network energy consumption and prolong the network lifetime effectively. We will extend our algorithm to three-dimensional and heterogeneous WSNs in the future.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Contract nos. 60373012, 11101243), the Natural Science Foundation of Shandong Province (Contract nos. ZR2012FM023, ZR2009GM009, and ZR2009AM013), the STPU of Shandong Province (Contract nos. J10LG09, J09LG34), and the Promotional Foundation for Middle-aged or Young Scientists of Shandong Province for contract (Contract nos. BS2009DX024, BS2010DX013).

References

[1] D. Zorbas, D. Glynos, P. Kotzanikolaou, and C. Douligeris, "BGOP: an adaptive coverage algorithm for wireless sensor networks," in *Proccedings of the 13th European Wireless Conference*, Paris, France, 2007.

[2] D. Zorbas, D. Glynos, P. Kotzanikolaou, and C. Douligeris, "Solving coverage problems in wireless sensor networks using cover sets," *Ad Hoc Networks*, vol. 8, no. 4, pp. 400–415, 2010.

[3] T. Erlebach, T. Grant, and F. Kammer, "Maximising lifetime for fault-tolerant target coverage in sensor networks," in *Proceedings of the 23rd ACM Symposium on Parallelism in Algorithms and Architectures, (SPAA '11)*, pp. 187–196, June 2011.

[4] H. M. Ammari, "Stochastic k-coverage in wireless sensor networks," in *Proceedings of the Wireless Algorithms Systems and Applications*, vol. 5682, pp. 125–134, 2009.

[5] H. M. Ammari and S. K. Das, "Centralized and clustered k-coverage Protocols for Wireless Sensor Networks," *IEEE Transactions on Computers*, vol. 61, pp. 118–133, 2012.

[6] W. C. Ke, B. H. Liu, and M. J. Tsai, "Efficient algorithm for constructing minimum size wireless sensor networks to fully cover critical square grids," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1154–1164, 2011.

[7] G. Xing, X. Wang, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration for energy conservation in sensor networks," *ACM Transactions on Sensor Networks*, vol. 1, pp. 36–72, 2005.

[8] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," *Wireless Networks*, vol. 8, no. 5, pp. 481–494, 2002.

[9] S. Shakkottai, R. Srikant, and N. Shroff, "Unreliable sensor grids: coverage, connectivity and diameter," in *Proceedings of*

the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '03), pp. 1073–1083, April 2003.

[10] K. Chakrabarty, S. S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks," *IEEE Transactions on Computers*, vol. 51, no. 12, pp. 1448–1453, 2002.

[11] S. Meguerdichian, F. Koushanfar, G. Qu, and M. Potkonjak, "Exposure in wireless ad-hoc sensor networks," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom '01)*, pp. 139–150, July 2001.

[12] Y. Zou and K. Chakrabarty, "A distributed coverage- and connectivity-centric technique for selecting active nodes in wireless sensor networks," *IEEE Transactions on Computers*, vol. 54, no. 8, pp. 978–991, 2005.

[13] S. Yang, F. Dai, M. Cardei, J. Wu, and F. Patterson, "On connected multiple point coverage in wireless sensor networks," *International Journal of Wireless Information Networks*, vol. 13, no. 4, pp. 289–301, 2006.

[14] C. Liu, K. Wu, Y. Xiao, and B. Sun, "Random coverage with guaranteed connectivity: joint scheduling for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 6, pp. 562–575, 2006.

[15] Y. Mao, Z. Xu, and Y. Liang, "An energy efficient connected coverage protocol in wireless sensor networks," *Lecture Notes in Computer Science*, vol. 4505, pp. 382–394, 2007.

[16] D. Zorbas, D. Glynos, and C. Douligeris, "Connected partial target coverage and network lifetime in wireless sensor networks," in *Proceedings of the 2nd IFIP Wireless Days, (WD '09)*, fra, December 2009.

[17] Y. Li, C. Vu, C. Ai, G. Chen, and Y. Zhao, "Transforming complete coverage algorithms to partial coverage algorithms for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 4, pp. 695–703, 2011.

[18] Y. Jin, L. Wang, J. Y. Jo, Y. Kim, M. Yang, and Y. Jiang, "EECCR: an energy-efficient m-Coverage and n-Connectivity routing algorithm under border effects in heterogeneous sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1429–1442, 2009.

[19] A. M. Khedr and W. Osamy, "Minimum perimeter coverage of query regions in a heterogeneous wireless sensor network," *Information Sciences*, vol. 181, no. 15, pp. 3130–3142, 2011.

[20] X. Han, X. Cao, E. L. Lloyd, and C. C. Shen, "Fault-tolerant relay node placement in heterogeneous wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 5, pp. 643–656, 2010.

[21] T. El Salti and N. Nasser, "Randomized 3-D routing in fully- and partially-covered sensor networks," *Mobile Networks and Applications*, vol. 15, no. 2, pp. 283–297, 2010.

[22] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low-cost outdoor localization for very small devices," *IEEE Personal Communications*, vol. 7, no. 5, pp. 28–34, 2000.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 345316, 14 pages
doi:10.1155/2012/345316

Research Article

Truthful Relay Assignment for Cooperative Communication in Wireless Networks with Selfish Source-Destination Pairs

Gang Liu,[1,2] Liusheng Huang,[1,2] Yu-e Sun,[3] Hongli Xu,[1,2] He Huang,[1,2] and Xueyong Xu[1,2]

[1] *Department of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China*
[2] *Suzhou Institute for Advanced Study, University of Science and Technology of China, Suzhou 215123, China*
[3] *School of Urban Rail Transportation, Soochow University, Suzhou 215100, China*

Correspondence should be addressed to Gang Liu, enoch@mail.ustc.edu.cn

Received 20 July 2012; Accepted 7 October 2012

Academic Editor: Fu Xiao

Copyright © 2012 Gang 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.

Cooperative communication is shown to be a promising technology to significantly increase the capacity of wireless networks. Due to the competition among multiple source-destination pairs for the same relay node set in the relay assignment problem, each pair may cheat others to achieve a more individual revenue. However, the cheating behavior may decrease the overall performance of the network greatly. Thus, there is a challenge for designing a truthful protocol that maximizes a pair's payoff only when this pair reveals its true individual information. In this paper, we propose a relay assignment protocol (RA-VCG) for cooperative communication to maximize the total social value (i.e., the total true value of all pairs) while guaranteeing truthfulness in an auction-theoretic sense by charging each pair an extra payment. Specially, RA-VCG implements a variation of the well-known VCG mechanism for the truthful relay assignment problem in the network with selfish source-destination pairs. Then, we prove the validity of this protocol and also show several surprising properties (such as no positive transfer and individual rationality) associated with this protocol. The simulation results show that the total social value achieved when each node takes untruthfully is about 23.3% less than that achieved when nodes behave truthfully.

1. Introduction

In recent years, there are growing interests in the idea of employing *cooperative communication (CC)* [1–4] to improve the performance of wireless networks. Spatial diversity has been shown to be very effective in copping fading in wireless channel by employing multiple transceiver antennas on a wireless node (i.e., MIMO). However, it may be infeasible for equipping multiple antennas on a wireless node in practice, as the footprint of multiple antennas may not fit on a wireless node, especially the wireless small-size device [5, 6]. In order to achieve spatial diversity without multiple antennas on the same node, the CC has been proposed in [1] by using distributed single antenna nodes. Under CC, each node is equipped with only a single antenna and can use the antennas on other (relay) nodes to achieve the spatial diversity.

There are two categories of CC modes: *amplify-and-forward* (AF) and *decode-and-forward* (DF) [1]. In AF mode, the relay node only performs a linear operation on the signal received from the source node, and then forwards it to the destination node. In DF mode, the relay node firstly decodes the signal received from the source node, and then forwards the reencoded signal to the destination node. Regardless of AF and DF, the choice of a relay node will affect the performance of CC greatly [3–5]. As shown in Section 3, an improperly relay node may make the performance of CC for a source-destination pair be worse than that of *direct transmission* (DT).

The previous works [2, 4–9] mainly studied the relay assignment problem for CC in the network where multiple source-destination pairs compete for the same pool of relay nodes without any cheating. However, for a source-destination pair, employing different relay nodes will achieve a different performance and the difference is very apparent. Moreover, the physical limitations of many transceivers prohibit them from transmitting data on multiple channels at the same time, so that each relay node can serve for at most one pair. These features promote the competition among all

pairs, where each pair will cheat the others to get a better relay node so that it can achieve a more individual revenue. However, these selfish source-destination pairs may make the total social value (i.e., the total true value of all pairs and the true value of each pair will be defined in Section 3) very low because of cheating.

Next, we illustrate the harm caused by cheating in a network example. As shown in Figure 1, the number in each triangle denotes the true value (which will be defined in Section 3) of CC for the source-destination pair employing the corresponding relay node. For example, if pair $s_1 - d_1$ denoted as t_1 employs relay node r_1 to use CC, the final true value is 10. From Figure 1, we can see that both pair t_1 and t_2 want to choose r_2 as a relay node, because both t_1 and t_2 can achieve the maximum individual true value by employing r_2 as relay node. However, due to the fact that there is only one pair can employ r_2 as relay node, then t_1 and t_2 will cheat each other. Assume that t_1 claims the values for employing relay nodes r_1 and r_2 are 5 and 30, respectively. Then t_2 may claim that the value for employing relay nodes r_2 and r_3 are 36 and 8. Since the objective is to maximize the total social value, t_1 will get r_1 and t_2 will get r_2. As a result, the real total social value is only $18 + 10 = 28$ because of cheating. However, if all pairs claim truthfully, then the maximal total social value is 36. Obviously, the cheating behavior decreases the performance of network.

Thus, there is a challenge for designing a truthful protocol that maximizes a source-destination pair's payoff (utility) only when this pair reveals its true individual value. So, we study the truthful relay assignment problem for CC in wireless networks where multiple selfish source-destination pairs compete for the same pool of relay nodes. Specially, our goal is to design a protocol that maximizes the total social value of all pairs while guaranteeing truthfulness.

The main contributions of this paper are the following.

(i) We point out that there are cheating behaviors among some pairs, and we also give several main cheating possibilities for each pair in relay assignment problem. So, we defined a truthful relay assignment problem in this paper where each source-destination pair is selfish and it can cheat each other to get a more individual revenue.

(ii) We design a protocol for truthful relay assignment problem called RA-VCG to maximize the total social value of all pairs while guaranteeing truthfulness. This protocol implements a variation of the well-known VCG mechanism. In order to guarantee truthfulness, RA-VCG will charge each pair a payment such that a source-destination pair's payoff will be maximized only when this pair reveals its true individual value.

(iii) We also prove that our protocol can achieve the maximum total social value for all source-destination pairs while guaranteeing truthfulness. Moreover, there are several another nice properties (such as no positive transfer and individually rationality) associated with RA-VCG.

Figure 1: Illustration of attractive cheating.

(iv) The simulation results show that the total social value achieved when each node takes untruthful behaviors is about 23.3% less than that achieved when nodes behave truthfully in RA-VCG.

The rest of this paper is organized as follows. We review the literature work in Section 2. Section 3 describes the system model and problem definition. In Section 4, we present RA-VCG protocol in detail. In Section 5, we prove the correctness of this protocol in terms of maximizing total social value and truthfulness. The simulation results are presented in Section 6. Section 7 concludes this paper.

2. Related Work

It is possible for a source-destination pair to employ multiple relay nodes to use CC, but the benefit of this approach is limited. In [3], Zhao et al. showed that it is sufficient to choose the "best" relay node to achieve full diversity order than to employ multiple relay nodes for a pair. Therefore, it paved the way for research on assigning at most one relay node to a pair and we will adopt this setting in this paper. However, this work [3] is limited to a single pair and cannot be easily extended to multiple source-destination pairs' situation.

Then, we will review the related works on the relay assignment problem with multiple source-destination pairs for CC. In [2], Ng and Yu gave a utility maximization problem for the joint optimization of relay assignment, resource allocation, and CC modes selection in a cellular network. Then, Kadloor and Adve [7] formulated a related convex optimization problem for the relay selection and power allocation problem. However, these two works assume that each relay node has infinite number of channels so as to help multiple pairs, but this setting will not always hold in practice where the number of a relay node's channels is limited.

There are also some works studied the relay assignment problem with the setting that each relay node can be assigned to at most one pair. Zhang et al. [8] investigate the relay

assignment problem in the wireless network where the number of relays equals or exceeds the number of source-destination pairs. However, this assumption may not always hold in practice where the number of relay nodes is less than that of source-destination pairs. In [4], Cai et al. studied the joint relay node selection and power allocation for AF mode so as to maximize the network throughput both for the network environment with a single pair and multiple pairs. Shi et al. [5, 6] proposed an optimal algorithm for the relay node assignment problem to maximize the minimum capacity among all pairs. In [9], the authors studied the joint relay assignment and power allocation problem aiming to minimize the total power consumption for all nodes. However, all the above works did not consider the selfishness of the source-destination pairs, and the selfishness can greatly decrease the network performance.

Specially, Li et al. [10] studied the relay assignment problem in cooperative wireless networks with self-interested nodes. They modeled the cooperative relationship among the nodes as an exchange market game where nodes trade transmission power between each other to get diversity gain. However, this market mechanism faces significant challenges due to the fear of market manipulation. Moreover, the authors in [10] assumed that there is no difference between any assisted nodes (relay nodes) from the point of each node (source-destination pair). But in fact the performance of each pair by employing different relay nodes will be different in practice. This feature causes the competition among multiple pairs, and each pair will cheat the others so as to get its favorite relay node.

The works about scheme design in [11–15] are most related to our work. Shastry and Adve [13] proposed a pricing-based system to stimulate the cooperation via payment, in order to ensure both the access point and the relay nodes benefit from cooperation. Wang et al. [14] employed a buyer/seller Stackelberg game, where a buyer announced its selection of relays and the required transmission power, then the relays asked proper prices to maximize their profits. In [15], Huang et al. proposed two repeated games. In each game, each user iteratively updated its bid to maximize its own utility function with the knowledge of others previous bids. However, all the above schemes cannot guarantee the optimal total social value. In [12], Yang et al. proposed a truthful auction scheme for cooperative communications. This scheme satisfies truthfulness, individual rationality, and budget balance properties in the same time. Similarly, this scheme also cannot guarantee the optimal total social value. Moreover, in this scheme, the authors assumed that all pairs knew the capacity of other pairs in the network so as to get an initial relay assignment solution, but each pair might cheat the others about its capacity to get its favorite relay node in practice. In [11], Yang et al. proposed a robust scheme to prevent selfish behavior of source nodes and cheating behavior of relay nodes while guaranteeing the social optimal system capacity. However, this paper assumed that the information of each source node was known by all other source nodes, and thus did not consider the truthfulness of source nodes. In fact, each source node might cheat the others to get its favorite relay node.

Figure 2: A simple example for cooperative communication.

Since VCG mechanism is effective in preventing cheating, there are many works to use VCG mechanism in wireless networks to guarantee truthfulness. Such as Anderegg and Eidenbenz [16] proposed a routing protocol which implemented a variation of VCG mechanism so as to find the most cost-efficient path. In [17], the authors investigated how to use VCG mechanism to allocate and price spectrum resource. Shu and Varaiya [18] proposed a VCG mechanism-based access control mechanism for packet traffic. Therefore, in order to prevent the cheating among some pairs, we implement a variation of VCG mechanism to guarantee truthfulness in this paper. Specially, we study the truthful relay assignment problem for CC in wireless networks where multiple selfish source-destination pairs compete for the same pool of relay nodes.

3. System Model

3.1. Cooperative Communication Model. The essence of CC can be shown by a simple example in Figure 2, where source node s potentially transmits to the destination node d with the help of relay node r. This transmission is done on a frame-by-frame basis and there are two time slots in a frame [1, 5, 6]. In the first time slot, source node s transmits a signal to destination node d, and this signal is also overheard by the relay node r. In the second time slot, the relay node r forwards the signal received in the first time slot to the destination node d. Now, node d can use a diversity combining technique on two copies of the data from two different paths to increase the capacity gain.

Assume source node s sends a data to destination node d with the transmitting power P_s^{tra}. Let h_{sd} capture the effect of path loss, shadowing and fading between nodes s and d, and then the received energy at the destination node d is

$$P_{s,d}^{\text{rec}} = P_s^{\text{tra}} * |h_{sd}|^2. \tag{1}$$

So, the signal noise ratio (SNR) [5, 6] at the destination d is

$$\text{SNR}_{s,d} = \frac{P_s^{\text{tra}} * |h_{sd}|^2}{\sigma_d^2} = \frac{P_{s,d}^{\text{rec}}}{\sigma_d^2}, \tag{2}$$

where σ_d^2 is the variance of the background noise at node d. Note that only destination node d knows the $\text{SNR}_{s,d}$, and node s have no knowledge about $\text{SNR}_{s,d}$.

In AF mode, relay r simply amplifies the signal received from node s in the first time slot and then forwards it to node d in the next slot. The achievable AF capacity is given by [1]

$$C_{\mathrm{AF}}(s,r,d) = \frac{W}{2}\log_2\left(1 + \mathrm{SNR}_{s,d} + \frac{\mathrm{SNR}_{s,r}\cdot\mathrm{SNR}_{r,d}}{\mathrm{SNR}_{s,r}+\mathrm{SNR}_{r,d}+1}\right), \tag{3}$$

where W is the bandwidth.

In DF mode, relay r decodes and estimates the signal from node s in the first time slot, and then forwards the reencoded estimated signal to node d in the next slot. The DF capacity shown in [1] is

$$C_{\mathrm{DF}}(s,r,d) = \frac{W}{2}\min\Big\{\log_2(1+\mathrm{SNR}_{s,r}), \log_2(1+\mathrm{SNR}_{s,d}+\mathrm{SNR}_{r,d})\Big\}. \tag{4}$$

When CC (i.e., relay node r) is not used, node s transmits directly to node d in both two time slots. Thus, the capacity of direct transmissions (DT) is

$$C_{\mathrm{DT}}(s,d) = W\log_2(1+\mathrm{SNR}_{s,d}). \tag{5}$$

From the above results, we have three observations.

(i) Comparing C_{AF} (or C_{DF}) to C_{DT}, it is hard to say that the capacity of CC is always larger than that of DT. In fact, a poor selection of relay node may cause the capacity of CC smaller than that of DT for a single source-destination pair. So, this fact indicates the significance of relay selection in CC and also causes the competition among multiple pairs for the same relay node set.

(ii) The capacity is a monotonic increasing function of the transmission power under all the communication modes (AF, DF, and DT). Therefore, both source node and relay node should transmit at the maximum power so as to achieve the maximum capacity. So, we let $P_s^{\mathrm{tra}} = P_r^{\mathrm{tra}} = P_{\max} = P$ in this paper.

(iii) The capacities for AF and DF have the same form (i.e., a function of $\mathrm{SNR}_{s,d}$, $\mathrm{SNR}_{s,r}$, and $\mathrm{SNR}_{r,d}$). Therefore, the algorithm designed for AF mode can be easily extended to DF mode scheme, vice versa.

3.2. Truthful Relay Assignment Problem. We consider a wireless network consisting of N nodes, where there are four subsets of nodes: (1) the set of source nodes, $\mathcal{S} = \{s_1, s_2, \ldots s_{N_s}\}$, with $|\mathcal{S}| = N_s$, (2) the set of destination nodes, $\mathcal{D} = \{d_1, d_2, \ldots d_{N_d}\}$, with $|\mathcal{D}| = N_d$, (3) the set of remaining potentially relay nodes except the assignment node, $\mathcal{R}' = \{r_1, r_2, \ldots r_{N_r-1}\}$, with $|\mathcal{R}'| = N_r - 1$, and (4) the set of an assignment node, $\mathcal{A} = \{A\}$, with $|\mathcal{A}| = 1$ (see Figure 3). Note that the assignment node A is a special potentially trusted relay node in charge of assigning relay node to each source-destination pair, and deciding the corresponding payment. Specially, we can denote the assignment node A as r_{N_r}. So, the set of available relay nodes can be denoted as $\mathcal{R} = \{r_1, r_2, \ldots r_{N_r-1},\ r_{N_r} = (A)\}$ with $|\mathcal{R}| = N_r$.

Figure 3: The wireless network in our model.

For clarity, we consider unicast where each source node s_i is paired with a destination node d_i, that is, $N_s = N_d$, and we also assume that all nodes are assumed to be within range of each other. Specially, we denote the source-destination pair from s_i to d_i $(1 \le i \le n)$ as t_i, and so the source-destination pair set is $\mathcal{T} = \{t_1, t_2, \ldots t_{N_s}\}$, with $|\mathcal{T}| = N_s$. Further, in order to avoid interference, we employ the orthogonal channels in the network (e.g., using OFDMA), which is also used in Laneman et al.'s [1] for CC. We also assume that the channel state is stable, that is, also assumed by the authors in [5, 6] for relay assignment problem. Moreover, each node is equipped with a single antenna and can transmit/receive within one channel at a time so that each relay node can serve for at most one pair. Specially, Zhao et al. [3] showed that it is sufficient to choose the "best" relay node to achieve full diversity order than to employ multiple relay nodes for a pair. Thus, it paved the way for research on assigning at most one relay node to a source-destination pair and we will adopt this setting in this paper.

More formally, we use $R(t_i)$ to denote the relay node employed by the pair t_i. In the case that t_i does not employ a relay node where t_i transmits with DT, we have $R(t_i) = \phi$. So, the capacity of pair t_i can be rewritten as

$$C_R(t_i, R(t_i)) = \begin{cases} C_{\mathrm{XF}}(s_i, R(t_i), d_i), & \text{if } R(t_i) \neq \phi, \\ C_{\mathrm{DT}}(s_i, d_i), & \text{if } R(t_i) = \phi, \end{cases} \tag{6}$$

with $C_{\mathrm{XF}}(\cdot) = C_{\mathrm{AF}}(\cdot)$ for AF and $C_{\mathrm{XF}}(\cdot) = C_{\mathrm{DF}}(\cdot)$ for DF.

We propose to model the true value of a source-destination pair through a parameter: the value-of-capacity parameter v_i of dollars per mbps (similar to the cost parameter: the cost-of-energy parameter c_i dollars per watt proposed in [16]) is individual for each pair in the network, and this parameter reflects the value of each pair for the unit capacity.

Therefore, the true value of pair t_i by employing relay node $R(t_i)$ is

$$V(t_i, R(t_i)) = \begin{cases} v_i * C_{\mathrm{XF}}(s_i, R(t_i), d_i), & \text{if } R(t_i) \neq \phi, \\ v_i * C_{\mathrm{DT}}(s_i, d_i), & \text{if } R(t_i) = \phi. \end{cases} \tag{7}$$

We assume that there is no collusion in the wireless network and that each pair wants to employ its favorite relay so as to maximize its individual revenue. However, the favorite relay node of some pairs may be the same one (e.g., in Figure 3, r_3 may be the favorite relay node of pairs t_1 and t_2; r_6 may be the favorite relay node of t_3, t_4, and t_5), and this feature will cause the competition among some pairs. Thus, each pair will try to cheat the others so as to get its favorite relay node, and this will significantly decrease the total social value. For clarity, we regard all relay nodes are truthful and this assumption is reasonable in the energy-abundance networks. Because there is no loss for each relay node to serve for a pair using CC except some energy consumption.

Specially, our objective is to design a relay assignment protocol to find a relay assignment result that maximizes the total social value:

$$\max \sum_{i=1}^{N_s} V(t_i, R(t_i)), \tag{8}$$

and causes all source-destination pairs to act truthfully, that is, to reveal their true values, where $R(t_i)$ is denoted as the corresponding relay node assigned to source-destination pair t_i by this protocol.

In order to guarantee truthfulness, the assignment node will charge each pair a payment (virtual money) in the protocol such that a source-destination pair's payoff will be maximized only when it reveals its true individual value. The payment determination for each pair will be presented in Section 4.3.

As with other truthful protocols in wireless networks, we assume a payment mechansm [19] that takes charge of accounting and transferring payments between nodes. There is also a tamper-proof hardware [20] which can protect virtual money from modification or other attacks. Note that, conceptually we could also think that the payments can be funded by the entities which own the wireless devices. However, we focus on effective truthful relay assignment problem in this paper. Other issues about the virtual money, securely crediting and transferring money are beyond the scope of this paper.

Note that this network model might correspond to a snapshot of a wireless network. For example, in the context of a Cognitive Radio Network (CRN) which consists of some second users and a primer user, each user also have a source node (device) and a paired destination node (device). When the primer user does not use its own spectrum in a period of time, it will divide its spectrum to some second users so as to maximize total social value. Thus, the assignment node corresponding to the destination node (or source node) of the primer node; the source-destination pairs correspond to the occupied second users that have data to transmit; the available relay nodes corresponding to the unoccupied second users' source node or destination node which do not have data to transmit. Moreover, every unoccupied second user is willing to participate in cooperative communication, so that the others can help this user when it has data to transmit. Specially, the primer node can also enforce each unoccupied user to participate in CC; otherwise, the primer user does not allow this pair to use its spectrum forever. Note that, in this paper, every node in the network only serve as one role in a period of time and we focus on the relay assignment in this period of time. Moreover, the role of each node may change in the next period of time, and it needs to reexecute the protocol designed in this paper. From above, when a primer user tries to divide its spectrum to some second users so as to maximize total social value in CRN, there are two important issues that need to be solved. One is spectrum allocation, where the primer user decides how to divide and allocate its spectrum. The other is relay assignment, where the primer user decides how to assign relay nodes (unoccupied second user) to source-destination pairs (occupied second users). Specially, in this paper, we focus on relay assignment problem and omit spectrum allocation problem. However, we will jointly consider spectrum allocation and relay assignment problem for cooperative communication in CRN in the future.

4. Truthful Relay Assignment Protocol

In order to achieve truthfulness, we design a truthful relay assignment protocol (RA-VCG) which implements a variation of the well-known VCG mechanism for CC in wireless networks. RA-VCG will charge each pair a payment (virtual money) such that a source-destination pair's payoff will be maximized only when it reveals its true individual value.

4.1. Basic Idea. Figure 4 shows the flow chart of RA-VCG. This protocol consists of the following three phases:

(1) relay discovery,

(2) relay assignment,

(3) data transmission.

In relay discovery phase, the assignment node will first broadcast a message to inform all the source nodes to select relay nodes. Then, each source node will send a message to find the corresponding relay node. Next, each relay node will forward the message from the source node to its corresponding destination node. After that, each destination node can compute the true value $V(t_i, R(t_i))$ by (7) and send this information to the assignment node (Algorithm 1).

In the relay assignment phase, the assignment node will determine the final relay assignment result according to all the information sent by the destination nodes. After that, the assignment node also decides the payments for each pair to guarantee truthfulness.

In data transmission phase, each pair pays the payment and starts transmitting data. Next, we will present our RA-VCG protocol in detail.

Figure 4: The flow chart of RA-VCG protocol.

4.2. Relay Discovery. We introduced a trusted node called assignment node to the network, which takes charge of assigning relay nodes and deciding the corresponding payments for each pair. For example, in the CRN, the assignment node corresponds to the primer user who shares his spectrum to some second users so as to maximize the total social value. Specially, if there is not a primer node in the network, we can designate a relay node as the assignment node.

In the beginning, the assignment node will broadcast a message START(A) with the maximum transmission power P, where A is the ID of the assignment node. If a node (source node, destination node, or potentially relay node) receives the message START(A), it will know that A is the assignment node. Specially, if this node is a source node s_i, it will broadcast a relay test message RTEST(s_i, d_i) with its maximum transmission power, where s_i is the ID of this source node and d_i is the ID of the destination node paired with s_i.

Each potentially relay node r_j (including the assignment node A) that receives a message RTEST(s_i, d_i) from a source node s_i executes the following operations:

(1) determine the received power $P^{\text{rec}}_{s_i,r_j}$ and compute the SNR_{s_i,r_j} by (2);

(2) send a relay reply message $\text{RREP}(s_i, d_i, r_j, \text{SNR}_{s_i,r_j})$ with its maximum transmission power to the corresponding destination node d_i in the RTEST message.

For each destination node d_i, it first creates a message queue Q_i and a message $\text{BID}(s_i, d_i, v_i)$, where s_i is the ID of its paired source node and v_i is the individual value-of-capacity parameter of this pair t_i. On receiving a message $\text{RREP}(s_i, d_i, r_j, \text{SNR}_{s_i,r_j})$ form potentially relay node r_j, the destination node d_i will execute the following operations:

(1) determine the received power $P^{\text{rec}}_{r_j,d_i}$ and compute the SNR_{r_j,d_i};

(2) if d_i has not received the message RTEST(s_i, d_i) from its paired source node s_i, then append SNR_{r_j,d_i} to this message as $\text{RREP}(s_i, d_i, r_j, \text{SNR}_{s_i,r_j}, \text{SNR}_{r_j,d_i})$. Then push the appended message to the message queue Q_i;

(3) if d_i has received the message RTEST(s_i, d_i) from source node s_i, it uses SNR_{s_i,r_j}, SNR_{r_j,d_i}, and SNR_{s_i,d_i} to computes the $C_{\text{XT}}(s_i, r_j, d_i)$ by (3) or (4). Append the BID packet by adding the relay node r_j and the corresponding capacity $C_{\text{XT}}(s_i, r_j, d_i)$.

Once the destination node d_i receives the message RTEST(s_i, d_i) from its paired source node s_i, it will execute the following operations:

(1) determine the received power $P^{\text{rec}}_{s_i,d_i}$ and compute the SNR_{s_i,d_i} by (2), and then compute the $C_{\text{DT}}(s_i, d_i)$ by (5). Append the BID packet by adding $C_{\text{DT}}(s_i, d_i)$;

(2) for each message $\text{RREP}(s_i, d_i, r_j, \text{SNR}_{s_i,r_j}, \text{SNR}_{r_j,d_i})$ in Q_i, use SNR_{s_i,r_j}, SNR_{r_j,d_i}, and SNR_{s_i,d_i} to compute $C_{\text{XT}}(s_i, r_j, d_i)$ by (3) or (4). Then append the BID message by r_j and $C_{\text{XT}}(s_i, r_j, d_i)$.

After the destination node d_i has already received all the RREP message from each potentially relay node r_j and the RTEST message from its paired source node s_i, it will derive a message whose format is $\text{BID}(s_i, d_i, v_i, C_{\text{DT}}(s_i, d_i), r_{j_1}, C_{\text{XT}}(s_i, r_{j_1}, d_i), \ldots, r_{j_{N_r}}, C_{\text{XT}}(s_i, r_{j_{N_r}}, d_i))$, and then send this BID message to the assignment node A. Note that, we assume that there is no packet loss in RA-VCG. This assumption is also reasonable because there is too many works providing the reliable packet transmitting.

1.Initialization step:
Each destination node d_i creates a message BID(s_i, d_i, v_i) and a message queue Q_i;
Assignment node A broadcasts a message START(A).
2. Process Step:
2.1. Source node s_i
On receiving the message START(A)
Broadcast a message RTEST(s_i, d_i).
2.2 Relay node r_j (**including assignment node** A)
On receiving the message RTEST(s_i, d_i)
Determine the received power $P^{\text{rec}}_{s_i,r_j}$. and compute the SNR_{S_i,r_j};
Send a message RREP$(s_i, d_i, r_j, \text{SNR}_{S_i,r_j})$ to the destination node d_i.
2.3. Destination node d_i
On receiving the message RREP$(s_i, d_i, r_j, \text{SNR}_{S_i,r_j})$
Determine the received power $P^{\text{rec}}_{r_j,d_i}$, and compute SNR_{r_j,d_i};
if d_i has not received the message RTEST(s_i, d_i) from
its paired source node s_i **then**
Append SNR_{r_j,d_i} to this RREP message;
Push the appended message to queue Q_i;
else
Compute $C_{\text{XT}}(s_i, r_j, d_i)$;
Append the BID packet by r_j and $C_{\text{XT}}(s_i, r_j, d_i)$.
end if
On receiving the message RTEST(s_i, d_i)
Determine the received power $P^{\text{rec}}_{s_i,d_i}$. and compute SNR_{s_i,d_i};
Compute $C_{\text{DT}}(s_i, d_i)$;
Append the BID packet by $C_{\text{DT}}(s_i, d_i)$;
for each message RREP$(s_i, d_i, r_j, \text{SNR}_{S_i,r_j}, \text{SNR}_{r_j,d_i})$ in Q_i **do**
Compute $C_{\text{XT}}(s_i, r_j, d_i)$;
Append the BID message by r_j and CXT(s_i, r_j, d_i).
end for
On receive all the RREP messages and the corresponding RTEST message
Send the BID message to the assignment node A.

Algorithm 1: Relay discovery phase.

4.3. Relay Assignment. After the relay discovery phase, all the destination nodes (source-destination pairs) have sent BID messages to the assignment node A that collects all the arriving BID messages. Then the assignment node A will perform an algorithm called truthful relay assignment (TRA) algorithm. In TRA, the assignment node A will first perform an optimal polynomial-time algorithm, called relay assignment (RA) to maximize the total social value by assigning relay nodes to each pair; after that, the VCG-payment of each pair is also determined by the assignment node A to guarantee truthfulness.

The essence of RA algorithm is the maximum matching of weighted bipartite graph. RA algorithm consists of three steps: bipartite graph construction, saturated matching, and relay node assignment, shown in Algorithm 2.

At first, RA algorithm builds up the following weighted bipartite graph $G = (V, E, w)$ based on the all BID messages and the input parameters: source node set $\mathcal{S}$, available relay node set $\mathcal{R}$, and destination node set $\mathcal{D}$. The graph G contains two independent vertex sets X and Y, where $|X| = |\mathcal{S}|$ and $|Y| = |\mathcal{R}| + |\mathcal{D}|$. Without loss of generality, assume that $X = \{x_i \mid s_i \in \mathcal{S},\ 1 \le i \le N_s\}$ where each vertex x_i represents the corresponding source node s_i, as well as $Y = \{y_j \mid r_j \in \mathcal{R},\ 1 \le j \le N_r\} \cup \{y'_i \mid d_i \in \mathcal{D},\ 1 \le i \le N_s\}$ where each vertex y_j represents the corresponding potentially relay nodes r_j and each vertex y'_i represents the corresponding destination node d_i. Based on the above, we will define the weights of the edges in the graph G. For each message BID$(s_i, d_i, v_i, C_{\text{DT}}(s_i, d_i), r_{j_1}, C_{\text{XT}}(s_i, r_{j_1}, d_i), \ldots, r_{j_{N_r}}, C_{\text{XT}}(s_i, r_{j_{N_r}}, d_i))$, we will define the corresponding weights as follows:

$$w(x_i, y'_i) = v_i * C_{\text{DT}}(s_i, d_i),$$
$$w\left(x_i, y_{j_k}\right) = v_i * C_{\text{XT}}\left(s_i, r_{j_k}, d_i\right), \quad 1 \le k \le N_r. \tag{9}$$

Now, a bipartite graph G has been constructed. Thus, the second step is to adopt the well-known KM algorithm [9, 21] to find out an optimal saturated matching for G, that is, the total weight of all the matched edges is maximized. After the second step, each vertex x_i will be matched to a vertex $M(x_i)$ in Y. Here, RA decides whether to assign relay nodes to each pair through the parameter *flag*. If *flag* is true, then assign relay nodes to each pair; otherwise, only return the total social value without relay assignment. Specially, when *flag* is true, if $M(x_i) = y_j$, then the source-destination pair t_i will

```
Step 1. Bipartite Graph Construction
  Create the vertex sets X and Y;
  for each message BID(s_i; d_i, v_i, C_DT(s_i,d_i), r_j1,
  C_XT(s_i, r_j1, d_i), ..., r_jN_r, C_XT(s_i, r_jN_r, d_i)) and s_i ∈ 𝒮
  do
    w(x_i, y'_i) = v_i * C_DT(s_i, d_i);
    for k = 1 to N_r do
      w(x_i, y_jk) = v_i * C_XT(s_i, r_jk, d_i);
    end for
  end for
Step 2. Saturated Matching
  Adopt KM algorithm to find a saturated matching in Graph
  G; Assume that vertex x_i is matched to vertex M(x_i);
Step 3. Relay Node Assignment
  for each s_i ∈ 𝒮 do
    if flag = true then
      if M(x_i) = y_j then
        R(t_i) = r_j;
      else
        R(t_i) = φ;
      end if
    end if
    result+ = w (x_i, M(x_i));
  end for
  Return result;
```

Algorithm 2: RA($\mathcal{S}$, $\mathcal{R}$, $\mathcal{D}$, flag).

```
RA(𝒮, ℛ, 𝒟, true);
for i = 1 to N_s do
  first = RA(𝒮 \ {s_i}, ℛ, 𝒟 \ {d_i}, false);
  second = RA(𝒮 \ {s_i}, ℛ \ {ℛ(t_i)}, 𝒟 \ {d_i}, false);
  pay(t_i) = first-second
end for
```

Algorithm 3: TRA.

employ relay node r_j to use CC; If $M(x_i) = y'_i$, then the pair t_i will use DT. Thus, we have

$$R(t_i) = \begin{cases} r_j, & \text{if } M(x_i) = y_j, \\ \phi, & \text{if } M(x_i) = y'_i. \end{cases} \tag{10}$$

RA algorithm is only a subroutine of TRA algorithm which has been shown in Algorithm 3. After performing the RA algorithm, the assignment node has assigned relay nodes for each pair. Then, it will also determine the VCG-payment for each pair to guarantee truthfulness. The VCG-payment for each pair is the harm that it causes to the rest of other pairs and we can formulate it as

$$\begin{aligned} \text{pay}(t_i) = {} & \text{RA}(\mathcal{S} \setminus \{s_i\}, \mathcal{R}, \mathcal{D} \setminus \{d_i\}, \text{false}) \\ & - \text{RA}(\mathcal{S} \setminus \{s_i\}, \mathcal{R} \setminus \{R(t_i)\}, \mathcal{D} \setminus \{d_i\}, \text{false}), \end{aligned} \tag{11}$$

where $\mathcal{S} \setminus \{s_i\}$ denotes that s_i is removed from $\mathcal{S}$. For simplicity, we will use the notations $M^{\mathcal{D}}_{\mathcal{S},\mathcal{R}}$ to denote RA($\mathcal{S}$, $\mathcal{R}$, $\mathcal{D}$, false). Thus, the payment of t_i can be rewritten as follows:

$$\text{pay}(t_i) = M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} - M^{\mathcal{R}\setminus\{R(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}}. \tag{12}$$

In (11) or (12), the first term is the total social value should the corresponding source node s_i and destination node d_i of pair t_i be removed from $\mathcal{S}$ and $\mathcal{D}$, respectively; the second term is the total social value should s_i, $R(t_i)$ and d_i be removed from $\mathcal{S}$, $\mathcal{R}$ and $\mathcal{D}$, respectively. Hence, the first term is the total social value received by all the other pairs when pair t_i did not participate in relay assignment, and the second term is the total social value received by all the other pairs when t_i participate in relay assignment. So pay(t_i) is the total harm caused by t_i.

As a result, the assignment node has completed relay assignment and also decides the VCG-payment for each pair

after performing TRA algorithm. So, it will broadcast the result by a message RA-RESULT$(t_1, R(t_1), \text{pay}(t_1), t_2, R(t_2), \text{pay}(t_2), \ldots, t_{N_s}, R(t_{N_s}), \text{pay}(t_{N_s}))$ to all nodes in the network.

4.4. Data Transmission. When the pair t_i receives the message RA-RESULT, if $R(t_i) \neq \phi$, it will pay the payment pay(t_i) to the relay node $R(t_i)$ (or the assignment node A where A can get this payment alone or divide it to all the relay nodes to encourage them to participate in RA-VCG), and then start to employ the relay node $R(t_i)$ to use CC. Otherwise, it will exploit DT.

4.5. Complexity Analysis. Now, we mainly analyze the computational complexity of TRA algorithm and the message complexity of RA-VCG.

First, let us focus on the time complexity of TRA algorithm. In the first step of RA algorithm, there are at most $N_s * N_r + N_s$ edges in the bipartite graph G. So, the time complexity for graph construction is $O(N_s * N_r)$. Moreover, the time complexity to find out a saturated matching for this bipartite graph G is $O(N_s^2 * N_r)$ shown in [9, 21]. Finally, it costs $O(N_s)$ to transform the matching result into the relay assignment in the third step. So, the total time complexity of RA is $O(N_s * N_r) + O(N_s^2 * N_r) + O(N_s) = O(N_s^2 * N_r)$. In TRA algorithm, it will call RA algorithm $2 * N_s + 1$ times and so the time complexity of TRA is $O(N_s^3 * N_r)$.

Next, we will focus on the message complexity of RA-VCG. In the relay discovery phase, the assignment node A sends a START message. Then each source node send a corresponding RTEST message after receiving the START message from assignment node A and so there are N_s RTEST messages in total. After that, each relay node will forward an RREP message for each source node to its paired destination node. Thus, the number of the total RREP messages is $N_s * N_r$. At last, each destination node sends a BID message to assignment node A and thus there are N_d BID messages in total. After the TRA algorithm, there is also one RA-RESULT sent by assignment node A to broadcast the relay assignment result. Thus, there are $1+N_s+N_s*N_r+N_d+1 = 2+2N_s+N_s*N_r$ messages in total. So, the message complexity of RA-VCG is $O(N_s * N_r)$.

5. Analysis

In this section, we show that RA-VCG meets the design requirements of maximizing the total social value and truthfulness in the presence of selfish source-destination pairs. We also show that our RA-VCG protocol has other nice properties.

Lemma 1. *If all source-destination pairs act truthfully, that is, reveal its true information in BID message, and then RA or TRA algorithm can find out an optimal relay assignment result which maximizes the total social value.*

Proof. The assignment node in RA-VCG collects all BID messages, and then converts the relay assignment problem to an optimal saturated matching problem in weighted bipartite graph based on these messages. If each pair acts truthfully, then the optimal saturated matching found by KM algorithm can also be converted to an optimal relay assignment result that maximizes the total social value. So, the lemma holds. □

From Lemma 1, we can see that maximizing the total social value of RA-VCG follows immediately from the description of RA-VCG, if we can guarantee truthfulness of each pair. Thus, now we will show that the RA-VCG can guarantee truthfulness. Specially, showing that RA-VCG is truthful is more involved. Truthfulness is given if and only if it is a dominant strategy for each pair t_i to always

(i) declare its true value-of-capacity parameter v_i in its BID message;

(ii) correctly compute and declare its capacity of DT;

(iii) correctly compute and declare its capacity of CC for all relay nodes (including both its desired relay node and undesired relay nodes) in BID message.

We call all these items "cheating possibilities".

In accordance with the definition of dominant strategies, we assume an adversarial pair to be omniscient: an omniscient pair knows all edge weight of the bipartite graph. Our protocol is truthful only if even such a powerful pair cannot use its knowledge to its own advantage. The reason for using such a strong adversarial model is to prevent some pairs from taking bets based on their partial knowledge of the network.

Lemma 2. *In RA-VCG, each source-destination pair t_i declares its true information in BID message.*

Proof. We use BID_{t_i} to denote the BID messages declared by pair t_i and use $\text{BID}_{\mathcal{T}\setminus t_i}$ to denote the BID messages declared by all the other pairs except t_i. Then, for pair t_i, if all other pairs declare $\text{BID}_{\mathcal{T}\setminus t_i}$, we show that t_i can maximize its utility only when it reveals its true individual value according to BID_{t_i}. So, the cheating behaviors will not increase the utility (payoff) of each pair. The utility (payoff) of the pair t_i is the true value received by exploiting the assigned relay node $R(t_i)$ minus the payment pay(t_i) charged by RA-VCG:

$$U(t_i) = V(t_i, R(t_i)) - \text{pay}(t_i). \tag{13}$$

There are only two cases to discuss in the following. Assume that t_i declares its true information in BID_{t_i}, and it gets its relay node $R(t_i)$. Thus, the true value, payment, and utility of t_i are $V(t_i, R(t_i))$, pay(t_i), and $U(t_i) = V(t_i, R(t_i)) - \text{pay}(t_i)$, respectively. Similarly, assume that t_i does not declare truthfully in BID message (assume that it declares BID'_{t_i} and $\text{BID}'_{t_i} \neq \text{BID}_{t_i}$), that is, does not satisfy all the "cheating possibilities," and t_i gets its relay node $R'(t_i)$. As a result, the true value, payment, and utility of t_i are $V(t_i, R'(t_i))$, pay$'(t_i)$, and $U'(t_i) = V(t_i, R'(t_i)) - \text{pay}'(t_i)$ in untruthful BID message, respectively.

In the first case, the source-destination pair t_i receives the same relay node by declaring BID_{t_i} and BID'_{t_i}. In this

case, the payment of t_i is also $\text{pay}'(t_i) = \text{pay}(t_i)$ (see (11)). So the utility of t_i is $U'(t_i) = V(t_i, R'(t_i)) - \text{pay}'(t_i) = V(t_i, R(t_i)) - \text{pay}(t_i) = U(t_i)$. Thus, the untruthful information in BID message has no effect on the utility of t_i in this case.

In the second case, the source-destination pair t_i receives a different relay node by declaring BID'_{t_i}, that is, $R(t_i) \neq R'(t_i)$. Hence, we must prove that $U(t_i) - U'(t_i) \geq 0$, which shows that the utility received from truthful information in BID message is always at least that received from untruthful information BID message in this case. Joining (12) and (13), we can get that

$$\begin{aligned} U(t_i) - U'(t_i) &= \left(V(t_i, R(t_i)) - \left(M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} - M^{\mathcal{R}\setminus\{R(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}}\right)\right) \\ &\quad - \left(V(t_i, R'(t_i)) - \left(M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} - M^{\mathcal{R}\setminus\{R'(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}}\right)\right) \\ &= \left(V(t_i, R(t_i)) + \text{M}^{\mathcal{R}\setminus\{R(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}}\right) \\ &\quad - \left(V(t_i, R'(t_i)) + M^{\mathcal{R}\setminus\{R'(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}}\right), \end{aligned} \tag{14}$$

where the definition of parameters in (14) can be seen from (12) in Section 4.3. Note that in (14), the first term is the maximum total social value when t_i is assigned relay node $R(t_i)$ by declaring its true information in BID message. The second term represents (equals) the maximum total social value in the situation that t_i declares truthful information, but for some reason t_i is assigned relay node $R'(t_i)$. However, when other pairs declare $\text{BID}_{\mathcal{T}\setminus t_i}$ and t_i declares truthful information in its BID message, we have already assumed that the relay assignment solution with the maximal social value assigns $R(t_i)$ to t_i (see Algorithm 3). Thus, in (14), the first term must be not smaller than the second one, and we get $U(t_i) - U'(t_i) \geq 0$. So, the lemma holds. □

Theorem 3. *RA-VCG is truthful and can maximize the total social value.*

Proof. From Lemma 2, we can see that each pair is truthful about not only the desired relay nodes but also other undesired relay nodes and DT. In other words, each pair is truthful about all the relay nodes and thus declares its true information in its BID message. Thus, RA-VCG is truthful. Since all source-destination pairs act truthfully (see Lemma 2) in RA-VCG, then we can get that RA-VCG (RA or TRA algorithm) can maximize the total social value from Lemma 1. Thus, the theorem holds. □

Our RA-VCG also has some other nice properties as follows.

Theorem 4. *If source-destination pair t_i uses DT in RA-VCG, that is, $R(t_i) = \phi$, the payment of t_i is 0.*

Proof. Note that if source-destination pair t_i is assigned the relay node $R(t_i) = \phi$, then the payment of t_i is

$$\begin{aligned} \text{pay}(t_i) &= M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} - M^{\mathcal{R}\setminus\{R(t_i)=\phi\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} \\ &= M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} - M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} = 0. \end{aligned} \tag{15}$$

Thus, the theorem holds. □

Theorem 5. *For each source-destination pair t_i in RA-VCG, the payment is nonnegative.*

Proof. In the first case $R(t_i) = \phi$, from Theorem 4, we can get $\text{pay}(t_i) \geq 0$.

In the second case $R(t_i) \neq \phi$, it is easy to see that $M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} \geq M^{\mathcal{R}\setminus\{R(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}}$, because the corresponding bipartite graph in the first term contains all the edges of the corresponding bipartite graph in the second term. So

$$\text{pay}(t_i) = M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} - M^{\mathcal{R}\setminus\{R(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} \geq 0. \tag{16}$$

Thus, the theorem holds. □

Theorem 6. *For each source-destination pair t_i in RA-VCG, the utility is no smaller than $V(t_i, \phi)$.*

Proof. The utility of t_i in RA-VCG is

$$\begin{aligned} U(t_i) &= V(t_i, R(t_i)) - \text{pay}(t_i) \\ &= V(t_i, R(t_i)) - M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} + M^{\mathcal{R}\setminus\{R(t_i)\}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} \\ &= M^{\mathcal{R}}_{\mathcal{S},\mathcal{D}} - M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} \\ &\geq M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} + V(t_i, \phi) - M^{\mathcal{R}}_{\mathcal{S}\setminus\{s_i\},\mathcal{D}\setminus\{d_i\}} \\ &= V(t_i, \phi). \end{aligned} \tag{17}$$

Thus, the theorem holds. □

From Theorems 4 and 5, we can see that RA-VCG has no positive transfer. It means that the assignment node does not need to pay anything to the source-destination pairs. Moreover, RA-VCG is individually rational. It means that the pairs are willing to participate in RA-VCG, that is, no pair is really being forced to participate, because the utility achieved by each pair in participating is not smaller than that achieved by using DT (see Theorem 6).

6. Simulation Results

In this section, we give some simulation results to evaluate the performance of RA-VCG in wireless networks with selfish source-destination pairs. Firstly, we present the simulation settings. Secondly, we will show that RA-VCG is indeed truthful (i.e., it can indeed prevent source-destination pairs from cheating each other). Thirdly, we will observe the total social value achieved by RA-VCG. Moreover, we consider both the AF and DF modes for CC.

Figure 5: Average utility versus source-destination pair number in case 1.

Figure 6: Average utility versus source-destination pair number in case 2.

6.1. Simulation Settings. We consider a wireless network generated by randomly placing all the nodes in the square area $800 * 800\,\mathrm{m}^2$. Under this network, we consider both the cases that the value-of-capacity parameter is the same or different for all pairs. In case 1, all the pairs have the same value-of-capacity parameter and we set this parameter as $v = 1$ for all pairs. While in case 2, each pair has different value-of-capacity parameter which is random in $[1, 4]$.

We assume the bandwidth $W = 22\,\mathrm{MHZ}$ for each channel and the maximum transmission power $P = 1\,\mathrm{W}$ at each node. Similar to [5, 6], we also assume that $h_{uv}^2 = d_{uv}^{-4}$, where d_{uv} is the Euclidean distance between nodes u and v, and 4 is the pass loss index. We assume the channel is AWSN channel and the variance of noise is $10^{-10}\,\mathrm{W}$ for all nodes. To verify the performance of RA-VCG, we assume each source-destination pair can declare its BID message truthfully and untruthfully. For untruthful (cheating) behavior, we assume the declared capacity of each pair according to each relay node (or DT) is random in $(0, 2C)$, where C is the capacity achieved by this pair. The data points represented in the following graphs are averaged over 100 simulation runs.

Figure 7: Average utility versus available relay node number in case 1.

6.2. Truthfulness of RA-VCG. We first observe the effect of the number of source-destination pairs on the average utility of each pair for RA-VCG. As shown in Figures 5 and 6, we have 15 available relay nodes in the network, and the number of source-destination pairs changed from 9 to 19. Then, We observe the effect of the number of available relay nodes on the average utility of each pair for RA-VCG. As shown in Figures 7 and 8, we have 15 source-destination pairs in the network, and the number of available relay nodes changes from 5 to 33.

Nevertheless AF and DF, the average utility of each pair by declaring untruthful BID message is less than the average utility obtained when each source-destination pair declares truthful BID message (i.e., see Figures 5–8). These simulation results show that RA-VCG maximizes a source-destination pair's payoff only when this pair reveals its true individual information. Hence, RA-VCG is truthful.

6.3. Impact of Untruthfulness on Total Social Value. Next, we focus on the impact of untruthful behaviors of source-destination pairs on the total social value. As shown in Figures 9 and 10, when the number of source-destination pairs increases, total social value increases for both cases 1 and 2. Moreover, the truthful behavior in AF mode and DF mode can increase 26.5% and 19.6% total social value than

Figure 8: Average utility versus available relay node number in case 2.

Figure 9: Total social value versus source-destination pair number in case 1.

Figure 10: Total social value versus source-destination pair number in case 2.

Figure 11: Total social value versus available relay node number in case 2.

untruthful behaviors both in case 1 and case 2. Figures 11 and 12 show that the total social value increases when the number of available relay nodes increases for both AF and DF in case 1 (see Figure 11) and 2 (see Figure 12). This is because with more relay nodes, there are more chances for each pair to employ relay nodes to use CC. However, since the number of source-destination pairs do not change, the total social value of DT is always the same when the number of relay nodes changes both in case 1 (which is 427.7 in Figure 11) and case 2 (which is 1043.7 in Figure 12). Moreover, we can see that AF mode and DF mode with truthful behavior can increase about 27.0% and 19.7% total social value comparing with untruthful behavior both in case 1 and case 2. Thus, from all these simulation results (Figures 9–12), we can see that the total social value achieved when each node takes untruthful behaviors is average of about 23.3% less than that achieved when nodes behave truthfully.

7. Conclusion

CC is a novel technology to achieve spatial diversity. Thus, this paper studies the truthful relay assignment problem in wireless network where each source-destination pair may

Figure 12: Payment ratio versus available relay node number in case 2.

cheat the others to achieve more individual revenue. Therefore, in order to prevent cheating, we propose RA-VCG, a relay assignment protocol for CC to maximize the total social value while guaranteeing truthfulness by charging each pair an extra payment in wireless networks with selfish source-destination pairs. Moreover, we prove the validity of this protocol and also show several surprising properties associated with this protocol. The simulation results show that the total social value achieved when each node takes untruthful behaviors is about 23.3% less than that achieved when nodes behave truthfully.

Acknowledgments

The work of this paper is supported by the National Grand Fundamental Research 973 Program of China under Grant no. 2011CB302-905, the National Science Foundation of China under Grant nos. 61170058, 61272133, and 51274202, National Science and Technology Major Project under Grant no. 2011ZX03005-004-04, the Natural Science Foundation of Jiangsu Province in China (Grant nos. BK2009150 and BK2012632), and Research Fund for the Doctoral Program of Higher Education of China no. 20103402110041.

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] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 328–339, 2007.

[3] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 1234–1238, Seattle, Wash, USA, July 2006.

[4] J. Cai, X. Shen, J. W. Mark, and A. S. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1348–1357, 2008.

[5] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing 2008 (MobiHoc '08)*, pp. 3–12, May 2008.

[6] S. Sharma, Y. Shi, Y. T. Hou, and S. Kompella, "An optimal algorithm for relay node assignment in cooperative Ad Hoc networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 879–892, 2011.

[7] S. Kadloor and R. Adve, "Relay selection and power allocation in cooperative cellular networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1676–1685, 2010.

[8] X. Zhang, A. Ghrayeb, and M. Hasna, "Relay assignment schemes for multiple source-destination cooperative networks," in *Proceedings of the 17th International Conference on Telecommunications (ICT '10)*, pp. 147–152, April 2010.

[9] H. Xu, L. Huang, G. Wang, T. Xu, and G. Liu, "Joint relay assignment and power allocation for cooperative communications," *Wireless Networks*, vol. 16, no. 8, pp. 2209–2219, 2010.

[10] D. Li, Y. Xu, J. Liu, and X. Wang, "Relay assignment and cooperation maintenance in wireless networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '10)*, April 2010.

[11] D. Yang, X. Fang, and G. Xue, "HERA: an optimal relay assignment scheme for cooperative networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 245–253, 2012.

[12] D. Yang, X. Fang, and G. Xue, "Truthful auction for cooperative communications," in *Proceedings of the 12th ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MOBIHOC '11)*, pp. 89–98, 2011.

[13] N. Shastry and R. S. Adve, "Stimulating cooperative diversity in wireless ad hoc networks through pricing," in *2006 IEEE International Conference on Communications (ICC '06)*, pp. 3747–3752, July 2006.

[14] B. Wang, Z. Han, and K. J. Ray Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using buyer/ seller game," in *Proceedings of the 26th IEEE International Conference on Computer Communications (IEEE INFOCOM '07)*, pp. 544–552, May 2007.

[15] J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based resource allocation for cooperative communications," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1226–1237, 2008.

[16] L. Anderegg and S. Eidenbenz, "Ad hoc-VCG: a truthful and cost-efficient routing protocol for mobile Ad hoc Networks with selfish agents," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom '03)*, pp. 245–259, September 2003.

[17] J. Jia, Q. Zhang, Q. Zhang, and M. Liu, "Revenue generation for truthful spectrum auction in dynamic spectrum access," in *Proceedings of the 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '09)*, pp. 3–12, May 2009.

[18] J. Shu and P. Varaiya, "Smart pay access control via incentive alignment," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1051–1060, 2006.

[19] S. Zhong, J. Chen, and Y. R. Yang, "Sprite: a simple, cheat-proof, credit-based system for mobile ad-hoc networks," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (Infocom '03)*, pp. 1987–1997, April 2003.

[20] L. Buttyan and J. Hubaux, "Enforce service availability in mobile as-hoc wans," in *Proceedings of the 1st ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc '00)*, pp. 87–96, 2000.

[21] J. Munkres, "Algorithm for the assignment and transportation problem," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, pp. 32–38, 1957.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 582403, 7 pages
doi:10.1155/2012/582403

Research Article

Novel Side Information Generation Algorithm of Multiview Distributed Video Coding for Multimedia Sensor Networks

Fu Xiao,[1,2,3] Jinkai Liu,[1] Jian Guo,[1,3,4] and Linfeng Liu[1,3,4]

[1] *School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2] *Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou 215006, China*
[3] *Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China*
[4] *Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing 210003, China*

Correspondence should be addressed to Fu Xiao, xiaof@njupt.edu.cn

Received 24 August 2012; Accepted 15 October 2012

Academic Editor: Ruchuan Wang

Copyright © 2012 Fu Xiao 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 traditional multiview distributed video coding scheme using regional unified coding may lead to distortion problem of decoding estimation of the intense motion region. This paper presents a novel multiview distributed video coding algorithm. In the main perspective, gaining the intense motion regions of Wyner-Ziv frame according to the criteria of ROI, the algorithm extracts their DCT low-frequency coefficients for the entropy coding, in order to generate the best temporal side information. For the nonintense motion regions, the algorithm utilizes motion compensation interpolation (MCI) to generate side information. Finally, side information based on fusion of temporal and spatial side information will be gained. Experimental results show that our proposed algorithm can gain more accurate motion estimation in the intense motion region. Quality of decoded image is improved with the same transmit rate; thus, energy consumption of sensor nodes will be decreased ultimately.

1. Introduction

In recent years, along with the rapid development of the wireless multimedia communication technology [1], digital video requirements are being increased. People would like to see natural characterization of objects clearer and more realistic: the traditional single-view video network can only provide two-dimensional visual, and three-dimensional visual senses cannot be provided better; so the multiview video network appears. In the multiview video network, which has the limits of low power, storage capacity, computational and communication ability, it does not only need low-complexity encoding, but also requires real-time video encoding and transmission. Traditional video coding standards, such as MPEG-x or H.26x, mainly rely on the hybrid architecture, encoder using motion estimation to fully exploit the video sequences of time, and spatial correlation information. Since the heavy computing burden of the motion estimation and compensation task in these video compression standards, the encoder is 5 to 10 times more complex than the decoder [2, 3]. Traditional video coding system is not suitable, novel coding methods are required. Wide attention has been focused on new video codec framework, distributed video coding (DVC) from scholars, which uses intraframe encoding and interframe decoding. Decoder explores the correlation of video signals for interframe prediction decoding; so DVC removes the complexity of the interframe prediction coding in encoder. Distributed Video Coding, which has the characteristics of low-complexity encoding and good robustness, can meet the needs of these new video applications very well.

There are some DVC frameworks that have been proposed, such as Girod and Aaron's Wyner-Ziv video coding [4, 5], PRISM (power-efficient robust high-compression syndrome-based multimedia) [6], Zixiang Xiong's layered DVC [7], Sehgal's state-free DVC [8], Wavelet-based DVC [9] and Multiview Distributed Video Coding [10–12]. Literature [10–12] proposed DVC algorithm based on turbo or LDPC, the Wyner-Ziv frame of Distributed Video Coding in all regions had not considered multiview [13, 14], motion

estimation techniques cannot accurately predict the area where there is more intense exercise. Then, the decoder cannot accurately generate the temporal side information (temporal SI). And the side information is merged by the temporal side information (temporal SI) and spatial side information (spatial SI). Therefore, decoder needs to request more feedback information, thus not only increases the rate, but the decoded portion of the image is still not accurate enough. Focus on this problem, an improved multiview distributed video coding algorithm is proposed in this paper. In the main perspective, we can get the intense motion region and the nonintense motion region according to the criteria of ROI. For the intense motion regions, the algorithm extracts their DCT low-frequency coefficients for the entropy coding. Decoder uses decoded DCT low-frequency coefficients for bidirectional hash motion estimation which enables the decoder to choose between past and future reference frames for frame interpolation to obtain the best temporal side information. For the nonintense motion regions, the algorithm utilizes motion compensation interpolation (MCI) to generate temporal side information. Finally, side information based on fusion of temporal and spatial side information can be gained. Simulation results show that this algorithm can enhance the efficiency of the intense motion region, thereby reducing the bit rate while improving the quality of decoded images, energy consumption of sensor nodes can be reduced ultimately.

The rest of the paper is organized as follows. Section 2 introduces the basic principles of multiview DVC. Section 3 proposes the Multiview DVC framework based on DCT hash in detail. In Section 4, the experimental results are given. Finally, Section 5 is the conclusion.

2. Multiview Distributed Video Coding System

The goal of distributed source coding (DSC) is that the complexity of the encoding side is transferred to the decoder in order to achieve efficient compression. In this particular field of multiview distributed video coding (MDVC), the side information is generated from the camera of the internal and external camera interpolation. Multiview video systems tend to produce large amounts of data which have strong correlation to eliminate redundant information in the data, how to improve the compression ratio is one of the key problem for MDVC, however power consumption of the camera sensor nodes are subject to this problem, therefore, we need a low-complexity encoder in order to avoid communication between the complex internal nodes.

Figure 1 depicts a multiview DVC system architecture [10–12], two views are adopted and assume that they are still. The camera of the first view as intracamera works in the traditional way, these video streams are independent of the other camera coding. The camera of the second view called Wyner-Ziv camera-independent encode, but utilizes the other video stream to decode. The frames of the first view are encoded using traditional intracoding, the second view utilizes DVC. The key frames (K) in the second view are encoded and decoded using a conventional intraframe codec. Between the key frames are Wyner-Ziv frames which are intra frame-encoded but interframe-decoded. Encoder applies a 8×8 DCT transform on W frame, then the coefficients are quantized using a uniform scalar quantizer. The Slepian-Wolf coder is implemented using a low-density parity-check code (LDPC). The parity bits produced by the LDPC encoder are stored in a buffer, which transmits a subset of these parity bits to the decoder upon request. If the decoder cannot reliably decode the bits, additional parity bits are requested from the encoder buffer through feedback. The request-and-decode process is repeated until an acceptable probability of bit error is guaranteed. The decoded bits are reconstructed as DCT coefficients, and are the finally generated by taking the inverse-DCT and inverse-quantization of the reconstructed DCT coefficients. The side information of multiview is fusion by temporal and spatial side information, so the side information is more accurate. In past research, Slepian-Wolf codec usually adopts Turbo code for error correction coding. While LDPC codes with its excellent performance, simple form and good prospects are gaining increasing interest recently.

3. Multiview Distributer Video Coding Scheme

3.1. Multiview DVC Framework Based on DCT Hash. Based on [10–12], the Wyner-Ziv frames (W frame) of multiview distributed video coding in all regions without distinction, motion estimation techniques cannot accurately predict the area which are more intense exercise. Then, the decoder cannot accurately generate the temporal side information (temporal SI). For this problem, this paper presents an improved multiperspective distributed encoding algorithm. In the main perspective, we select intense motion macroblock as the ROI macroblock, the low-frequency DCT coefficients of the ROI macroblock are selected to help the side information creation at the decoder and then improve the coding efficiency and the decoded image quality. Figure 2 illustrates the multiview distributed video coding (MDVC) framework based on the DCT hash, the first view uses conventional intracoding scheme, the decoded video stream is transformed into the spatial side information for W frame of the second view by homograph. The K frames of the second view are encoded and decoded by conventional intraframe scheme, while for W frames of the second view, using the scheme of the combination of LDPC coding and entropy coding. At the encoder, according to ROI discrimination algorithm, W frame is divided into ROI macroblock (8×8 macroblock) and non-ROI macroblock (8×8 macroblock). Then, for the ROI macroblock, the low-frequency DCT coefficients are selected as DCT hash, which uses the entropy coding. The residual of ROI macroblock and non-ROI macroblock are encoded and decoded by LDPC coding. At the decoder, if a DCT hash is available to guide the motion estimation process, bidirectional hash-based interpolation is performed; otherwise, MCI is used. Then, the temporal side information and spatial side information are fused to generate the best side information. Side information is used in the reconstruction to obtain the decoded DCT coefficients; finally, IDCT and IQ (inverse quantization) are applied to generate the W decoded frame W'.

Figure 1: Multiview distributed video coding framework.

Figure 2: Multiview distributed video coding framework based on the DCT hash.

3.2. ROI Macroblock Selection and Temporal Side Information Generation [13]. Similar to our previous work [13], in hash-based motion estimation, which is described previously, the hash bits of W frame are sent for all macroblocks to assist the decoder to generate the side information. However, motion vector of many macroblocks is equal to zero or very small in most video sequences; so only little part of blocks has large displacement. For many macroblocks, MCI can make a good estimation. So, it is not necessary to send DCT hash for these macroblocks. Therefore, encoder uses SAD criteria to distinguish ROI, then get ROI and non-ROI macroblock. Similar to our previous work [13], current frame is X_w, and the previous reference frame is X_P; so SAD criteria can be obtained from the following equation:

$$\mathrm{SAD} = \sum_{(x,y)\in B_i} \left| X_w(x,y) - X_p(x,y) \right|, \tag{1}$$

where B_i represents each macroblock, and (x, y) is pixel coordinates inside 8×8 macroblock. If SAD $\geq T$, we get ROI macroblock, otherwise, non-ROI macroblock. An adequate threshold T has been found experimentally.

As our previous work [13], the temporal side information B'_R of ROI macroblock B_R is generated by bidirectional hash motion estimation:

$$B'_R = \frac{|mv_2(x,y)| \times B_P(x+mvx_1(x,y), y+mvy_1(x,y))}{|mv_1(x,y)| + |mv_2(x,y)|} + \frac{|mv_1(x,y)| \times B_F(x+mvx_2(x,y), y+mvy_2(x,y))}{|mv_1(x,y)| + |mv_2(x,y)|}, \quad (2)$$

where B_P, B_F are the best matching macroblocks of ROI macroblock B_R in the past and future reference frames K_P, K_F; $mv_1(x,y)$ is the motion vector of B_R relative to B_P, $mv_2(x,y)$ is the motion vector of B_R relative to B_F; $mvx(x,y)$ represents the macroblock motion vector in the x direction, $mvy(x,y)$ is the macroblock motion vector in the y direction.

3.3. Spatial Side Information Generation. Adjacent video sensor node monitors the same target scene from different locations and different angles; because of this multi-angle correlation, we can use the video sequence of the adjacent video sensor nodes at the same moment to generate spatial side information by homograph. We assume that all nodes are time synchronized in the wireless multimedia sensor networks. Since the position and perspective of the video sensor node is fixed, so only a homographic matrix transformation is needed. The homograph is a 3×3 matrix that relates video sensor node V_1 to another one V_2 in the homogenous coordinates system. As in [3, 15], each point is from $V_1(x_1, y_1)$ is mapped to a point (x_2, y_2) of V_2 up to a scale μ such that

$$\mu \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_3 & a_4 & a_5 \\ a_6 & a_7 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix},$$
$$x_2 = \frac{(a_2 + a_0 x_1 + a_1 y_1)}{(a_6 x_1 + a_7 y_1 + 1)}, \quad (3)$$
$$y_2 = \frac{(a_5 + a_3 x_1 + a_4 y_1)}{(a_6 x_1 + a_7 y_1 + 1)},$$

where $a_0, a_1, \ldots, a_7$ are 8 sport parameters. When $a_6 = a_7 = 0$, the model is an affine geometry transformation; $a_0 = a_4 = 1$, $a_1 = a_3 = a_6 = a_7 = 0$, the model is a pure transformation; $a_0 = a_4$, $a_1 = -a_3$, $a_6 = a_7 = 0$, the model is transformation—zoom—rotation. To compute the model parameters, we can use gradient descent method as proposed in [3, 15].

3.4. Fusion Side Information Generation. Combining the temporal and spatial side information, fusion side information can be gained. We get a binary fusion mask in which 0 indicates that the pixel is taken from the spatial side information; 1 indicates that the pixel is taken from the temporal side information. Fusion process can be simply described as follows: the temporal side information and spatial side information with the previous key frame were compared, if the spatial side information is closer to the pixel value of the key frame, we set binary mask to 0; if the temporal side information is closer to the pixel value, we set binary mask to 1. We perform the same processing with future key frame. Thus, binary mask for side information can be obtained. Finally, we perform an OR logic operation between both binary masks to obtain the binary fusion mask. The fusion process is shown in Figure 3.

4. Experiments and Analysis

Similar to our previous work [13], in our simulation experiments, LDPC codes are generated by PEG algorithm [16], and the rate of LDPC code is 7/8. To change the rate, we varied the number of quantization levels, LDPC encoder produces different output bit rate, and then different compression ratio can be obtained. After several experimental analyses and comparisons, 64 is the ideal threshold of ROI criteria in our experiments, DC + 8AC of low-frequency DCT coefficients are selected as DCT hash. The proposed scheme is tested with exit and ballroom [17] which are QCIF (176×144) video sequences (at 25 fps and with a total of 100 frames). Video streams of Camera 1 uses the H.263 coding scheme; the Camera 0 uses distributed video coding, Key frames and Wyner-Ziv frames coding sequence for "K-W-K-W"; the key frames K and Wyner-Ziv frames W are alternative coding. We compare the rate-distortion performance of multiview distributed video coding only uses temporal side information, only uses spatial side information, uses fusion side information, H.263 intraframe coding (I-I-I-I), H.263 interframe coding (I-P-P-P), and JPEG coding. H.263 + codec uses TMN8. The experimental results are shown in Figure 4. multiview video coding has significantly better performance (2 to 3 dB) than that of H.263 intraframe coding. However, multiview video coding system is less than the overall complexity of the H.263 intraframe coding. The gap from H.263 interframe coding is a certain distance. The proposed approach can bring improvements up to 0.2–0.5 when compared to the multiview distributed video coding only uses temporal side information. The proposed algorithm can gain more accurate motion estimation in the intense motion region to obtain the best side information. Decoder utilizes the belief-propagation (BP) over cycle-free Tanner graphs in the iterative decoding process.

Figure 5 shows the decoded images (the 17st frame) of "exit" and "ballroom" sequences using the fusion side information. Figure 6 shows the decoded images (the 17st frame) of "exit" and "ballroom" sequences only using temporal side information. The subjective effects of the decoded image have been improved.

5. Conclusion

In this paper, a novel multiview distributed video coding algorithm is presented. In the main perspective, we select ROI macroblocks based on SAD criteria, a bidirectional hash-based interpolation is used to generate side information macroblock for motion intense area; however, motion nonintense area utilizes motion compensation interpolation

Figure 3: Fusion-based side information.

(a) "Exit" rate versus PSNR

(b) "Ballroom" rate versus PSNR

Figure 4: The simulation results of two sequences.

(a) 17st frame of exit (PSNR = 36.08)

(b) 17st frame of ballroom (PSNR = 34.87)

Figure 5: Decoded frame of "exit" and "ballroom" using our proposed algorithm.

(a) 17st frame of exit (PSNR = 35.78) (b) 17st frame of ballroom (PSNR = 34.59)

Figure 6: Decoded frame of "exit" and "ballroom" using side information.

(MCI). Finally, we gain the side information based on fusion of temporal and spatial side information. Experimental results demonstrate the algorithm's effeciency.

Acknowledgments

The paper is sponsored by the National Natural Science Foundation of China (61003236, 61170065), the Natural Science Foundation of Jiangsu (BK2011755), Scientific Technological Support Project of Jiangsu (BE2012755, BE2012183), Project sponsored by Jiangsu Provincial Research Scheme of Natural Science for Higher Education Institutions (11KJB520016), Scientific Research and Industry Promotion Project for Higher Education Institutions (JHB2012-7), Doctoral Fund of Ministry of Education of China (20103223120007), Nature Science Foundation of NUPT (KJS1022) and Priority Academic Program Development of Jiangsu Higher Education Institutions (Information and Communication Engineering, yx002001).

References

[1] J. Huang, L. Sun, R. Wang, and H. Huang, "Improved virtual potential field algorithm based on probability model in three-dimensional directional sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 942080, 9 pages, 2012.

[2] G. Jiang, F. Shao, M. Yu, K. Chen, and T. Y. Choi, "Efficient block matching for Ray-Space predictive coding in Free-Viewpoint television systems," in *Proceedings of the 6th International Conference on Computational Science and Its Applications*, pp. 307–316, 2006.

[3] W. Yufei, W. Ruchuan, H. Haiping, and S. Lijuan, "Multi-model sensors information based distributed video processing for wireless multimedia sensor networks," *Journal of Image and Graphics*, vol. 15, no. 1, pp. 161–166, 2010.

[4] A. Aaron, D. Varodayan, and B. Girod, "Wyner-Ziv residual coding of video," 2006, http://ivms.stanford.edu/~dsc/wzcodingvideo.

[5] D. Varodayan, A. Aaron, and B. Girod, "Exploiting spatial correlation in pixel-domain distributed image compression," in *Proceedings of the 25th Picture Coding Symposium (PCS '06)*, pp. 1–4, April 2006.

[6] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: a video coding paradigm with motion estimation at the decoder," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436–2448, 2007.

[7] Q. Xu and Z. Xion, "Layered Wyner-Ziv video coding," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3791–3803, 2006.

[8] A. Sehgal, A. Jagmohan, and N. Ahuja, "A state-free causal video encoding paradigm," in *Proceedings of the International Conference on Image Processing (ICIP'03)*, pp. 65–72, September 2003.

[9] B. Wu, X. Ji, D. Zhao, and W. Gao, "Wavelet based distributed video coding with spatial scalability," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '08)*, pp. 3458–3461, Seattle, Wash, USA, May 2008.

[10] X. Artigas, E. Angeli, and L. Torres, "Side information generation for multiview distributed video coding using a fusion approach," in *Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG '06)*, pp. 250–253, Reykjavik, Iceland, June 2006.

[11] J. Zhong, X. Hu, W. B. Kleijn, and E. Kozica, "Constructive camera pose control for optimizing multiview distributed video coding," in *Proceedings of the 47th IEEE Conference on Decision and Control (CDC '08)*, pp. 3372–3379, Cancún, Mexico, December 2008.

[12] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Fusion-based multiview distributed video coding," in *Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks (VSSN '06)*, pp. 139–144, New York, NY, USA, October 2006.

[13] F. Xiao, J. Liu, L. Sun, and R. Wang, "A new side information estimate based distributed video coding algorithm for wireless multimedia sensor networks," *International Journal of Advancements in Computing Technology*, vol. 4, no. 1, pp. 320–326, 2012.

[14] S. Lijuan, M. Ziping, X. Fu, and W. Ruchuan, "A new video compress algorithm for wireless multimedia sensor network," *Journal of Image and Graphics*, vol. 16, no. 7, pp. 1276–1282, 2011.

[15] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 497–501, 2000.

[16] X.-Y. Hu, E. Eleftheriou, and D. M. Arnold, "Regular and irregular progressive edge-growth tanner graphs," *IEEE*

Transactions on Information Theory, vol. 51, no. 1, pp. 386–398, 2005.

[17] "Mitsubishi Electric Research Laboratories, MERL multi-view video sequences," 2012, ftp://ftp.merl.com/pub/avetro/mvc-testseq.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 763015, 10 pages
doi:10.1155/2012/763015

Research Article

A Virtual-Ring-Based Data Storage and Retrieval Scheme in Wireless Sensor Networks

Xingpo Ma, Jianliang Gao, Weiping Wang, and Jianxin Wang

School of Information Science and Engineering, Central South University, Hunan, Changsha 410083, China

Correspondence should be addressed to Jianliang Gao, gaojianliang@csu.edu.cn and Weiping Wang, wpwang@csu.edu.cn

Received 21 July 2012; Accepted 29 September 2012

Academic Editor: Limin Sun

Copyright © 2012 Xingpo 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-network data storage and retrieval is one of the most important problems in wireless sensor networks. Many schemes have been proposed to solve this problem. However, most of them do not take the frequencies of event and query into consideration. In fact, the frequencies of event and query are very important factors in real applications of in-network data storage and retrieval. In this paper, we introduce a virtual-ring-based data storage and retrieval scheme, which is called VRS, to solve the problem. VRS divides the whole sensor network field into some virtual rings. According to the frequencies, one of the virtual rings is selected out as the rendezvous ring, which plays the role of a bridge between the information consumers and the information producers. Extensive experiments have been done to evaluate the performance of the proposed scheme, VRS. Simulation results show that VRS outperforms the existing work in load balance, delay of data storage and retrieval, and the lifetime of the sensor networks.

1. Introduction

In wireless sensor networks, full data collection [1, 2] is a traditional method for information consumers to get the useful information. However, this method triggers at least two problems. On the one hand, not all the data collected by the sink node are needed by the users, and it is a waste of energy to send the useless event data to the sink node. On the other hand, the method leads to bad load balance performance, because sensor nodes around the sink node have to take much more load than other sensor nodes.

In this paper, we focus on the in-network data storage and retrieval problem in wireless sensor networks, where information consumers (sinks, users) search for the desired event data produced by information producers (sources). There are many such applications. For example, a sensor network might be deployed in a battlefield for the soldiers to know the battlefield situation better. A soldier in the battlefield might be interested in whether there are some hostile tanks nearby, but it is hard for the soldier to be aware of when visibility is low (either due to smoke or at night). However, the sensor nodes that detect the tanks can transmit the information to the soldier. The problem is how can the sensors that detect the tanks send the information to the soldier without knowing where the soldier is. There are mainly two challenges for the problem of in-network data storage and retrieval. First, neither information producer nor information consumer knows where the others are. Second, the sensor nodes are resource constrained and their communication radii are short.

Recently, a lot of in-network data storage and retrieval schemes [3–13] have been proposed. In most of the existing schemes, the frequencies of event and query are less considered, while in fact they play an important role on the efficiency of the schemes. Intuitively, if the frequency of event is high and the frequency of query is low, the data replicas of event should be stored near the information producers. On the contrary, the data replicas of event should be stored near the consumers. Recently, Yu et al. [3] has proposed an in-network storage and retrieval scheme, in which the frequencies of event and query are taken into consideration when selecting the optimal storage node. However, there is only one storage node in the scheme, which leads unbalanced load distribution and can shrink the lifetime of the sensor networks.

There are mainly three contributions in the paper. Firstly, we propose a virtual ring model for the wireless sensor networks. The virtual ring model can be formed in a totally

distributed way. Secondly, we propose a novel in-network data storage and retrieval scheme based on the virtual ring model for wireless sensor networks, where the frequencies of event and query are considered. Thirdly, we have done a lot of experiments to evaluate the performance of our scheme and compare it with the scheme presented in [3], which also considers the event and query frequencies. Simulation results show that VRS outperforms the existing work in load balance, delay of data storage and retrieval, and the lifetime of the sensor networks. Besides, the results also show that VRS can ensure stable lifetime of a sensor network as the sensor nodes increase when the effective time of event data is short. The lifetime of the sensor networks in this paper is defined as the time before the first node dies.

The rest of this paper is arranged as follows. In Section 2, we introduce the related work. Section 3 presents the sensor network model as well as the assumptions of our work. In Section 4, we provide our in-network data storage and retrieval scheme named VRS in detail. The simulation results and analysis are given in Section 5. Section 6 makes a conclusion.

2. Related Work

Existing in-network data storage and retrieval schemes mainly follow three basic approaches. The first one is based on the transmission trail [4–7], the second one is based on the deterministic storage node/nodes [3, 8–13], and the last one is the combination of the former two approaches [14–16].

The main idea of the transmission-trail-based schemes is as follows. Find two intersecting trails: one is for data replication and the other one is for transmitting the queries. Thus the desired data can be retrieved at the crossover point/points of the two trails if they intersect with each other. Rumor routing [4] is an early work of such kind of schemes. Rumor routing creates paths leading to each event. When a query is generated, it will be transmitted in the manner of random walk until it finds the event path. Rumor routing cannot guarantee good successful retrieval rate and is not energy effective, since the directions of the forwarding route are random and dynamic. In double ruling [5], the data replication trail as well as the query transmitting trail is a closed curve, which is projected from a circle on a sphere with the same center as the sphere to a plane. It is obvious that any two circles on the sphere with the same center must intersect, and their projections also intersect. In GLIDER [6], all sensor nodes are divided into tiles using a number of landmarks. The event data are first hashed into a tile, and then they are replicated along three curves towards three neighboring landmarks. The event data in a tile are retrieved by transmitting a query either following a zigzag curve that visits the tile's boundary with each guide or toward the next landmark on the globally planned combinatorial path [6]. In RDRIB [7], a cycle of four axes as the virtual boundary is constructed. All event data are firstly transmitted into the interior of the virtual boundary, and then the data are replicated and retrieved along the path towards the axes using the double ruling technique.

The main idea of the schemes is straight-forward. In those schemes, one or more sensor nodes is/are selected as the storage node/nodes, which is/are known by all the consumers and producers. Thus, data storage and retrieval can be achieved on the storage node/nodes. GHTs [8] are a classic work of this category. One of the shortcomings in GHTs is that it brings a hotspot problem [9]. To solve the hotspot problem, a lot of works were done later. DBAS [10] and DCAAR [11] hash an event type or an attribute of the event type to a grid region, instead of a point. EHS [12] hashes an event type to a point in a face, and the nodes in the face are responsible for storing the event data of that type. Another problem in GHTs is that data retrieval may fail when the nodes nearest to the hashed points move. To solve this problem, C-DCS [13] proposed a cluster-based approach.

Some other schemes were proposed combining the former two basic ideas. A data storage method based on rings was proposed in [14]. Although the load among sensor nodes is balanced using this method, it cannot be independent of time synchronization technique. Besides, multiple sink nodes need to be deployed, which increases the cost of sensor networks. Xie et al. [15] proposed a ring-based multiresolution data storage scheme, where n index trees were built over the virtual rings, and each tree kept several replicas of each data item with different data resolution. It is evident that the cost of replicating data items is huge if data generating frequencies are high. The idea of hashing an event type to a point was also used in [16]. The difference between the scheme proposed in [16] and others of the same kind is that the index instead of the event data is stored on the nodes in the virtual ring around the hashing point. It is easy to see that if the data generating rate is high and the query rate is low, the scheme in [16] is efficient. However, if the data generating rate is low and the query rate is high, the scheme in [16] is not efficient any longer. This is because event data are stored near the detecting nodes, and each query must be sent to the index nodes first and then sent to the storage nodes to get the event data.

The literatures mentioned above do not consider the frequencies of event and query. Yu et al. [3] proposed an adaptive data storage and retrieval scheme, where the frequencies of event and query were considered. However, the scheme only chooses one node as the storage node, thus the hotspot problem still exists.

3. Network Model and Assumptions

In our sensor network model, N sensor nodes as well as one sink node are deployed randomly in a nonbrand field. We assume that the outline of the sensor network field is convex polygon, and the other kind of field shapes is beyond the scope of this paper. We also assume that the approximate location of the central point of the field and the approximate shortest distance from the central point to the boundary of the field are preloaded on each sensor node. We hold that it is a reasonable assumption, because those data can be got by measuring the field before the sensor nodes are deployed. Let L denote the approximate shortest distance from the central

point to the boundary of the sensor network field in the following of this paper.

Packet loss is allowed in our network model. We assume that sensor nodes know their geographic locations. This can be achieved by means of the GPS [17] or some other location service methods [18, 19]. We assume that every sensor node has the same communication radius R and the network is connective. The whole sensor network formulates a connected unidirectional graph $G = \{V, E\}$, in which V is the set of nodes and E is the set of edges. V and E can be described as follows:

$$V = \{v_0, v_1, \ldots, v_N\},$$
$$E = \left\{\left(v_i, v_j\right)\right\} \quad \left(\mathrm{DIS}\left(v_i, v_j\right) < R\right), \tag{1}$$

where $|V| = N + 1$, v_0 is the sink node, and $\mathrm{DIS}(v_i, v_j)$ denotes the distance between v_i and v_j. We assume that any sensor node can be a producer or a consumer, and the role of a sensor node can convert between a producer and a consumer. In the following sections, f_d^i denotes the event data packet generating frequency of producer i, and f_q^i denotes the query frequency of consumer i.

4. Virtual-Ring-Based In-Network Data Storage and Retrieval Scheme

In this section, we present a novel in-network data storage and retrieval scheme, named VRS, in detail.

4.1. Overview of VRS. The main idea of VRS is given as follow. As shown in Figure 1, the sensor network field is divided into many virtual rings, one of which is selected out as the rendezvous ring to mainly take the responsibility of data storage and retrieval. Specifically, the event data will be sent to the rendezvous ring for storing if they are generated outside the rendezvous ring, and the event data will be stored locally if they are produced inside the rendezvous ring. Data retrieval can be achieved by querying the nodes in the rendezvous ring.

The virtual rings can be formed in a totally distributed way. For ease of presentation, we give each of the virtual rings an ID. From the central point of the sensor field, the farther the virtual ring is, the larger the ID of the virtual ring is. ID of the innermost virtual ring is set to zero, and the value deviation between the IDs of two adjacent virtual rings is one. For each node, it first computes the straight-line distance from itself to the central point of the sensor network field, and then it can decide which virtual ring it belongs to by simple calculation. The method for a sensor node S to determine the ID of the virtual ring which it belongs to is as follows.

(i) If the distance from S to the central point of the field is not bigger than $L - L\%R$, S ascertains itself in the virtual ring with ID k if the following inequality holds:

$$R^* k < \mathrm{DIS}(S, \mathrm{CENTER}) < R^*(k+1) \quad (k \geq 0,\ k \in Z), \tag{2}$$

where $\mathrm{CENTER} = (x_{\mathrm{center}}, y_{\mathrm{center}})$ denotes the location of the central point in the sensor network field.

(ii) If the distance from S to the central point of the field is bigger than $L - L\%R$, the virtual ring ID that S locates in is $\lfloor L/R \rfloor$.

Figure 1: Example of the rendezvous ring in the virtual ring model. (The colored one is the rendezvous ring).

In the following parts, we will introduce the way to determine the rendezvous ring and the procedure of data storage and retrieval in detail.

4.2. Rendezvous Ring Determination. Before determining the rendezvous ring, the sink node first collects some information of each sensor node from the sensor network. The information includes node locations, the event frequencies, and the query frequencies. Then the sink node determines the rendezvous ring in a centralized way. Finally, the sink node broadcasts the ID of the rendezvous ring to all the sensor nodes in the sensor network. The strategy for the sink node to determine the rendezvous ring is as follows. The virtual ring with ID k is determined as the rendezvous ring if it satisfies the following equation:

$$\frac{E_{\mathrm{total}}^k}{n^k} = \min\left\{\frac{E_{\mathrm{total}}^i}{n^i},\ 0 \leq i < \left\lfloor \frac{L}{R} \right\rfloor,\ i \in \mathrm{Z}\right\}, \tag{3}$$

where E_{total}^i denotes the total cost of data storage and retrieval for all the sensor nodes in a certain time period T when the ring with ID i is selected as the rendezvous ring and n^i denotes the number of nodes in the ring whose ID is i. Obviously, two metrics have been considered by using this strategy: the total energy consumption and the load balance.

To calculate E_{total}^i, we have

$$E_{\mathrm{total}}^i = \sum_{j=1}^{N} E_j^i \quad \left(0 \leq i < \frac{(L - L\%R)}{R},\ i \in \mathrm{Z}\right), \tag{4}$$

where E_j^i denotes the energy consumption of data storage and retrieval for node j in time period T when the virtual ring with ID i is selected as the rendezvous ring. Thus, the main job to calculate E_{total}^i is to calculate E_j^i, $(0 \le i < (L - L\%R)/R,\ 0 < j \le N,\ i, j \in Z)$.

Since we assume that a sensor node can be an information producer or an information consumer at different time, E_j^i must contain two aspects: $E_{j\text{-storage}}^i$ and $E_{j\text{-retrieval}}^i$, where $E_{j\text{-storage}}^i$ denotes the energy consumption for node j to transmit the event data to ring i when node j plays the role of an information producer during time period T, and $E_{j\text{-retrieval}}^i$ denotes the energy consumption for node j as a consumer to retrieve the desired event data during time period T. Thus we have the following equation:

$$E_j^i = E_{j\text{-storage}}^i + E_{j\text{-retrieval}}^i. \tag{5}$$

$E_{j\text{-retrieval}}^i$ also contains two aspects, which are $E_{j\text{-query}}^i$ and $E_{j\text{-result}}^i$. $E_{j\text{-query}}^i$ denotes the energy consumption of query transmission, and $E_{j\text{-result}}^i$ denotes the energy consumption of transmitting the query results to node j. Then the following equation holds:

$$E_j^i = E_{j\text{-storage}}^i + E_{j\text{-query}}^i + E_{j\text{-result}}^i. \tag{6}$$

In Section 4.4, we will present the way to calculate $E_{j\text{-storage}}^i$, $E_{j\text{-query}}^i$, and $E_{j\text{-result}}^i$ in detail.

4.3. Event Data Storage. In our data storage strategy, at least one copy of each event data should be stored by the nodes in the rendezvous ring. To implement this, all the event data produced by the producers outside the rendezvous ring should be sent to the rendezvous ring. Besides, any sensor node which the event data pass by keeps a replica of the event data. If a producer locates in the rendezvous ring, it just needs to save the data in its own cache.

Now we pay attention to the event data that are produced by producers outside the rendezvous ring. Since we use GPSR [20] as our routing protocol, the destinations of the event data packets should be carefully designed. In this paper, the destinations are designed as follows. If the ID of the ring which any producer is located in is bigger than the ID of the rendezvous ring, the destination of each event data packet should be the central location of the sensor network field. Otherwise, the destination should satisfy the following condition Ω:

$$\text{condition } \Omega\text{: DIS(DEST, CENTER)} > \text{ID}_r * R, \tag{7}$$

where DEST denotes the destination of the event data packet, DIS(DEST, CENTER) denotes the distance between DEST and the central point of the sensor network field, and ID_r denotes the ID of the rendezvous ring. The event data packet must be intercepted by the nodes in the rendezvous ring in this way. In VRS, to reduce and balance the energy consumption, the destination of an event data packet generated in the virtual ring whose ID is smaller than ID_r should satisfy another two conditions as follows.

Condition Ψ. It should be in the same line with the producer and the central point of the sensor network field.

Condition $\Re$. The central point of the sensor network field should not be between the destination and the producer in the line.

When the destination of the event data packet is worked out, the producer sends the event data packet to the destination using GPSR routing protocol, and the event data packet must pass the rendezvous ring. When the event data packet is received by the first node in the rendezvous ring, it will be stored by the node, and stop being transmitted.

4.4. Event Data Retrieval. In the paper, we are interested in real-time event information discovery in wireless sensor networks. If there are the desired event data in a sensor network, they can be found in the rendezvous ring, because at least one replica of each event data is kept on some node/nodes in the rendezvous ring.

Two cases should be considered when executing the query procedure. One case is that the consumers are located outside the rendezvous ring, while the other case is that the consumers are located inside the rendezvous ring. In fact, if the query procedure in the former case is known, it is also clear in the latter case. We only present the query procedure in the former case because of the limited space.

The query procedure for the consumers who are located outside the rendezvous ring consists of three stages. The first stage is to transmit queries towards the rendezvous ring, the second stage is to process the queries in and/or outside the rendezvous ring, and the third stage is to send query results back to the consumers. The way to transmit queries to the rendezvous ring is just the same as the way to transmit the event data to the rendezvous ring, so we will not repeat it. Here, we mainly answer the question how the queries are processed in and outside the rendezvous ring.

The query procedure outside the rendezvous ring is not complicated. When a node outside of the rendezvous ring receives a query, it only needs to check whether the desired event data in its own cache exist. If there are, it sends the event data back to the consumer. Otherwise, the node continues routing the query to the rendezvous ring.

The query procedure inside the rendezvous ring is a little complex. To process queries inside the rendezvous ring efficiently, we have established a polar coordinate system in the sensor field. The polar coordinate system takes the center of the sensor field as the pole, takes the ray, which originates from the center and points to the right-hand side horizontally, as the polaraxis, and takes a counterclockwise direction as the positive direction of the polar angle.

Now, we describe how the queries are processed in the rendezvous ring. Every node in the rendezvous ring has two roles. One role is a leading node and the other role is an ordinary node. These two roles can exchange with each other. Suppose a consumer C who is located outside the rendezvous ring creates a query QRY. C sends QRY towards the rendezvous ring. Suppose that node U in the rendezvous ring receives QRY. Then node U becomes a leading node, and it checks its own cache to see whether there are event data that

meet the requirement of the query. If there are such event data, node U sends the event data to the consumer using GPSR routing protocol [14]. If there are no such event data, U generates a copy of QRY, and broadcasts the copy to its one hop neighbors. Suppose that a node V, which is U's neighbor in the rendezvous ring, receives the copy. It checks its own cache to see whether there are the event data that satisfy the requirement of the copy. If there are such event data, node V sends the event data to node U. Otherwise, V sends a null message to U. We can make the nodes in the rendezvous ring distinguish the query QRY with its copies by setting different message types.

If node U discovers the desired event data from its one hop neighbors, it sends them to the consumer. Otherwise, node U chooses one of its one-hop neighbors as the next hop of the query QRY and sends QRY to the next hop. Suppose that node W is the next hop of QRY. To improve the efficiency of query processing, we specify that node W should satisfy the following two conditions:

(i) it should be in the rendezvous ring;

(ii) it should be as far as possible to node U counterclockwise.

We can select out the node that satisfies the second condition with the help of the polar coordinate system. Node W then becomes a leading node and does the same thing as node U. This procedure continues until the desired event data are discovered or all the nodes in the rendezvous ring are accessed. Figure 2 illustrates the procedure of event data retrieval in VRS.

4.5. Detailed Energy Calculation. We have explained the way to determine the rendezvous ring in Section 4.1. The key step in the procedure of determining the rendezvous ring is to calculate E_j^i, which consisted of $E_{j\text{-storage}}^i$, $E_{j\text{-query}}^i$, and $E_{j\text{-result}}^i$. Suppose that the time period T is a unit time interval. Then $E_{j\text{-storage}}^i$ can be calculated as follows:

$$E_{j\text{-storage}}^i = f_d^j \times s_d \times abs\left(\mathrm{ID}_{\text{ring}}^j - i\right) \times (\delta_{\text{send}} + \delta_{\text{receive}}), \tag{8}$$

where $\mathrm{ID}_{\text{ring}}^j$ denotes the ID of the ring in which node j is located, δ_{send} denotes the energy consumption when sending one byte of data, and δ_{receive} denotes the energy consumption when receiving one byte of data.

To calculate $E_{j\text{-query}}^i$, we need to estimate the average hop count H_{query} for the queries to travel in the rendezvous ring. This depends on where the desired event data will be discovered. Let θ denote the average angle that the queries have rotated around the center of the field in the rendezvous ring before meeting the desired event data. H_{query} can be calculated as follows:

$$H_{\text{query}} = \theta \times \left(i + \frac{1}{2}\right) \times \frac{R}{R} = \theta \times \left(i + \frac{1}{2}\right). \tag{9}$$

We assume that each event data item has an effective time interval τ when it is stored on a node. When the effective time expires, the event data item will be cleared. Let P^l denote the probability that there are some effective event data on node l. Then we have

$$P^l = \begin{cases} f_d^l \times \tau, & \left(f_d^l \times \tau < 1\right), \\ 1, & \left(f_d^l \times \tau \geq 1\right). \end{cases} \tag{10}$$

The average probability that there are some effective event data on a node is

$$P_{\text{average}} = \frac{\sum_{l=1}^{N} P^l}{N}. \tag{11}$$

Let P denote the probability that there are some effective event data among n nodes, and the following equations hold:

$$\begin{aligned} P &= 1 - \left(1 - P_{\text{average}}\right)^n, \\ n &= \left\lceil \log_{1-P_{\text{average}}}^{1-P} \right\rceil. \end{aligned} \tag{12}$$

Let $P = P'$, where $1 - P' < \varepsilon$ (ε is an arbitrarily small positive integer). Then we can obtain an integer n':

$$n' = \left\lceil \log_{1-P_{\text{average}}}^{1-P'} \right\rceil. \tag{13}$$

Thus we can obtain the approximate value of θ as follows:

$$\theta \approx \frac{2\pi n'}{N}. \tag{14}$$

Based on Equations (9) and (14), the value of H_{query} can be worked out.

Then $E_{j\text{-query}}^i$ can be worked out using the following equation:

$$\begin{aligned} E_{j\text{-query}}^i = {} & f_q^j \times s_q \times \left(abs\left(id_{\text{ring}}^j - i\right) + H_{\text{query}}\right) \\ & \times (\delta_{\text{send}} + \delta_{\text{receive}}) + \left(H_{\text{query}} + 1\right) \\ & \times f_q^j \times s_q \times \left(\delta_{\text{send}} + nb_{\text{avg}} \times \delta_{\text{receive}}\right), \end{aligned} \tag{15}$$

Figure 2: Example of the procedure of data retrieval. C is a consumer. The black dots denote the leading nodes, and other dots denote the ordinary nodes.

where nb_{avg} denotes the average number of one hop neighbors of each node. Suppose that the polar angle of node j is θ_j, and the approximate polar coordinates of the point where the query meets the event data are $(\theta_{mt}, (i + 1/2)R)$. Then we have

$$\theta_{mt} = \begin{cases} \theta_j + \theta, & \left(\theta_j + \theta < 2\pi\right), \\ \theta_j + \theta - 2\pi, & \left(\theta_j + \theta \geq 2\pi\right). \end{cases} \tag{16}$$

Let the Cartesian coordinates of the point where the query meets the event data is approximately (x_{mt}, y_{mt}). Then we can make a transformation between the polar coordinates and the Cartesian coordinates:

$$\begin{aligned} x_{mt} &= \left(i + \frac{1}{2}\right) \times R \times \cos(\theta_{mt}) + x_{\text{center}}, \\ y_{mt} &= \left(i + \frac{1}{2}\right) \times R \times \sin(\theta_{mt}) + y_{\text{center}}, \end{aligned} \tag{17}$$

where $(x_{\text{center}}, y_{\text{center}})$ denotes the Cartesian coordinates of the central point of the sensor network field. The approximate hop count H_{result} between the meetingpoint and node j is as follows:

$$H_{\text{result}} = \frac{\sqrt{\left(x_{mt} - x_j\right)^2 + \left(y_{mt} - y_j\right)^2}}{R}, \tag{18}$$

where (x_j, y_j) denotes the Cartesian coordinates of node j.

We can work out $E^i_{j\text{-result}}$ With H_{result} and some other parameters as follows:

$$E^i_{j\text{-result}} = f^j_q \times s_d \times \alpha \times (H_{\text{result}} + 1) \times (\delta_{\text{send}} + \delta_{\text{receive}}), \tag{19}$$

where α is the data compression ratio. When $E^i_{j\text{-storage}}$, $E^i_{j\text{-query}}$, and $E^i_{j\text{-result}}$ are all worked out, E^i_j in Formula (6) can be worked out.

5. Simulation Results and Analysis

We use OMNET++ as our simulator to evaluate the performance of VRS on the application of real-time event data discovery. We compare VRS with ODS [3], which also takes the frequencies of event and query into consideration, mainly on the metrics of network lifetime, load balance, and the delay of data storage and retrieval. In ODS, data are stored and retrieved on the optimal storage node.

5.1. Simulation Scenario. The test scenario is a 400 m × 400 m square field, in which 400–900 sensor nodes are deployed randomly. Other parameters of our simulation are listed in Table 1. Each node chooses two numbers from range (0, $f_{\max}$) randomly as its event frequency and query frequency, respectively.

5.2. Evaluation the Algorithm of Selecting the Rendezvous Ring in VRS. We evaluate the algorithm of selecting the rendezvous ring in VRS by simulation. The initial energy is set to 1 J, and the node number is configured to 500. We take each of the virtual rings as the rendezvous ring, respectively, and test the value of $EN = E^i_{\text{total}}/n^i$, where $0 \leq i < (L - L\%R)/R$ and the lifetime of the network, respectively.

Table 1: The simulation parameter settings.

Parameter	Value
Size of an event data package (s_d)	20 byte
Size of a query package (s_q)	10 byte
Event data compression ratio (α)	0.5
Maximum packet generating frequency ($f_{\max}$)	20 packets/100 s
Radio transmission range (R)	40 m
Energy cost of sending one byte data	0.0144 mJ/byte
Energy cost of receiving one byte data	0.00864 mJ/byte

Figure 3 shows the value of E^i_{total}/n^i when taking different virtual rings as the rendezvous ring, respectively. From Figure 3, we can see that the values of EN decrease as the ring IDs increase. This is because the nodes are deployed randomly and a virtual ring may contain more nodes as the virtual ring becomes farther to the center of the sensor field. Thus, if the value of E^i_{total} does not change too much, the values of E^i_{total}/n^i must decrease as the ring IDs increase.

Figure 4 shows the simulation result of the lifetime of the sensor network when taking different virtual rings as the rendezvous ring. It is clear that the lifetime of the sensor network increases as the ring IDs increase. From Figures 3 and 4, we can see that the smaller the value of EN is, the longer is the lifetime of the sensor network. This reflects that our method on selecting the rendezvous ring is correct to a certain degree.

5.3. Lifetime Evaluation. In this section, we show the simulation results of the lifetime of the sensor network in ODS and VRS. In this experiment, the initial energy is set to 1 J, and the number of nodes is varied from 500 to 900.

Figure 5 shows the lifetime of the sensor network in ODS and VRS with different data effective time. We can see that the lifetime of both two schemes decreases as the node number increases in the most cases. As the number of nodes in the sensor network increases, the load of the storage node in ODS and that of the nodes in the rendezvous ring in VRS increase. The node dying first is often the storage node in ODS or a node in the rendezvous ring in VRS. However, we also note that the lifetime of the sensor network in VRS does not change too much as the node number increases when the effective time is set to 0.2 s. The reason is that, when the data effective time is small enough, almost all the queries meet the desired event data in the rendezvous ring. As the number of nodes increases, the probability for a query to meet the desired event data with a small hop count in the rendezvous ring increases, which can decrease the energy consumption on query processing in the rendezvous ring. However, the cost of transmitting the event data to the rendezvous ring increases as the node number increases. Thus the total energy

Figure 3: The value of EN, which is equal to E^i_{total}/n^i, when taking different virtual rings as the rendezvous ring, respectively.

Figure 4: The lifetime of the sensor network when taking different virtual rings as the rendezvous ring, respectively.

consumption of the sensor nodes in the rendezvous ring may not change too much, which also makes the lifetime of the sensor network not change too much in VRS.

We observe that, as the data effective time decreases, the lifetime of the sensor network in both two schemes decreases. As the data effective time decreases, the queries need to go much farther to meet the desired event data, which increases the energy consumption of data retrieval. However, the energy consumption of data storage does not change as the data effective time changes. Thus, as the data effective time decreases, the total energy consumption increases, and the lifetime of the sensor network becomes shorter.

It is clear that the lifetime of the sensor network in VRS is much longer than that in ODS, because VRS can balance the load much better than ODS. Later on, we will show the load distribution of the sensor nodes in the two schemes, respectively.

Figure 5: Network lifetime in ODS and VRS with different effective time.

5.4. Delay of Data Storage and Retrieval. In this section, we test the performance of VRS on the average delay of data storage and retrieval. We take the hop count that the data have traveled before they rest on a node in the rendezvous ring as the delay of data storage. The delay of data retrieval is represented as the hop count that the query travels from the consumer to the meeting point, plus the hop count that the query result travels from the meeting point to the consumer. In this test, we set the initial energy to 3 J, and set the simulation time to 100 s.

Figure 6 shows the average delay of data storage in ODS and VRS when the data effective time is set to 0.5 s. In fact, when testing the performance on the delay of data storage, it does not matter what value the data effective time is set to, because the data effective time does not affect the track of the event data at all. We can see that the average delay of data storage in VRS is much less than that in ODS. In VRS, the event data only need to be routed into the rendezvous ring, which is much easier than routing the event data to the storage node using GPSR routing protocol. Besides, there are many nodes located in the rendezvous ring, and they just need to store the data generated by themselves locally, which can greatly decrease the average delay of data storage.

Figure 7 gives the average delay of data retrieval in ODS and VRS with different data effective time. As shown in Figure 7, the average delay of data retrieval in VRS is less than that in ODS. In VRS, there are many nodes located in the rendezvous ring, and they discover the desired event data by broadcasting queries to their neighbors. The probability to meet the desired event data in a short hop count is very high in such a way. Besides, the nodes that are located outside the rendezvous ring are much easier to route queries to the rendezvous ring in VRS than most of the nodes in ODS to route the queries to the rendezvous nodes.

Figure 6: Delay of data storage.

Figure 7: Delay of data retrieval with different effective time.

In Figure 7, as the data effective time decreases, the average delay of data retrieval increases both in ODS and VRS. As the data effective time decreases, the probability for a query to meet the desired data in a short hop count decreases, and the query needs to go much farther to hit the desired event data both in ODS and VRS.

5.5. Load Balancing Evaluation. This experiment tests the performance of ODS and VRS on load balance. The initial energy of each sensor node is set to 3 J, the simulation time is set to 100 s, and the data effective time is set to 1 s.

Figure 8 shows the energy consumption of the node that consumes the most energy among all the nodes in the sensor network. The energy consumption by the node that consumes the most energy in ODS is much higher than that in VRS. Theoretically speaking, the node that consumes the most energy must be the storage node in ODS or a node in the rendezvous ring in VRS. Because there is only one storage node in ODS, the load of the storage node in ODS must be much higher than that of any node in the rendezvous ring in VRS. Thus, the energy consumption of the storage node in ODS must be much higher than that of any node in VRS.

Figure 8: Energy consumption of the node that is the biggest energy consumer.

Figure 9 shows the distribution of the energy consumption of all the sensor nodes in ODS and VRS, respectively, when the node number is set to 500. It is obvious that VRS can balance the load much better than ODS.

6. Conclusion

In this paper, we propose a novel in-network data storage and retrieval scheme, named VRS, in wireless sensor networks. VRS divides the whole sensor network field into virtual rings and selects the rendezvous ring from all the virtual rings to mainly take responsibility of event data storage and query processing. VRS takes the frequencies of event and query into consideration when determining the rendezvous ring, which makes it more efficient. We have done a lot of experiments to evaluate the performance of VRS, and compare it with another existing scheme called ODS. The simulation results show that VRS not only outperforms ODS in terms of load balance and network lifetime but also exhibits comparable latency to ODS.

With regards to future works, we plan to design a distributed algorithm of the rendezvous ring selection and make VRS a totally distributed scheme. Another direction is to combine VRS with the virtual coordinate mechanism [21], which can make VRS a GPS-free scheme.

Figure 9: Distribution of energy consumption: (a) ODS and (b) VRS.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grants no. 61173169 and no. 61106036 and the Program for New Century Excellent Talents in University of Ministry of Education under Grant no. NCET-10-0798.

References

[1] M. Liu, J. N. Cao, G. H. Chen, L. J. Chen, X. M. Wang, and H. G. Gong, "EADEEG: an energy-aware data gathering protocol for wireless sensor networks," *Journal of Software*, vol. 18, no. 5, pp. 1092–1109, 2007.

[2] J. B. Liang, J. X. Wang, T. S. Li, and J. E. Chen, "Maximum lifetime algorithm for precise data gathering based on tree in wireless sensor networks," *Journal of Software*, vol. 21, no. 9, pp. 2289–2303, 2010.

[3] Z. Yu, B. Xiao, and S. Zhou, "Achieving optimal data storage position in wireless sensor networks," *Computer Communications*, vol. 33, no. 1, pp. 92–102, 2010.

[4] D. Braginsky and D. Estrin, "Rumor routing algorithm for sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA '02)*, pp. 22–31, September 2002.

[5] R. Sarkar, X. Zhu, and J. Gao, "Double rulings for information brokerage in sensor networks," in *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MobiCom '06)*, pp. 286–297, September 2006.

[6] F. Qing, J. Gao, and J. G. Leonidas, "Landmark-based information storage and retrieval in sensor networks," in *Proceedings of the 25th Conference of the IEEE Communication Society (INFOCOM '06)*, pp. 1–12, April 2006.

[7] C. H. Lin, J. J. Kuo, and M. J. Tsai, "Reliable GPS-free double-ruling-based information brokerage in wireless sensor networks," in *Proceedings of the 29th Conference on Information Communications (INFOCOM '10)*, pp. 1–5, March 2010.

[8] S. Ratnasamy, B. Karp, L. Yin et al., "GHT: a geographic hash table for data-centric storage," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA '02)*, pp. 78–87, September 2002.

[9] K. T. Ma, K. Y. Cheng, K. S. Lui, and V. Tam, "Improving data centric storage with diffuse caching in wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 9, no. 3, pp. 347–356, 2009.

[10] Y. Lai, H. Chen, and Y. Wang, "Dynamic balanced storage in wireless sensor networks," in *Proceedings of the 4th Workshop on Data Management for Sensor Networks (DMSN '07)*, pp. 7–12, September 2007.

[11] R. Biswas, K. Chowdhury, and D. P. Agrawal, "Attribute allocation and retrieval scheme for large-scale sensor networks," *International Journal of Wireless Information Networks*, vol. 13, no. 4, pp. 303–315, 2006.

[12] W. H. Liao and W. C. Wu, "Effective hotspot storage management schemes in wireless sensor networks," *Computer Communications*, vol. 31, no. 10, pp. 2131–2141, 2008.

[13] H. C. Le, H. Guyennet, and N. Zerhouni, "Mobile effect reduction in data-centric storage for wireless sensor networks," in *Proceedings of the 3rd IET International Conference on Intelligent Environments (IE '07)*, pp. 304–311, September 2007.

[14] G. L. Li and H. Gao, "A load balance data storage method based on ring for sensor networks," *Journal of Software*, vol. 18, no. 5, pp. 1173–1185, 2007.

[15] L. Xie, L. J. Chen, D. X. Chen, and L. Xie, "Ring-based multi-resolution data storage for sensor networks," *Journal of Software*, vol. 20, no. 12, pp. 3163–3178, 2009 (Chinese).

[16] W. S. Zhang, G. H. Cao, and T. La Porta, "Data dissemination with ring-based index for wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 7, pp. 832–847, 2007.

[17] E. D. Kaplan, *Understanding GPS: Principles and Applications*, Artech House, Boston, Mass, USA, 1996.

[18] B. Li, Y. He, F. Guo, and L. Zuo, "A novel localization algorithm based on isomap and partial least squares for wireless sensor

networks," *IEEE Transactions on Instrumentation and Measurement*, no. 99, pp. 1–11, 2012.

[19] N. Zhang, X. Cheng, M. Song, and D. Chen, "Time-bounded essential localization for wireless sensor networks," *IEEE/ACM Transactions on Networking*, no. 99, pp. 1–13, 2012.

[20] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom '00)*, pp. 243–254, August 2000.

[21] M. J. Tsai, H. Y. Yang, and W. Q. Huang, "Axis-based virtual coordinate assignment protocol and delivery-guaranteed routing protocol in wireless sensor networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 2234–2242, May 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 386093, 10 pages
doi:10.1155/2012/386093

Research Article

Target Q-Coverage Problem with Bounded Service Delay in Directional Sensor Networks

Deying Li,[1] Hui Liu,[1] Xianling Lu,[1] Wenping Chen,[1] and Hongwei Du[2]

[1] *School of Information, Renmin University of China, Beijing 100872, China*
[2] *Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, China*

Correspondence should be addressed to Deying Li, deyingli@ruc.edu.cn

Received 2 August 2012; Accepted 4 September 2012

Academic Editor: Qun Li

Copyright © 2012 Deying 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.

Maximizing network lifetime is an important objective for the target-coverage problem. With practicable manufacture and cost reduction, directional sensor has been widely used in wireless sensor networks to save energy. In this paper, we address the target Q-coverage (TQC) problem to prolong the network lifetime with bounded service delay constraint in directional sensor networks. We propose a protocol to find a collection of coverage sets that satisfy the coverage quality requirement and the bounded service delay constraint, where the target in each coverage set may not be served continuously but can be served with tolerant service delay. By steering some sensors' directional antennas, our protocol could deal with the changes of network topology or monitoring tasks. Simulation results show that the performance of our protocol is close to the upper bound of the optimal solution.

1. Introduction

Wireless sensor networks (WSNs) could be applied to fire detection, battle field surveillance, environmental monitoring, and so on [1]. In these applications, one of the important tasks is to monitor a set of targets and collect the relevant data in the geographical region. Since sensor nodes are often deployed randomly, target coverage, which aims at covering the specified targets by a subset of the deployed sensor nodes, is a fundamental issue for target monitoring. In practical application, the importance of each target is often different, which requires various qualities of coverage. In other words, targets should be covered with different numbers of sensors to satisfy their QoS requirements, this is called *target Q-coverage* problem. Usually, sensor nodes use omnidirectional antennas to sense the targets. However, directional antennas can switch to a particular direction to transmit farther distances than omnidirectional antennas. With the manufacturing technology improvement and cost reduction, sensors with directional antennas are widely employed in many network models.

There have been a number of works on target coverage problem in directional sensor network [2–7]. Most of them assume that there are sufficient sensors deployed such that all targets always can be covered within the network lifetime. However, the assumption may not hold if some working sensors happen to fail or new targets need to be served. Thus, the previous coverage scheduling protocols may be not feasible in such flexible and scalable network environment. To solve the problem, the authors of [8] steer sensors periodically to serve more targets. Since targets may not be covered continuously, a protocol was proposed to minimize the maximum service delay, which determines how fast an interested message can be detected for a target. In their protocol, the network lifetime is not an optimization objective.

In this paper, we address the *target Q-coverage* problem with maximizing the network lifetime in directional sensor network and design a target coverage schedule to satisfy a bounded service delay. We suppose that at any time, each sensor serves for at most one sector to sense targets. Thus, all targets may not be covered simultaneously. The service delay is defined as the time interval from an event occurring to the event being detected. In target coverage problem, the service delay can be regarded as the tolerant maximal time duration of completing target coverage.

The *target Q-coverage* problem addressed in this paper is to construct a collection of coverage sets and schedule them

to work alternately to prolong the network lifetime, which satisfies the target coverage requirement under given service delay bound $\tilde{\varepsilon}$. We first model the problem as a mixed-integer programming problem. Then, we present the upper bound of the problem and propose a centralized schedule protocol. Finally, we simulate our algorithm. The simulation results demonstrate that the performance of our algorithm is close to the upper bound of the optimal solution.

The rest of this paper is organized as follows. In Section 2, we briefly present the related works in the literature. Then, we introduce the TQC problem under a service delay bound and prove it is NP-hard in Section 3. In Section 4, we show the formulation of this problem, the upper bound of network lifetime and the definition of sensor weight. Moreover, we present the relationship between the service delay bound and the total sector numbers of a sensor. In Section 5, we propose a centralized protocol to solve this problem. The simulation results presented in Section 6 demonstrate the high performance of our protocols in prolonging the network lifetime. We conclude this paper in Section 7.

2. Related Works

Target coverage is a fundamental problem in wireless sensor networks for environment monitoring and surveillance purposes. To maximize the lifetime of a network, a large number of energy-efficient scheduling algorithms have been proposed for omnidirectional sensor networks [9–12]. These works study the 1-coverage, k-coverage, connected coverage, and multiattributes coverage. By organizing sensors into a number of subsets and elaborately managing the duty cycle of subsets in which some sensor nodes are scheduled to sleep or enter a power saving mode while the remaining active nodes keep working, significant energy savings can be achieved.

There are also many strategies to deal with target coverage problem for directional sensor networks [2–7]. The authors of [2] study the partial coverage with directional antenna and firstly propose the *maximum coverage with minimum sensors* (MCMSs) problem, in which coverage, in terms of the number of targets to be covered, is maximized whereas the number of sensors to be activated is minimized. The authors of [3] address the *multiple directional cover sets* (MDCSs) problem by organizing the directions of sensors into a group of nondisjoint cover sets to extend network lifetime. Wang et al. approximate network lifetime for target coverage scheduling problem by randomized approach and genetic algorithm, respectively [4, 6]. In [5], a *weighted centralized greedy algorithm* (WCGA) is proposed to improve the coverage rate for targets. In [7], the authors study the target-oriented scheduling in directional sensor networks, where each target has different coverage quality requirement based on the distance.

Since different targets may need different numbers of sensors to cover to satisfy the various sensing quality requirements, the *target Q-coverage* problem is first defined in [13], and a column generation-based approach is proposed. In [14, 15], a heuristic algorithm and an energy-efficient greedy algorithm are proposed to solve the *target Q-coverage* problem.

There are some existing works which study how to schedule sensors "on" and "off" in omnidirectional sensor network, such that each target may be covered within a bounded time interval and the sensors would not drain their energy too fast [16–18]. Cao et al. [16] propose a distributed protocol to schedule node sleeping which guarantees a bounded-delay sensing coverage while maximizing network lifetime. Gu et al. [17] propose a *unified sensing coverage architecture* (uSense) to address flexible and efficient target coverage problem by three novel ideas: asymmetric architecture, generic switching, and global scheduling. The authors in [18] present a sleep-wake sensor system based on cyclic cellular automata, which is verified to be highly effective in the case of frequent communication failures and sensor failures. Wang and Cao [8] consider the coverage problem in directional sensor network, where sensors are periodically steered to cover more targets. Since the targets may be not covered continuously, the authors propose a centralized protocol and a distributed protocol to minimize the maximum service delay, respectively. However, it does not take into account the network lifetime.

This paper is inspired by the above works and makes a tradeoff between coverage delay and network lifetime in traditional sensor networks. We focus on maximizing network lifetime for *target Q-coverage* problem with a bounded service delay constraint in directional sensor network.

3. Network Model and Problem Description

Given a directional sensor network with m targets $\{a_1, a_2, \ldots, a_m\}$ and n directional sensor $\{s_1, s_2, \ldots, s_n\}$. Each sensor has p ($p \geq 2$) sectors which are centered at the sensor node with a sensing radius r and the sensing angle $\theta = 2\pi/p$. Let $d_{i,j}$ denote the jth sector of the ith sensor, and let $S_{i,j}$ denote the set of targets covered by sector $d_{i,j}$. Target a_l can be covered by sector $d_{i,j}$, if and only if a_l is in the sensing range of sector $d_{i,j}$. We suppose at any time a sensor can serve for at most one sector.

The tolerant service delay is the time interval from an event occurring at a target to the event being detected by a nearby sensor. It determines how fast an interested message can be detected for a target. In the target coverage problem, the service delay can be regarded as the tolerant time duration of completing a target coverage. In directional sensor network without considering the rotation delay, the service delay depends on two parameters: the number of sectors used by each sensor and the time duration of serving a sector at one time. We give an example to show the service delay in Figure 1, where the time duration of serving a sector at one time is set as 1. There are three full target coverage schedules as follows: (1) $s_1 \rightarrow \{a_1, a_2, a_3\}$, $s_2 \rightarrow \{a_4, a_5\}$, $s_3 \rightarrow \{a_6\}$; (2) $s_1 \rightarrow \{a_2, a_3\}$, $s_2 \rightarrow \{a_1, a_5\}$, $s_3 \rightarrow \{a_4, a_6\}$; (3) $s_1 \rightarrow \{a_3\}$, $s_2 \rightarrow \{a_1, a_2, a_5\}$, $s_3 \rightarrow \{a_4, a_6\}$. The maximal service delay of the schedule (1) occurs at sensor s_1 which has 3 sectors for serving target. The delay is $3 \cdot 1 = 3$. With schedule (2), the maximal service delay is 2. With

Figure 1: An example of service delay in directional sensor network.

schedule (3), the maximal service delay is 3, as sensor s_2 has 3 sectors to serve.

We use $\tilde{\varepsilon}$ to denote the service delay. By the above analysis, we can approximate it as the product of the maximal number of the serving sectors of sensors ε and the time duration that a sensor serves each sector w.

Suppose that each sensor has a unified lifetime E if a sensor serves for all the sectors simultaneously. Then if at any time a sensor serves only one sector, the lifetime of the sensor is $p \cdot E$. We use an integer vector $Q = \{q_1, q_2, \ldots, q_m\}$ to represent coverage requirements of targets, where q_l represents the minimum number of sectors that target a_l needs to be covered by.

In the following, we show the definition of the *target Q-coverage* problem bounded with a service delay $\tilde{\varepsilon}$.

Definition 1. A subset D of sectors is called *SDQ*-covset if it can satisfy the coverage requirement, that is, for each target a_l ($l = 1, 2, \ldots, m$), there are at least q_l different sectors in D to cover a_l, and the number of selected sectors from each sensor cannot exceed $\lfloor \tilde{\varepsilon}/w \rfloor$, where w denotes the duration time that a sensor works on each sector.

Definition 2. Bounded service delay for target Q-coverage (SDTQC) problem: given m targets and n sensors in an energy-constrained directional sensor network, where each sensor has p sectors, ($p \geq 2$). The SDTQC problem is to find a collection of *SDQ*-covsets, where there are at least q_l ($q_l \geq 1$) different sectors to periodically cover a_l target ($l = 1, \ldots, m$), such that the network lifetime is maximized, meanwhile ensure that the time duration for each *SDQ*-covset cannot exceed the service delay bound $\tilde{\varepsilon}$. Here, at any time each sensor serves at most one sector.

Now, we show how to calculate the network lifetime. In Figure 1, each sensor has 6 sectors and we suppose the lifetime of each sensor is 2 if a sensor serves for the 6 sectors simultaneously. So if each sensor serves at most one sector at any time, the lifetime of each sensor is $6 \times 2 = 12$.

Supposing the given service delay bound $\tilde{\varepsilon} = 2$ in Figure 1, the network lifetime is computed as follows.

(1) Suppose $w = 1$, so the maximal number of sectors that each sensor served in each coverage set is $2/1 = 2$. By checking the schedules (1)(2)(3), only schedule (2) satisfies the service delay bound. As each coverage is accomplished in 2 time slots and each sensor serves only one sector each time, the lifetime that each sensor consumes with schedule (2) is 2 for every coverage (each coverage task). So if each sensor can construct $12/2 = 6$ coverage sets, then the network lifetime is $T_1 = 6 \times 2 = 12$.

(2) Suppose $w = 0.5$, then the maximal number of sectors that each sensor can serve in each coverage set is $2/0.5 = 4$. Thereby, we can use the schedules (1)(2)(3) to cover the targets in turns, and suppose the schedule order is $(1)(2)(3)(1)\cdots$. Then, with schedule (1) the sensors s_1, s_2, and s_3 cost lifetime 1.5, 0.5, and 1, the working time and service delay of schedule (1) is 1.5; sensors s_1, s_2, and s_3 in schedule (2) cost 1 lifetime for each coverage, the working time and service delay of schedule (2) is 1; the sensors s_1, s_2 and s_3 in schedule (3) cost lifetime 0.5, 1.5, and 1, so the working time and service delay of schedule (3) is 1.5. For one round of schedules as (1)(2)(3), each of the three sensors costs 3 lifetime so they can finish $12/3 = 4$ rounds of schedules and $4 \times 3 = 12$ schedules, that is (1)(2)(3)(1)(2)(3)(1)(2)(3)(1)(2)(3). Then, the network lifetime is $T_2 = (1.5 + 1 + 1.5) \times 4 = 16$.

Theorem 3. *The SDTQC problem is NP-hard.*

Proof. The multiple directional cover sets (MDCSs) problem [3] is a special case of the SDTQC problem. This is because when $q_l = 1$ for each target a_l and the delay bound is 0, the SDTQC problem becomes the MDCS problem. Since the MDCS problem is NP-hard [3], the SDTQC problem is NP-hard. □

4. The SDTQC Problem

In this section, we first formulate the SDTQC problem into a mixed integer programming problem. Secondly, we show the relationship of the given service delay bound and the number of the sectors used by each sensor in each coverage. Thirdly, we present the upper bound of the network lifetime for the SDTQC problem.

4.1. The Integer Programming Problem for the SDTQC Problem. Firstly, we describe the notations for the SDTQC problem as follows.

(i) m: the number of targets in the network.

(ii) n: the number of directional sensors in the network.

(iii) K: the number of *SDQ*-covsets.

(iv) r: the sensing radius of sensor.

(v) p: the number of sectors for each sensor. The angle of sector is $\alpha = 2\pi/p$.

(vi) E: the initial lifetime of sensor if all its sectors are served together, so the lifetime of a sensor is $p \cdot E$ if only one sector is served at any time.

(vii) $\tilde{\varepsilon}$: the bounded service delay.

(viii) s_i: the ith directional sensor, $1 \le i \le n$.

(ix) a_l: the lth target, $1 \le l \le m$.

(x) $d_{i,j}$: the jth sector of the ith sensor.

(xi) S_i: the sector set of sensor s_i.

(xii) A: the set of targets. $A = \{a_1, a_2, \ldots, a_m\}$.

(xiii) S: the set of sensors. $S = \{s_1, s_2, \ldots, s_n\}$.

(xiv) Q: targets' coverage requirement. $Q = \{q_1, q_2, \ldots, q_m\}$, where q_l represents the number of sectors needed to serve for the lth target simultaneously.

(xv) $S_{i,j} = \{a_l \mid a_l \text{ is covered by sector } d_{i,j}\}$, which represents the target set covered by sector $d_{i,j}$, $1 \le i \le n$, $1 \le j \le p$.

(xvi) $C_l = \{d_{i,j} \mid d_{i,j} \text{ covers } a_l\}$, which represents a set of sectors that can cover target a_l, $1 \le l \le m$.

(xvii) D_k: the kth SDQ-covset.

(xviii) t_k: the working time allocated to each working sector of each sensor in SDQ-covset D_k, $k = 1, \ldots, K$.

(xix) Boolean variable $x_{i,j,k}$, $x_{i,j,k} = 1$ if the sector $d_{i,j}$ is in the SDQ-covset D_k, otherwise $x_{i,j,k} = 0$.

The SDTQC problem can be formulated to a mixed integer programming problem as follows:

$$\max \sum_{k=1}^{K} \left(t_k \cdot \max_i \sum_{j=1}^{p} x_{i,j,k} \right), \tag{1}$$

subject to

$$\sum_{d_{i,j} \in C_l} x_{i,j,k} \ge q_l, \quad \forall a_l, \forall D_k, \tag{2}$$

$$\sum_{k=1}^{K} \sum_{j=1}^{p} x_{i,j,k} \cdot t_k \le p \cdot E, \quad \forall s_i \in S, \tag{3}$$

$$\sum_{j=1}^{p} x_{i,j,k} \cdot t_k \le \tilde{\varepsilon}, \quad \forall s_i, \forall D_k, \tag{4}$$

where

$$x_{i,j,k} = 0 \text{ or } 1, \quad t_k \ge 0. \tag{5}$$

Remarks. Constraint (2) guarantees that each target a_l is covered by at least q_l sectors in each *SDQ-covset* D_k.

Constraint (3) guarantees that for each sensor, the sum of lifetimes costed in all *SDQ-covsets* cannot exceed its initial lifetime $p \cdot E$.

Constraint (4) indicates that for each *SDQ-covset* D_k, its working time, which is also its service delay, cannot exceed the given service delay bound $\tilde{\varepsilon}$.

The objective function (1) is to maximize the sum of all *SDQ-covsets*' lifetime. For each *SDQ-covset* D_k, the lifetime should be $t_k \cdot \max_i \sum_{j=1}^{p} x_{i,j,k}$, which is the maximal serving time among all sensors.

4.2. The Upper Bound of Network Lifetime. In this subsection, we present the upper bound of the SDTQC problem.

We know that the maximum lifetime of a sensor network is the smallest lifetime among all sensors. Since the time and the number of serving sectors of each sensor is uncertain, we consider the minimal time of serving each target in the network. Thus, the upper bound of network lifetime is as follows:

$$u = \min_l \left[\frac{|C_l|}{q_l} \cdot \frac{p \cdot E}{w} \cdot \tilde{\varepsilon} \right]. \tag{6}$$

The maximal number of *SDQ-covsets* which each sensor can join is $((p \cdot E/(w)))$. For each target, there are at most $|C_l|$ sectors to cover it. So, if we select q_l sensors(sectors) to cover target a_l each time, there could be $(|C_l|/q_l)$ groups of sensors to compose covsets. If we do not consider the service delay, the upper bound of the network lifetime is

$$\min_l \left[\frac{|C_l|}{q_l} \cdot \frac{p \cdot E}{w} \cdot w \right] = \min_l \left[\frac{|C_l|}{q_l} \cdot p \cdot E \right]. \tag{7}$$

Given a service delay bound $\tilde{\varepsilon}$ for each *SDQ-covset* D_k, the time that *SDQ-covset* D_k completes a coverage is $t_k \cdot \max_i \sum_{j=1}^{p} x_{i,j,k}$ (the notations of t_k and $x_{i,j,k}$ are shown in Section 4.1). Obviously, $\max_k (t_k \cdot \max_i \sum_{j=1}^{p} x_{i,j,k}) \le \tilde{\varepsilon}$.

Then, all sensor's lifetime for covering target a_l to complete the target Q-coverage problem with the service delay bound $\tilde{\varepsilon}$ is calculated as follows:

$$\frac{|C_l|}{q_l} \cdot \frac{p \cdot E}{w} \cdot \max_k \left(t_k \cdot \max_i \sum_{j=1}^{p} x_{i,j,k} \right) \le \frac{|C_l|}{q_l} \cdot \frac{p \cdot E}{w} \cdot \tilde{\varepsilon}. \tag{8}$$

Therefore, the upper bound of network lifetime of the target Q-coverage problem with the service delay bound $\tilde{\varepsilon}$ is shown in (6).

5. The Scheduling Protocol

In this section, we first present the sector weight. Secondly, we propose a centralized scheduling protocol based on the sector weight for the SDTQC problem.

5.1. The Sector Weight. In this subsection, we define the weight W_s of a sector as: $W_s = W_n + W_t + W_c$, where

W_n represents the contribution of the sector to the network, W_t represents the contribution of the sector to each target covered by the sector, and W_c represents the validity of the coverage area by sector.

Let $\widetilde{D}$ denote the temporary set of the sectors which have been selected as *SDQ-covset* at the current serving time, let $\widetilde{A}$ denote the temporary set of targets whose requirements have been satisfied by $\widetilde{D}$, and let $\widetilde{q}_l$ denote the current number of sectors in $\widetilde{D}$, which serve target a_l.

The contribution of a sector to network is defined as follows:

$$W_n\left(d_{i,j}\right) = \frac{\left|S_{i,j} - \widetilde{A}\right|}{\left|A - \widetilde{A}\right|}, \tag{9}$$

where $|S_{i,j} - \widetilde{A}|$ indicates the number of targets which are covered by sector $d_{i,j}$ and whose coverage requirements are not satisfied by $\widetilde{D}$. $|A - \widetilde{A}|$ indicates the number of targets whose coverage requirements are not satisfied.

W_t is defined as follows:

$$W_t\left(d_{i,j}\right) = \max_{a_l \in S_{i,j}} \frac{q_l - \widetilde{q}_l}{|C_l| - \widetilde{q}_l}. \tag{10}$$

W_c is defined as the maximal proportion of the unsatisfied coverage requirement targets in the coverage target set of $d_{i,j}$

$$W_c\left(d_{i,j}\right) = \frac{\left(\left|S_{i,j} - \widetilde{A}\right|\right)}{\left|S_{i,j}\right|}. \tag{11}$$

Thereby, the sector weight is presented as follows:

$$W_s\left(d_{i,j}\right) = \frac{\left|S_{i,j} - \widetilde{A}\right|}{\left|A - \widetilde{A}\right|} + \max_{a_l \in S_{i,j}} \frac{q_l - \widetilde{q}_l}{|C_l| - \widetilde{q}_l} + \frac{\left(\left|S_{i,j} - \widetilde{A}\right|\right)}{\left|S_{i,j}\right|}. \tag{12}$$

Now, as shown in Figure 2(a), we give an example that constructs target Q-coverage based on the sector weight.

There are 9 targets and 4 sensors, each of which has 4 sectors which are labeled in counterclockwise order and the first one begins with the X(positive)-axis shown in Figure 2(b). We discuss 3 sectors' weights: $d_{1,2}$, $d_{1,3}$, and $d_{4,2}$. The coverage requirements of targets are as follows: $q_1 = 1$, $q_3 = 1$, $q_4 = 1$, $q_5 = 2$, and $q_8 = 1$. Suppose the current *SDQ-covsets* $D_k = \varnothing$.

(1) We compute the sector weights of $d_{1,2}$ and $d_{1,3}$. $W_s(d_{1,2}) = (1/9) + \max(1) + 1 = 2(1/9)$ and $W_s(d_{1,3}) = (3/9) + \max(1/2, 1/2, 2/3) + (3/3) = 2$. So, we add sector $d_{1,2}$ into D_k. This indicates the importance of giving priority to the sector that satisfies the requirements of the targets.

(2) We compute the sector weight of $d_{1,3}$ and $d_{4,2}$ as $D_k = \{d_{1,2}\}$. $W_s(d_{1,3}) = (3/8) + \max(1/2, 1/2, 2/3) + (3/3) = 2(1/24)$ and $W_s(d_{4,2}) = (2/8) + \max(1/2, 2/3) + (2/2) = 1(22/24)$. So, we add sector $d_{1,3}$ into D_k because $W_n(d_{4,2}) + W_t(d_{4,2})$ is smaller than $W_n(d_{1,3}) + W_t(d_{1,3})$. This indicates the importance of giving priority to the sector which covers more uncovered targets.

Figure 2: An example of a network with 4 sensors and 9 targets, each sensor has 4 sectors.

5.2. A Centralized Scheduling Protocol with the Sector Weight. In the following, based on the sector weight, we propose a centralized scheduling protocol for the SDTQC problem: (the CSSW-SDTQC protocol). The main idea of the CSSW-SDTQC protocol for selecting a *SDQ-covset* is that each time we choose the sector which has the maximum available lifetime. If there are some sensors with the same available lifetime, we will select the sector with the maximum sector weight. Here, the number of selected sectors from each sensor cannot exceed ε. Then we update the correlated variables and the value of each sector weight. We repeat these steps till we find a *SDQ-covsets* D_K and check whether it is minimal. Then we assign a maximum realizable lifetime t_K to the *SDQ-covsets*. Finally, we repeat this process until the network lifetime cannot be increased.

(1) Let d_i' denote the selected sector number of sensor s_i in the current *SDQ-covset* D_k, S_i denote the total sector set of sensor s_i and $d_{\max}$ denote the maximal sector number from each sensor in the current D_k
(2) Set $E[i] = p \cdot E$ $(1 \le i \le n)$, $K = 0$.
(3) **while** (each target a_j is covered by at least q_j available sectors) **do**
(4) $S_{\text{temp}} = S$, $A_{\text{temp}} = A$, $D_K = \varnothing$, $Q[l] = q_l$ $(1 \le l \le m)$, $d_i' = 0$ $(1 \le i \le n)$
(5) Initialize the sector weight of each sector. $K++$
(6) **while** $A_{\text{temp}} \ne \varnothing$ **do**
(7) Sort sectors in S_{temp} in descending order based on the number of available lifetime. If some sectors have the same value, give the priority to the sector with the maximal value of the sector weight.
(8) Choose the first sector $d_{i,j}$ from the sort set.
(9) $D_K = D_K \cup \{d_{i,j}\}$, $d_i' = d_i' + 1$, $S_{\text{temp}} = S_{\text{temp}} - \{d_{i,j}\}$
(10) **if** $d_i' == \varepsilon$ **then** $S_{\text{temp}} = S_{\text{temp}} - S_i$
(11) **for all** target $a_l \in A_{\text{temp}}$ **do**
(12) **if** a_l is covered by $d_{i,j}$ **then** $Q[l] = Q[l] - 1$
(13) **if** $Q[l] == 0$ **then** $A_{\text{temp}} = A_{\text{temp}} - \{a_l\}$
(14) Update the sector wight of each sector.
(15) **end while**
(16) Minimize D_K, check each $d_{i,j}$ $(d_{i,j} \in D_K)$ whether it is redundant.
(17) Compute $MAX_Lifetime(D_K)$.
(18) $d_{\max} = \text{MAX}(d_i')$
(19) $w' = \min(w, (MAX_Lifetime(D_K)/\varepsilon))$, $t_K = d_{\max} \cdot w'$, $T = T + t_K$
(20) **for all** sectors $d_{i,j} \in D_K$ **do**
(21) $E[i] = E[i] - w'$
(22) **if** $E[i] == 0$ **then** $S = S - S_i$
(23) **end while**
(24) **return** K *SDQ-covsets*: $D_1, D_2, \ldots, D_K$ and the network lifetime T

Algorithm 1: CSSW-SDTQC$(S, A, E, Q, C_l, S_{i,j}, \varepsilon, p, w)$.

The algorithm is formally represented as follows.

The input parameters for this algorithm is below: S is the set of sectors, A is the set of targets, E is the initial lifetime of a sensor, Q is the targets' coverage requirement, and the definitions of $S_{i,j}$, C_l, $\tilde{\varepsilon}$, p, and w are described as in Section 4. $\varepsilon = \lfloor \tilde{\varepsilon}/w \rfloor$. The set S_{temp} records sectors whose corresponding sensors still have residual energy such that these sectors can be used in additional *SDQ-covset*. The set A_{temp} contains the targets that still need to be covered by the current *SDQ-covset* D_K. In Algorithm 1, the step 16 minimizes D_K to make sure that there is no redundant sector in it. In step 17, the *MAX_Lifetime* (D_K) is to compute the maximal lifetime of *SDQ-covset* D_K, which is the minimal available lifetime among all sensors. In step 18, we find the maximal value $d_{\max}$ by comparing each d_i', which is the maximal number of the serving sectors among all sensors, $d_{\max}$ is used to compute the maximal service delay of the current *SDQ-covset* D_K to construct a target Q-coverage. Therefore, for each D_K, $t_K = d_{\max} \cdot w'$.

The centralized scheduling protocol returns k *SDQ-covsets*: $D_1, D_2, \ldots, D_k$ and the network lifetime T.

6. Performance Evaluation

In this section, we will compare our algorithm (CSSW) with the upper bound (UB) of optimal solution in Section 4.2.

6.1. Simulation Setup. In our simulation, we deploy the sensor nodes and targets randomly in a region of 300 m × 300 m area. The initial lifetime of each sensor E is set as 5, and the service delay bound $\tilde{\varepsilon}$ is set as 0.5. To evaluate the performance of our protocol under the different situations, we respectively set the service time of each sector w as 0.25, 0.1, and 0.05.

We evaluate the impact of the following three parameters on the performance of our algorithm: the number of sensors N, the number of targets M, and the sensing range of sensor R. The coverage requirement q_l of each target is randomly set from 1 to 3. By keeping the service delay bound $\tilde{\varepsilon}$ as 0.5, we test our algorithm under the following three cases:

(1) $w = 0.25$, $\varepsilon = 2$, and $p = 2$,

(2) $w = 0.1$, $\varepsilon = 5$, and $p = 5$,

(3) $w = 0.05$, $\varepsilon = 10$, and $p = 10$.

6.2. Performance Comparison. Firstly, we test the impact of the number of sensors on our algorithm and the upper bound. We set the sensing range of each sensor (R) as 30, 30, 20, and the number of targets (TN) as 30, 50, and 30, respectively. The number of sensors increases from 150 to 270 with an increment of 30. As Figure 3 shows, the network lifetime increases with the number of sensors increasing.

Figure 3: Network lifetime versus sensor number.

Secondly, we change the number of targets and compare the network lifetime of our algorithms with the upper bound. We set the sensing range of sensors (R) as 20, 20, and 30, and randomly deploy 200, 150, and 200 sensors (SN) in the simulation area, respectively. The number of targets is increased from 40 to 80 with a step width of 10. As Figure 4 illustrates, the network lifetime decreases with the number of targets increasing.

Thirdly, we evaluated the impact of sensing range on our algorithm and upper bound. We set the number of sensors as 120, 120 and 180, and the number of targets as 50, 30, 50, respectively. The sensing range is changed from 25 to 45 with an increment of 5. As Figure 5 shows, the network lifetime increases with the sensing range increasing.

According to the results of the above simulations, we find that the smaller the working time of each sector is, the larger

Figure 4: Network lifetime versus target number.

the network lifetime is. These results validate the correctness of our example for the network lifetime in Section 3.

Moreover, from the three figures, we can find two phenomena: (1) when fixing the service delay bound, the smaller the serving time w of each sector is, the closer the gap between our algorithm and the upper bound; (2) with the increasing of the sensor number and the sensing range or the decreasing of the number of targets, the network lifetime of our algorithm is farther away from the upper bound.

The reason for the first phenomenon is as follows: when the serving time of each sector is decreased, the maximal number of sectors that each sensor can use is increased, so that the total lifetime of each sensor is increased and is closer to the upper bound.

Figure 5: Network lifetime versus sensing range.

For the second phenomenon, when the sensors become denser, each sector can cover more targets, the upper bound gets larger while the network lifetime of our algorithm cannot increase any more, since the selected *SDQ-covsets* in each round cannot always use the minimal number of sectors.

From the simulation results, we observe that our algorithm is very close to the upper bound, which shows that our algorithm is efficient.

7. Conclusion

In this paper, we study the bounded service delay for *target Q-coverage* problem in directional sensor networks. This network model can be used in various applications, especially when the network topology changes or some nodes maybe fail. We rotate some correlative sensors to make them serve the targets periodically. This will prolong the service delay

which affects the quality of service. Hence, we study the *target Q-coverage* problem with a bounded service delay in directional sensor networks. We first formulate the problem into a mixed integer programming problem. A centralized scheduling protocol is proposed to address this problem based on assigning the sectors to weight. At last, we compare our algorithm with the upper bound to evaluate our algorithm's efficiency.

Acknowledgments

Dr. D. Li was supported in part by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China under Grant 10XNJ032. Ms. X. Lu was jointly supported in part by National Natural Science Foundation of China under Grants 61070191 and 91124001, and Research Fund for the Doctoral Program of Higher Education of China under Grant 20100004110001.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[2] J. Ai and A. A. Abouzeid, "Coverage by directional sensors in randomly deployed wireless sensor networks," *Journal of Combinatorial Optimization*, vol. 11, no. 1, pp. 21–41, 2006.

[3] Y. Cai, W. Lou, M. Li, and X.-Y. Li, "Target-oriented scheduling in directional sensor networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 1550–1558, May 2007.

[4] J. Wang, C. Niu, and R. Shen, "Randomized approach for target coverage schedulingin directonal sensor network," in *Proceedings of the Conference of Engineering Student Society (ICESS '07)*, vol. 4523 of *Lecture Notes in Computer Science*, pp. 379–390.

[5] U.-R. Chen, B.-S. Chiou, J.-M. Chen, and W. Lin, "An adjustable target coverage method in directional sensor networks," in *Proceedings of the IEEE Asia-Pacific Services Computing Conference*, pp. 174–180, 2008.

[6] J. Wang, C. Niu, and R. Shen, "Priority-based target coverage in directional sensor networks using a genetic algorithm," *Computers and Mathematics with Applications*, vol. 57, no. 11-12, pp. 1915–1922, 2009.

[7] H. Yang, D. Li, and H. Chen, "Coverage quality based target-oriented scheduling in directional sensor networks," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, pp. 1–5, May 2010.

[8] Y. Wang and G. Cao, "Minimizing service delay in directional sensor networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '11)*, pp. 1790–1798, Shanghai, China, April 2011.

[9] M. Cardei, M. T. Thai, Y. Li, and W. Wu, "Energy-efficient target coverage in wireless sensor networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '05)*, pp. 1976–1984, March 2005.

[10] H. Liu, P.-J. Wan, and X. Jia, "Maximal lifetime scheduling for sensor surveillance systems with K sensors to one target," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 12, pp. 1526–1536, 2006.

[11] A. I. Erzin, V. V. Zalyubovskiy, and H. Choo, "Maximizing lifetime for a sensor network," in *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication (ICUIMC '08)*, pp. 436–440, February 2008.

[12] L. Wang, D. Li, and Z. Li, "Maximizing lifetime for multi-attributes of target coverage problem in wireless sensor networks," in *Proceedings of the China Wireless Sensor Network Conference (CWSN '08)*, 2009.

[13] Y. Gu, Y. Ji, J. Li, and B. Zhao, "QoS-aware target coverage in wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 9, no. 12, pp. 1645–1659, 2009.

[14] M. Chaudhary and A. Pujari, "Q-coverage problem in wireless sensor networks," in *Proceedings of the International Conference on Distributed Computing and Networking (ICDCN '09)*, vol. 5408 of *Lecture Notes in Computer Science*, pp. 352–330.

[15] H. Liu, W. Chen, H. Ma, and D. Li, "Energy-efficient algorithm for the target Q-coverage problem in wireless sensor networks," in *Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications (WASA '10)*, pp. 21–25, 2010.

[16] Q. Cao, T. Abdelzaher, T. He, and J. Stankovic, "Towards optimal sleep scheduling in sensor networks for rare-event detection," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN '05)*, pp. 20–27, April 2005.

[17] Y. Gu, J. Hwang, T. He, and D. H.-C. Du, "USense: a unified asymmetric sensing coverage architecture for wireless sensor networks," in *Proceedings of the International Conference on Distributed Computing Systems (ICDCS '07)*, pp. 8–18, 2007.

[18] Y. M. Baryshnikov, E. G. Coffman Jr., and K. J. Kwak, "High performance sleep-wake sensor systems based on cyclic cellular automata," in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN '08)*, pp. 517–526, April 2008.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 868943, 9 pages
doi:10.1155/2012/868943

Research Article

Two-Layer Storage Scheme and Repair Method of Failure Data in Wireless Sensor Networks

Yulong Shen,[1,2] Xiaowei Dang,[1] Min Shu,[1] Ning Xi,[1] and Jianfeng Ma[1]

[1] *School of Computer Science and Technology, Xidian University, Shaanxi, Xi'an 710071, China*
[2] *Department of Computer Science, Wayne State University, Detroit, MI 48202, USA*

Correspondence should be addressed to Yulong Shen, ylshen@mail.xidian.edu.cn

Received 20 July 2012; Accepted 8 September 2012

Academic Editor: Fu Xiao

Copyright © 2012 Yulong Shen 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.

Distributed data storage is a key technology in the data collection in wireless sensor networks. The storage scheme based on network coding is applied to data collection in wireless sensor networks because of its high reliability and low overhead. However, it is an open problem to reduce data repair communication overhead caused by the failure of storage nodes. This paper focuses on this issue and presents a two-layer distributed data storage scheme. The lower-layer nodes store the encoded data blocks and the upper-layer nodes store the re-encoded blocks that are responsible for failure data recovery. Based on the two-layer data storage scheme, a data repair method is proposed to decrease the repair communication overhead with only sacrificing lower storage overhead. Compared with MSR, interference alignment-based scheme and group interference alignment scheme, the proposed method has lower repair communication overhead. We prove that the proposed method can reduce the repair communication overhead to $o(1/\sqrt{k})$ times and it is suitable to resource-constrained distributed wireless sensor networks.

1. Introduction

Wireless sensor networks, Internet of things and M2 M make the computer networks extended to things. The data collection in wireless sensor networks perform environmental monitoring, information transmission, data storage, and independent provided service. The collected data are distributed stored at the data storage nodes. In the case of node failure, high reliability and low-overhead distributed data storage is an open problem which has received widespread attention in recent years [1, 2]. With the consideration of resource-constrained property of the storage nodes in the data collection in wireless sensor networks, some effective distributed storage methods are proposed in [3–8]. The distributed storage scheme based on network coding is one of them and it has been researched extensively. It encodes the original data into a number of encoded data blocks and then stores them at different storage nodes. To reconstruct the original data, users only need to get a reasonable number of encoded data blocks (not less than the original data amount). Compared with the data backup method, the network coding technique has advantages of low storage overhead and robustness.

But it also brings the repair problem: When a storage node fails, the data on nonfailure nodes are used to repair the failure data to keep the same level reliability. To repair the failure encoded data blocks, the traditional method is that the newly added storage node collect sufficient encoded data (not less than the original data amount), then decode the encoded data to get the original data, and reencode the original data blocks to recover the failure data. As a result, traditional method causes great repair communication overhead which makes it not optimal for the wireless sensor networks because of the strict limitation of energy. Therefore, communication overhead becomes the primary factor of designing data repair algorithm. In order to reduce the repair communication overhead, some researchers focus their attention on designing data repair algorithms. Among these repair algorithms, the interference alignment repair algorithm presented in [9] and regenerating codes repair algorithm introduced in [7, 8] are outstanding.

For regenerating codes repair algorithm, there are two interesting points on its optimal tradeoff curve: minimum-bandwidth regenerating (MBR) codes and minimum-storage regenerating (MSR) codes [7, 8]. For MSR codes, its core

constructive techniques are interference alignment and the network coding. With the interference alignment technique, MSR codes can reduce the repair communication overhead. The more interference elements are aligned, the more communication overhead is reduced. In order to decrease the interference dimension to a maximum extent, a common eigenvector should be computed and used. But when the interference dimension is more than 3, the complexity of computing the common eigenvector will increase greatly and that will become a significant challenge to the wireless network nodes with limited calculation ability. Dimakis et al. [7] have proved that the repair communication overhead of exact-MSR codes repair algorithm can be achieved with interference alignment only when the code rate is not higher than 1/2. For MBR codes, it can decrease the repair communication overhead to the minimum by sacrificing significant storage overhead. Its storage overhead is about 2 times of MSR. So for the given redundancy, the MBR codes are no longer optimal in terms of reliability.

The interference alignment repair algorithm is based on a hybrid storage model. In this model, the data consist of two parts. One is systematic part which is composed of the original data blocks. The other is nonsystematic part which is composed of the linear combination of the systematic part. Using the interference alignment algorithm to repair the failure data, the newly added node should first collect sufficient data to reduce the interference factors to 3 and then use the interference alignment technique to repair the failure data. The method proposed in [9] can reduce the repair communication overhead, but it is still high.

This paper focuses on optimizing the repair communication overhead of distributed storage. Having analyzed the tradeoff between storage overhead and repair communication overhead and turned the flat storage structure into hierarchical storage structure, this paper proposes a distributed storage method, which is based on two-layer storage structure, and a data repair algorithm. The proposed algorithm decreases the repair communication overhead by sacrificing lower storage overhead. The two-layer storage structure has two kinds of encoded data. They are the encoded network coding data and reencoded data. The lower-layer nodes are responsible for the original data reconstruction. The upper-layer nodes are responsible for failure data recovery. The data repair algorithm based on two-layer storage structure can ensure the restored data and the original encoded data have the recoverability property. Moreover, the two-layer storage structure keeps the storage system in a dynamic steady state all the time. That is, the data reliability of the entire system is stable. The analysis shows that the proposed method can greatly reduce the repair communication overhead by sacrificing lower storage overhead. This paper also proves that the proposed method can reduce the repair communication overhead to $o(1/\sqrt{k})$ times of the traditional method at least and satisfies the basic requirements of sensor networks.

This paper is organized as follows. Section 2 is related works. Section 3 proposes two-layer data storage scheme and data repair method. Section 4 evaluates the repair communication overhead of the proposed method. The conclusion is in Section 5 .

2. Related Work

There is a tradeoff between the communication overhead of repairing failure data and the storage overhead in a distributed data storage system. To ensure the availability of the stored data, some methods to balance the storage overhead and the communication overhead of repairing the failure data are proposed in [1, 7, 9–14]. There are three kinds of recovery algorithms: regenerating codes recovery algorithm [1, 7, 11–14], interference alignment recovery algorithm [1, 9], and tree-structured recovery algorithm based on network topology [10].

For the regenerating codes recovery algorithm, MSR codes and MBR codes [7] can be its representation because of most researchers having a huge interest in them. With the consideration of the minimum storage overhead, MSR codes have minimal storage overhead on a single storage node. MSR codes use the interference alignment technique to reduce the repair communication overhead and the repair process can be seen in [7]. Compared with the traditional backup method, MSR codes repair algorithm can significantly reduce the repair communication overhead by interference alignment technique. But MSR codes have some deficiencies. (1) Using interference alignment technique to reduce the interference dimension to 1 should use the common eigenvectors of all the interfering elements. But the complexity of computing the common eigenvectors will greatly increase with the increase of interference elements. It will be a great challenge to the wireless storage nodes with limited calculation ability. (2) Repair communication overhead of exact-MSR codes repair algorithm can be achieved only when the coding rate is at most 1/2 [7], otherwise its desired results cannot be guaranteed.

MBR codes consider the repair problem with the view of minimum repair communication overhead. Repair communication overhead of MBR codes is equal to its single-node storage overhead. And its repair communication overhead is minimal among all the known data repair algorithms. The meticulous process of building MBR codes is in [7]. Nevertheless, the shortcoming of MBR codes is its great storage overhead. That is because each data block is stored twice. So for the given redundancy, the reliability of MBR codes is no longer optimal.

With minimal storage overhead for a single storage node, the data repair method based on interference alignment reduces repair communication overhead by merging interference elements. The interference alignment technique can reduce the interference elements to 1 by collecting sufficient data. Then the failure data can be repaired by solving linear equations. The details are shown in [9]. The shortcomings of the interference alignment repair algorithm are as follows: (1) Interference alignment repair algorithm is mainly acted on the systematic data. When non-systematic data are failure, they are turned into systematic. (2) Compared with the traditional method, the interference alignment repair algorithm does not decrease the repair communication overhead significantly.

Data repair algorithm based on network topology is named tree-structured data regeneration. This method views

the data repair as recoding of the encoded data blocks. This method is based on the random network coding theory. And it is chiefly used to reduce the repair time. Compared with the traditional method, its repair communication overhead is not reduced.

For these shortcomings of the available data repair algorithms, this paper analyzes the tradeoff between the storage overhead and repair communication overhead, transforms the flat node storage method into hierarchical network coding storage method which is inspired by the tree-structured data regeneration, and then proposes the distributed network coding storage method and repair algorithm.

3. Two-Layer Storage Structure and Data Repair Method

In this section, two-layer storage structure and data repair method are proposed. In the two-layer storage structure, the encoded data blocks are organized into two layers.

3.1. The Construct of Two-Layer Data Storage Structure. There are two types of encoded data blocks in the two-layer storage structure and they constitute the lower-layer and the upper-layer of the two-layer data storage structure, respectively. The network coding data blocks of the original data consist of the lower-layer data blocks; while the upper-layer encoded data blocks are the linear combination of the lower-layer data blocks. The construction process is as follows.

The original data of size M bits are divided equally into k blocks (of size M/k bits each), represented by a k-dimension vector $E = [e_1, e_2, \ldots, e_{k-1}, e_k]^T$. These k data blocks are expended into n encoded blocks by linear network coding, represented by a n-dimension vector $B = [b_1, b_2, \ldots, b_{n-1}, b_n]^T$, that is, $B = AE$:

$$B = AE = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,k} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,k} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}, \quad (1)$$

where A denotes the $n * k$ encoding matrix. The n encoded data blocks are allocated to the lower-layer storage nodes. The corresponding nodes that stored these encoded data blocks are the lower-layer nodes of the two-layer storage structure. From the property of the network coding, we know that the data at lower layer can reconstruct the original data.

The traditional data repair method illustrates that the data repair is essentially the process of solving linear equations. The communication overhead of the data recovery depends on the required data blocks, which are also the number of linear equations. Thus, reducing the communication overhead of data recovery is to decrease the number of equations.

To reduce the number of equations, the upper-layer encoded data scheme is proposed. The upper-layer nodes store the reencoded data from the lower-layer data by

Figure 1: Two-layer encoded data structure with (3, 11, 6) code.

(m, n) code, where m is the number of the upper-layer encoded data blocks. Similar to the original data blocks, the upper-layer encoded data can also be denoted by $C = [c_1, c_2, \ldots, c_{m-1}, c_m]^T$, that is, $C = FB'$:

$$C = FB' = \begin{bmatrix} f_{1,1} & f_{1,2} & \ldots & f_{1,mp} \\ f_{2,1} & f_{2,2} & \ldots & f_{2,mp} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1} & f_{m,2} & \ldots & f_{m,mp} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{mp} \end{bmatrix}, \quad (2)$$

where F is the $m * mp$ encoding matrix of the upper layer encoded data. p is the number of the lower layer data blocks for encoding a data block of the upper layer and $p < k$. Each row of F has p non-zero elements and every column of F has only one nonzero element. Moreover, $m \leq \lceil n/p \rceil$ and $n - mp$ are the number of the lower layer data blocks that are not involved in re-encoding the upper layer data blocks. The upper layer encoded data blocks can be expressed by $c_i = f_i B'$. c_i denotes the upper layer encoded data. f_i is a mp-dimension row vector of F_{m*mp}. B', which is the subvector of B, is a mp-dimension row vector and represents the lower-layer encoded data participated in re-encoding the upper-layer data. Afterwards, the upper layer data blocks are stored at different nodes which are different from the lower-layer nodes, and these nodes are the upper-layer nodes of the two-layer storage structure. For convenience, the two-layer encoded data structure is represented by triple (m, n, k) code. Figure 1 is an example and shows a two-layer encoded data structure with (3, 11, 6) code.

3.2. The Methods of Repairing the Failure Data. The exact repair means that the recovered data are exactly the same as the failures. For all of the upper-layer nodes and the lower-layer nodes involved in re-encoding process, when any of them is failure, the failure data can be exactly repaired. For the rest of the lower-layer nodes, when one of them fails, there are two ways, exact repair and functional repair, to recover the failure data. The functional repair is that the newly generated block can contain the different failure data as long as the system maintains the network coding (n, k)

property. Two types of repairs are presented as follows: exact repair and hybrid repair.

3.2.1. Exact Repair. In exact repair, when the upper-layer nodes failed, the new joined node should only collect the lower-layer data blocks which are responsible for the previous re-encoded block to exact repair the failure data. For the lower-layer node participated in re-encoding, such as node a_1 in Figure 1, if it failed to repair the failure data exactly, the data stored at a_2, a_3, and b_1 are only required.

For the rest of the lower-layer nodes, if one of them failed, the upper-layer data blocks are used to realize the exact repair. The subsets of the upper-layer encoded blocks can exactly repair the failure data blocks stored at the lower layer nodes that are not involved in re-encoding process and the number of elements in each subset is p' (p' is the number of the upper-layer data blocks that are used for data repair. To maintain (n, k) network coding property, $pp' \geq k$). Moreover, any two of the subsets are required to have no intersection to ensure the failure data can be accurately repaired. As a result, m should satisfy $m/p' \geq n - mp$. The left of the inequality represents the number of the subsets, while the other side of the inequality represents the number of the lower layer nodes that are not involved in re-encoding process. Then with the help of the well-designed repair matrix, the failure data blocks can be repaired exactly by collecting the corresponding data blocks of the upper layer and combining them linearly afterwards.

3.2.2. Hybrid Repair. Hybrid repair is a hybrid model of the exact repair and functional repair. The hybrid model is: If any of the upper-layer nodes or the lower-layer nodes involved in reencoding process failed, the data stored at them are exactly repaired; however, if the lower-layer nodes which are not involved in the reencoding process failed, the data stored at them are functionally repaired. The functional repair is actually the linear combination of p' ($p' \geq 2$) blocks of the upper-layer data. For example, as show in Figure 1, if a_{10} failed, the data stored at a_{10} can be functionally repaired by the data stored at b_1 and b_2. With the help of the accurately calculated repair coefficients, the repaired data can preserve the network coding (n, k) property. For convenience, the new joined node storing the recovered data is still named a_{10}. To keep the same level reliability of the data, the functional repair must make sure that the repaired lower layer storage system maintains (n, k) network coding property.

The functional repair of the data stored at the lower layer nodes which are not involved in re-encoding process is a linear combination of data blocks of the upper layer. Therefore, the functional repair can be represented by $r_i = \xi_i C$. r_i is the recovered data and ξ_i is a m-dimension row vector for repair. Each row of $S = [\xi_1 \xi_2 \cdots \xi_{j-1} \xi_j]^T$ has p' nonzero elements, where$j$ values at least $n - mp$ and each column of S has only one nonzero element. To ensure the repaired lower layer storage system maintains (n, k) network coding property, we should work out all the encoding vectors that have (n, k) network coding property with that of the data stored at the lower layer in advance and their number is set to n_1 which is not less than $n - mp$. Then strictly calculate the encoding coefficients of the upper layer data and make the final repair vectors equal to the vectors we calculated beforehand. At this time j is equal to n_1. Therefore, the key problem of functional repair is to calculate these coefficients. Their number is $mp + (n - mp)p'$. Among them mp coefficients are used for encoding the upper-layer data and the others are used to repair the failure data. The repair vector of the data repaired can be expressed in the form of the summation of the product of these coefficients. These coefficients can be calculated with the help of repair vectors of the ultimate repaired data. Data repair can also be represented by $R = S[F(AE)]$ and Rank $(R) \leq j$. To maintain (n, k) network coding property, the rank of R must be greater than $n - mp$. The solution of these coefficients exists when the rank of R equals to that of its augmented matrix by choosing the value of the non-zero elements reasonable. Because of $n - mp < mp + (n - mp)p'$, the solution will not be unique.

4. Evaluation

In this section, we analyze the proposed data repair method and evaluate the communication overhead for data repair, namely, repair communication overhead. We also compare it with the existing data repair method. The communication overhead for data repair is represented by the amount of communication data in the data repair process.

4.1. Repair Communication Overhead Evaluation of the Exact Repair

4.1.1. Repair Communication Overhead. Whether the lower-layer data blocks that are involved in re-encoding process or not brings the difference of the repair schemes and even makes the repair communication overhead not the same. From the repair methods mentioned above, we know that the repair communication overhead of the upper-layer data blocks and the lower-layer data blocks involved in re-encoding process is p, and the repair communication overhead of the rest of the lower layer data blocks is p'. p is not necessarily equal to p'. When we compute the communication overhead of data repair, it is not logical to use an accurate value to represent the repair communication overhead of the entire storage system. Therefore, the expectation value of the repair communication overhead provides a good idea to represent the repair communication overhead of the entire storage system. Let m be x, let p be y, and let p' be z. We assume that each storage node has the same failure probability in the entire storage system. When a storage node failed, the expectation of its repair communication overhead is

$$E(x, y, z) = \frac{(x + xy)y}{n + x} + \frac{(n - xy)z}{n + x}, \tag{3}$$

where x, y, and z are subject to

$$k \leq yz \leq \left\lceil \sqrt{k} \right\rceil^2, \tag{4}$$

$$2 \leq y \leq \left\lceil \frac{k}{2} \right\rceil, \tag{5}$$

$$2 \le z \le \left\lceil \frac{k}{2} \right\rceil, \tag{6}$$

$$x \ge z(n - xy), \tag{7}$$

$$n - xy \ge 0, \tag{8}$$

$$\begin{aligned} E(x, y, z) &= \frac{(x + xy)y}{n + x} + \frac{(n - xy)z}{n + x} \\ &= \frac{xy^2 - xyz + nz + xy}{n + x}. \end{aligned} \tag{9}$$

Set $t = n - xy$, then $n/x = y + t/x$. E will be turned into

$$\begin{aligned} E(x, y, z) &= \frac{xy(y - z + 1) + nz}{n + x} \\ &= \frac{y(y - z + 1) + (y + t/x)z}{1 + y + t/x} \\ &= \frac{y^2 + y + (t/x)z}{1 + y + t/x} \\ &= \frac{y(y + 1 + tz/xy)}{1 + y + t/x}. \end{aligned} \tag{10}$$

From the equation, we know that the value of E is related to x, y, and z. x is similarly related to y and z. As a result, the value of E is determined by y and z. According to the relationship between y and z, we will have the following 3 cases.

Firstly, if $y = z$, then $E = y$. From (4), we know that $y_{\min} = \lceil \sqrt{k} \rceil$. At this time, we can calculate that $E = \lceil \sqrt{k} \rceil$.

Secondly, if $y < z$, then $E > y$. For the given value of z, the range of y is determined by the formula (4), and at the same time the range of x is determined by the formula (8). Within the range of x and y, it can be thought that x, y, and z are independent of each other. For formula (3), the partial derivative of x is:

$$\frac{\partial E}{\partial x} = \frac{n(y^2 - yz + y - z)}{n + x} = \frac{n(y - z)(y + 1)}{(n + x)^2}. \tag{11}$$

When x is maximal, E will be minimal. From (8), we can see that $x_{\max} = n/y_{\min}$ and $y_{\min} = 2$. At this time, $z = \lceil k/2 \rceil$, then $E = 2$.

Thirdly, if $y > z$, then $E < y$. Similar to the $y < z$ situation, when x is minimal, E is minimal. From (7), we can see that $x \ge zn/(1 + yz) = n/(y + 1/z)$. From (4), (5), and (6), we know that if $y = \lceil k/2 \rceil$ and $z = 2$, x will be minimal and at this time $x_{\min} = 2n/(1 + k)$, $E = (1 + k)^2/(6 + 2k)$.

Compare the value of E at these 3 situations and we can see that if $k = 3, 4$, the minimal repair communication overhead is $(1 + k)^2/(6 + 2k)$; if $k \ge 5$, the minimal repair communication overhead is 2.

4.1.2. Evaluation of the Repair Communication Overhead

Theorem 1. *If $k \ge 3$ and the relationship between n and k is $k + 1 \le n \le 2k - 1$, the repair communication overhead of exact repair based on the two-layer storage structure is lower than that of MSR.*

Proof. If $k \ge 5$, the repair communication overhead of exact repair based on the two-layer storage structure is 2, while the MSR is $d/(d - k + 1)$ [7], where d is the number of nodes that are involved in data repair and $k \le d \le n - 1$. Let $f(d) = d/(d - k + 1)$ and its derivative of d is a monotone decreasing function. So $f(d)$ will be minimal when $d = n - 1$ and $f(d)_{\min} = (n - 1)/(n - k) = 1 + (k - 1)/(n - k)$. From the condition that $k + 1 \le n \le 2k - 1$, we can know $f(d)_{\min} = 1 + (k - 1)/(n - k) \ge 2$. Therefore, it turns to be correct that the repair communication overhead of the exact repair based on two-layer storage structure is lower than that of MSR when $k \ge 5$. Moreover, when $k = 3, 4$, the repair communication overhead of the exact repair based on two-layer storage structure is lower than 2. That is to say when n and k satisfy $k+1 \le n \le 2k-1$, the conclusion is also correct. Hence, theorem 1 proves to be correct. □

Theorem 2. *If $k \ge 3$, the repair communication overhead of exact repair based on two-layer storage structure is lower than that of repair method based on interference alignment which is proposed in [9].*

Proof. The repair communication overhead of the basic interference alignment repair algorithm proved by [9] is $(qk - q + 1)/q$ (q is the number of data pieces stored at a single storage node). Let $f(q) = (qk - q + 1)/q = k - 1 + 1/q$. If $k \ge 3$, then $f(q) > 2$. The repair communication overhead of exact repair based on two-layer storage structure is 2 if $k \ge 5$. For $k \ge 5$, the conclusion is correct. When k are 3, 4, the repair communication overhead of exact repair based on two-layer storage structure is 4/3 and 25/14, respectively. Both of them are smaller than 2, so the conclusion is also correct when k are 3, 4.

For the repair algorithm based on group interference alignment, the repair communication overhead is $p + (k - p)/q$ and it is higher than p (p is the number of storage nodes that a data group contains, $2 \le p < k$). If $k \ge 3$, the repair overhead of group interference alignment is $p + (k - p)/q$ and it is higher than 2. Therefore, the Theorem 2 is obviously correct for the repair method based on group interference alignment. □

4.1.3. Numeral Result. In order to compare the repair overhead of these data repair method above and verify the correctness of the conclusions, the numeral result is shown in Figure 2. The histogram in Figure 2 shows, respectively, the repair overhead of MSR repair, exact repair based on two-layer storage structure, repair based on basic interference alignment and group interference alignment, where (n, k) values are as follows: (5, 3), (6, 4), (9, 5), (12, 6), (15, 7), and (17, 8). Comparing the conclusions of this paper with the numeral results displayed in Figure 2, it can be seen apparently that they are consistent.

4.2. Repair Overhead Evaluation of the Hybrid Repair

4.2.1. Repair Communication Overhead. Similar to the exact repair, the communication overhead of hybrid repair can also be represented by its expectation value. Let m be x, let p be y,

FIGURE 2: The repair communication overhead evaluation of exact repair.

and let p' be z, and assume that in the entire storage system, the failure probability of each storage node is exactly the same. When a storage node failed, the expectation of repair overhead is

$$E(x,y,z) = \frac{(x+xy)y}{n+x} + \frac{(n-xy)z}{n+x}, \tag{12}$$

where x, y, and z are subject to

$$k \le yz \le \left\lceil \sqrt{k} \right\rceil^2, \tag{13}$$

$$xy < n, \tag{14}$$

$$z < x, \tag{15}$$

$$2 \le y \le \left\lceil \frac{k}{2} \right\rceil, \tag{16}$$

$$2 \le z \le \left\lceil \frac{k}{2} \right\rceil, \tag{17}$$

$$C(x,z) \ge n - xy, \tag{18}$$

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \tag{19}$$

$$n - xy \le n_1, \tag{20}$$

$$n > n_1. \tag{21}$$

We can gain

$$\begin{aligned} E(x,y,z) &= \frac{(x+xy)y}{n+x} + \frac{(n-xy)z}{n+x} \\ &= \frac{xy^2 - xyz + nz + xy}{n+x}. \end{aligned} \tag{22}$$

Let $t = n - xy$. From (19), we can draw $n! \to \sqrt{2\pi n}(n/e)^n$ [15]. So, $C(x,z)$ is

$$\begin{aligned} C(x,z) &\approx \frac{x^{x+1/2} e^{1/12n}}{\sqrt{2\pi} z^{z+1/2} (x-z)^{x-z+1/2}} \\ &\approx \frac{x^{x+1/2}(1+1/12n)}{\sqrt{2\pi} z^{z+1/2} (x-z)^{x-z+1/2}}. \end{aligned} \tag{23}$$

When $t \to C(x,z)$, the storage overhead will be minimal, and the value of t is $\mathrm{Min}(n_1, C(x,z))$.

From (12), we can see that the value of E is related to x, y, and z. The relationship between x, y, and z can be seen from (18). Formula (13) gives the condition that y, z should be satisfied. When any of x, y, and z is determined, the range of the others will be known. Within the range of them, it can be thought that x, y, and z are independent of each other. For the given value of$y$, the range of x and z can respectively be determined and within the range of x and z, the three variables are independent of each other. For formula (12), the partial derivative of x is

$$\frac{\partial E}{\partial x} = \frac{n(y^2 - yz + y - z)}{(n+x)^2} = \frac{n(y-z)(y+1)}{(n+x)^2}. \tag{24}$$

And the partial derivative of z is

$$\frac{\partial E}{\partial z} = \frac{n - xy}{n+x}. \tag{25}$$

Combine (23) and (18), we can have

$$f(x,y,z) = \frac{x^{x+1/2}(1+1/12n)}{\sqrt{2\pi} z^{z+1/2} (x-z)^{x-z+1/2}} - (n - xy). \tag{26}$$

Formula (18) can be rewritten as $f(x,y,z) \ge 0$. To compute convenient, by the Taylor formula, (26) can be simplified to

$$f(x,y,z) = \frac{A(1+(1/12n))}{B\sqrt{2\pi}} - (n - xy), \tag{27}$$

$$\begin{aligned} A &= 1 + 105.9(x-3) + 126.5(x-3)^2, \\ B &= \left[1 + 11(z-2) + 11.7(z-2)^2\right] \\ &\quad \times \left[1 + 1.5(x-z-1) + 1.375(x-z-1)^2\right]. \end{aligned} \tag{28}$$

Now, we discuss the value of E. According to the relationship between y and z, we will have the following 3 situations.

First, if $y = z$, then $E = y$. From (13), we will know that $y_{\min} = \lceil \sqrt{k} \rceil$. At this time, we can calculate that $E = \lceil \sqrt{k} \rceil$.

Second, if $y > z$, formulas (24) and (25) show us the value of E increases with the increasing of x, z and the values of x, z decrease with the increase of y which are known from (14) and (20). So for the given y, to make E minimal, x and z should be minimal within their range. When $y = a(\lceil \sqrt{k} \rceil + 1 \le a \le \lceil k/2 \rceil)$, $z = \lceil k/a \rceil$ and the minimum of x can be drawn from (27).

Theorem 3. *Under the condition that the relationship between n and k is $n/k \le \sqrt{e}/(\sqrt{e}-1)$, for the given y and z, when $f(x,y,z) \ge 0$, it makes E minimal.*

Proof. x, y are known, so (27) is a function on x and

$$\frac{df}{dx} = y + \left[\ln\frac{x}{x-z} - \frac{0.5z}{x(x-z)}\right]x^{x+1/2}(x-z)^{-(x-z+1/2)}. \tag{29}$$

For $y = a$, $z = \lceil k/a \rceil$, let $g(x) = x/(x-z) - e^{0.5z/x(x-z)}$, $x/(x-z)$ decreases with the increase of x. Formula (14) shows that $x \le [n/a]$, so

$$\frac{x}{x-z} \ge \frac{[n/a]}{[n/a]-\lceil k/a \rceil} > \frac{\lceil n/a \rceil}{\lceil n-k/a \rceil} = \frac{n}{n-k} \ge \sqrt{e},$$

$$\sqrt{e} < \frac{x}{x-z} \le \left\lceil \frac{k}{a} \right\rceil + 1, \tag{30}$$

$$\frac{0.5z}{x(x-z)} < 0.5,$$

then

$$e^{0.5z/x(x-z)} < \sqrt{e}. \tag{31}$$

Therefore, $g(x) > 0$ and $f(x,y,z)$ is an increasing function. To make $(x,y,z) \ge 0$, the value of x must be not less than that of making $f(x,y,z) = 0$. Then, the minimal value of x can be goten on the condition that $f(x,y,z) = 0$.

The minimum value of x can be drawn from Theorem 3 when the relationship between n and k is $n/k \le \sqrt{e}/(\sqrt{e}-1)$. For the given y and z, $f(x,y,z)$ is a function of a, so x can be set as $x = h(a)$. Substitute the value of x into (12) and then compute the derivative of a. Within the range of$a$, if the derivative is greater than 0, the value of a will be $\lceil \sqrt{k} \rceil + 1$, and at this time $z = [\sqrt{k}]$, $x = h(\lceil \sqrt{k} \rceil + 1)$, then the minimal E can be computed. When the derivative is less than 0, the value of a will be $\lceil k/2 \rceil$, and at this time $z = 2$, $x = h(\lceil k/2 \rceil)$. Therefore, the minimum of E can be gained. If it cannot make sure whether the derivative is greater than 0 or not, we can firstly compute the value of a which makes the derivative equal to 0. Substitute the value of a into (12), then we can get the minimum E.

If $y < z$, formulas (24) and (25) show us that the value of E increases with the increase of z and decreases with the increase of x. At the same time, the values of x and z are both decreased with the increase of y. That can be seen from (14) and (20). As a result, for the given y, to make E minimal, x should be maximal within its range, while z should be minimal within its range. When $y = a$ $(2 \le a \le [\sqrt{k}])$, $z = \lceil k/a \rceil$. And when the relationship between n and k is $n/k \le \sqrt{e}/(\sqrt{e}-1)$, f will be an increasing function and the maximum of x will be $[n/a]$.

Then, formula (12) turns into

$$E(a) = \frac{[n/a]a^2 + [n/a]a + n\lceil k/a \rceil - [n/a]\lceil k/a \rceil a}{n + [n/a]}. \tag{32}$$

We can compute the derivative of a by the equality (32). If the derivative is greater than 0, set $a = 2$, and substitute the value of a into (32) afterwards, then the minimum of E can be gained. If the derivative is less than 0, set $a = [\sqrt{k}]$. Then substitute the value of a into (32). The minimum of E can be gained. If it cannot make sure whether the derivative is greater than 0 or not, we can compute the value of a firstly, which makes the derivative equal to 0. Then put the value of a into (32), we can get the minimum E.

The repair communication overhead of hybrid repair based on two-layer storage structure is $\lceil \sqrt{k} \rceil$ when $y = z$. If the repair communication overhead is greater than $\lceil \sqrt{k} \rceil$ in both cases, $y > z$ and $y < z$, the minimal repair communication overhead of hybrid repair is $\lceil \sqrt{k} \rceil$. If the repair communication overhead is smaller than $\lceil \sqrt{k} \rceil$ in any of the two cases: $y > z$ and $y < z$, the minimal repair communication overhead of hybrid repair is at most $\lceil \sqrt{k} \rceil$. In one word, the repair communication overhead of hybrid repair is at most $\lceil \sqrt{k} \rceil$. □

Theorem 4. *The hybrid repair based on two-layer storage structure can reduce the repair communication overhead to $o(1/\sqrt{k})$ times of the traditional data recovery algorithm.*

Proof. From the analysis mentioned above, it can be concluded that the repair communication overhead of hybrid repair based on two-layer storage structure is at most $\lceil \sqrt{k} \rceil$. Compared with the traditional method whose repair communication is k, the hybrid repair reduces the repair communication overhead to $\lceil \sqrt{k} \rceil / k \approx 1/\sqrt{k}$ times of the traditional method. □

4.2.2. Evaluation of the Repair Communication Overhead

Theorem 5. *If the relationship between n and k is $k + 1 \le n \le k + \sqrt{k}$, it can make sure that the repair communication overhead of hybrid repair based on two-layer storage structure is lower than that of MSR.*

Proof. The repair communication overhead of hybrid repair is at most $\lceil \sqrt{k} \rceil$. While the repair over head of MSR is $d/(d-k+1)$, where d is the number of storage nodes that involved in data repair and $k \le d \le n-1$. Let $f(d) = d/(d-k+1)$, and from its derivative, we can know that $f(d)$ is a decreasing function, and the minimum of $f(d)$ can be gotten at $d = n-1$ and at this time $f(d)_{\min} = (n-1)/(n-k) = 1+(k-1)/(n-k)$. With the condition $k+1 \le n \le k+\sqrt{k}$, we can gain $f(d)_{\min} = 1 + (k-1)/[\sqrt{k}] \ge \lceil \sqrt{k} \rceil$, so if $k + 1 \le n \le k + [\sqrt{k}]$, the repair communication overhead of hybrid repair based on two-layer storage structure is lower than that of MSR. □

Theorem 6. *For the repair method based on group interference, if $p \ge \lceil \sqrt{k} \rceil$ (p is the number of storage nodes that a data group contains), its repair communication overhead is higher than that of hybrid repair based on two-layer storage structure.*

Proof. For the repair method based on group interference, its repair communication overhead is $p + (k-p)/q$ which is higher than p (q is the number of data pieces stored at a single storage node). That has been proved by [9]. If $p \ge \lceil \sqrt{k} \rceil$, the repair communication overhead of it is higher

Figure 3: The repair communication overhead evaluation of the hybrid repair.

than $\lceil \sqrt{k} \rceil$. However, the repair communication overhead of hybrid repair based on two-layer storage structure is at most $\lceil \sqrt{k} \rceil$ which is known from the analysis above. Therefore, the theorem is proved. □

For the repair method based on basic interference alignment, since its repair communication overhead is $(qk - q + 1)/q = k - 1 + 1/q > k - 1$, the repair communication overhead of hybrid repair based on two-layer storage structure is lower.

4.2.3. Numeral Result. To verify the correctness of the conclusions, the numeral result is shown in Figure 3. The histogram in Figure 3 shows, respectively, the repair communication overhead of MSR repair, exact repair based on two-layer storage structure, repair based on basic interference alignment and group interference alignment, where (n, k) values are as follows: (5, 3), (6, 4), (7, 5), (8, 6), (10, 7), and (12, 8).

Comparing the conclusions of this paper with the numeral results displayed in Figure 3, it can be seen apparently that they are consistent.

5. Conclusion

This paper analyzes the tradeoff between storage overhead and repair communication overhead in the distributed data storage. We turn the flat storage structure into hierarchical storage structure and present a two-layer distributed data storage scheme to improve the repair communication overhead. Based on the two-layer data storage scheme, a data recovery method is proposed to decrease the repair communication overhead with sacrificing lower storage overhead. The proposed method has lower repair communication overhead than that of MSR, basic interference alignment and group interference alignment schemes. We prove this method reduces the repair communication overhead to $o(1/\sqrt{k})$ times. The proposed scheme is suitable for resource-constrained and node frequent failure distributed wireless sensor networks.

Acknowledgments

The authors gratefully acknowledge inspiring discussions with Weisong Shi and Xiaohong Jiang. This work was supported by the National Natural Science Foundation of China (61100153, 61172068) and the Aviation Science Foundation of China (2009198101, 2010ZC31001, 2010ZC31002, 20101981015, and 2011ZC31006).

References

[1] C. Suh and K. Ramchandran, "Exact-repair MDS codes for distributed storage using interference alignment," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '10)*, pp. 161–165, Dublin, Ireland, June 2010.

[2] N. B. Shah, K. V. Rashmi, P. V. Kumar et al., "Interference alignment in regenerating codes for distributed storage: necessity and code constructions," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2134–2158, 2012.

[3] H. Weatherspoon and J. D. Kubiatowicz, "Erasure coding versus replication: a quantitiative comparison," in *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, pp. 328–338, Cambridge, Mass, USA, March 2002.

[4] J. Kubiatowicz, D. Bindel, Y. Chen et al., "OceanStore: an architecture for global-scale persistent storage," in *Proceedings of the 9th Internatinal Conference Architectural Support for Programming Languages and Operating Systems (ASPLOS '00)*, pp. 190–201, Boston, Mass, USA, November 2000.

[5] I. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[6] S. Rhea, G. Wells, P. Eaton et al., "Maintenance-free global data storage," *IEEE Internet Computing*, vol. 5, no. 5, pp. 40–49, 2001.

[7] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A survey on network codes for distributed storage," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 476–489, 2011.

[8] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, 2010.

[9] Y. Wu and A. G. Dimakis, "Reducing repair traffic for erasure coding-based storage via interference alignment," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '09)*, pp. 2276–2280, Seoul, Republic of Korea, July 2009.

[10] J. Li, S. Yang, X. Wang, X. Xue, and B. Li, "Tree-structured data regeneration with network coding in distributed storage systems," in *Proceedings of the 17th International Workshop on Quality of Service (IWQoS '09)*, pp. 1–5, Shanghai, China, July 2009.

[11] D. Cullina, A. G. Dimakis, and T. Ho, "Searching for minimum storage regenerating codes," in *Proceedings of the Allerton Conference on Control, Computing, and Communication*, 2009.

[12] K. V. Rashmi, N. B. Shah, P. Vijay Kumar, and K. Ramchandran, "Explicit construction of optimal exact regenerating codes for distributed storage," in *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1243–1249, Urbana-Champaign, Ill, USA, October 2009.

[13] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Explicit codes minimizing repair bandwidth for distributed storage," in *Proceedings of the IEEE Information Theory Workshop (ITW '10)*, pp. 1–5, Bangalore, India, January 2010.

[14] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5227–5239, 2011.

[15] P. R. Beesack, "Improvements of Stirling's formula by elementary methods," University of Beograd Publications, Elektrotehničkog fakulteta, Serija: Matematika i Fizika, 1969.