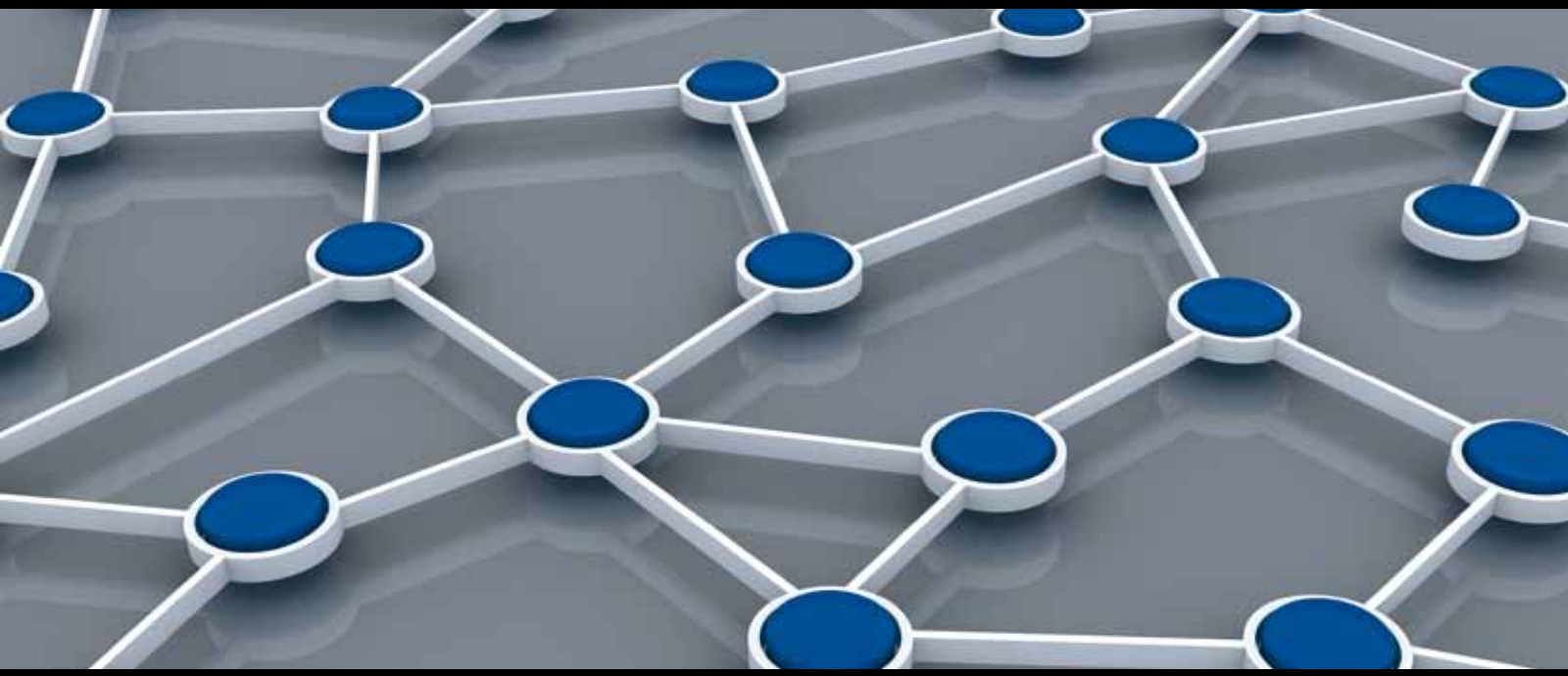
International Journal of Distributed Sensor Networks

Mass Cooperative Transmission and QoS Supported Mechanism in Wireless Sensor Networks

Guest Editors: Gelan Yang, Deguang Le, Yong Jin, and Su-Qun Cao



Mass Cooperative Transmission and QoS Supported Mechanism in Wireless Sensor Networks

International Journal of Distributed Sensor Networks

Mass Cooperative Transmission and QoS Supported Mechanism in Wireless Sensor Networks

Guest Editors: Gelan Yang, Deguang Le, Yong Jin, and Su-Qun Cao

Editorial Board

Contents

Mass Cooperative Transmission and QoS Supported Mechanism in Wireless Sensor Networks, Gelan Yang, Deguang Le, Yong Jin, and Su-Qun Cao
Volume 2014, Article ID 363584, 3 pages

Energy Efficient and High Speed Error Control Scheme for Real Time Wireless Sensor Networks, Ali Barati, Ali Movaghar, and Masoud Sabaei
Volume 2014, Article ID 698125, 9 pages

Distributed Cooperative Algorithm for k-M Set with Negative Integer k by Fractal Symmetrical Property, Gelan Yang and Shuai Liu
Volume 2014, Article ID 398583, 7 pages

Multisensor Based Neutral Function Identification of Solenoid Valve, Yanqing Guo, Yongling Fu, Xiaoye Qi, and Chun Cao
Volume 2014, Article ID 384973, 9 pages

A Topology Construct and Control Model with Small-World and Scale-Free Concepts for Heterogeneous Sensor Networks, Lifang Liu, Xiaogang Qi, Jilong Xue, and Mande Xie
Volume 2014, Article ID 374251, 8 pages

Network-Coding-Based Energy-Efficient Data Fusion and Transmission for Wireless Sensor Networks with Heterogeneous Receivers, Lei Wang, Yuwang Yang, Wei Zhao, Lei Xu, and Shaohua Lan
Volume 2014, Article ID 351707, 13 pages

A Proactive Complex Event Processing Method for Large-Scale Transportation Internet of Things, Yongheng Wang and Kening Cao
Volume 2014, Article ID 159052, 8 pages

Fractal Cross-Layer Service with Integration and Interaction in Internet of Things, Bing Jia, Shuai Liu, and Yongjian Yang
Volume 2014, Article ID 760248, 11 pages

Design Approach Based on EtherCAT Protocol for a Networked Motion Control System, Lei Wang, Muguo Li, Junyan Qi, and Qun Zhang
Volume 2014, Article ID 750601, 15 pages

Cooperative Transmission Mechanisms in Next Generation WiFi: IEEE 802.11ac, Baofeng Ji, Kang Song, Ying Hu, and Hongjun Chen
Volume 2014, Article ID 927192, 12 pages

An Improved Multipoint Relaying Scheme for Message Propagation in Distributed Peer-to-Peer System, Zhiping Liao, Song Liu, and Shengfeng Xi
Volume 2014, Article ID 792814, 10 pages

Optimal Deployment and Scheduling with Directional Sensors for Energy-Efficient Barrier Coverage, Lu Zhao, Guangwei Bai, Yanhui Jiang, Hang Shen, and Zhenmin Tang
Volume 2014, Article ID 596983, 9 pages

Optimal Placement of Wireless Sensor Nodes for Bridge Dynamic Monitoring Based on Improved Particle Swarm Algorithm, Ting You, Huixia Jin, and Peijiang Li
Volume 2013, Article ID 390936, 9 pages

Random Walk Based Location Prediction in Wireless Sensor Networks, Zhaoyan Jin, Dianxi Shi, Quanyuan Wu, and Huining Yan
Volume 2013, Article ID 691042, 9 pages

Predictive Analytics by Using Bayesian Model Averaging for Large-Scale Internet of Things, Xinghui Zhu, Fang Kui, and Yongheng Wang
Volume 2013, Article ID 723260, 10 pages

Human Moving Pattern Recognition toward Channel Number Reduction Based on Multipressure Sensor Network, Zhaoqin Peng, Chun Cao, Jiaoying Huang, and Wentao Pan
Volume 2013, Article ID 510917, 10 pages

Peer Selection Strategy Using Mobile Agent and Trust in Peer-to-Peer Streaming Media System, He Xu, Han-Chen Huang, Ruchuan Wang, and Jun Dong
Volume 2013, Article ID 791560, 15 pages

An Improved Centralized Cognitive Radio Network Spectrum Allocation Algorithm Based on the Allocation Sequence, Jianli Zhao and Jinsha Yuan
Volume 2013, Article ID 875342, 13 pages

A Microseismic/Acoustic Emission Source Location Method Using Arrival Times of PS Waves for Unknown Velocity System, Longjun Dong and Xibing Li
Volume 2013, Article ID 307489, 8 pages

Monte Carlo Based Personalized PageRank on Dynamic Networks, Zhang Junchao, Chen Junjie, Jiancheng Song, and Rong-Xiang Zhao
Volume 2013, Article ID 829804, 8 pages

Distributional Fractal Creating Algorithm in Parallel Environment, Shuai Liu, Weina Fu, Huimin Deng, Caihe Lan, and Jiantao Zhou
Volume 2013, Article ID 281707, 8 pages

High Reliable Relay Selection Approach for QoS Provisioning in Wireless Distributed Sensor Networks, Yong Jin
Volume 2013, Article ID 216478, 11 pages

A Novel PROMSIS Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks, Shengmei Liu, Zhongjiu Zheng, and Su Pan
Volume 2013, Article ID 846791, 6 pages

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 363584, 3 pages
http://dx.doi.org/10.1155/2014/363584

Editorial

Mass Cooperative Transmission and QoS Supported Mechanism in Wireless Sensor Networks

Gelan Yang,[1] Deguang Le,[2] Yong Jin,[2] and Su-Qun Cao[3]

[1] *College of Information Science and Engineering, Hunan City University, Yiyang 413000, China*
[2] *School of Computer Science & Engineering, Changshu Institute of Technology, Changshu 215500, China*
[3] *Faculty of Mechanical Engineering, Huaiyin Institute of Technology, Huai'an 223003, China*

Correspondence should be addressed to Gelan Yang; glyang@mail.ustc.edu.cn

Received 16 April 2014; Accepted 16 April 2014; Published 28 May 2014

Many new massive services cause the data scale and type of WSNs (wireless sensor networks) to grow so fast, which include industrial, military, and natural environmental monitoring, real-time multimedia communication, and target tracking system. However, traditional data transmission protocols and architecture may not be suitable for all applications of WSNs for the diversity of QoS guarantees, mass transmission mechanism, and mobility support. These bring about significant challenges and design considerations for WSNs in complicated wireless environment. We have accepted a few papers that address the state of the art and the future directions of these research and application areas for mass cooperative transmission and QoS supported mechanism in ubiquitous wireless sensor networks in this issue.

These papers mainly focus on providing the recent advances on cooperative communication and diversity of QoS guarantee schemes based on massive data of wireless sensor networks or wireless multimedia sensor networks. Authors have described the novel mechanisms, as well as state-of-the-art contributions on key issues like detecting similar duplicate records, dissemination, big data management, big data processing, reliable and real-time transmission, and so on.

"*Energy efficient and high speed error control scheme for real time wireless sensor networks*" by A. Barati et al. proposes a novel energy-efficient and high speed error control scheme based on the redundant residue number system (RRNS) allowing real-time application of WSNs. These advantages therefore make the proposed method more desirable for real-time wireless sensor network applications where reliability, low energy consumption, and high speed operations are absolute necessities.

"*A novel PROMSIS vertical handoff decision algorithm for heterogeneous wireless networks*" by S. Liu et al. proposes the novel preference ranking organization method by similarity to ideal solution (PROMSIS) vertical handoff algorithm for heterogeneous wireless networks. Its key idea includes the preference structure of the PROMETHEE and the concept of Euclid distance of the TOPSIS.

"*Distributional fractal creating algorithm in parallel environment*" by S. Liu et al. transforms the classic fractal creating algorithm to a distributional algorithm and applies it into the parallel environment. This paper uses a generalized $3x + 1$ function and an approximate generalized $3x + 1$ function to experiment and validate the effectiveness of this method.

"*Fractal cross-layer service with integration and interaction in Internet of Things*" by B. Jia et al. creates a novel fractal intelligent cross-layer service platform, which contains the architecture, foundational models, and algorithms such as service ontology model, semantic description language, and management strategy. Experimental result shows its benefit in interoperability and well-understanding.

"*Distributed cooperative algorithm for k-M set with negative integer k by fractal symmetrical property*" by G. Yang and S. Liu uses the computed points' value to avoid computational waste and uses fractal theory to divide Julia set with different

iteration times in order to decrease the computational time of escape time algorithm. This paper uses k-M set with negative integer to experiment and validates the effectiveness of the novel algorithm by using the fractal symmetrical properties, which is given by strict proofs.

"*Monte Carlo based personalized PageRank on dynamic networks*" by Z. Junchao et al. designs the Monte Carlo based incremental method for personalized PageRank computation. The authors do a random walk starting from each node and save the performed walks into a fingerprint database. Then, the fingerprint database was updated in a fixed time interval. When a query is issued by a user, the personalized PageRank vector was estimated by the proposed approximation algorithm.

"*A microseismic/acoustic emission source location method using arrival times of PS waves for unknown velocity system*" by L. Dong and X. Li describes MS/AE (microseismic/acoustic emission) source location method using P-wave and S-wave arrivals for unknown velocity system (PSAFUVS), in order to eliminate the location error of MS/AE monitoring systems.

"*Random walk based location prediction in wireless sensor networks*" by Z. Jin et al. studies the problem of future location prediction and assumes the location histories of mobile objects as a rating matrix and then uses a random walk based social recommender algorithm, which was used to predict the future locations of mobile objects.

"*Optimal deployment and scheduling with directional sensors for energy-efficient barrier coverage*" by L. Zhao proposes the efficient sensor deployment (ESD) problem and energy-efficient barrier coverage (EEBC) problem for directional sensor networks. The deployment model for the distribution of sensor locations to analyze whether a target area can be barrier covered was described. In addition, the relationship between the probability of barrier coverage and network deployment parameters was examined.

"*High reliable relay selection approach for QoS provisioning in wireless distributed sensor networks*" by Y. Jin studies the reliable adaptive relay selection approach and adaptive QoS supported algorithm based on Markov chain model, which considered the different packet states and error control algorithm assignment. The analyses results show that this proposed relay selection approach could perform better in terms of saturation throughput, reliability, and energy efficiency than other traditional approaches. Specially, the quality of real-time multimedia streaming could be improved significantly such as the decodable frame ratio and delay.

"*Multisensor based neutral function identification of solenoid valve*" by Y. Guo et al. investigates the importance of the solenoid valve identification for condition monitoring need. Therefore, a multisensor based detecting system is set up for output pressure sensing of solenoid valves. Aiming at simplifying the recognition process, linear correlation analysis is employed for feature extraction while support vector machine classifier is employed for automatic identification. With different types of valves applied to the measurement, experiment outcome reveals the high identification accuracy, which shows the feasibility of this system for industrial application.

"*Human moving pattern recognition toward channel number reduction based on multipressure sensor network*" by Z. Peng et al. emphasizes on moving state acquisition based on foot pressure sensing. An insole-based plantar pressure measuring method is developed based on multisensing measurement. For establishing the foot pressure acquisition system, gait recognition algorithm based on physiological parameter feature extraction and human motion recognition algorithm based on support vector machine are studied. Based on experimental outcomes, a faster and simplified system is obtained. Experiments have been conducted on foot pressure sensing system to confirm the validity of the algorithms.

"*A proactive complex event processing method for large-scale transportation Internet of Things*" by Y. Wang and K. Cao proposes a proactive complex event processing method named PROCEP for large-scale transportation Internet of Things. A parallel Markov decision processes model is designed based on a novel multilayered adaptive dynamic Bayesian model. The experimental evaluations show that this method can support proactive complex event processing well in large-scale transportation Internet of Things.

"*Predictive analytics by using Bayesian model averaging for large-scale Internet of Things*" by X. Zhu et al. proposes a dynamic Bayesian model averaging method for high-accuracy prediction analytics in large-scale IoT applications. Gaussian mixture models and expectation-maximization inference are used for basic Bayesian prediction. A novel dynamic Bayesian model averaging method is proposed based on event context clustering. Simulation experiments show that the proposed prediction analytic method has better accuracy compared with traditional methods.

"*An improved multipoint relaying scheme for message propagation in distributed peer-to-peer system*" by Z. Liao et al. proposes an algorithm of LMPR. It is a methodology for improving MPR for message propagation in distributed unstructured P2P system. And a mechanism of specified relay-list to further reduce the number of its retransmitting of message is added. Simulation results show that the scheme of LMPR can improve the traditional system in the aspect of message redundancy, network overhead, and fluctuation.

"*Design approach based on EtherCAT protocol for a networked motion control system*" by L. Wang et al. presents a new design approach in order to improve a special networked motion control system based on EtherCAT protocol. And details of the testing platform including software and hardware are also given. A set of experiments are conducted to evaluate the crucial performance. Experimental results show that the design approach can be widely extended to other automation application scenarios.

"*Optimal placement of wireless sensor nodes for bridge dynamic monitoring based on improved particle swarm algorithm*" by T. You et al. proposes a methodology for optimal placement of wireless sensor networks for dynamic monitoring of bridges. The optimizing sensor location is achieved by using the improved wavelet particle swarm algorithm, based on the relationship between the bridge modal and the energy consumption of wireless network. According to

the experimental results, the proposed method can reflect the dynamic performance of the test bridge well, which can be used for dynamic monitoring of wireless sensors in various bridge types.

"*Peer selection strategy using mobile agent and trust in peer-to-peer streaming media system*" by H. Xu et al. describes a peer selection strategy aiming at improving the quality of service (QoS) for peer-to-peer (P2P) streaming media system, on the basis of mobile agent and trust. Also a new architecture to simulate P2P network is presented, which uses agent technology to OverSim platform, simulation results show that the proposed strategy improves the streaming QoS.

"*Network-coding-based energy-efficient data fusion and transmission for wireless sensor networks with heterogeneous receivers*" by L. Wang et al. proposes a multirate network coding scheme for improving the energy efficiency of wireless sensor networks. A network topology based on a five-layer network model is established from the given sensor networks. A simulated instance is conducted to evaluate the performance of the proposed scheme. Experimental outcomes show that 9% of transmission overhead can be saved.

"*A topology construct and control model with small-world and scale-free concepts for heterogeneous sensor networks*" by L. Liu et al. introduces a topology model with small-world and scale-free concepts for heterogeneous sensor network. And the topology evolution algorithm is designed as well. The simulation results show that the proposed topologies have small-world, scale-free feathers and a better performance in improving energy efficiency and enhancing network robustness.

"*An improved centralized cognitive radio network spectrum allocation algorithm based on the allocation sequence*" by J. Zhao and J. Yuan establishes the mathematical model for the cognitive radio spectrum allocation based on the graph theory and the existing spectrum allocation algorithms based on the distributed cognitive radio networks and the centralized cognitive radio networks. This paper proposed the central cognitive radio network spectrum allocation algorithm based on the allocation sequence. The simulation results show that the improved algorithm can significantly reduce the allocation time of the algorithm compared to the existing graph theory algorithms.

"*Cooperative transmission mechanisms in next generation WiFi: IEEE 802.11ac*" by B. Ji et al. proposes an enhanced NAV transmission mechanism, which not only needs little modifications to current standard draft but also can achieve performance improvement significantly. Theoretical analysis is also performed on the proposed mechanism and the formulas of the achieved gains are also derived. Numerical results confirm the validation of our theoretical analysis and further substantiate that the proposed scheme obtains obvious throughput gain over the conventional mechanism.

Acknowledgments

Thanks are due to the authors of the special issue for their contributions and to the reviewers for their valuable comments on the submissions. Dr. Gelan Yang acknowledges the support by the Scientific Research Fund of Hunan Provincial Education Department (Grant no. 12B023).

Gelan Yang
Deguang Le
Yong Jin
Su-Qun Cao

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 698125, 9 pages
http://dx.doi.org/10.1155/2014/698125

Research Article

Energy Efficient and High Speed Error Control Scheme for Real Time Wireless Sensor Networks

Ali Barati,[1] Ali Movaghar,[2] and Masoud Sabaei[3]

[1] *Department of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran*
[2] *Department of Computer Engineering, Sharif University of Technology, Tehran, Iran*
[3] *Computer Engineering and Information Technology Department, Amirkabir University of Technology, Tehran, Iran*

Correspondence should be addressed to Ali Barati; abarati@iaud.ac.ir

Received 5 October 2013; Revised 24 January 2014; Accepted 2 February 2014; Published 26 May 2014

Academic Editor: Yong Jin

Reliability and energy consumption are two of the main constraints in wireless sensor networks (WSNs). In this paper, a novel energy efficient and high speed error control scheme is introduced that is based on the Redundant Residue Number System (RRNS) allowing real-time application of WSNs. The proposed approach employs a new 3-moduli set $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1\}$ and an efficient reverse converter which relies on the Mixed Radix Conversion (MRC) algorithm and achieves significant improvements both in terms of conversion delay and hardware design. In order to obtain error controllability, two-redundant-moduli set $\{2^{3n+1}-1, 2^{4n+1}-1\}$ is also added to the main 3-moduli set. The theoretical results backed by simulation tests confirm that the solution put forward in this paper outperforms popular error control methods for WSNs in terms of error controllability, energy efficiency, and reduction of end-to-end delay.

1. Introduction

Wireless sensor networks with a large number of multifunctional low power sensor nodes have a broad range of applications, such as battlefield surveillance, environmental monitoring, disaster relief, and healthcare [1–5]. The main task of wireless sensor networks is to collect the sensed data from environment and send them back to the sink node for further processing. Because of the dynamic, lossy, and low-power nature of sensor networks, two of the most important constraints that designers have to overcome in these networks are energy consumption and reliability [6]. Moreover, high speed operation is another critical issue in real time wireless sensor networks applications. In-network data aggregation which is proposed for energy conservation [7] and error control schemes such as Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC) is used to improve reliability. In ARQ-based schemes, the receiver must detect lost packets and then request the sender to retransmit packets [8]. In FEC codes, to transmit data from sensor to the sink node, each aggregator must decode the received data, aggregate them with new data, and encrypt them again before sending to the sink node. These operations cause a significant end-to-end delay and high energy consumption, which leads to a decrease in the network lifetime. Therefore, it is necessary to develop a new solution that offers both low power consumption and guaranteed reliability as well as low end-to-end delay in real-time wireless sensor networks.

In this paper, authors utilize RRNS to design a reliable, energy-efficient, and high speed algorithm for real time applications of wireless sensor networks. The proposed RRNS-based model offers desirable features including robust security, an enhanced error detection, and correction capability. Furthermore, the solution allows parallel operations and thus making real-time functionality as well as the use of low-power components possible. By employing a 3-moduli set $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1\}$ and exploiting the MRC algorithm, an efficient reverse converter is designed. Two redundant-moduli set $\{2^{3n+1}-1, 2^{4n+1}-1\}$ is subsequently added to the mentioned three main RNS moduli to create a 5-moduli RRNS set $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1, 2^{3n+1}-1, 2^{4n+1}-1\}$. This

inclusion provides the new system with an effective error control capability.

The rest of the paper is organized as follows: a few of related works are discussed in Section 2. In Section 3 a full description of the proposed scheme is presented. This is followed in Section 4 by a detailed explanation of an efficient reverse converter design and derivation of the necessary expressions. In Section 5, error controllability of the proposed scheme is evaluated through simulation. Finally, the paper is concluded in Section 6.

2. Related Works

As was discussed in the introduction section, error control methods in WSNs are divided into two general categories, namely, Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC). The ARQ method is used only for error detection and once a corrupted data is sensed, a request for information retransmission is made. This approach is based on Cyclic Redundancy Check (CRC). To set specifications for a CRC code it requires a generator of polynomial to be defined. This polynomial resembles the divisor in a polynomial long division, which takes the message as the dividend and in which the quotient is discarded and the remainder becomes the result, with the important distinction that the polynomial coefficients are calculated according to the carry-less arithmetic of a finite field. In order to make checksum, attach a polynomial generator to the packet before it is transmitted. At the receiver, if the division of the bit stream by $G(x)$ generates no remainder, then this is proof that transmission has been free of error. For each error-free packet received a positive acknowledgment (ACK) is sent back by the receiver and for each corrupted packet received a negative acknowledgment (NACK) is sent back by the receiver [9, 10].

By adding redundancy to the sent data, the FEC codes let the receiver detect the errors and correct them by itself. One of these codes is BCH which is represented by (n, k, t), where "n" refers to the length of the total sent bits, "k" denotes the length of the main data, and "t" stands for the maximum number of the errors that could be corrected.

In the age of modern wireless communications, the need for end-to-end reliable data transfer is growing incessantly. Fortunately, to ensure error free transmission, a variety of popular error control techniques devoted to error detection and correction have been published in the literature. In what follows, a number of these reported methods are examined to ascertain how they achieve their objectives. In [6, 11] for example, we note that error control schemes in Bluetooth sensor networks were explored, that is, a low cost wireless technology designed to facilitate the formation of ad hoc networks. In [11], authors studied an energy efficient model for adaptive and custom error control scheme and also investigated energy consumption and reliability constraints of WSNs. For custom coding, they introduced new packet types using some error control strategies. These new packets exploited CRC for error detection (without ARQ), BCH code with and without CRC, and the Hamming code with and without CRC. They also unveiled two adaptive techniques and a packet selection strategy that was based on channel state. The result of their analysis showed that error recovery was energy efficient provided that the channel conditions were known and a suitable control scheme was employed. This meant that with good channel conditions, there was no need to vary energy consumption and the use of packets with little or no error protection resulted in an efficient scenario. However, for low values of signal to noise ratio (SNR), the BCH packet offered the most effective tool due to its ability to correct more errors, in spite of higher energy consumption. Thus, under the circumstances where the error occurrence probability was high, a more robust error recovery strategy had to be employed for the network.

Notice that the results obtained in [11] were only applicable to networks, where designers had access to information on the channel conditions. Moreover, power consumption of the WSN had to be taken into account. It is clear that in situations where channel conditions are very vulnerable a robust error control scheme is the best choice. But the feasibility of this choice must be considered against energy constraints of the given WSN. Another work that also focused on reliability in Bluetooth sensor networks was presented by Khodadoustan et al. [6] which included an attempt to establish reliability/energy tradeoff in Bluetooth error control scheme. The paper investigated a combination of ARQ and a particular CRC code known as CRC-CCITT and also carried out an analysis of the data packets by using different BCH codes. Furthermore, by conducting simulations, authors examined the proposed coding techniques' performance as a result of changes in the number of hops and their effect on SNR. The results of this study can help network designers to decide on suitable packet types and error control schemes. Turning to [8, 20], we find that researchers' main interests were in reliability for underwater wireless sensor networks (UWSNs). Sensors in underwater wireless sensor networks are often mobile and the channels are acoustic. Such WSNs have low available bandwidth, large propagation delays, high error probability, and highly dynamic network topology [20]. Thus, error recovery is a major challenge in underwater wireless sensor networks.

Guo et al. in [8] suggested the use of network coding in multipath for UWSNs to achieve efficient error recovery in the presence of high error probability. An analytical evaluation of the performance of this scheme against several other error-recovery systems revealed that this method was very efficient in terms of error recovery and energy consumption.

Segmented Data Reliable Transfer protocol (SDRT) was another solution that authors reported for UWSNs in [20]. SDRT was a hybrid approach based on ARQ and FEC techniques which adopted efficient erasure codes whereby the packets were transmitted block-by-block and hop-by-hop. Moreover, a mathematical model was subsequently put forward which achieved further energy saving and enabled designers to set an appropriate size for the block for each packet. By using this model, the number of needed packets could be accurately predicted.

Iyer et al. in [21] offered STCP protocol which is a reliable end-to-end transport protocol. This protocol provides different levels of reliability in accordance with application; that is,

it provides two levels of reliability for data flows for event-driven and continuous data flow applications. When packets are transmitted, for event-driven applications and continuous data flow applications, ACK packets and NACK packets are used to indicate that the packets have reached the destination, respectively. Before transporting packets over STCP, sensor nodes transmit a Session Initiation packet to the base station. This Session Initiation packet sends data such as the type of data flow, transmission rate, and number of flows to the base station. The base station then stores this packet and sets timer and other parameters of this data flow. When data is received, the base station sends a ACK message to make connection. When sensor node receives ACK, it starts to transmit data. The base station transmits ACK or NACK message, based on the type of the flow. STCP transfers packets according to the size of the window. Also, this method has a congestion detection mechanism and employs occupancy monitoring of the queue of intermediate nodes. The assumption in this paper is that all network intermediate nodes have clocks which are synchronized by sink.

Marchi et al. in [22] suggested a reliable transport protocol called DTSN. This method provides two levels of reliability and is based on full reliability and differential reliability. If it is required that all packets are received at the destination accurately, full reliability is used; otherwise differential reliability is employed. DTSN relies on Selective Repeat ARQ and uses ACK and NACK to provide full reliability. Packets transmit data packets to destination in accordance with the size of Acknowledgment Window (AW). Then Explicit Acknowledgement Request is transmitted to destination. If the destination has received all the packets, ACK packet will be sent; otherwise NACK packet will be sent. DTSN protocol adds Forward Error Correction (FEC) mechanism to provide differential reliability. In this method, a session is a source/destination relationship univocally identified by the tuple ⟨source address, destination address, application identifier, session number⟩ designated the session identifier. The number of the session is selected randomly and the packets of a session are numbered continuously. Intermediate nodes store packets in their buffers and then transmit them, so they are able to transmit them repeatedly if necessary. However, when a large volume of data is transmitted, a large memory is needed, which is a problem.

In all of the aforesaid schemes, there is the need for data decryption and encryption during the aggregation phase using secret key. Thus, these algorithms are very energy and time deficient. Moreover, the security and confidentiality of data are in great peril because of decryption of the received packet in each aggregator, aggregation of the recovered message with the new data and, finally, encryption of the new data by the aggregator that is to be sent to the sink. In the proposed scheme, security is preserved in each aggregator and there is no need to decrypt and then encrypt the data again in each aggregator, which results in notable conservation of energy and reduction of transmission time.

In addition to these, none of the above schemes considered the "delay" parameter. Although WSNs are immensely popular but for real-time applications, end-to-end delay becomes a critical design issue. In such circumstances, if delay turns out to be more than a fixed value, the results and information will be useless. It is clear, therefore, that methods presented in [1, 6, 8, 20] could not be adopted effectively and again the technique explored in this paper is the one that meets the requirement for high speed operations.

3. The Proposed Scheme

Minimizing energy dissipation and maximizing network lifetime are among the central concerns when designing applications and protocols for sensor networks. Clustering has been proven to be energy efficient in sensor networks since data routing and relaying are only operated by cluster heads. Besides, cluster heads can process, filter, and aggregate data sent by cluster members, and thus reducing network load and alleviating the bandwidth.

Residue number system is a nonweighted system that employs residue of numbers divided by several specific moduli to represent them [23–25]. Residue number system can be used for computational applications that need real-time processing such as digital signal processing [26], digital filtering [27], image processing [28], RSA encryption algorithm [29], and digital communications [30].

A residue number system is defined in terms of relatively prime modulus set $\{P_1, P_2, \ldots, P_n\}$ whose $\gcd(P_i, P_j) = 1$ for $(i \neq j)$. In this system a weighted number X can be represented as $(x_1, x_2, \ldots, x_n)$, where

$$x_i = X \mod P_i = |X|_{P_i}, \quad 0 \leq x_i < P_i. \tag{1}$$

Such a representation is unique for any integer X in the range $[0, \mathrm{M})$, where $M = \prod_{i=1}^{n} P_i$ is the dynamic range of the modulus set $\{P_1, P_2, \ldots, P_n\}$ [31].

Redundant Residue Number system is specified by moduli set $\{m_1, m_2, m_3, \ldots, m_h, m_{h+1}, \ldots, m_{h+r}\}$ and $m_i > m_{i-1}$ for $i = 2, 3, \ldots, h + r$. If all the moduli are set pair wise prime, the dynamic range of the system is $[0, M = \prod_{i=1}^{h+r} m_i)$.

In the Redundant Residue Number system with $h + r$ modulo, X where $\alpha \leq X < \alpha + M$ with $h + r$ residue, is represented as $(x_1, x_2, \ldots, x_h, x_{h+1}, \ldots, x_{h+r})$ [32–34].

Having studied a variety of literature on the issue of error control in WSNs and the challenges it poses, the proposed scheme has come up with a solution that is based on Redundant Residue Number System (RRNS). The algorithm offers many advantages and overcomes the weakness seen in other approaches. Here, each sensor in a cluster computes remainders of the sensed data using a predefined modulus set and sends the result, that is, data which is a smaller number compared to original data, to the cluster head where the received numbers are then aggregated. It is clear that since aggregation is being performed on remainders rather than on the entire sensed data, it allows for more compact processing units to be utilized at reduced power consumption in cluster heads and also given that all aggregation operations take place in parallel; this has the benefit of increasing the speed noticeably. It should also be noted that unlike other methods studied, where preaggregation decoding and postencoding are required, in the proposed scheme these processes are unnecessary and only cluster head aggregation operation is

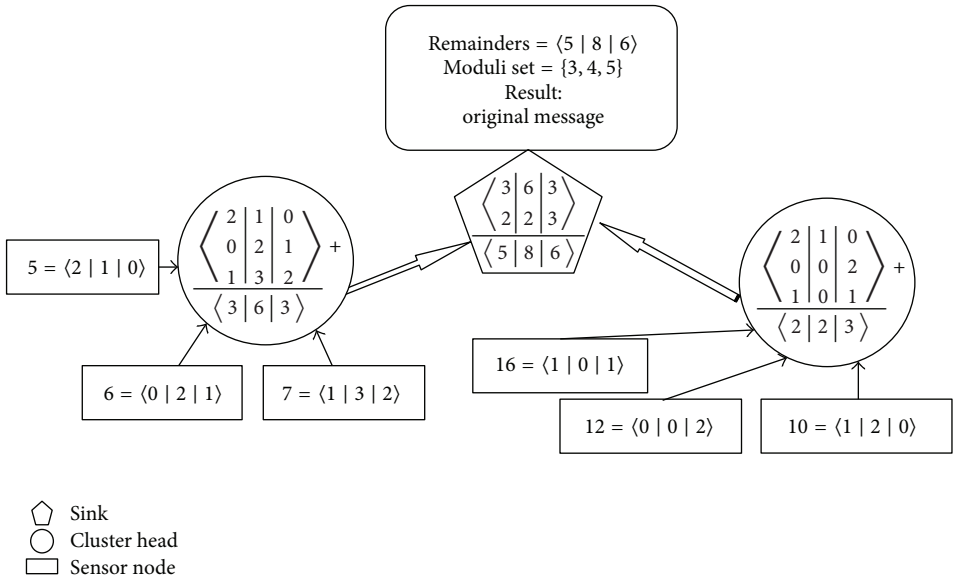


Figure 1: An example of data aggregation functionality using the proposed scheme.

required. It is therefore seen that this solution results in a significant drop in end-to-end delay which makes it suitable for real-time wireless sensor network applications.

The proposed method for error control could be applied to tree-like network topologies and cluster based network topologies. In tree-like network, children transmit the residue of their sensed data, that has been divided by the moduli set, to their parent and their parent will aggregate the received residues and his own residues. Next, the parent transmits the result to his own parent. This will continue until the residues reach sink. In a cluster-based network, the members of the cluster transmit the residue of their sensed data divided by moduli set to head cluster. Then the head cluster aggregates the received residues and its own residues and transmits the result to the next head clusters.

In this study, we considered 5 moduli, 3 main and 2 redundant. Since wireless sensor networks are limited in terms of energy, processing power, size, memory,. . .., selection of the number and value of moduli was made in a way that their reverse converters led to the lowest energy consumption and delay. How we select the main moduli determines the value of dynamic range (obtained by product of main moduli). In fact, if we increase the dynamic range, it will allow us to represent larger numbers using the proposed method. On the other hand, increasing the number of redundant moduli results in higher capability of error detection and correction. Thus, if t redundant modulo is considered, t modulo containing error can be detected and $\lfloor t/2 \rfloor$ residue can be corrected. However, the increase in the number of the moduli leads to an increase in reverse converter cost.

An example illustrating the basic theme of the proposed scheme is shown below, consisting of six sensors located in two clusters. Each cluster has three sensors and a cluster head. In this example, aggregation is a summation operation where the remainders of 5 based on the moduli $\{3, 4, 5\}$ are $\{2, 1, 0\}$ and the remainders of other sensed data are calculated in the same way. In the aggregators, where summation takes place, all of the results received from the first modulo are added together also as all of the results from the second modulo and so on for the third modulo. This feature offers the added advantage of requiring lower space for the arithmetic block. As can be seen, in the sink node, the final decrypted remainders are $\{5, 8, 6\}$ and the modulus set is $\{3, 4, 5\}$.

Now, in order to achieve a lower power consumption and as compact a design as possible, a new moduli set $\{2^{2n+1}, 2^{2n+1} - 1, 2^n - 1\}$ replaces $\{3, 4, 5\}$ in Figure 1. Moreover, the sink node using a reverse converter can decode the message and recover the original data.

Hence an efficient reverse converter is needed. There are two basic approaches that convert a number from RNS to its binary equivalent. These are the Chinese Reminder Theorem (CRT) and Mixed Radix Conversion (MRC) [6], with MRC algorithm being the preferred choice in this paper. Considering also that reliability in data delivery is an important issue in almost all wireless sensor network applications, two-redundant-moduli set $\{2^{3n+1} - 1, 2^{4n+1} - 1\}$ is added to the new 3-moduli set $\{2^{2n+1}, 2^{2n+1} - 1, 2^n - 1\}$, to further enhance error control capability of the proposed scheme. It should be pointed out that the proposed moduli set acts as secret key which is used by the sink node to decrypt the received remainders and obtain the original message. If an adversary that has no knowledge of the moduli set could acquire the transmitted packet, it would not be able to decrypt the message. It is clear that redundant residue number allows transmit data to be encrypted by employing symmetric key and thus provides a powerful tool for error bit detection as well as an effective means for data correction.

4. Design and Implementation of Proposed Residue to Binary Converter

Various methods can be used to design a reverse converter. The most popular methods are MCR and CRT. The use of CRT or MRC depends on selected moduli since multiplicative inverse should be computed in both methods. Computation of multiplicative inverse is the most difficult part of reverse converter. Therefore, it is important to look for multiplicative inverses that allow the formulation of the most simple equations. Certain moduli set can be designed well and at less cost by employing MCR since their multiplicative inverse is computed more easily through simplified (less complex) equations. As a result, the volume of required hardware and consumption of energy as well as network delay are all reduced. In this study we applied MRC. Using this method the values of multiplicative inverse become so simpler, and the design of the reverse converter costs so much less.

The residue to binary conversion can be performed using the MRC algorithm as follows:

$$X = V_n \prod_{i=1}^{n} P_i + \cdots + V_3 P_2 P_1 + V_2 P_1 + V_1. \tag{2}$$

The coefficients $V_i P$ can be obtained from residues by

$$\begin{aligned} V_1 &= x_1, \\ V_2 &= \left| (x_2 - x_1) \left| P_1^{-1} \right|_{P_2} \right|_{P_2}, \end{aligned} \tag{3}$$

$$V_3 = \left| \left((x_3 - x_1) \left| P_1^{-1} \right|_{P_3} - V_2 \right) \left| P_2^{-1} \right|_{P_3} \right|_{P_3}. \tag{4}$$

In the general case, we have

$$\begin{aligned} V_n = {} & \left(\left((x_n - V_1) \left| P_1^{-1} \right|_{P_n} - V_2 \right) \left| P_2^{-1} \right|_{P_n} \right. \\ & \left. - \cdots - V_{n-1} \right) \left| P_{n-1}^{-1} \right|_{P_n} \Big|_{P_n}, \end{aligned} \tag{5}$$

where $|P_i^{-1}|_{P_j}$ denotes the multiplicative inverse of P_i modulo P_j. In mathematics, a multiplicative inverse for a number X is a number that when multiplied by X yields the multiplicative identity, 1. The modular multiplicative inverse of a modulo m can be derived from the extended Euclidean algorithm.

According to (2)–(5), the proposed reverse converter can be designed for the new 3-moduli set $\{2^{2n+1}, 2^{2n+1} - 1, 2^n - 1\}$ as follows.

Consider the 3-moduli set $\{P_1, P_2, P_3\} = \{2^{2n+1}, 2^{2n+1} - 1, 2^n - 1\}$ with three corresponding residues (x_1, x_2, x_3). In order to design a residue to binary converter, we first need to obtain the multiplicative inverse values and substitute these values with the modulus set in the conversion algorithm formulas. The resultant equations should then be simplified by using the arithmetic properties. The final conversion stage involves the implementation of these simplified equations using hardware components such as full adders and logic gates. The following propositions are used to obtain the closed form expressions, which will be the means to compute the multiplicative inverses based on the MRC algorithm.

Proposition 1. *The multiplicative inverse of* (2^{2n+1}) *modulo* $(2^{2n+1} - 1)$ *is* $k_1 = 1$.

Proof. One has

$$\left| 2^{2n+1} \right|_{2^{2n+1}-1} = 1. \tag{6}$$

□

Proposition 2. *The multiplicative inverse of* (2^{2n+1}) *modulo* $(2^n - 1)$ *is* $k_2 = 2^{n-1}$.

Proof. One has

$$\left| 2^{n-1} \times 2^{2n+1} \right|_{2^n-1} = \left| 2^{3n} \right|_{2^n-1} = 1. \tag{7}$$

□

Proposition 3. *The multiplicative inverse of* $(2^{2n+1} - 1)$ *modulo* $(2^n - 1)$ *is* $k_3 = 1$.

Proof. One has

$$\left| 2^{2n+1} - 1 \right|_{2^n-1} = 1. \tag{8}$$

Therefore, let the values $< k_1 = 1$, $k_2 = 2^{n-1}$, $k_3 = 1$, $P_1 = (2^{2n+1})$, $P_2 = (2^{2n+1} - 1)$, and $P_3 = (2^n - 1) >$ in (2)–(4) and thus we have

$$\begin{aligned} X &= x_1 + P_1 (V_2 + V_3 P_2) \\ &= x_1 + \left(2^{2n+1}\right) \left(V_2 + \left(2^{2n+1} - 1\right) V_3\right), \end{aligned} \tag{9}$$

$$V_1 = x_1, \tag{10}$$

$$V_2 = \left| (x_2 - x_1) \left| P_1^{-1} \right|_{P_2} \right|_{P_2} = \left| x_2 - x_1 \right|_{2^{2n+1}-1}, \tag{11}$$

$$\begin{aligned} V_3 &= \left| \left((x_3 - x_1) \left| P_1^{-1} \right|_{P_3} - V_2 \right) \left| P_2^{-1} \right|_{P_3} \right|_{P_3} \\ &= \left| 2^{n-1} \times ((x_3 - x_1) + (-V_2)) \right|_{2^n-1}. \end{aligned} \tag{12}$$

In order to design an efficient reverse converter, expressions (9) and (11)-(12) can be rewritten as follows:

$$\begin{aligned} V_2 &= \left| (x_2 - x_1) \left| P_1^{-1} \right|_{P_2} \right|_{P_2} = \left| x_2 - x_1 \right|_{2^{2n+1}-1} \\ &= \left| x_2 \right|_{2^{2n+1}-1} + \left| -x_1 \right|_{2^{2n+1}-1} = V_{21} + V_{22}, \end{aligned} \tag{13}$$

where

$$\begin{aligned} V_{21} &= \left| x_2 \right|_{2^{2n+1}-1} = \underbrace{x_{2,2n}\, x_{2,2n-1} \cdots x_{2,0}}_{(2n+1)\text{bits}}, \\ V_{22} &= \left| -x_1 \right|_{2^{2n+1}-1} = \underbrace{\bar{x}_{1,2n}\, \bar{x}_{1,2n-1} \cdots \bar{x}_{1,0}}_{(2n+1)\text{bits}}. \end{aligned} \tag{14}$$

Now, in order to implement V_3 on the basis of (12), we have

$$\begin{aligned} V_3 &= \left| \left((x_3 - x_1) \left| P_1^{-1} \right|_{P_3} - V_2 \right) \left| P_2^{-1} \right|_{P_3} \right|_{P_3} \\ &= \left| 2^{n-1} \times (x_3 - x_1) - V_2 \right|_{2^n-1} \\ &= V_{31} + V_{32} + V_{33}, \end{aligned} \tag{15}$$

where

$$V_{31} = \left|2^{n-1} \times x_3\right|_{2^n-1} = \underbrace{x_{3,0}}_{1\text{bit}}\ \underbrace{x_{3,n-1} \cdots x_{3,1}}_{(n-1)\text{bits}}$$

$$V_{32} = \left|-2^{n-1} \times x_1\right|_{2^n-1} = \left\{ \begin{array}{c} \bar{x}_{1,0}\ \bar{x}_{1,n-1} \cdots \bar{x}_{1,1} \\ \bar{x}_{1,n}\ \bar{x}_{1,2n-1} \cdots \bar{x}_{1,n+1} + \\ \underbrace{\bar{x}_{1,2n}}_{1\text{bit}}\ \underbrace{1 \cdots 1 \cdots 1 \cdots 1}_{(n-1)\text{bits}} \end{array} \right\} \quad (16)$$

$$V_{33} = \left|-V_2\right|_{2^n-1} = \left\{ \begin{array}{c} \bar{V}_{2,n-1} \cdots \bar{V}_{2,1} \bar{V}_{2,0} \\ \bar{V}_{2,2n-1} \cdots \bar{V}_{2,n+1} \bar{V}_{2,n} + \\ \underbrace{1 \cdots 1 \cdots 1 \cdots 1}_{(n-1)\text{bits}} \underbrace{\bar{V}_{2,2n}}_{1\text{bit}} \end{array} \right\}.$$

Finally, to obtain X based on (9), we have

$$\begin{aligned} X &= x_1 + P_1 \left(V_2 + V_3 P_2\right) \\ &= x_1 + \left(2^{2n+1}\right) \underbrace{\left(V_2 + \left(2^{2n+1} - 1\right) V_3\right)}_{C} \\ &= x_1 + \left(2^{2n+1}\right) C \\ C &= V_2 + \left(2^{2n+1} - 1\right) V_3 \end{aligned}$$

$$C = \left\{ \begin{array}{cc} \overbrace{V_{3,n-1} \cdots V_{3,0}}^{n\text{bits}} & \overbrace{V_{2,2n} V_{2,2n-1} \cdots V_{2,n}}^{(n+1)\text{bits}} \quad \overbrace{\bar{V}_{3,n-1} \cdots \bar{V}_{3,0}}^{n\text{bits}} + \\ \underbrace{1 \cdots 1 \cdots 1 \cdots 1 \cdots 1 \cdots 1 \cdots 1 \cdots 1}_{(2n+1)\text{bits}} & \underbrace{V_{2,n-1} \cdots V_{2,0}}_{n\text{bits}} \end{array} \right\}$$

$$X = x_1 + \left(2^{2n+1}\right) C = \text{Concatenation}\left(x_1, C\right). \quad (17)$$

□

The area and delay specifications for the proposed reverse converter are shown in Table 1.

From this Table, we have

$$\begin{aligned} \text{Total area} &= (7n+3)\, A_{\text{FA}} + (5n+2)\, A_{\text{NOT}} \\ &\quad + (4n-1)\, A_{\text{OR}} + (4n-1)\, A_{\text{XNOR}} \quad (18) \\ \text{Total delay} &= (9n+8)\, t_{\text{FA}} + 3t_{\text{NOT}}. \end{aligned}$$

Now let us discuss comparison of the proposed reverse converter's performance for the new moduli set $\{2^{2n+1}, 2^{2n+1} - 1, 2^n - 1\}$ against competing models that have either the same or a lower dynamic range in terms of hardware requirement and speed of operations. It should be noted that in any reverse converter most of the delay is associated with full adders which also occupy the largest area allocated to components assembly. Thus for the sake of comparison it is sufficient to investigate only A_{FA} and t_{FA}; see Table 2.

5. Performance Evaluation

In this section, the performance of the proposed moduli set $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1, 2^{3n+1}-1, 2^{4n+1}-1\}$ is evaluated in terms of error detection and error correction capabilities. Evaluation of the error control capability requires an investigation of the effectiveness of the employed "error correction" approach. In general, an error control technique is divided into two parts "error detection" and "error detection and correction". For simplicity's sake, here we refer to "error correction" to mean both error detection and correction noting that error bits must first be detected and located before they can be corrected. Referring to the residue number system, one of its features is that it can correct "burst errors."

This means that if error bits are residing in one modulo, the error control algorithm can easily detect and correct them. However, when error bits are located in more than one modulo, they cannot be corrected although it may be possible to detect them. So, the aim now is to discover what the average percentage of error detection is when error bits reside in more than one modulo by employing the proposed RNS based solution.

In the proposed moduli set, $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1, 2^{3n+1}-1, 2^{4n+1}-1\}$, $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1\}$ are the main three and $\{2^{3n+1}-1, 2^{4n+1}-1\}$ are the redundant moduli, respectively. Now, the first, second, and third main moduli are $(2n+1)$, $(2n+1)$, and (n) bits, respectively. Given this architecture, the error bits can only be detected and corrected if (a) they are

Table 1: Characterization of each part of the proposed reverse converter.

Parts	FA	NOT	AND/XOR	OR/XNOR	Delay
OPU 1	—	$(2n+1)$	—	—	t_{NOT}
CPA 1	$(2n+1)$	—	—	—	$(4n+2)t_{\text{FA}}$
CSA 1	N	—	—	—	t_{FA}
CSA 2	1	—	—	$(n-1)$	t_{FA}
OPU 2	—	$(2n+1)$	—	—	t_{NOT}
CSA 3	N	—	—	—	t_{FA}
CSA 4	N	—	—	—	t_{FA}
CSA 5	1	—	—	$(n-1)$	t_{FA}
CPA 2	N	—	—	—	$(2n)t_{\text{FA}}$
OPU 3	—	N	—	—	t_{NOT}
R– CPA	N	—	—	$(2n+1)$	$(3n+1)t_{\text{FA}}$

Table 2: Area and Delay comparison.

Reverse converter	Area (A_{FA})	Delay (t_{FA})
[12]	$8n+2$	$12n+5$
[13]	$(5n^2+43n)/6+\ 16n-1$	$18n+7$
[14]	$10n+5$	$13n+1$
[15]	$12.5n+6$	$12n+6$
[16]	$10n+5$	$12n+1$
[17]-1	$9n+5$	$11.5n+6$
[17]-2	$n^2+12n+12$	$16n+22$
[17]-3	$9n+10$	$11n+14$
[18]	$n^2/2+\ 11n+14$	$11n+8$
[19]-CICE	$2.5n^2+25.5n+12$	$18n+23$
[19]-CIHS	$2.5n^2+\ 37.5n+28$	$12n+15$
[19]-C2CE	$20n+17$	$13n+22$
[19]-C3CE	$23n+11$	$16n+14$
Proposed	$7n+3$	$9n+8$

Table 3: Area and delay comparison.

N	Main modulus	Redundant Modulus	Maximum error correction	Error detection capability	
				Error bits in one modulo	Error bits in two modulus
2	{32, 31, 3}	{127, 511}	5 bits	100%	99.37%
3	{128, 127, 7}	{1023, 8191}	7 bits	100%	100%
4	{512, 511, 15}	{8191, 131071}	9 bits	100%	100%

placed either in the first, the second, or the third modulo (b); maximum error bit count does not exceed $(2n+1)$, $(2n+1)$, and (n), respectively. Hence, the maximum achievable error correction capability is equal to $(2n+1)$ error bits occurring in the first or the second modulo. On the other hand, the minimum correction capability that can be achieved using the proposed modulus set is equal to one error bit within any modulo of the received packet.

In the proposed error control method, first sensed data are divided by the proposed moduli set and then the residues are computed. Next, the nodes transmit the residues instead of sensed data. After receiving residues, sink computes the original data using 3 main moduli and their received residues. When the residue of the computed number has been divided by redundant moduli and the received residues for those redundant moduli are not equal, an error is detected and error correction procedure is called for it. Sink can reconstruct the original data using 3 residues out of the 5 residues and MRC reverse converter. There are only 10 possible conditions for selecting 3 residues out of 5 residues ($C_3^5 = \binom{5}{3} = (5!/(2! * 3!)) = 10$). Sink computes the original data for all of these possible conditions. Given the fact that during data transfer some residues may have contained errors, not all sink computed numbers are accurate and equal. Thus, among these 10 numbers those out of dynamic range are removed. After a voting process, those numbers that have received the greatest votes are selected as correct numbers.

In order to gauge the error detection capability of the proposed set in a situation where errors occur in more than one modulo and, where therefore, no correction is possible, we simulated the behavior of the error control algorithm using C++ programming language. The experiment runs $30'000'000$ different error bit states by setting $n = \{2, 3, 4\}$. Substituting these values for n in the set we obtain $\{32, 31, 3, 127, 511\}$, $\{128, 127, 7, 1023, 8191\}$, and $\{512, 511, 15, 8191, 131071\}$, respectively. The test results are displayed in Table 3. Note that the maximum error correction is equal to $(2n+1)$ bits, as was mentioned before, and the three error states considered are bits which are located in one of the main modulus (correction is possible and detection is 100%), bits which are located in two moduli, and finally the case where bits are resided in three moduli.

With reference to Table 3 there are several individual conditions for error detection: among the 5 received residues only 1 residue contains error; 2 residues contain error; 3 residues contain error. Since 2 redundant moduli have been considered for this moduli set, error correction of a received residue is possible. It makes no difference whether just one or all bits of the modulo are damaged; this can be corrected. Based on the damaged bits of the residues we will have different conditions (e.g., it is possible that the first and second bits, the first and third bits, or any other combination of bits are damaged); simulation in C++ language considers all these conditions. All possible conditions are simulated for $n = 2$ (i.e., moduli set $\{32, 31, 3, 127, 511\}$), for $n = 3$ (i.e., moduli set $\{128, 127, 7, 1023, 8191\}$), and for $n = 4$ (i.e., moduli set $\{512, 511, 15, 8191, 131071\}$ and each time a part of data was damaged intentionally to investigate detection and correction capability of the method. Table 3 displays error detection percentage for different values of n. It also gives the most number of bits of a modulus that are corrected for various n values.

We evaluated the proposed scheme through NS-2 simulation tool [35]. We considered a randomly deployed network, that is, one hundred nodes that were initially spread over an area of 100 square meters in a random manner. The radio coverage of each sensor node and sink was assumed to be a circular area of 100 m in diameter. The antenna model deployed was omnidirectional. Also, the initial energy of sensor nodes set to 0.1 Joule.

The proposed scheme was analyzed in terms of energy consumption, end-to-end delay, and network lifetime using STCP [21], DTSN [22], TCP, and Reed Solomon. Table 4 shows the simulation parameters.

All simulations were repeated 10 times before calculating the average of the results. As illustrated in Figure 2, the simulation results confirm that the proposed scheme consumes significantly less energy than the other schemes investigated. In Figure 3 end-to-end delay is examined where it can be seen that the proposed scheme generates less network delay. Finally, in Figure 4 it is easy to see that the proposed scheme outperforms its counterparts in terms of network lifetime.

6. Conclusion

In this paper, authors unveiled an innovative energy-efficient algorithm for real time wireless sensor networks that was based on a new 3-moduli set $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1\}$. By exploiting the MRC technique, an efficient reverse converter was also designed which overcame two-key constraints, namely, component space and high speed operation. In order to take full advantage of the error controllability property of the proposed solution, two-redundant-moduli sets were then added to the initial three main sets resulting in a combined set $\{2^{2n+1}, 2^{2n+1}-1, 2^n-1, 2^{3n+1}-1, 2^{4n+1}-1\}$.

The new method's error correction capability was found to be $(2n+1)$ bits under the ideal condition and error detection improvement reached a figure of 99.01% under the worst case scenario condition and 100% when $(n \geq 3)$.

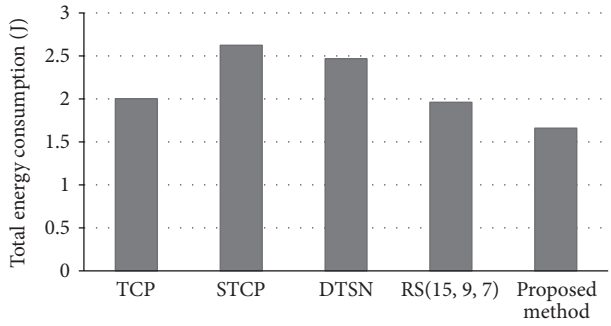


Figure 2: Total energy consumption.

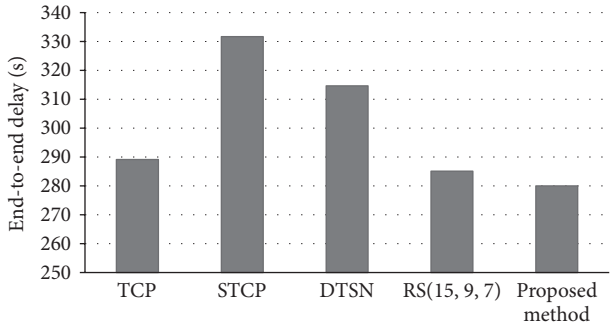


Figure 3: End-to-end delay.

Table 4: Simulation parameters.

Area of sensor field	$100 * 100\,m^2$
Antenna model	Omnidirectional
Simulation time	1000 S
Radio range of a sensor node	100 m
Number of sensor nodes	100
Initial energy	0.1 J
Interface queue type	Droptail
Interface queue (IFQ) length	50 Packets
Energy model	Battery

By simulating the proposed error control scheme, using an ns-2 network simulation tool and making comparisons against some popular methods, outperformance in terms of reductions of energy consumption and end-to-end delay as well as extension of network lifetime was clearly evident. These advantages therefore make the proposed method more desirable for real-time wireless sensor network applications where reliability, low energy consumption, and high speed operations are absolute necessities.

Conflict of Interests

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

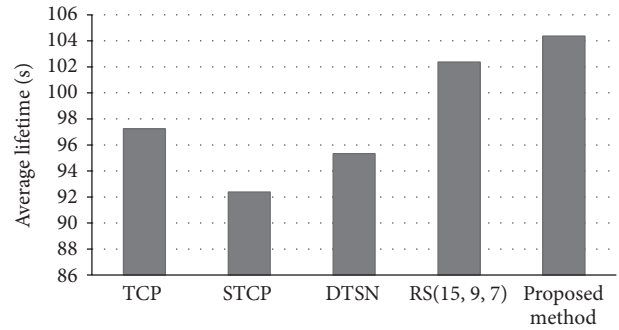


Figure 4: Average network lifetime.

References

[1] G. Song, Y. Zhou, F. Ding, and A. Song, "A mobile sensor network system for monitoring of unfriendly environments," *Sensors*, vol. 8, no. 11, pp. 7259–7274, 2008.

[2] J. Lloret, M. Garcia, D. Bri, and S. Sendra, "A wireless sensor network deployment for rural and forest fire detection and verification," *Sensors*, vol. 9, no. 11, pp. 8722–8747, 2009.

[3] M. Dunbabin and L. Marques, "Robots for environmental monitoring: significant advancements and applications," *IEEE Robotics and Automation Magazine*, vol. 19, no. 1, pp. 24–39, 2012.

[4] P. Kumar, S.-G. Lee, and H.-J. Lee, "E-SAP: efficient-strong authentication protocol for healthcare applications using wireless medical sensor networks," *Sensors*, vol. 12, no. 2, pp. 1625–1647, 2012.

[5] Y.-M. Hong, H.-C. Lin, and Y.-C. Kan, "Using wireless sensor network on real-time remote monitoring of the load cell for landslide," *Sensor Letters*, vol. 9, no. 5, pp. 1911–1915, 2011.

[6] S. Khodadoustan, F. Jalali, and A. Ejlali, "Reliability/energy trade-off in Bluetooth error control schemes," *Microelectronics Reliability*, vol. 51, no. 8, pp. 1398–1412, 2011.

[7] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[8] Z. Guo, B. Wang, P. Xie, W. Zeng, and J.-H. Cui, "Efficient error recovery with network coding in underwater sensor networks," *Ad Hoc Networks*, vol. 7, no. 4, pp. 791–802, 2009.

[9] I. F. Akyildiz and M. C. Vuran, *Wireless Sensor Networks*, John Wiley & Sons, Chichester, UK, 2010.

[10] E. Çayırcı and C. Rong, *Security in Wireless Ad Hoc and Sensor Networks*, John Wiley & Sons, Chichester, UK, 2009.

[11] J. H. Kleinschmidt, W. C. Borelli, and M. E. Pellenz, "An energy efficiency model for adaptive and custom error control schemes in Bluetooth sensor networks," *International Journal of Electronics and Communications (AEU)*, vol. 63, no. 3, pp. 188–199, 2009.

[12] A. S. Molahosseini, K. Navi, C. Dadkhah, O. Kavehei, and S. Timarchi, "Efficient reverse converter designs for the new 4-moduli sets $\{2^n - 1, 2^n, 2^n + 1, 2^{2n+1} - 1\}$ and $\{2^n - 1, 2^n + 1, 2^{2n}, 2^{2n}+1\}$ based on new CRTs," *IEEE Transactions on Circuits and Systems I*, vol. 57, no. 4, pp. 823–835, 2010.

[13] B. Cao, C.-H. Chang, and T. Srikanthan, "A residue-to-binary converter for a new five-moduli set," *IEEE Transactions on Circuits and Systems I*, vol. 54, no. 5, pp. 1041–1049, 2007.

[14] A. S. Molahosseini, C. Dadkhah, and K. Navi, "A new five-moduli set for efficient hardware implementation of the reverse

converter," *IEICE Electronics Express*, vol. 6, no. 14, pp. 1006–1012, 2009.

[15] M. Esmaeildoust, K. Navi, and M. Taheri, "High speed reverse converter for new five-moduli set $\{2^n, 2^{2n+1} - 1, 2^{n/2} - 1, 2^{n/2} + 1, 2^n + 1\}$," *IEICE Electronics Express*, vol. 7, no. 3, pp. 118–125, 2010.

[16] A. S. Molahosseini and M. K. Rafsanjani, "An improved five-modulus reverse converter," *World Applied Sciences*, vol. 11, no. 2, pp. 132–135, 2010.

[17] P. V. Ananda Mohan and A. B. Premkumar, "RNS-to-binary converters for two four-moduli sets $\{2^n - 1, 2^n, 2^n + 1, 2^{n+1} - 1\}$ and $\{2^n - 1, 2^n, 2^n + 1, 2^{n+1} + 1\}$," *IEEE Transactions on Circuits and Systems I*, vol. 54, no. 6, pp. 1245–1254, 2007.

[18] B. Cao, T. Srikanthan, and C. H. Chang, "Efficient reverse converters for four moduli sets $\{2^n - 1, 2^n, 2^n + 1, 2^{n+1} - 1\}$ and $\{2^n - 1, 2^n, 2^n + 1, 2^{n-1} - 1\}$," *IEE Proceedings on Computers and Digital Techniques*, vol. 152, no. 5, pp. 687–696, 2005.

[19] P. V. Ananda Mohan, "New reverse converters for the moduli set $\{2^n - 3, 2^n - 1, 2^n + 1, 2^n + 3\}$," *International Journal of Electronics and Communications (AEU)*, vol. 62, no. 9, pp. 643–658, 2008.

[20] P. Xie, Z. Zhou, Z. Peng, J.-H. Cui, and Z. Shi, "SDRT: a reliable data transport protocol for underwater sensor networks," *Ad Hoc Networks*, vol. 8, no. 7, pp. 708–722, 2010.

[21] Y. G. Iyer, S. Gandham, and S. Venkatesan, "STCP: a generic transport layer protocol for wireless sensor networks," in *Proceedings of the 14th International Conference on Computer Communications and Networks (ICCCN '05)*, pp. 449–454, San Diego, Calif, USA, October 2005.

[22] B. Marchi, A. Grilo, and M. Nunes, "DTSN: distributed transport for sensor networks," in *Proceedings of the 12th IEEE International Symposium on Computers and Communications (ISCC '07)*, pp. 165–172, Aveiro, Portugal, July 2007.

[23] M. Hosseinzadeh, A. S. Molahosseini, and K. Navi, "An improved reverse converter for the moduli set $\{2^n - 1, 2^n, 2^n + 1, 2^{n+1} - 1\}$," *IEICE Electronics Express*, vol. 5, no. 17, pp. 672–677, 2008.

[24] H. Garner, "The residue number system," *IEEE Transactions Electronic Computer*, vol. 8, pp. 140–147, 1959.

[25] K. Navi, A. S. Molahosseini, and M. Esmaeildoust, "How to teach residue number system to computer scientists and engineers," *IEEE Transactions on Education*, vol. 54, no. 1, pp. 156–163, 2011.

[26] R. Chokshi, K. S. Berezowski, A. Shrivastava, and S. J. Piestrak, "Exploiting residue number system for power-efficient digital signal processing in embedded processors," in *Proceedings of the International Conference on Compilers, Architecture, and Synthesis for Embedded Systems (CASES '09)*, pp. 19–28, October 2009.

[27] R. Conway and J. Nelson, "Improved RNS FIR filter architectures," *IEEE Transactions on Circuits and Systems II*, vol. 51, no. 1, pp. 26–28, 2004.

[28] M. Wnuk, "Remarks on hardware implementation of image processing algorithms," *International Journal of Applied Mathematics and Computer Science*, vol. 18, no. 1, pp. 105–110, 2008.

[29] J.-C. Bajard and L. Imbert, "A full RNS implementation of RSA," *IEEE Transactions on Computers*, vol. 53, no. 6, pp. 769–774, 2004.

[30] J. Ramírez, A. García, U. Meyer-Baese, and A. Lloris, "Fast RNS FPL-based communications receiver design and implementation," in *Proceeding of the 12th International Conference Field Programmable Logic*, pp. 472–481, 2002.

[31] F. J. Taylor, "Residue arithmetic: a tutorial with examples," *IEEE Computer*, vol. 17, no. 5, pp. 50–62, 1984.

[32] F. Barsi and P. Maestrini, "Error correcting properties of redundant residue number systems," *IEEE Transactions on Computers*, vol. 22, no. 3, pp. 307–315, 1973.

[33] E. Kinoshita and K.-J. Lee, "A residue arithmetic extension for reliable scientific computation," *IEEE Transactions on Computers*, vol. 46, no. 2, pp. 129–138, 1997.

[34] N. Z. Haron and S. Hamdioui, "Redundant residue number system code for fault-tolerant hybrid memories," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 7, no. 1, article 4, 2011.

[35] "Network simulator—(ns-2)," http://www.isi.edu/nsnam/ns/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 398583, 7 pages
http://dx.doi.org/10.1155/2014/398583

Research Article

Distributed Cooperative Algorithm for k-M Set with Negative Integer k by Fractal Symmetrical Property

Gelan Yang[1,2] and Shuai Liu[2,3]

[1] *College of Information Science and Technology, Hunan City University, Yiyang 413000, China*
[2] *College of Computer Science, Inner Mongolia University, Hohhot 010012, China*
[3] *School of Physical Science and Technology, Inner Mongolia University, Hohhot 010012, China*

Correspondence should be addressed to Shuai Liu; cs_liushuai@imu.edu.cn

Received 6 December 2013; Accepted 19 February 2014; Published 8 May 2014

Academic Editor: Yong Jin

In recent years, fractal is widely used everywhere and escape time algorithm (ETA) became the most useful fractal creating method. However, ETA performs not so well because it needs huge computations. So, in this paper, we first present an improved fractal creating algorithm by symmetrical radius of k-M set. Meanwhile, we use distributed cooperative method to improve classic ETA into parallel system, which is called distributed cooperative ETA (DCETA). Secondly, we present the proof of fractal property in k-M set $f_c(z) = z^k + c$ with exponent k ($k < 0$), which concludes its threshold and symmetrical property. Finally, computational result shows correctness of the novel DCETA, which shows better computational effectiveness and lower waste.

1. Introduction

Since Mandelbrot has presented M set (Mandelbrot's set), fractals have been soon used everywhere [1]. Nowadays, generalized M sets become representatives of chaotic dynamics. We present M set in Definition 1 and k-M set in Definition 2 as follows [2]. In Definition 1 $f_c^n(z)$ is defined in (1).

Definition 1. M-set $= \{c \mid \lim_{n\to\infty} f_c^n(0) \nrightarrow \infty \ (c \in C, n = 1 \cdots \infty)\}$

$$f_c^n(z) = \left(f_c^{n-1}(z)\right)^2 + c \quad \text{where } f_c^1(z) = z^2 + c. \tag{1}$$

Then, when we reform $F_c^n(z)$ to $F_c^n(z)$ in (2), definition of k-M set is presented in Definition 2. Consider

$$F_c^n(z) = \left(F_c^{n-1}(z)\right)^k + c \quad \text{where } F_c^1(z) = z^k + c. \tag{2}$$

Definition 2. k-M set $= \{c \mid \lim_{n\to\infty} f_c^n(0) \nrightarrow \infty \ (c \in C, n = 1 \cdots \infty)\}$.

Nowadays, many scholars research deeply in k-M set with many kinds of exponents. In fact, we know that the dynamics system k-M set is generalization of M set, and classic M set is actually 2-M set.

Then, because the structure of k-M set is complex, we use escape time algorithm (ETA) to create its fractal. The algorithm ETA compares $|f_c^n(c)|$ to a given large number N of all points c. Admittedly, a point c is in k-M set when the iteration time is more than another given number M. To explain it, we present the escape time algorithm in Algorithm 1.

In Algorithm 1, if $I_z > I$, the iterations of z are convergent; else are divergent.

In fact, Algorithm 1 is an approximate algorithm because M and N cannot be determined. For example, if there exists a point c makes $|f_c^k(z)(0)| < T$ and $|f_c^{k+1}(z)(0)| > T$ or $|f_c^k(z)(0)| > T$ and $|f_c^{k+1}(z)(0)| < T$, we will find different results when $I = k$. So we cannot make sure the point c is convergent or not for ETA in these two conditions. Another problem is that when the iteration exponent is negative or complex number with negative real part, the iteration value is often changed from greatly large to greatly low or opposite. It is hardly to use ETA in this condition. So a strict proof is necessary to be given, which is used to prove convergence and divergence of k-M set in complex plane with $k < 0$.

Then, there is another problem in fractal creating algorithms, that is, the huge time cost. Assuming the displayed area contains $M \times N$ points, which need to compute I times

Step 1.
Let T = escape threshold, I = max iterated number.
Assuming $I_c = 0$ is iterated times of points c.
Set $T_c = c$ as $f_c^n(z)$ (0) of points c.
Step 2.
For all points c in complex plane,
While $I_c \leq I$ and $|T_c| < T$
$I_c = I_c + 1$. $T_c = f(T_c)$.
Step 3.
Color all points c by different I_z.

Algorithm 1: Escape time algorithm (ETA).

in the worst condition (I is maximum iteration times), we know that the computational times $C = M \times N \times I$. This is a huge number and it needs huge time to be processed. So we have to use a distributed cooperative algorithm to decrease the computational times.

For years, distributed cooperative algorithms are widely used in many areas, which indeed decrease time cost by using multiprocessors instead of single processor. Today, researchers usually use multicomputers to construct a cooperative system. Furthermore, parallel and distributed system is always used for transportation [3], estimation [4], optimization [5], and localization [6] in engineering area. Admittedly, we call them distributed cooperative transportation/estimation/optimization/localization algorithms.

In fractal area, there are many initial iterated functions with complex structure [7, 8]. So we have to use novel iterated method to process them [9]. Admittedly, distributional method is an effective method to decrease the computational time, just like parallel and cloud environment [10, 11].

In this paper, in order to decrease huge time cost which comes from fractal creating algorithms of k-M set, we present a distributed cooperative algorithm under a parallel system at first. We create k-M set with both classic ETA and novel DCETA to validate the correctness of DCETA. Besides, in order to validate the positive of DCETA, the parallel experimental results are also presented and analyzed.

Second, we prove that k-M set has strongly symmetrical properties with phase angles. In this case, we decrease the iterated time further. Meantime, we analyze these results and find the parallel properties of this distributional algorithm. It is also experimented in DCETA and the experimental results also validate our conclusions.

In conclusion, in this paper, we firstly present a separate method to reform ETA into distributional environment. Meanwhile, we analyze this novel algorithm with experimental results. Secondly, we research in properties of k-M set and find self-symmetrical regions with phase angles. Then, we present a separate method to reform ETA into distributional environment. Meanwhile, we analyze this novel algorithm with experimental results. The results also validate properties of k-M set. Finally, some Julia sets are created to validate containable properties between k-M set and the corresponding Julia sets.

The remainder of the paper is organized as follows. We proved properties of k-M set when $k < 0$ in Section 2. Moreover, we have their parallel experimental results and analysis in Section 3. Finally, Section 4 summarizes the main results of the paper.

In this paper, k is a negative integer, i and n are positive integers, c is a complex number, T and I are positive real numbers.

2. Distributed Cooperative ETA with Fractal Symmetrical Property

In order to run ETA in distributed system, we use SIMC in the distributional method in Algorithm 2.

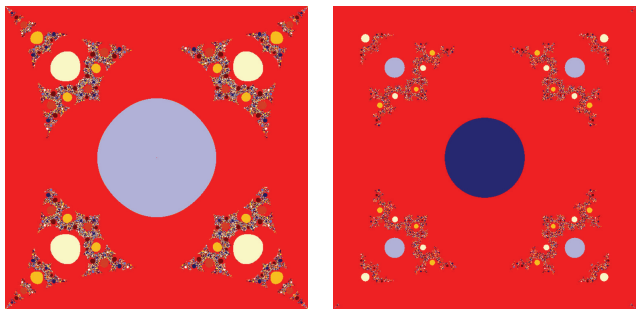
Not as usual as other cooperative algorithm, DCA creates the final image with the hidden cooperative multitask nodes, which provides a part of the final image. So we cannot find the cooperative relation in DCT. In our experiment, by these two distributed cooperative methods which are presented, the parallel system is constructed with 3 same PCs. In this system, one PC is the primary node and other two are task nodes. Then, all subresults are connected in the primary node as the final result.

In this paper, we use $k = -3$ and -4 to process the experiment and compute a sector with center $(0, 0)$ and radius $(\sqrt{2}/2)s$ (s is a side of displayed area). For example, when $k = -3$, the original and improved fractal is presented in Figure 1 (with no larger than threshold 10 and no lower than threshold 0.1). Similarly, the original and improved fractal is presented in Figure 2 (with no larger than threshold 10 and no lower than threshold 0.1) when $k = -4$. In these figures, Theorems 4 and 6, Lemma 5, and Inference 2 are validated. In detail, we find the symmetrical property from the general views of Figures 1-2, and they validate Lemma 5 and Inference 2. Then, the escape areas are all similar-round with the escape points in Figures 1-2. They validate Theorem 4. Additionally, we have the farthest fractal area to validate Theorem 6.

In this paper, the two worker nodes compute results of the area (ρ, θ) with modality of polar coordinates. In the area, $\rho = (\sqrt{2}/2)s$, and $\theta = [2p\pi/|k|, 2(1 + p)\pi/|k|)$, where $p \in \{0, 1\}$ denotes number of the two worker nodes. We present Figures 3-4 to validate the conclusion we given. In Figure 3, we present the two fractal image created by two worker nodes.

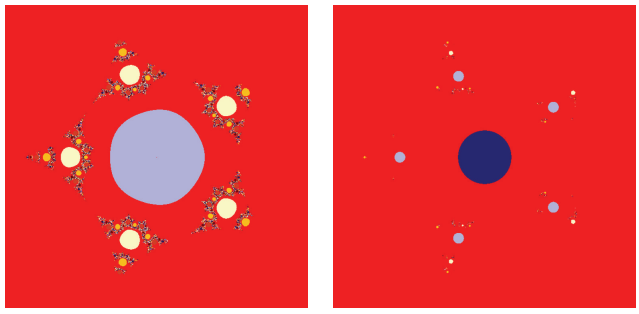
Primary node
(i) For all task nodes
Send group message ($area_i$, f) to ith task nodes;
in the message, $area_i = (x_{start}, x_{end}) \times (y_{start\_i}, y_{end\_i})$ denotes iterated area, and function f dotes the iterated mapping
(ii) While all task nodes send its result
Connect all sub-images and display it;
(iii) Finished;
Task nodes
(i) If primary node send message ($area_{ii}$, f)
Use ETA to create sub-image by compute all points in this area with iterated function f;
(ii) Send sub-image to primary node;
Finished;

Algorithm 2: DCA of ETA.



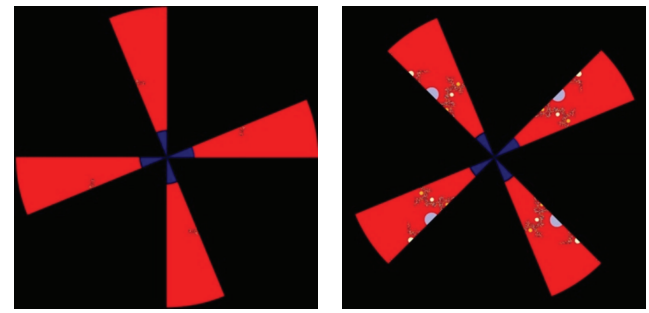
(a) (b)

Figure 1: Fractal of k-M set by original (10) and improved (0.1) threshold ($k = -3$).



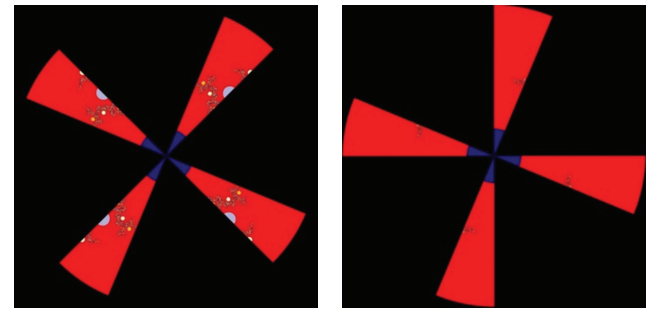
(a) (b)

Figure 2: Fractal of k-M set by original (10) and improved (0.1) threshold ($k = -4$).



(a) (b)

Figure 3: Fractal of k-M set by the two worker nodes ($k = -3$).



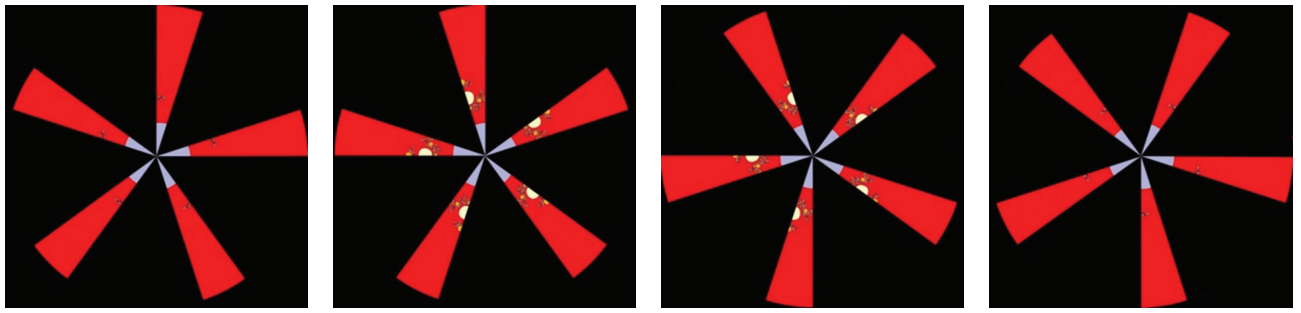
(a) (b)

Figure 4: Corresponding fractal of k-M set in the two worker nodes ($k = -3$).

In Figure 4, we present the remainder of whole k-M set. Then, we find that the pseudoescape lines are the symmetry axes of k-M set. Moreover, Figure 5 is similar to Figures 3-4. It is presented to validate our conclusion with $k = -4$.

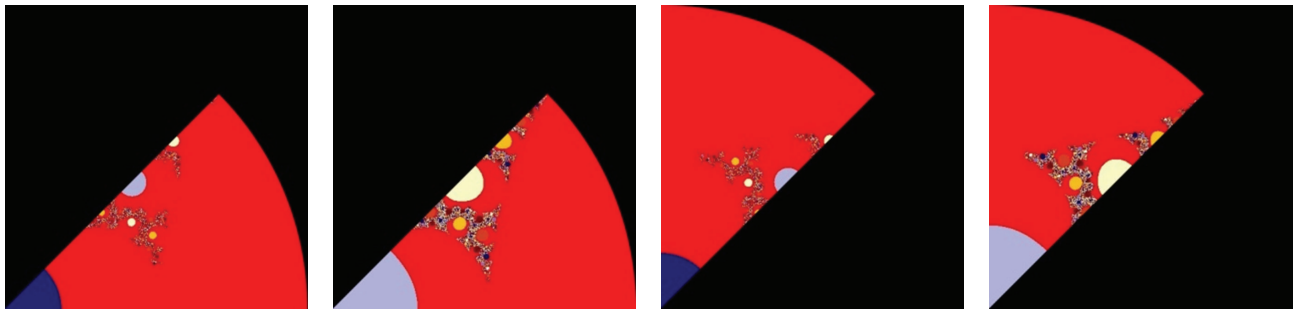
Then, the basic computational area is presented in Figures 6(a)-6(b) and 7(a)-7(b), which are corresponding to $k = -3$ and $k = -4$. Moreover, their symmetrical fractals of pseudoescape line are presented in Figures 6(c)-6(d) and Figures 7(c)-7(d). In these figures, (a) is the original computational area with original threshold, (b) is the improved computational area with improved threshold, and (c) and (d) are the symmetrical areas with corresponding axis pseudoescape lines. Meantime, in these two examples, the expected displayed area is $(-1.2, 1.2) \times (-1.2, 1.2)$. So radius of the computational sector is about 1.697.

Finally, Figure 8 presents the symmetrical properties of k-M set by the two thresholds, and we validate Theorems 6 and 7 with Figures 6, 7, and 8.



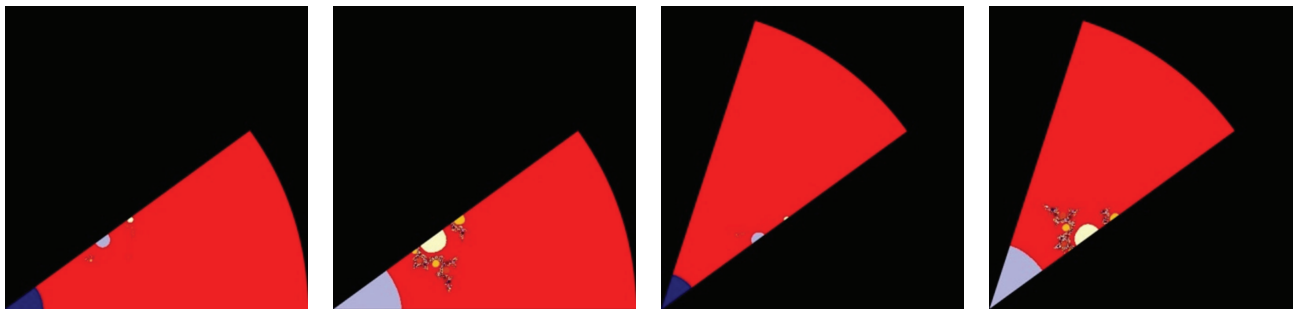
(a) (b) (c) (d)

Figure 5: Original and corresponding fractal of k-M set by the two worker nodes ($k = -4$).



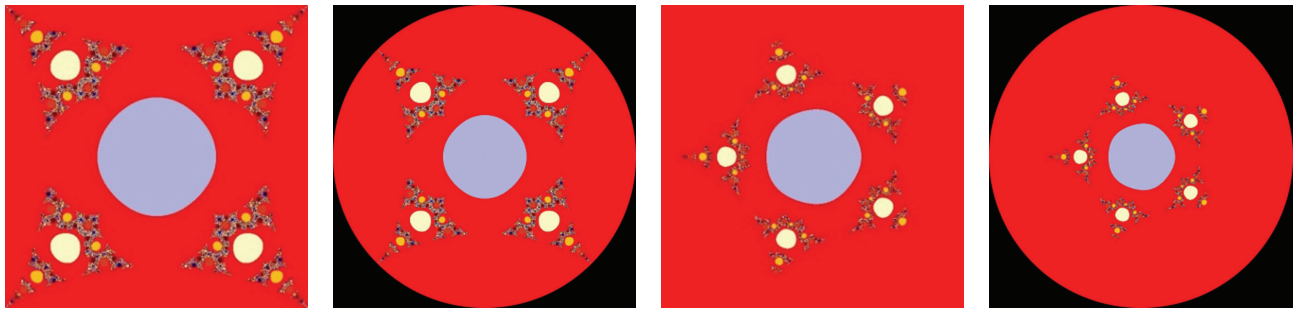
(a) (b) (c) (d)

Figure 6: Fractal of the computational area by original and improved threshold ($k = -3$).



(a) (b) (c) (d)

Figure 7: Fractal of the computational area by original and improved threshold ($k = -4$).



(a) (b) (c) (d)

Figure 8: Fractal of k-M sets by xOy and polar coordinates with two thresholds ($k = -4$).

3. Proof of Fractal Property in k-M Set

3.1. Correctness of the New Threshold. In this paper, we study properties when k is a negative integer from (2). At first, we give threshold and convergent properties of k-M set by Theorem 3.

Theorem 3. *k-M set contains whole complex plane except countable points.*

Furthermore, when $k = -1$, we reach Inference 1, which presents escape points of k-M set.

Inference 1. *Origin is the only escape point of k-M set when $k = -1$.*

In this way, we can analyze these countable escape points of k-M set. So when we let $E = \{\text{Set of escape points}\}$, we know that $S = \{S_i\}$, $i = 1 \cdots \infty$ is a partition of E, where $S_i = \{c \mid c \text{ is a solution of } r_i^* = 0\}$. It is because of the following:

$$\begin{gathered} S_i \neq \emptyset, \\ S_i \cap S_j = \emptyset \quad \text{when } i \neq j, \\ \bigcup_{i=1}^{\infty} S_i = E. \end{gathered} \tag{3}$$

Then, when we reach Theorem 4 to find all escape points which are attracting, we will found the structure of k-M set.

Theorem 4. *All escape points of k-M set are attracting.*

Proof. Firstly, (4) is applied to compute all escape points of k-M set:

$$F_c^i(0) = 0. \tag{4}$$

Generalizing (4) to infinite, we have (5). It means that c is an escape point. However, using $1/\infty = 0$, we have (6). It means that c is similar to a periodic point. Meantime, in this paper, *a* point c is called pseudoperiodic point when it reaches (4)–(6):

$$F_c^{i+1}(0) = \infty, \tag{5}$$

$$F_c^{i+2}(0) = c. \tag{6}$$

So when applying (7), we have (8) to reach its derivative:

$$f^i(z) = f^{i-1}\left(z^k + c\right), \tag{7}$$

$$\left(f^i(z)\right)' = \prod_{j=0}^{i-1}\left(f\left(z_j\right)\right)' \quad \left(z_0 = c,\ z_j = f^j(z_0)\right). \tag{8}$$

Then we have (9) to reform (8):

$$\begin{aligned} \left|\left(f^i(z)\right)'\right| &= \prod_{j=0}^{i-1}\left|\left(f\left(z_j\right)\right)'\right| \\ &= \prod_{j=0}^{i-1}\left(k\left|z_j\right|^{k-1}\right) = |k|^i \cdot \left(\left|\prod_{j=0}^{i-1} z_j\right|\right)^{k-1}. \end{aligned} \tag{9}$$

Admittedly, infinite is attracting fixed point, and zero is pseudo-2 periodic point.

So when applying (10) into (9), we have (11) when $i-1$ is even or (12) when $i-1$ is odd:

$$z_i \cdot z_{i+1} = z_i \cdot \left(z_i^k + c\right) = z_i^{k+1} + cz_i, \tag{10}$$

$$\left|\left(f^i(z)\right)'\right| = |k|^i \cdot \left(\prod_{j=0}^{(i-2)/2}\left|z_{2j}^{k+1} + cz_{2j}\right|\right)^{k-1}, \tag{11}$$

$$\left|\left(f^i(z)\right)'\right| = |k|^i \cdot \left(z_{i-1} \cdot \prod_{j=0}^{(i-3)/2}\left|z_{2j}^{k+1} + cz_{2j}\right|\right)^{k-1}. \tag{12}$$

So when $z_i = 0$ or $z_i = \infty$, (10) is infinite, and computational result of (11) is zero. Moreover, it is admittedly that computational result of (12) is zero. So we know these pseudoperiodic points are all attracting.

Theorem 4 is proved. □

Moreover, we know that escape points of k-M set are all attracting from Theorem 4. So when we use a threshold T as the threshold, we know that $1/T$ can also be used as a threshold in the fractal generation.

When assuming that radius of the approximately circle is ε. So we have (13) from the newer threshold:

$$\left|f^i(c+\varepsilon)\right| < \frac{1}{T} \quad \text{when } f^i(c) = 0. \tag{13}$$

Then, it reforms to (14) because ε is tiny and $(f^i(c))' = 0$:

$$\left|\frac{\varepsilon \cdot \left(\left(f^i(c)\right)' + \left(f^i(c+\varepsilon)\right)'\right)}{2}\right| = \left|\frac{\varepsilon \cdot \left(f^i(c+\varepsilon)\right)'}{2}\right| < \frac{1}{T}. \tag{14}$$

So, we have (15) to present attracting area of the ith escape point c:

$$\varepsilon < \frac{2}{T\left|\left(f^i(c+\varepsilon)\right)'\right|} \leq \frac{2}{T}. \tag{15}$$

In this case, we prove the novel threshold of k-M set. Then, the symmetrical fractal property of k-Mandelbrot set is presented in the following section.

3.2. Fractal Symmetrical Property of k-M Set. At first, when we analyze S_i in (9)–(11), we have Lemma 5 as follows.

Lemma 5. *$c^* = c \cdot e^{2n\pi i/(1-k)} \in S_i$ when $c \in S_i$ $(n = 1 \cdots 1-k)$.*

Proof. Using mathematical induction, first, let $i = 1$; using c^* in $r_i^* = r_0^* + (r_{i-1}^*)^k$, we have

$$\begin{aligned} r_1^*(c^*) &= r_0^*(c^*) + (r_0^*(c^*))^k \\ &= c \cdot e^{2n\pi i/(1-k)} \cdot \left(1 + c^{k-1}\right) = e^{2n\pi i/(1-k)} \cdot \left(c + c^k\right) \\ &= e^{2n\pi i/(1-k)} \cdot r_1^*(c). \end{aligned} \tag{16}$$

Then, let $i = n$; assuming

$$r_n^*(c^*) = e^{2n\pi i/(1-k)} \cdot r_n^*(c) \tag{17}$$

we have

$$\begin{aligned} r_{n+1}^*(c^*) &= r_0^*(c^*) + (r_n^*(c^*))^k \\ &= e^{2n\pi i/(1-k)} \cdot c + e^{2kn\pi i/(1-k)} \cdot (r_n^*(c))^k \\ &= e^{2n\pi i/(1-k)} \left(c + e^{2n\pi i/(1-k)} \cdot r_n^*(c)\right) \\ &= e^{2n\pi i/(1-k)} r_1^*(c). \end{aligned} \tag{18}$$

Summarizing (16)–(18), we have (19). So when $c \in S_i$, which is equal to $r_1^*(c) = 0$, we know that $r_1^*(c^*) = 0$:

$$\left|r_n^*(c^*)\right| = \left|r_n^*(c)\right|. \tag{19}$$

It means that $c^* \in S_i$. Lemma 5 is proved □

In this way, we separate all S_i to $1-k$ parts, which differ from each other with phase angle $2\pi/(1-k)$.

Then, we reach Inference 2 from (19).

Inference 2. *$c^* = c \cdot e^{2n\pi i/(1-k)} \in k$-$M$ set when $c \in k$-M set $(n = 1 \cdots 1-k)$.*

In this case, we know that k-M sets also can be separated to $1-k$ parts, which differ from each other with phase angle $2\pi/(1-k)$.

From Lemma 5 and Inference 2, we find the separation of S_i clearly by fractal symmetry of k-M set. In our distributional strategy, S_i is divided into $1-k$ parts, which differ from each other with phase angle $2\pi/(1-k)$. In this case, we separate k-M set to n nodes. The ith node computes area with phase angle $(2\pi(i-1)/n(1-k),\ 2\pi i/n(1-k))$ and we only compute area with phase angle $(0, 2\pi/(1-k))$. The other area is same as the computational area and can be collaged with it.

Then, we have Theorem 6 to present the farthest "escape point" of origin, which denotes the displayed area of k-M set.

Theorem 6. *In Algorithm 1, the farthest pseudoescape point of origin is at the line $\theta = (2n+1)\pi/(1-k)$, $n = 0 \cdots 1-k$.*

Proof. Without loss of generality, we assume a point $c = r \cdot e^{ix}$ $(x > 0)$ as a farthest escape point. Then, we find that $|f(c)|$ has the lowest modulus than any other points with modulus r in the following by using mathematical induction.

(a) Let $i = 1$; $|f^i(c)|$ is lower than other points with modulus r, where $x = (2n+1)\pi/(1-k)$ and phase angle of $f(c)$ is same as c. Its proof is in the following.

Applying $c = r \cdot e^{ix}$ in (11), we have

$$f(c) = c \cdot \left(c^{k-1} + 1\right) = c \cdot \left(r^{k-1} \cdot e^{ix(k-1)} + 1\right), \tag{20}$$

$$\left|f(c)\right| = r \cdot \left|r^{k-1} \cdot e^{ix(k-1)} + 1\right|. \tag{21}$$

Admittedly, $|f(c)|$ reach its lowest value when $x(k-1) = (2n+1)\pi$; applying it into (20), we find the lowest value is

$$\left|f(c)\right| = r \cdot \left|1 - r^{k-1}\right|. \tag{22}$$

(b) Let $i = n$ and assuming that $|f^n(c)|$ is lower than other points with modulus r, where $x = (2n+1)\pi/(1-k)$ and phase angle of $f^n(c)$ is same as c.

(c) Let $i = n+1$; $|f^i(c)|$ is lower than other points with modulus r, where $x = (2n+1)\pi/(1-k)$ and phase angle of $f^{n+1}(c)$ is same to c. Its proof is in the following.

Assuming that

$$f^n(c) = R \cdot c, \tag{23}$$

we have

$$f^{n+1}(c) = c \cdot \left(R^k \cdot c^{k-1} + 1\right). \tag{24}$$

So, when $x = (2n+1)\pi/(1-k)$, we have $|f^{n+1}(c)| = r \cdot |R^k - 1|$ which is the lowest of all points with modulus r.

Summarizing (a)–(c), we prove that $f^n(c)$ is lowest when its modulus is lower than T in Algorithm 1. Then when we apply Theorem 3, we have threshold of these pseudoescape points which are determined by $(f^n(c))^k$. So c is farther from origin when $f^n(c)$ is lower.

It means Theorem 6 is proved. □

In this way, we know that k-M set has $1-k$ farthest pseudoescape points, which are at lines $\theta = (2n+1)\pi/(1-k)$ for $n = 1 \sim 1-k$. In this paper, we use radii to name these $1-k$ lines. Then, we use Theorem 7 to present their symmetrical properties.

Theorem 7. *$|f^i(c)| = |f^i(c')|$ $(i = 1, 2, \ldots)$ for all points $c = r \cdot e^{i\alpha}, c' = r \cdot e^{i(2\theta-\alpha)}$ with $\theta = (2n+1)\pi/(1-k)$, $n = 0 \cdots -k$.*

Proof. Using mathematical induction, first, let $i = 1$; we have

$$f^1(\overline{c}) = \overline{c} = \overline{f^1(c)}. \tag{25}$$

Then, let $i = n$, and assuming

$$f^n(\overline{c}) = \overline{f^n(c)}, \tag{26}$$

we have

$$\begin{aligned} f^{n+1}(\overline{c}) &= \left(f^n(\overline{c})\right)^k + \overline{c} \\ &= \left(\overline{f^n(c)}\right)^k + \overline{c} = \overline{\left(f^n(c)\right)^k} + \overline{c} \\ &= \overline{\left(f^n(c)\right)^k + c} = \overline{f^{n+1}(c)}. \end{aligned} \tag{27}$$

Summarizing (25)–(27), we have that $f^n(\overline{c}) = \overline{f^n(c)}$ for all n. In other words, k-M set is symmetry with the real axis.

Then, when applying Lemma 5 and Inference 2, we know that k-M set is symmetric with phase angle $2n\pi/(1-k)$ ($n = 1 \cdots 1-k$). It means

$$c' = r \cdot e^{i(2\theta-\alpha)} = \overline{c} \cdot e^{2\theta i}. \tag{28}$$

Moreover, it is admittedly that $2\theta = 2(2n+1)\pi/(1-k)$ ($n < [k/2]+1$) or $2\theta = 2(2n+k)\pi/(1-k) + 2\pi$ ($n > [k/2]$). So we have (29) by (19), (26), and (28):

$$\left|f^n\left(c'\right)\right| = \left|f^n\left(\overline{c}\right)\right| = \left|f^n\left(c\right)\right|. \tag{29}$$

It means Theorem 7 is proved. □

In this case, we know that these pseudoescape lines are all symmetrical lines of k-M set. So the ratio of computational area is reduced to $1/2(1-k)$ with original displayed area. In conclusion of Section 3.2, we know the correctness of symmetrical property in k-M set.

4. Conclusion

In this paper, we analyze k-M set with formula $f_c(z) = z^k + c$ ($k = -1, -2, \ldots$) by using distributed cooperative algorithm. We present an algorithm DCETA to improve classic ETA to a distributed parallel system, and this novel fractal creating method has lower computations. With experimental results, we validate our conclusions.

Then, we proved the fractal property of k-M set, which contains the correctness of the novel threshold and the fractal symmetrical properties. These novel properties are improvements of [12, 13].

Since we have already studied some special k-M set with many methods [14], and used k-M set in many other areas, in the next step, first, we will find some properties of k-M set, which can enhance understanding of k-M set. Then, we will use properties of k-M set into distributed parallel system to enhance its effectiveness and robustness.

Conflict of Interests

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

Acknowledgments

The authors wish to thank the anonymous reviewers for their helpful comments in reviewing this paper. This work is supported by grants from Programs of Higher-level talents of Inner Mongolia University (nos. 125126, 115117), Scientific Projects of Higher School of Inner Mongolia (no. NJZY13004), Scientific Research Fund of Hunan Provincial Education Department (no. 12B023), and National Natural Science Foundation of China (nos. 61261019, 61262082).

References

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, San Francisco, Calif, USA, 1982.

[2] J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2003.

[3] A. C. Sanderson, "A distributed algorithm for cooperative navigation among multiple mobile robots," *Advanced Robotics*, vol. 12, no. 4, pp. 335–349, 1997.

[4] X. C. Hao, J. Z. Wu, C. F. Chien, and M. Gen, "The cooperative estimation of distribution algorithm: a novel approach for semiconductor final test scheduling problems," *Journal of Intelligent Manufacturing*, pp. 1–13, 2013.

[5] I. Necoara and D. Clipici, "Efficient parallel coordinate descent algorithm for convex optimization problems with separable constraints: application to distributed MPC," *Journal of Process Control*, vol. 23, no. 3, pp. 243–253, 2013.

[6] S. Liu, X.-J. Che, and Z.-X. Wang, "Existence domain analysis and numerical algorithm of fixed point for generalized 3x + 1 function T(x)," *Acta Electronica Sinica*, vol. 39, no. 10, pp. 2282–2287, 2011 (Chinese).

[7] S. Liu, W. Fu, W. Zhao, J. Zhou, and Q. Li, "A novel fusion method by static and moving facial capture," *Mathematical Problems in Engineering*, vol. 2013, Article ID 503924, 6 pages, 2013.

[8] S. Liu and Z. Wang, "Fixed point and fractal images for a generalized approximate 3x + 1 function," *Journal of Computer-Aided Design and Computer Graphics*, vol. 21, no. 12, pp. 1740–1744, 2009 (Chinese).

[9] S. Liu, X. Che, and Z. Wang, "Improvement of escape time algorithm by no-escape-point," *Journal of Computers*, vol. 6, no. 8, pp. 1648–1653, 2011.

[10] M. Liu, S. Liu, W. Fu et al., "Distributional escape time algorithm based on generalized fractal sets in cloud environment," *Chinese Journal of Electronics*. In press.

[11] S. Liu, W. Fu, H. Deng, C. Lan, and J. Zhou, "Distributional fractal creating algorithm in parallel environment," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 281707, 8 pages, 2013.

[12] N. Chen and W. Zhu, "Constructed general high-order Mandelbrot fractal images and symmetrical escape time algorithm," *Journal of Northeastern University: Natural Science*, vol. 17, no. 3, pp. 225–229.

[13] X. Liu and W. Zhu, "Constructed M-set and filled J-set by accelerative escape time algorithm of combination," *Journal of Northeastern University: Natural Science*, vol. 18, no. 4, pp. 413–416.

[14] S. Liu, X. Cheng, C. Lan et al., "Fractal property of generalized M-set with rational number exponent," *Applied Mathematics and Computation*, vol. 220, pp. 668–675, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 384973, 9 pages
http://dx.doi.org/10.1155/2014/384973

Research Article

Multisensor Based Neutral Function Identification of Solenoid Valve

Yanqing Guo,[1] Yongling Fu,[1] Xiaoye Qi,[1] and Chun Cao[2]

[1] *School of Mechanical Engineering and Automation, Beihang University, No. 37 Xueyuan Road,Haidian District, Beijing 100191, China*
[2] *School of Automation Science and Electrical Engineering, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing 100191, China*

Correspondence should be addressed to Yongling Fu; fuyongling@126.com

Received 27 November 2013; Revised 9 February 2014; Accepted 16 March 2014; Published 16 April 2014

Academic Editor: Gelan Yang

Condition monitoring of hydraulic systems has been using automatic control in industrial system. In this paper, a sensor network based intelligent control is proposed for efficient solenoid valve identification. The detection system learns to detect the change of output pressure of multipoints that represent a more complicated task. Linear correlation analysis is introduced for feature extraction, which allows for a significant reduction in the dimension of original data without compromising the change detection performance. Implemented as an agent identifying the valve types under measurement, the support vector machine classifier achieves a significant high accuracy in identification and an increase in deployment efficiency. Experimental results prove that the system is feasible for application designs and could be implemented on technological platforms.

1. Introduction

In industrial field, the automatic control of hydraulic system could be introduced to facilitate the working process [1]. However, failures inevitably occur at the most inconvenient time creating both technical and financial restrictions. Condition monitoring of hydraulic hardware shows great promise as an effective approach to manage the dynamic performance of hydraulic components [2, 3]. Generally, condition monitoring of hydraulic systems contains the real-time measurement of pump, cylinder, valve, and so forth. Currently, condition monitoring is slowly replacing the common practice of regular preventative maintenance whereby the component is replaced at predetermined intervals before it breaks down [4]. It clearly helped if the performance of the components and hence the hydraulic circuit can be predicted, particularly in the presence of fault conditions [5]. Thereby, researchers are ongoing to maintain the working process of fluid power systems.

For current hydraulic systems, electromagnetic directional valves (solenoid valves), which are involved with leak reduction technology, miniaturized mechanical elements, and fast electronic components, are faster and more accurate than the former valves [6]. Particularly, it is a critical control component in hydraulic control systems [7, 8]. The working parameter of a solenoid valve contains neutral function, working pressure, action mode, and so forth. The neutral function often forms the direction flowing and it is the area which is currently receiving the least attention. From the monitoring point of view, the working performance of solenoid valves would be detected from the desire to initially concentrate on the neutral function. On the other hand, wrong recognition or use of neutral function may lead to severe operation inefficiency as well as huge replacement costs. Therefore, identification of neutral function can be introduced to transform the characteristic into to quantitative state.

For all types of condition monitoring systems, online multipoint monitoring is the most advanced one [9]. The ability to deploy a recognition process and then clarify the internal structure and deduce faults is the ultimate goal of an engineer. For identifying the neutral function, data acquisition and signal processing systems are crucial. Online data acquisition provides the greatest flexibility since the output pressure of different ports can be rapidly known to the operator. It is quite common for data to be obtained using

hand-held instruments that are connected to appropriate test points [10]. Further, multisensing systems are now commonplace and make the transition to online monitoring relatively easy to accommodate [11]. In addition, more systematic and comprehensive methods also include failure modes and effects analyses, risk and criticality analyses, data mining methods, and prognosis techniques [12–14].

According to the aforementioned issues, aiming at conditioning the working performance of solenoid valves, we have developed a neutral function recognition system based on multisensing. In this system, sensors for output pressure acquisition are distributed at multipoints. With parameters of interests extracted, recognition model was set up for matching neutral function of different types of solenoid valves. The model also serves as a useful training package whereby existing or potential neutral functions can be inserted and the change in performance can be noted. Accordingly, the determination of neutral function would thereby provide a path to working performance monitoring.

The paper is organized as follows: A model for neutral function identification is developed in Section 2. Section 3 illustrates the working procedure of the proposed multisensing scheme and preprocessing algorithm. In Section 3, we briefly review the existing condition monitoring methods and propose a new algorithm. The methodology within this study is based on machine learning. In Section 4, the experiment is depicted, which highlight the feasibility of the method. We draw the conclusion in Section 5.

2. System Description

For electromagnetic directional valves, several different performances could be detected or monitored, like reversing performance, leakage performance, pressure loss, median function identification, and so forth [15]. In this study, we thus focus on the recognition of three-position four-way valve (a kind of most commonly used electromagnetic directional valve); through it, we can acquire some universal test methods to monitor electromagnetic reversing valve.

A sensor network is built up in our test. The hardware frame diagram of the condition monitoring system is shown in Figure 1. In this study, Zigbee based communication modules are employed. We obtain the pressure data of the four ports and the limit the information of the cylinder using the sensor node. Each sensing unit consists of pressure sensors and one router for signal sending; all these units could be hard wired to the testing target around the plant. For each testing point, working parameters are collected through one coordinator and processed afterward. To indicate different types of valves in practical use, we employ two kinds of solenoid valves with four sensing units within this network. We consider a sensing region that contains multiple detecting nodes, which deliver information based on wireless sensor network. Output of the system is the movement of the solenoid. All the data can be transferred to the computer terminal for display, storage, and further analysis.

Within this multisensing system, we totally have nine parameters to monitor (Table 1). According to Figure 2, sensing devices are connected to the monitoring points to cover the complete system. Each side of the target valve has a solenoid "a" and a solenoid "b" fixed. The output pressure of port P, T, A, and B can be monitored by four pressure sensors P_P, P_A, P_B, and P_T, respectively. High pressure oil enters the system through port P while low pressure oil returns tank through port T. Port A and port B are both work ports for driving the load work. In this system, the test cylinder is employed to indicate movement. The load force is adjusted by the throttle valve group. The left and right limit positions rely for the switch "B_L" switch and "B_R" fixed on each side of the cylinder. The leakage of valve is monitored by flow meter Q. Hence, the recognition of the neutral function based on multipoint measurements can be deployed for working performance monitoring.

Table 1: Sensors distribution.

No.	Name	Function	Range
s1	P_P	P port pressure test	0– 40 Mpa
2	P_A	A port pressure test	0– 40 Mpa
3	P_B	B port pressure test	0– 40 Mpa
4	P_T	T port pressure test	0– 40 Mpa
5	Q	Leakage test	0– 10 L/min
6	B_L	Left limit switch	Digital input
7	B_R	Right limit switch	Digital input
8	a	Valve position solenoid	Digital input
9	b	Valve position solenoid	Digital input

3. Preprocessing Based on Linear Correlation Analysis

The schematic of the neutral function identification scheme is shown in Figure 3. It consists of three subsystems. Raw data were exactly obtained from pressure sensors, which were combined with solenoids condition as well as the position of the cylinder for data fusion. Therefore, segmented data could be applied to feature selection algorithm. The internal features were extracted for characterizing different working conditions. For the purpose of optimizing the processing procedure, we tend to simplify the data set through the linear feature transformation algorithm. For on line condition monitoring, machine learning concept could be implemented; hence, data were then analyzed using statistical bases established within the recognition model. Various parameters would be interpreted using a set of rules gained from mathematical analysis.

Feature extraction is a significant process of obtaining distinctiveness from time series data sets. Aiming at discriminating the reference class from other classes, feature selection can be considered as a data-compression process which removes irrelevant information and preserves relevant information from the raw data [16]. Typically, feature extraction approach applied to raw signals precedes the classification procedure. In this case, the condition of the valve can be divided into six phases for feature selection (Figure 4): (1) solenoid on left engaged; right unengaged, (2) solenoid on left engaged; left/right limit reached, (3) left/right limit reached; solenoid

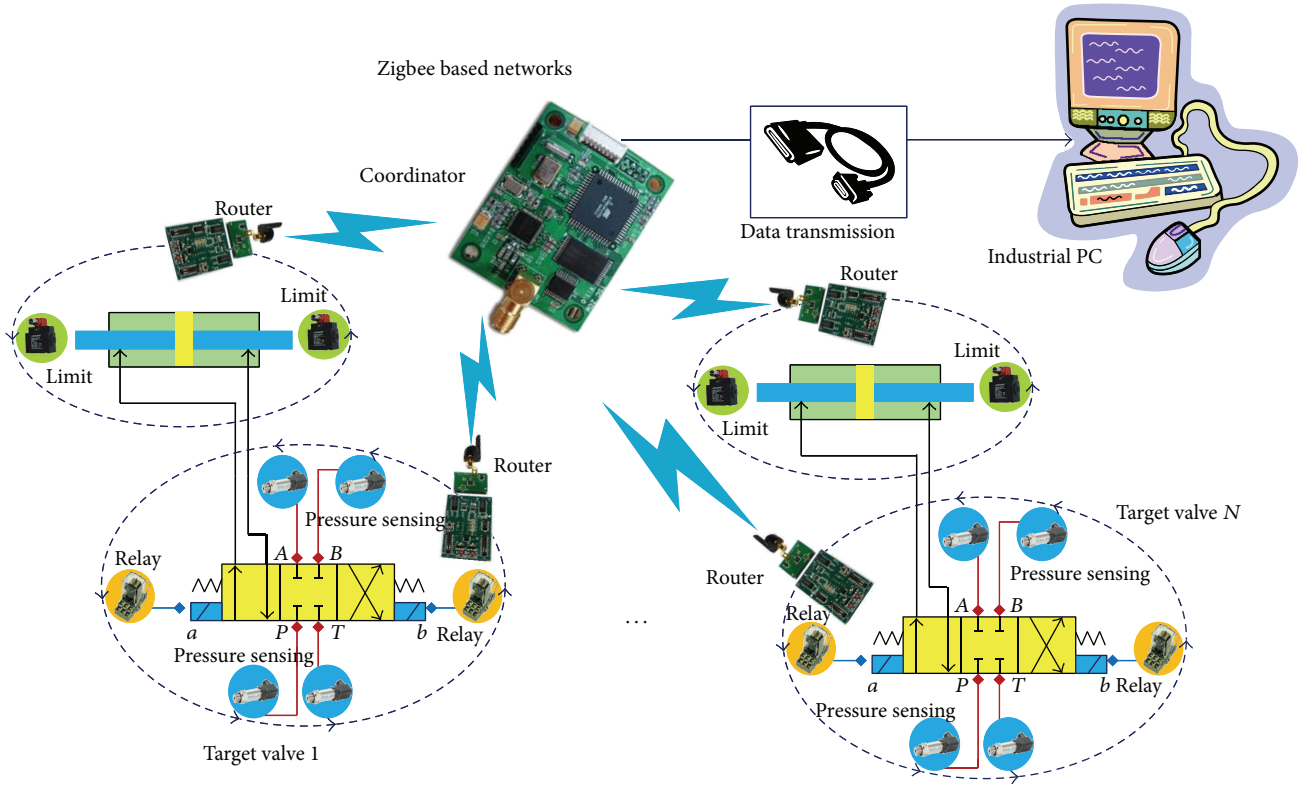


Figure 1: Hardware frame diagram of condition monitoring system.

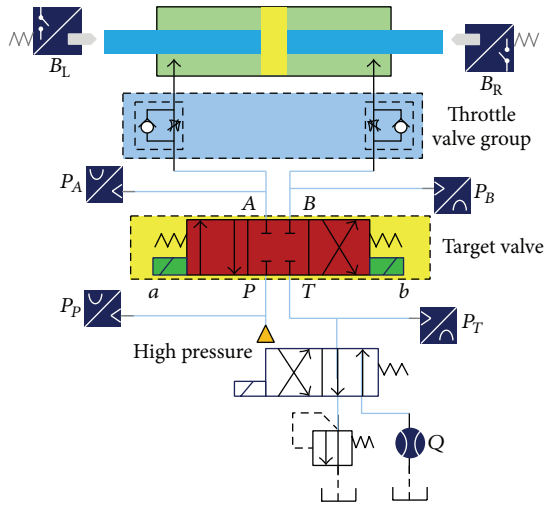


Figure 2: Monitoring points of the system.

on left unengaged, (4) solenoid on right engaged; left unengaged, (5) solenoid on right engaged; left/right limit reached, and (6) left/right limit reached; solenoid on right unengaged.

Due to the curse of dimensionality, a robust classifier is hard to be built and the computational cost is prohibitive. So, dimensionality reduction is proposed to remove the irrelevant information and extract relevant features. The commonly used methods to reduce dimension can be categorized into linear (e.g., PCA, MDS, and LDA) and nonlinear (e.g., LLE, ISOMAP, Laplacian Eigenmap, KPCA, and KDA) methods [17]. Since we prefer to gain the natural relationship of different features, we introduce the linear correlation analysis [18]. Therefore, the method to compute maximal linearly independent subset of a matrix is presented as follows.

Supposing $\alpha_1, \alpha_2, \ldots, \alpha_m$ are of n dimension ($m \geq n$) while $\alpha_i = (v_{i1}, v_{i2}, \ldots, v_{in})(i = 1, 2, \ldots, m)$, if there exist $\lambda_1, \lambda_2, \ldots, \lambda_m$ (not all zero) as well as

$$\lambda_1\alpha_1 + \lambda_2\alpha_2 + \cdots + \lambda_m\alpha_m = 0, \quad (1)$$

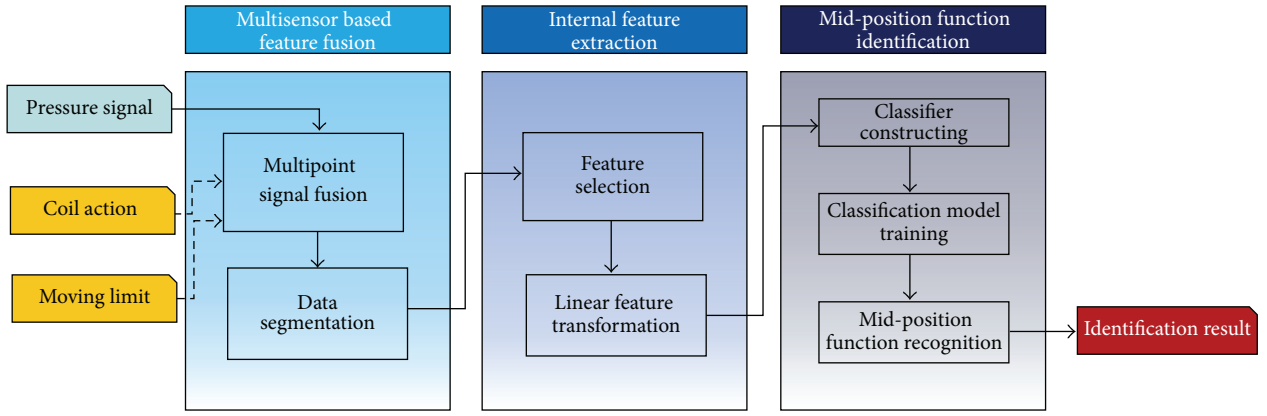
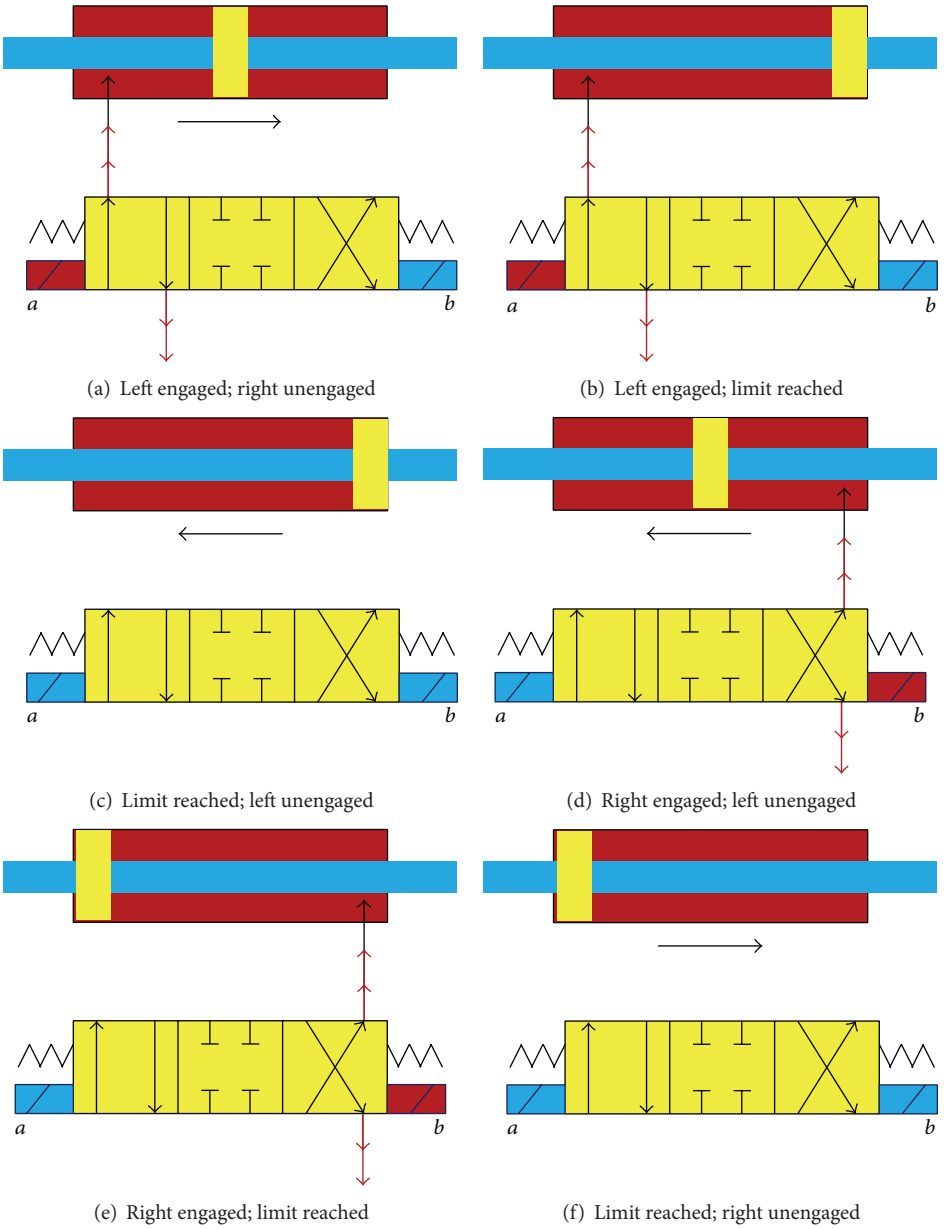


Figure 3: Neutral conditions identification process.



(a) Left engaged; right unengaged

(b) Left engaged; limit reached

(c) Limit reached; left unengaged

(d) Right engaged; left unengaged

(e) Right engaged; limit reached

(f) Limit reached; right unengaged

Figure 4: Working position of solenoid valve.

then the set $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ is called linearly dependent, where $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ is linearly independent if and only if the above expression holds for $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$.

Meanwhile, $\{b_1, b_2, \ldots, b_r\}$ is a subset of vector set $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}(r \leq n \leq m)$, if $b_1, b_2, \ldots, b_r$ is linearly independent and every vector in $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ can be linearly expressed by $b_1, b_2, \ldots, b_r$, and $\{b_1, b_2, \ldots, b_r\}$ is called a maximal linearly independent subset [19].

Obviously, a maximal linearly independent subset of $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ is a basis of a vector space for which $\{b_1, b_2, \ldots, b_r\}$ forms a basis. Then, vector set $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ can be replaced by its maximal linearly independent subset in the linear system, and the dimension of vector α_i will be reduced to r.

Three elementary row operations are usually used to find a maximal linearly independent subset of vector set $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ as follows:

(1) transform the transpose of two row vectors;

(2) multiply a row vector by a constant k;

(3) regulate one row vector by plus another row vector c times.

After getting the maximal linearly independent subset, r eigenvectors of vector α_i in set $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ can be known. When processing the signals, the corresponding eigenvectors can be used to substitute for vector α_i. Consequently, in order to decrease the time required to train models, to prevent overfitting, and to facilitate real-time implementation, an optimal feature set would be available by linear regression analysis.

4. Neutral Conditions Identification with SVM

The features of solenoid valves can be extracted by computing the maximal linearly independent subset, and then they were breathed into a classifier. In this paper, we employed support vector machine (SVM) to recognize different neutral function of valves.

Support vector machines, developed from statistical learning theory, are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis [20, 21]. SVM had successfully been used in many fields and received satisfactory results, such as bioinformatics recognition, regression analysis, and function approximation [22, 23]. In this paper, we aim at pattern identification of valves. A function, deduced from available examples, is used as a classifier to identify neutral conditions of solenoid valves.

The basic SVM takes a set of input data and predicts, for each given input, which of the two possible classes forms the output. Intuitively, a good separation has the largest distance to the nearest training data point of any class, since, in general, the larger the margin, the lower the generalization error of the classifier (Figure 5).

Given a set of training vectors $x_i \in R^n$, $i = 1, \ldots, N$ and an indicator $y_i \in \{-1, +1\}$, an SVM training algorithm builds a linear model as follows:

$$f(x) = w \cdot x + b. \tag{2}$$

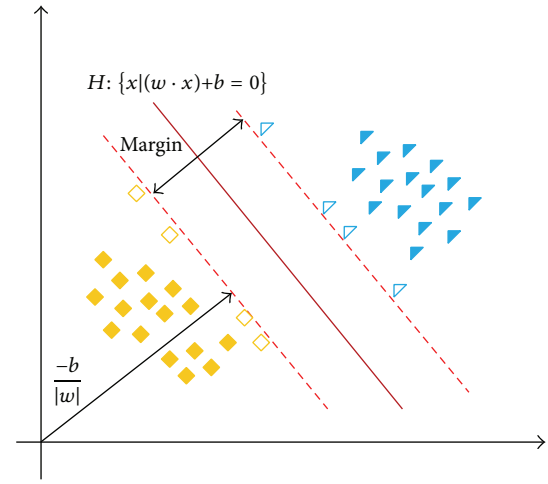


Figure 5: Classification of two classes by SVM.

The parameters w and b are defined on the basis of a primal optimization problem:

$$\begin{aligned} \min \quad & P(w, \xi) = \min\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i\right) \\ \text{subject to:} \quad & y_i(w \cdot x_i) + b \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, \ldots, N, \end{aligned} \tag{3}$$

where C is for controlling the tradeoff between the model complexity and empirical risk [23]. Another parameter is set for the classifier G representing the kernel trick, which can be expressed as

$$k(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j), \tag{4}$$

where $\varphi(x_i)$ is a nonlinear mapping. Input vector x_i can be mapped into a richer feature space through $\varphi(x_i)$ by kernel function (4).

By introducing Lagrange multipliers $\alpha_i \geq 0, i = 1, 2, \ldots, N$, the constrained problem, the possible high dimensionality of the vector variable ω, will be solved, and the Lagrangian can be deduced as (5)

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} y_i \alpha_i k(wx_i - b) + \sum_{i=1}^{N}\alpha_i. \tag{5}$$

Then, the task reduces to the following optimization problem to maximize (5) in α_i and to minimize it in w and b and w can be computed thanks to the α terms:

$$w = \sum_{i=1}^{N} y_i \alpha_i x_i \tag{6}$$

while

$$\sum_{i=1}^{N} \alpha_i y_i = 0. \tag{7}$$

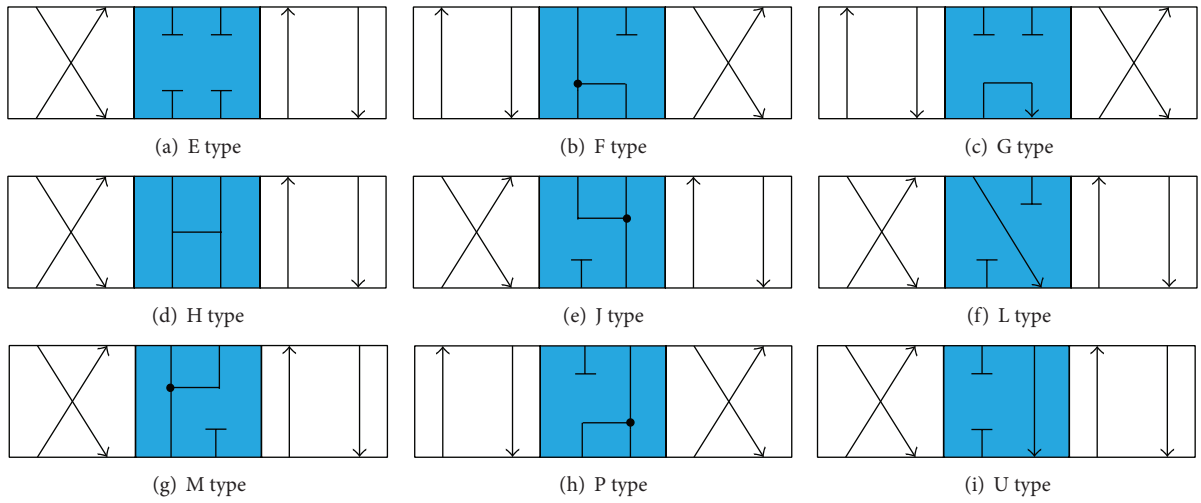


Figure 6: Experimental types of valves.

Table 2: Dimensionality reduction outcome.

Feature	Phase					
	1	2	3	4	5	6
P_P average value	√	√	√	√	—	√
P_P effective value	√	—	√	√	—	√
P_P kurtosis	√	√	—	√	√	—
P_A average value	√	√	√	√	—	√
P_A effective value	√	—	√	√	—	√
P_A kurtosis	√	√	√	√	√	√
P_B average value	√	—	√	√	—	√
P_B effective value	√	—	√	√	√	√
P_B kurtosis	√	√	√	√	√	√
P_T average value	—	—	√	—	—	√
P_T effective value	—	—	—	—	—	—
P_T kurtosis	√	—	√	√	—	—

* √: nonlinear correlation, —: strong linear correlation.

By substituting (6) and (7) into (5), the decision function of SVM model can be derived as

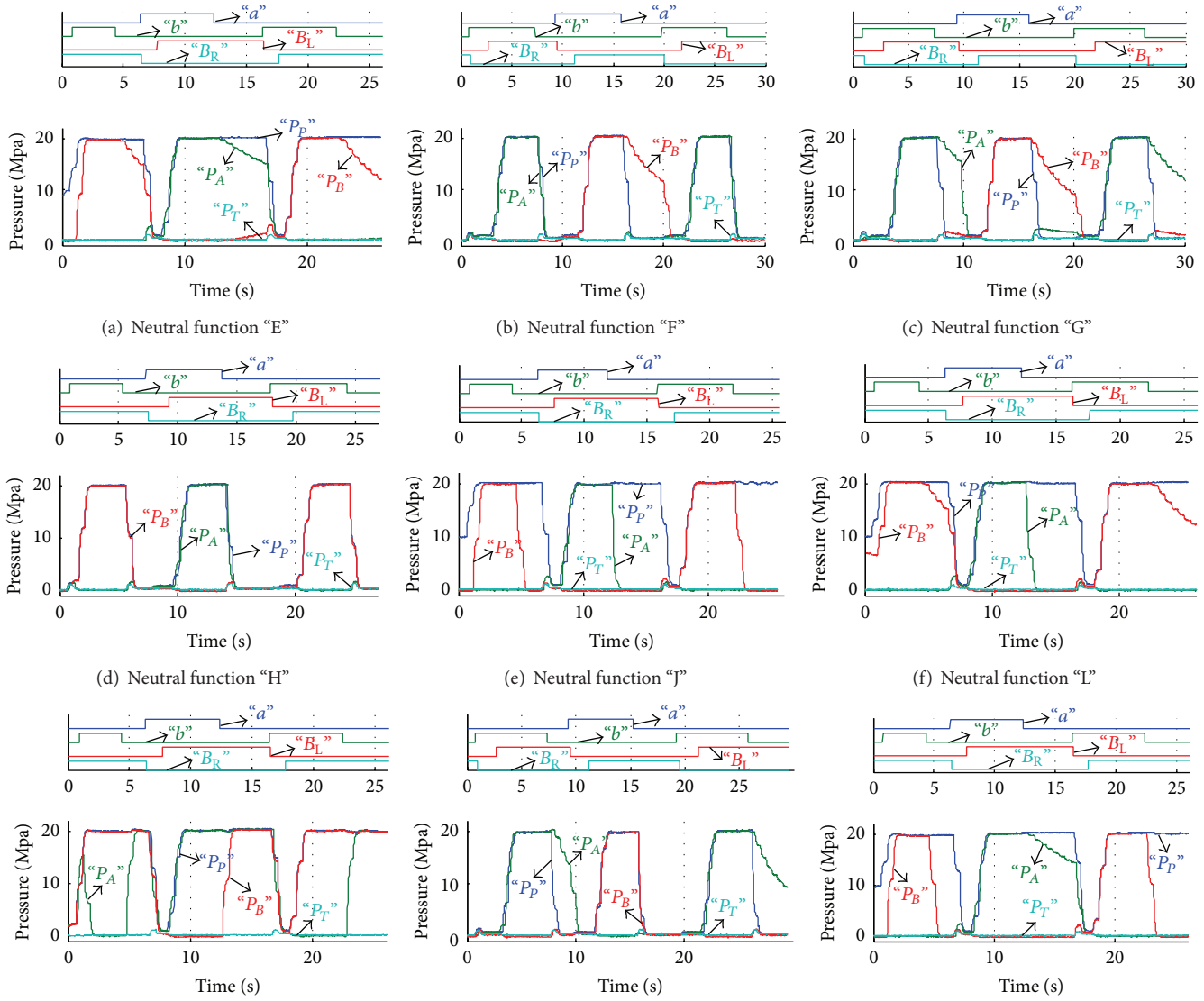
$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i (x_i, x) + b\right). \quad (8)$$

5. Experimental Results

It is essential to establish that the model is sufficiently accurate. The conditioning model can be satisfactorily identifying the neutral function of the target valve. Experiments were conducted in order to validate the model developed in this research. In the present study, we monitor the effect of multipoints pressure in real time of nine types of solenoid valves (Figure 6). The pressure of each type of valve within one working stroke is shown in Figure 7. According to different acquisition units in Figure 2, the working phases can be indicated by line "a," "b," "B_R," and "B_L" while output values of the four sensing points of "P_A," "P_B," "P_T," and "P_P" vary compiling to the moving phases of the valve. Periodic changes in pressure can be observed for all types of valves. In this study, 193 sets of data samples were picked up. According to the data segmentation mode in Section 3, we picked up three parameters as features of each pressure sensor, which are effective value, kurtosis, and average value, so we have twelve eigenvalues of every phase. For data fusion, we totally get seventy-two subvectors representing one working process.

Considering the complexity of the data from one solenoid valve, the dimensionality reduction algorithm was therefore employed. We applied the linear correlation analysis to feature variables to distill the essential characteristics. For this application, new sample sets were obtained to represent original pressure information. According to Table 2, features of strong correlation were excluded; thus, we got an array of fifty within it.

An autonomic neutral conditions identification framework is typically made up of a data training part and a testing part. For all the 193 sets of samples, the former 100



(a) Neutral function "E" (b) Neutral function "F" (c) Neutral function "G"

(d) Neutral function "H" (e) Neutral function "J" (f) Neutral function "L"

(g) Neutral function "M" (h) Neutral function "P" (i) Neutral function "U"

Figure 7: Pressure outputs of different valves.

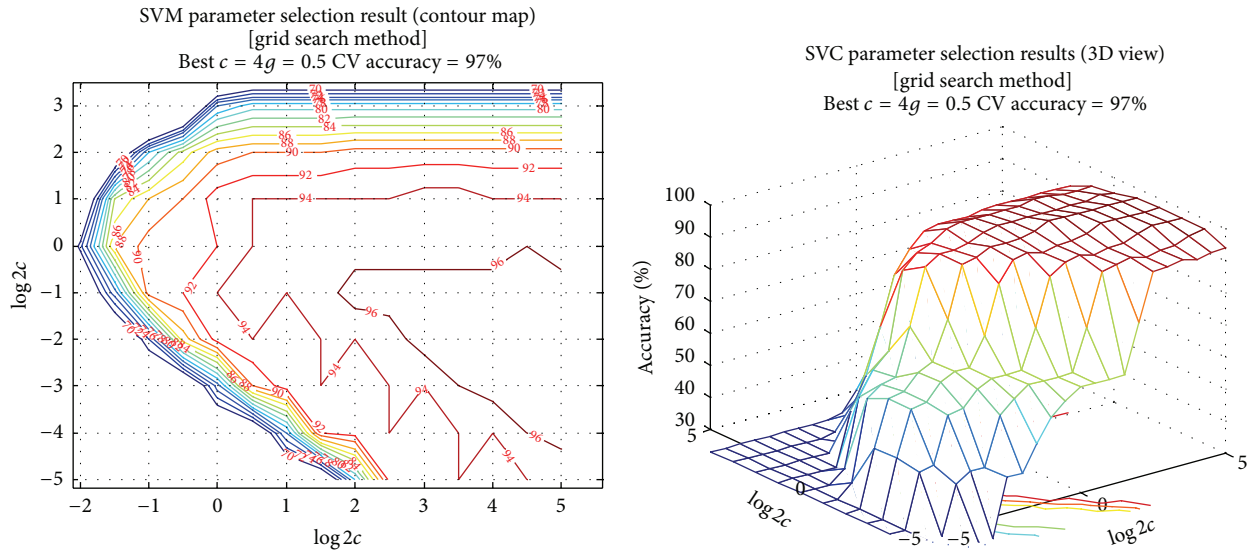
were for classifier training and the latter for testing. The objective of our work is to construct a classifier for valve type identification. Features representing the pressure of multipoints were memorized in matrix and sent to the classifier. For the training part, we got an optimal c of 4 and g of 0.5 through cross-validation (Figure 8).

In Figure 9, we summarize the classification performance results achieved by trained SVM classifier. The testing accuracy is as high as 100%. The running time of the identification program is 4.9 ms, which can definitely meet the need of condition monitoring system.

The success of our system in achieving these requirements can be attributed to two strategies: (1) multisensor fusion improved classification accuracy; producing stable decisions for quick transition recognition and (2) dimensionality reduction algorithm during feature extraction process categorized data into the optimal sample sets for the construction of the classifier. Such design, although strict, allows the control engineers to make use of well-established multisensing theory to perform solenoid valve identification.

6. Conclusion

In this study, we developed a multipoint sensing based condition monitoring system that could automatically identify neutral function of solenoid valve types. This methodology is able to identify the valve types from each port's pressure. To summarize, this paper has suggested a multisensing methodology of linear correlation analysis based preprocessing of input and a SVM based automatic classifying of working parameters. The results have demonstrated this system's precision. Further, the model can be easily implemented and cause lower maintenance cost of the complex electromechanical system. With the architecture of sensor networks,



(a) Contour map of parameter selection results

(b) 3D view of parameter selection results

Figure 8: SVM parameter selection results.

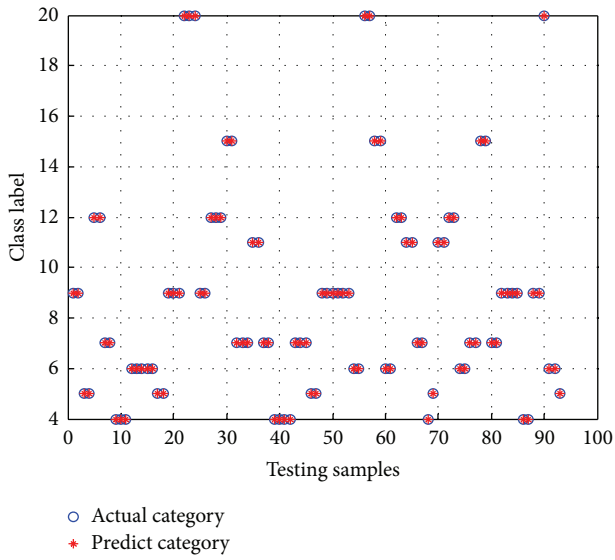


Figure 9: Actual and predict modes of testing samples.

the proposed system offers a convenient and cost-effective approach to identify the neutral functions of solenoid valves.

In our system, output of multipoint is applied to the monitoring module. The dimensionality reduction method simplified the recognition based on linear correlation analysis. We were also able to decide the classifier with the optimization of parameters that can identify the type of solenoid valves automatically. In a neutral function recognition experiment, the accuracy rate reached 100%. Therefore, this methodology can successfully be applied to other condition monitoring procedures.

However, this method is not able to test the neutral function under other moving conditions because it is applied only when the action of the valve follows the six-phase motion sequences in our experiment. As such, this method has limitations. Hereafter, a more comprehensive methodology for multiworking modes could be required for monitoring moving conditions in general.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by Aviation Fund no. 2012ZD51055 and Fundamental Research Funds for the Central Universities no. YWF-13-T-RSC-100. Special appreciation should be given to Research Centre of Fluid Power Transmission and Control, Beihang University, for the sake of their selfless help.

References

[1] L. Chen, Y. Hang, T. Laifa, and L. Hongmei, "Performance assessment of hydraulic servo system based on bi-step neural network and autoregressive model," *Journal of Vibroengineering*, vol. 15, pp. 1546–1559, 2013.

[2] J. Watton, *Modelling, Monitoring and Diagnostic Techniques for Fluid Power Systems*, Springer, London, UK, 2007.

[3] M. Werlefors and A. Medvedev, "Observer-based leakage detection in hydraulic systems with position and velocity feedback," in *Proceedings of the 17th IEEE International Conference on Control Applications (CCA '08)*, pp. 948–953, San Antonio, Tex, USA, September 2008.

[4] V. Mahulkar, H. Mcginnis, M. Derriso, and D. E. Adams, "Fault identification in an electro-hydraulic actuator and experimental

validation of prognosis based life extending control," in *Proceedings of the 5th European Workshop on Structural Health Monitoring*, pp. 961–967, 2010.

[5] M. Quyen Le, M. Tu Pham, R. Moreau, J. P. Simon, and T. Redarce, "Bilateral control of nonlinear pneumatic teleoperation system with solenoid valves," *JIEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1463–1471, 2013.

[6] Q. Liu, H. Bo, and B. Qin, "Design and analysis of direct action solenoid valve based on computational intelligence," *Nuclear Engineering and Design*, vol. 240, no. 10, pp. 2890–2896, 2010.

[7] Q. Wang, F. Yang, Q. Yang, J. Chen, and H. Guan, "Experimental analysis of new high-speed powerful digital solenoid valves," *Energy Conversion and Management*, vol. 52, no. 5, pp. 2309–2313, 2011.

[8] R. M. Atkinson, M. R. Montakhab, K. D. A. Pillay et al., "Automated fault diagnosis for hydraulic systems. Part 1. Fundamentals," *Proceedings of the IMechE I: Journal of Systems and Control Engineering*, no. 206, pp. 207–214, 1992.

[9] D. Mustafa, G. Nilgun, K. Fevzi, and G. Abdulaziz, "Remote sensing of tea plantations using an SVM classifier and pattern-based accuracy assessment technique," *International Journal of Remote Sensing*, vol. 34, no. 23, pp. 8549–8565, 2013.

[10] M. Irfan, N. Saad, R. Ibrahim, and V. S. Asirvadam, "Development of an intelligent condition monitoring system for AC induction motors using PLC," *Business Engineering and Industrial Applications Colloquium*, pp. 789–794, 2013.

[11] ISO, 13379:2003, Condition Monitoring and Diagnostics of Machines—General Guidelines on Data Interpretation and Diagnostics Techniques, ISO, Geneva, Switzerland, 2003.

[12] ISO, 17359:2011, Condition Monitoring and Diagnostics of Machines - General Guidelines, ISO, Geneva, Switzerland, 2011.

[13] A. Steinboeck, W. Kemmetmuller, C. Lassl, and A. Kugi, "Model-based condition monitoring of an electro-hydraulic valve," *Journal of Dynamic Systems, Measurement, and Control*, vol. 135, pp. 061010/1–061010/9, 2013.

[14] L. G. Soo, S. H. Jin, and K. H. Chul, "Multiphysics analysis of a linear control solenoid valve," *Journal of Fluids Engineering-Transactions of the ASME*, vol. 135, no. 1, pp. 011104.1–011104.10.

[15] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.

[16] Y. Pang, L. Zhang, Z. Liu, N. Yu, and H. Li, "Neighborhood Preserving Projections (NPP): A novel linear dimension reduction method," in *Proceedings of the International Conference on Intelligent Computing (ICIC '05)*, pp. 117–125, August 2005.

[17] W. Huang and H. Yin, "Linear and nonlinear dimensionality reduction for face recognition," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 3337–3340, November 2009.

[18] Y. Musheng, *Higher Algebra*, Fudan University Press, Shanghai, China, 2004.

[19] O. Ivanciuc, "Applications of support vector machines in chemistry," *Reviews in Computational Chemistry*, vol. 23, pp. 291–400, 2007.

[20] X. Peng and D. Xu, "Twin Mahalanobis distance-based support vector machines for pattern recognition," *Information Sciences*, vol. 200, pp. 22–37, 2012.

[21] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[22] Y. X. Lai, C. F. Lai, Y. M. Huang, and H. C. Chao, "Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home," *Information Sciences*, vol. 230, pp. 39–55, 2013.

[23] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 374251, 8 pages
http://dx.doi.org/10.1155/2014/374251

Research Article

A Topology Construct and Control Model with Small-World and Scale-Free Concepts for Heterogeneous Sensor Networks

Lifang Liu,[1] Xiaogang Qi,[2,3] Jilong Xue,[4] and Mande Xie[5]

[1] *School of Computer Science and Technology, Xidian University, Xi'an 710071, China*
[2] *School of Mathematics and Statistics, Xidian University, Xi'an 710071, China*
[3] *The State Key Laboratory of Complex Electromagnetic Environmental Effects on Electronics and Information System, Luoyang 471003, China*
[4] *Department of Computer Science, Peking University, Beijing 100083, China*
[5] *School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China*

Correspondence should be addressed to Mande Xie; xiemd@zjgsu.edu.cn

Received 26 December 2013; Accepted 15 February 2014; Published 27 March 2014

Academic Editor: Su-Qun Cao

Topology construction and control is a vital technique in wireless sensor networks. In this paper, based on small-world and scale-free concepts of complex network theory and considering the characteristics of wireless sensor network, a topology model with small-world and scale-free concepts for heterogeneous sensor network is presented. This work is achieved by applying heterogeneous sensors and preferential attachment mechanism. Furthermore, the topology evolution algorithm is designed. Finally, we simulate the network characteristics, and simulation results are consistent with the theoretic analysis and show that topologies of wireless sensor network built by this model have small-world and scale-free feathers and can significantly improve energy efficiency as well as enhance network robustness, leading to a crucial improvement of network performance.

1. Introduction

Wireless sensor network (WSN) consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion, or pollutants, and to cooperatively transmit their data through the network to a sink node. Today WSNs are more and more widely used in a variety of industrial and consumer applications, such as industrial process monitoring and control, machine health monitoring, environment and habitat monitoring, health care applications, home automation, and traffic control [1].

In military and warfare applications, WSNs are deployed in hostile monitoring environment and the sensor node is with limited energy support. Energy exhaustion and natural damage of some sensor nodes often lead to the failure of the whole network. Construction and control of topology as a vital technique plays an important role to conquer these problems in WSNs. The main purpose of construction and control of topology is to achieve a higher communication quality, have efficient energy use and strong robust topology through power control, have important network node selection, remove unnecessary communication links, and so on.

Most topology models are based on some network theories such as random graph theory and complex network theory. The complex networks widely exist in real world such as electrical power grids, global transport networks, coauthorship and citation networks, and so on. As an interdisciplinary research area, complex networks arouse worldwide attention [2–4]. Two most typical network models in complex network theory are small-world network and scale-free network [4, 5]. The small-world network has two independent structural features: (i) a small average shortest path length and (ii) a large clustering coefficient [6]. By applying small-world theory in topology construction of WSNs, the network performance will be improved in querying data efficiency, energy efficiency, network lifetime, and so forth [7]. A scale-free network is a network whose node's

degree follows a power law distribution, and the scale-free topology characteristics have a higher robustness to endure the random failure [8]. However, few people study how to construct WSN topology with small-world and scale-free characteristics at the same time.

In this paper, considering the characteristics of WSNs, such as residual energy, degree saturation, and maximum communication radius of sensor, a topology model with small-world and scale-free concept (TMSSC) is proposed based on wireless heterogeneous sensor networks, and the topology evolution algorithm is designed. Topology constructed by this model not only has higher energy efficiency and transmission efficiency but also has higher robustness to endure the random failure.

The rest of the paper is organized as follows. Section 2 in this paper introduces the background of small-world networks and scale-free networks and their application in WSNs. Section 3 describes the TMSSC model and its algorithm, and also the network characteristic is analyzed. Finally, the simulation results analysis and conclusion are, respectively, presented in Sections 4 and 5.

2. Related Works

2.1. Small-World Networks in WSNs. Two main characteristics of small-world network are small average shortest path length and large clustering coefficient, which are the most important factors affecting the network performance. Helmy proposed a small-world concept wireless network through randomly adding a little of logical links to WSNs [9]; this leads to a small average shortest path length of networks, and then he proved that small-world network phenomenon also exists in wireless networks with spatial properties.

Based on the results of Helmy [9], Cavalcanti et al. improved the connectivity of wireless ad hoc networks by using small-world characteristics [10]. A few of sensor nodes with high energy and strong communication capability called H-sensor are introduced in this paper. Results show that H-sensors can improve the connectivity of networks significantly.

Research results by Chitradurga and Helmy [11] show that the average path length can significantly be decreased by adding shortcuts in the network. Specially, the length of shortcut just needs to be 25% to 40% of the diameter of a network with 1000 uniform distribution of sensor nodes; the average shortest path length of the modified network can reach 60% to 70% as that of original one.

Hawick paid attention to network coverage, fault tolerance, and sensor network lifetime by applying small-world theory [12]. Results show that adding a few of random links among sensor nodes can not only decrease the average path length but also decrease the number of isolated clusters, which lead to a great improvement of network coverage and lifetime.

2.2. Scale-Free Networks in WSNs. As the recent research focus, scale-free networks in WSNs are mostly improved by BA network model, which was proposed by Barabási and Albert [4]. This model is based on two important characteristics:

(i) growth: the scale of the network is expanding;

(ii) preferential attachment: the new node is more inclined to join with those nodes with higher degree.

BA scale-free network did not take spatial properties into consideration, so many scholars improved this model. For example, Saffre et al. [13] proposed an algorithm to build network topology with scale-free concept using local geographic information of sensor. Literature [8] proposed a topology evolution of WSNs among cluster heads by random walkers. Topology built by this model has a scale-free characteristic and better robustness, but this paper takes little consideration in communication features of WSNs, which might lead to a limited utility.

3. TMSSC Topology Evolution Model

In this section, we firstly introduce the topology model with small-world and scale-free concept (TMSSC) in heterogeneous wireless sensor networks, which consist of large number of sensors (L-sensor) and a small portion of super sensors (H-sensor). Firstly, some model assumptions are introduced. Then, we define some concepts and notations which may be used. Finally, the TMSSC topology evolution model is proposed and the algorithm is designed. Also, the dynamic characteristics of topology analysis are shown in this section.

3.1. Model Assumption. In our model, we assume the sensor networks consist of large portion of L-sensors and small portion of H-sensors. The H-sensors are equipped with high energy and strong communication capability and can communicate with H-sensor and L-sensor by using the different radios, respectively, in which one radio is responsible for long-distance communication and one is responsible for short-distance communication as equipped on the L-sensors.

In order to keep the symmetry of the network after adding H-sensor in WSNs, we assume that if an H-sensor v_i connects to another node v_j; it should at least satisfy the following conditions:

$$\begin{aligned} d\left(v_i, v_j\right) &\le L_{\text{high}}, \quad \text{if } v_j \in \text{H-sensor}, \\ d\left(v_i, v_j\right) &\le L_{\text{low}}, \quad \text{if } v_j \in \text{L-sensor}, \end{aligned} \tag{1}$$

where $d(v_i, v_j)$ is the Euclidean distance between node v_i and node v_j and L_{high} and L_{low} are communication radius of H-sensor and L-sensor, respectively. With the situation of v_i being an L-sensor, the condition is that the distance between 2 nodes just is less than the communication radius L_{low}.

Finally, we assume that the network topology management mechanism is achieved by using clustering. Clustering topology has the advantages in topology management, energy use, data fusion, and so on. This paper employs the HEED clustering algorithm which takes into consideration both the residual energy and communication cost [14]. So we assume

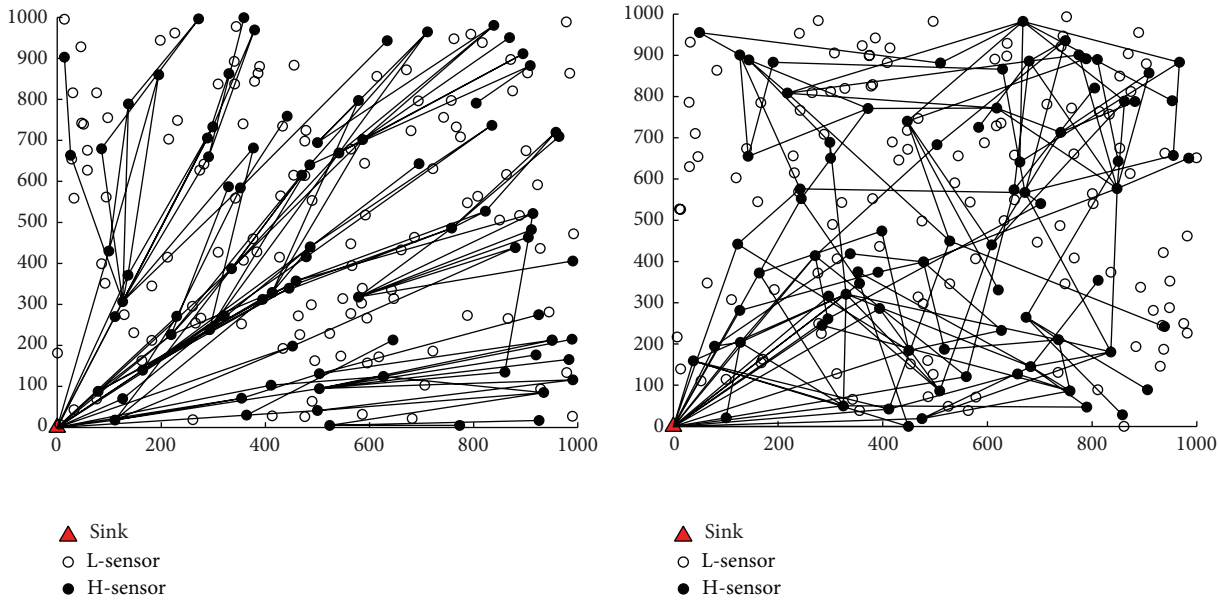


Figure 1: Topology construction of H-sensors with different connection mechanism: (a) connect with probability Φ_{ij} and (b) connect randomly.

that TMSSC topology model is built by using the cluster head nodes.

3.2. Definitions and Notions

3.2.1. Same Directed Angulation towards Sink Φ_{ij}. Same directed angulation towards sink Φ_{ij} defines the similarity between the directed angulation of node v_i towards sink and the directed angulation of node v_i towards node v_j. Since the goals of all the clustering head nodes are to transmit the collected data to the sink, the waste of communication energy is greatly reduced by taking into account the direction of edges between H-sensors [7]. We assume that the coordinates of node v_i are (x_i, y_i), the coordinates of sink node are (x_0, y_0), the straight line from v_i to sink is $y = m_1x + b_1$, and the straight line from v_i to v_j is $y = m_2x + b_2$; note the angle between two lines as θ; thus, $\tan\theta = (m_1 - m_2)/(1 + m_1m_2)$. The definition of Φ_{ij} between v_i and v_j is

$$\Phi_{ij} = \frac{1}{\tan\theta} = \frac{1 + m_1m_2}{m_1 - m_2}. \tag{2}$$

If node v_i connects with another node v_j with the probability of Φ_{ij}, all edges in the constructed topology will have the tendency towards sink node. Figures 1(a) and 1(b), respectively, show the topology in which edges between H-sensors are connected with probability Φ_{ij} and in random mode.

3.2.2. Other Definitions

(i) Degree k_i is the number of edges connected with node i.

(ii) Communication radius L: the communication radius of L-sensor is L_{low} and the value of H-sensor is L_{high}.

(iii) Residual energy E_i is the residual energy of node i; a node is death when $E_i = 0$.

3.3. Design of Model and Algorithm. Based on the above definitions and assumptions, we propose the TMSSC topology evolution model according to "preferential attachment" mechanism. The model is described as follows.

3.3.1. Growth. Initially the network just contains a small scale topology which consists of m_0 L-sensors, m_0 H-sensors, e_0 edges, and sink node. Then, choose a new node to join the topology every round. If the node is H-sensor, then add m_2 edges between this new node and current topology, and this new node only can connect to H-sensors. If it is L-sensor, add m_1 edges between this new node to current topology, and it can connect with any other nodes.

3.3.2. Preferential Attachment. If the new node is an L-sensor, what should be taken into account are energy balance, the degree saturation, and communication radius. Thus, the probability Π_j^l of new node v_i connecting to node v_j in the current topology is formulated as

$$\Pi_j^l = \frac{E_jk_j}{\sum_{i=1}^{N_l} E_ik_i} \cdot P\left\{d\left(v_i, v_j\right) \le L_{\text{low}}\right\}, \tag{3}$$

where E_i and k_i, respectively, indicate residual energy and degree of node v_i, N_l is the number of L-sensors, n is number of nodes in which degree is saturated, $d(v_i, v_j)$ is the distance of two nodes, and $P\{d(v_i, v_j) \le L_{\text{low}}\} = \pi L_{\text{low}}^2/S$.

Algorithm of TMSSC
Require: Initial topology with m_0 H-sensors and L-sensors;
(1) V = RandomGenerate(N);
(2) $N_{\text{in}} = 2 \times m_0$;
(3) L_{in} = initial $2m_0$ nodes;
(4) **for all** $v_i \in V$ **do**
(5) broadcastHelloPacket(v_i);
(6) receiveNeighborInformation();
(7) **end for**
(8) **while** $N_{\text{in}} < N$ **do**
(9) v_j = randomSelect($V - L_{\text{in}}$);
(10) **for all** $v_i \in$ Neighbor of v_j **do**
(11) $\Pi_i = \begin{cases} \Pi_i^l, v_j \in \text{L-sensor} \\ \Pi_i^h, v_j \in \text{H-sensor} \end{cases}$
(12) **if** random() $\le \Pi_i$ **then**
(13) v_i connect with v_j;
(14) $L_{\text{in}} \leftarrow v_j$;
(15) $N_{\text{in}} \leftarrow N_{\text{in}} + 1$;
(16) **end if**
(17) **end for**
(18) **end while**

Algorithm 1: Algorithm of TMSSC topology evolution model.

If the new node v_i is an H-sensor, it should consider the angulation towards sink Φ_{ij} besides above factors. Because of the longer distance between H-sensors than L-sensors, if the deviation of transmission direction is too large, it will waste energy greatly. Thus, the probability Π_j^l of new node v_i connecting to node v_j is formulated as

$$\Pi_j^h = \frac{\Phi_{ij} k_j}{\sum_{i=1}^{N_h} \Phi_{ij} k_i} \cdot P\left\{d\left(v_i, v_j\right) \le L_{\text{high}}\right\}, \tag{4}$$

where k_i indicate degree of node v_i and N_h is the number of H-sensors in current topology. The n is number of nodes in which degree is saturated, $d(v_i, v_j)$ is the distance of two nodes, and $P\{d(v_i, v_j) \le L_{\text{high}}\} = \pi L_{\text{high}}^2 / S$.

In order to implement this model, we design the algorithm pseudo code of TMSSC topology evolution model, which is shown in Algorithm 1. This algorithm eliminates the clustering process. More details about clustering algorithm are presented in [14].

3.4. Dynamic Analysis of Topology Evolution. According to mean-field theory [15], we assume that k_i is continuous, and thus the probability Π_i can be considered as a continuous rate of change of k. Consequently, we have the equation for node i as follows:

$$\frac{\partial k_i}{\partial t} = (1 - p) m_1 \Pi_i^l + p \cdot m_2 \Pi_i^h, \tag{5}$$

where p is the probability of H-sensor connecting to the current topology and $1 - p$ is the probability of L-sensor connecting to the current topology; Π_i^l and Π_i^h are formulated as (3) and (4).

Since the continuity of E_i, we have $E_{\min} \sum_{i=1}^{N_l} k_i \le \sum_{i=1}^{N_l} E_i k_i \le E_{\max} \sum_{i=1}^{N_l} k_i$ and $\exists E_\xi \in [E_{\min}, E_{\max}]$ such that

$$\sum_{i=1}^{N_l} E_i k_i = E_\xi \sum_{i=1}^{N_l} k_i. \tag{6}$$

Similarly, $\exists \Phi_\xi \in [\Phi_{\min}, \Phi_{\max}]$ such that

$$\sum_{i=1}^{N_h} \Phi_i k_i = \Phi_\xi \sum_{i=1}^{N_h} k_i. \tag{7}$$

Here we assume that a few nodes have reached the saturation value of degree $K_{\max}$. That is, n is very minimal so that it can be ignored here. At time t, taking into account that $\sum_{i=1}^{N_l} k_i = 2m_1 t$ and $\sum_{i=1}^{N_h} k_i = 2m_2 t$, we have

$$\frac{\partial k_i}{\partial t} = \beta \frac{k_i}{t}, \tag{8}$$

where

$$\beta = (1 - p) \frac{E_i \pi L_{\text{low}}^2}{2 S E_\xi} + p \frac{\Phi_i \pi L_{\text{high}}^2}{2 S \Phi_\xi}. \tag{9}$$

According to the assumptions, the initial condition of above equation is that node i is added to the topology at time t_i with connectivity $k_i(t_i) = (1 - p)m_1 + p \cdot m_2$; note that $m = (1 - p)m_1 + p \cdot m_2$. The solution of this equation is

$$k_i(t) = m\left(\frac{t}{t_i}\right)^\beta. \tag{10}$$

Using (10), the probability that a node has a connectivity $k_i(t)$ less than k, $P(k_i(t) < k)$, can be written as

$$P\{k_i(t) < k\} = P\left\{t_i > \frac{m^{1/\beta}}{k^{1/\beta}}t\right\}. \quad (11)$$

Also, we assume that node is added to the system at equal time intervals; the probability density of t_i is

$$pd(t_i) = \frac{1}{2m_0 + t}. \quad (12)$$

By substituting this into (11), we obtain the following equation:

$$P\{k_i(t) < k\} = 1 - P\left\{t_i < \frac{m^{1/\beta}}{k^{1/\beta}}t\right\} = 1 - \frac{m^{1/\beta}t}{k^{1/\beta}(t + 2m_0)}. \quad (13)$$

The probability density for $pd(k)$ can be obtained using

$$pd(k) = \frac{\partial P\{k_i(t) < k\}}{\partial k} = \frac{m^{1/\beta}t}{\beta(t + 2m_0)} \cdot \frac{1}{k^{\gamma}}, \quad (14)$$

where $\gamma = 1 + 1/\beta$. Accordingly, the degree distribution of this topology agrees with power-law distribution; in other words, this topology has scale-free characteristics.

4. Network Features and Simulation Results

4.1. Simulation Assumptions. In the simulation, we use a network with a total of 1000 nodes (L-sensors and H-sensors) randomly deployed in a sensor field of 1000 × 1000 m^2. There is only one sink node located at (0, 0). The communication radii of L-sensors and H-sensors are 60 m and 500 m, respectively. Initial number of nodes $m_0 = 3$, every time an edge is added into topology. All results are obtained by n times simulation, each time the simulation network topologies are different; here $n = 30$.

4.2. Small-World Features Evaluation. In this section, we evaluate the average minimum effective path length (L) and the clustering coefficient (C). The average minimum effective path length is defined as arithmetic average of the length of the shortest path from each node to sink node. The clustering coefficient of node v_i is defined as $C_i = n/(k_i(k_i - 1)/2)$, where n is the number of edges linked between node v_i's k_i neighbor nodes. Thus, the clustering coefficient of network is $C = \overline{C_i}$.

In each generated topology, we evaluate the average minimum effective path length and the average clustering coefficient of the network. After n simulations, we obtain the average of these n results. The values $C(p)$ and $L(p)$ are, respectively, defined as the clustering coefficient and the average minimum effective path length of the WSN topology with H-sensor probability p. The values $C(0)$ and $L(0)$ are defined as the same parameters but for no H-sensor addition. According to simulation results, the ratios of $C(p)/C(0)$ and $L(p)/L(0)$ are calculated. These ratios clearly show the influence of adding a fraction of H-sensors in the WSN.

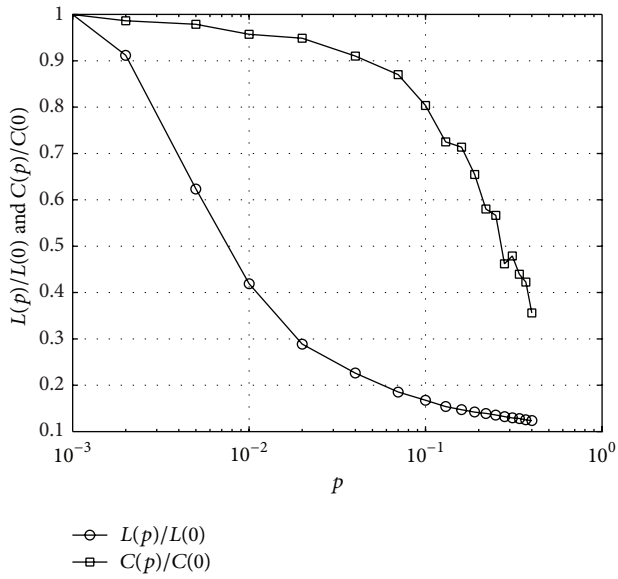


Figure 2: Small-world features of topology built by TMSSC in the simulation, where $N = 1000$, $L_{low} = 60$ m, and $L_{high} = 500$ m.

Figure 2 illustrates the average minimum effective path length and the clustering coefficient of the network when changing the value of the probability p. We find that the average minimum effective path length and clustering coefficient are not changed when probability $p = 0.001$, because those few H-sensors do not contribute too much to reduce the above parameters. When p is close to 0.01, a network with small-world characteristics can be generated, the average minimum effective path length is reduced to 16.7% of that in the original network, and clustering coefficient is 80.4% of that in the original network. When $p > 0.1$, the average minimum effective path length does not reduce significantly. However, the clustering coefficient reduces quickly. All these results show that when $p = 0.01 \sim 0.1$, the average minimum effective path length is reduced quickly, the clustering coefficient changed a little, and the network has the small-world features.

4.3. Scale-Free Features Evaluation. We find that when $p = 0.01 \sim 0.1$ the network built by TMSSC has small-world features. In this section, we will evaluate the degree distribution function $P(k)$ of networks with H-sensor probabilities $p = 0.01$ and $p = 0.1$. In each generated topology, we evaluate the degree distribution of the network. After n times simulation, we obtain the average values of these n times results. We can get Figures 3(a), 3(b), 3(c), and 3(d) and use them to evaluate the scale-free characteristics.

Figures 3(a) and 3(b) illustrate the degree distribution of the network, which show that most of the nodes in the network have small degree. For example, the degree of 82.5% nodes is less than 4 in Figure 3(a) and that of 81.9% nodes is less than 3 in Figure 3(b). This characteristic makes the network robust enough to resist the random node failures.

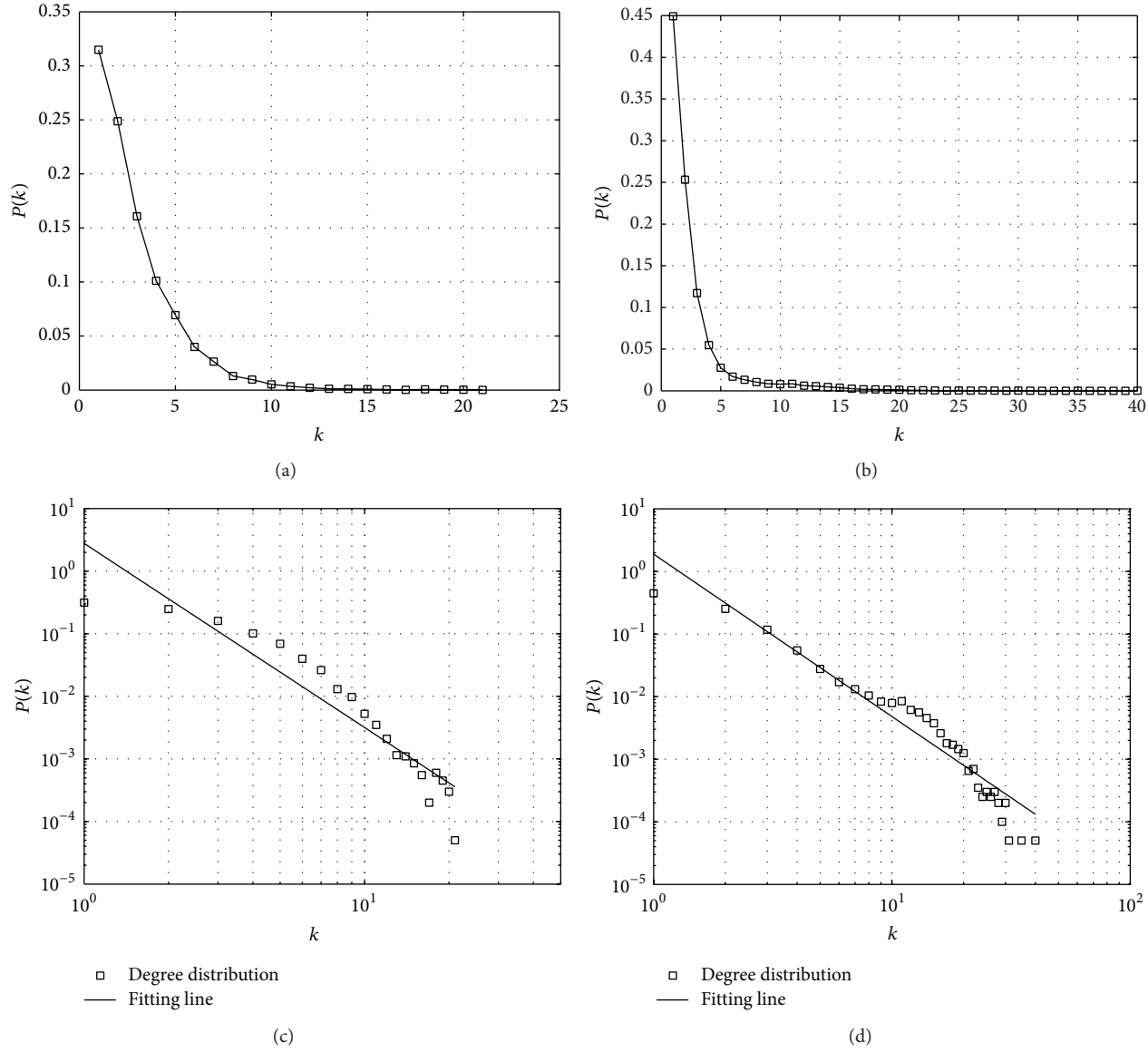


FIGURE 3: Degree distribution of topology built by TMSSC in the simulation: (a) degree distribution with $p = 0.01$, (b) degree distribution with $p = 0.1$, (c) degree distribution with $p = 0.01$ in Log-Log coordinates, and (d) degree distribution with $p = 0.1$ in Log-Log coordinates.

Owing to its scale-free structure, the random node failures cannot affect the network's performance significantly.

Figures 3(c) and 3(d) illustrate the above degree distribution in double logarithmic coordinates; we find that the degree distribution of networks built by TMSSC approximately obeys the power law distribution, and this agrees with the dynamic analysis results. In other words, the networks built by TMSSC have scale-free features.

4.4. Network Performances

4.4.1. Energy Efficiency. Compared with the BA model, TMSSC takes into consideration the energy balance. So, WSNs built by TMSSC perform a higher energy efficiency than the networks built by BA model under the same conditions. Under same assumptions with Section 4, we evaluate the energy efficiency by the survival nodes percentage by the BA and TMSSC model. In each deployed network, a random initial energy is given to each node. We build the two different topologies by TMSSC and BA model and run the network simulation, respectively. Every 5 rounds, we rebuild the networks and record the number of survival nodes. After n simulations, we obtain the average of these n results; see Figure 4(a). Another simulation is used to evaluate the impact of different number of H-sensors on the network lifetime; see Figure 4(b).

According to Figure 4(a), we illustrate that the survival nodes percentage changes over running time. We find that the TMSSC is better than the BA model. When it runs at 200 rounds, the survival nodes percentages of TMSSC and BA are 69.9% and 65.2%, respectively, because of the contribution of energy balance in TMSSC. Furthermore, in order to evaluate

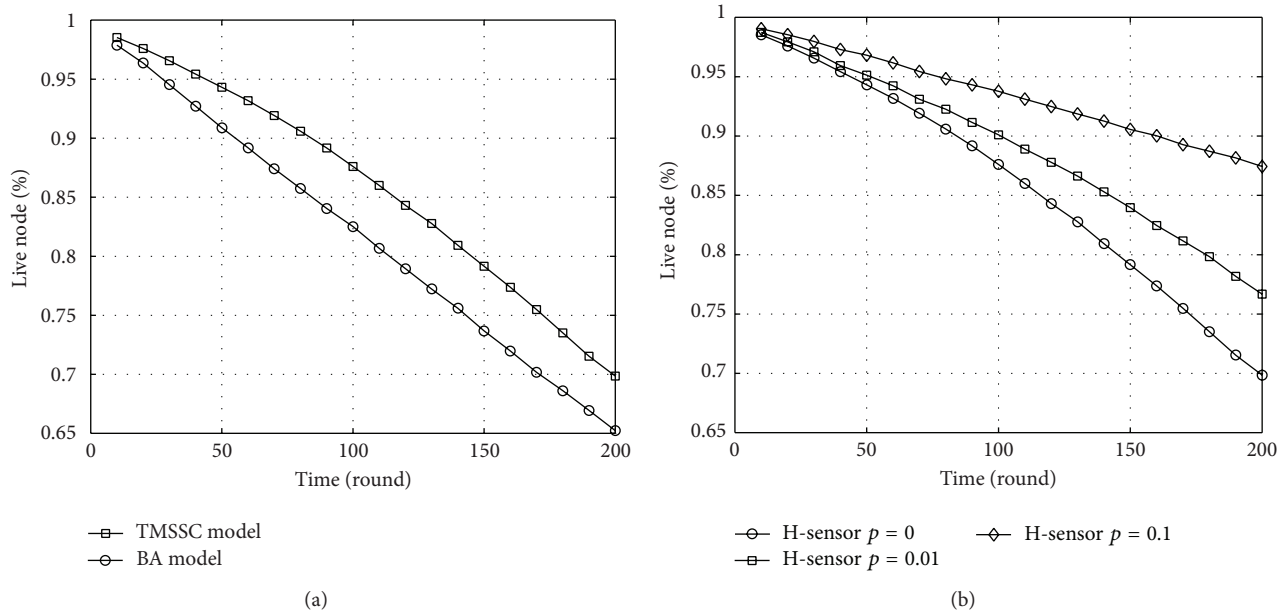


Figure 4: The contrast of energy efficiency in simulation: (a) energy efficiency of networks built TMSSC and BA model and (b) energy efficiency of networks built by TMSSC with $p = 0$, $p = 0.01$, and $p = 0.1$.

the energy efficiency with different probability of H-sensors of TMSSC, we run another simulation by adding $p = 0, 0.001, 0.1$ H-sensors into the same network, respectively, and the comparison under these three situations is shown in Figure 4(b). By adding H-sensor probability $p = 0.01$, the survival nodes percentage is improved by 9.8% than that in the original network when simulation runs at 200 rounds. When $p = 0.1$, the value is increased to 25.2%. Above results show that the network life time and the network energy efficiency can be highly improved by adding a fraction of H-sensors into WSNs.

4.4.2. Network Robustness. According to the analysis in Section 3, the topology built by TMSSC is with a scale-free feature. Networks with this characteristic have a higher robustness to resist the random node failures. In this section, we investigate the network robustness by using the concept of largest effective component S, which defines the largest component containing sink node in a WSN. We use the ratio $S(f)/S(0)$ to analyze the network robustness of randomly removing a fraction f of nodes in sensor network, where $S(0)$ is the original network with no node removed. We compare the robustness of three network topologies which are, respectively, built by BA model and TMSSC model with $p = 0.01$ and $p = 0.1$; the results are shown in Figure 5.

Figure 5 shows that networks built by TMSSC have a better performance on robustness against random node failures than BA model, and the robustness is improved by increasing H-sensors. In addition, this curve shows that the generated network has the characteristic of scale-free network. After randomly removing a little fraction of nodes, the network works well. When $p = 0.01$, there have been 14% of nodes removed from the network and the largest effective component is 80.1% of the original size. So, network built by TMSSC model has the scale-free properties and can resist the random node failures.

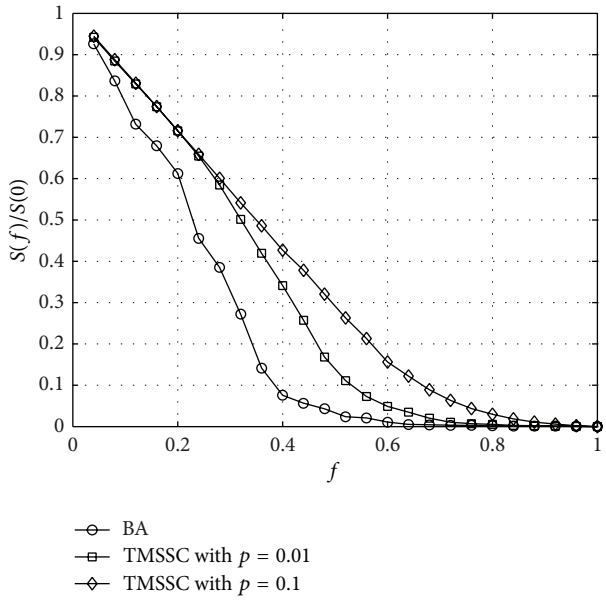


Figure 5: The robustness of networks built by BA and TMSSC model; the probability of H-sensor in TMSSC is $p = 0.01$ and $p = 0.1$.

5. Conclusion

In this paper, we firstly present a topology model with small-world and scale-free concepts for heterogeneous sensor

networks, and this model is achieved by the preference attachment mechanism and the heterogeneous sensor networking. Secondly, we design the topology evolution algorithm and give a dynamic analysis of topology evolution. Finally, we give an extensive simulation to evaluate the topology characteristics and network performances. Simulation results agree with theoretic analysis and show that WSNs topology built by this model has the small-world and scale-free characteristics. Also, the generated sensor networks by this method have higher energy efficiency and the stronger robustness against the random node failures and can highly improve the network performance.

Conflict of Interests

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

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under Grant nos. 71271165 and 61373174, the State Key Laboratory of Complex Electromagnetic Environmental Effects on Electronics and Information System under Grant nos. CEEME2012k0207B and CEEME2014k0302A, Grant no. 2011C14024 from Science and Technology Department of Zhejiang Province Program, and Grant no. 2010R50041 from Key Innovation Team of Science and Technology Department of Zhejiang Province.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.

[3] M. E. J. Newman and D. J. Watts, "Scaling and percolation in the small-world network model," *Physical Review E*, vol. 60, no. 6, pp. 7332–7342, 1999.

[4] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[5] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[6] M. Gitterman, "Small-world phenomena in physics: the Ising model," *Journal of Physics A*, vol. 33, no. 47, pp. 8373–8381, 2000.

[7] D. L. Guidoni, R. A. F. Mini, and A. A. F. Loureiro, "On the design of heterogeneous sensor networks based on small world concepts," in *Proceedings of the 11th International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiIM '08)*, pp. 309–314, Vancouver, Canada, 2008.

[8] L.-J. Chen, M. Liu, D.-X. Chen, and L. Xie, "Topology evolution of wireless sensor networks among cluster heads by random walkers," *Chinese Journal of Computers*, vol. 32, no. 1, pp. 69–76, 2009.

[9] A. Helmy, "Small worlds in wireless networks," *IEEE Communications Letters*, vol. 7, no. 10, pp. 490–492, 2003.

[10] D. Cavalcanti, D. Agrawal, J. Kelner, and D. Sadok, "Exploiting the small-world effect to increase connectivity in wireless ad hoc networks," in *Telecommunications and Networking*, vol. 3124 of *Lecture Notes in Computer Science*, pp. 388–389, 2004.

[11] R. Chitradurga and A. Helmy, "Analysis of wired short cuts in wireless sensor networks," in *Proceedings of the IEEE/ACS International Conference on Pervasive Services (ICPS '04)*, pp. 167–177, July 2004.

[12] K. A. Hawick and H. A. James, "Small-world effects in wireless agent sensor networks," *International Journal of Wireless and Mobile Computing*, vol. 4, no. 3, pp. 155–164, 2010.

[13] F. Saffre, H. Jovanovic, C. Hoile, and S. Nicolas, "Scale-free topology for pervasive networks," *BT Technology Journal*, vol. 22, no. 3, pp. 200–208, 2004.

[14] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004.

[15] A.-L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A*, vol. 272, no. 1-2, pp. 173–187, 1999.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 351707, 13 pages
http://dx.doi.org/10.1155/2014/351707

Research Article

Network-Coding-Based Energy-Efficient Data Fusion and Transmission for Wireless Sensor Networks with Heterogeneous Receivers

Lei Wang, Yuwang Yang, Wei Zhao, Lei Xu, and Shaohua Lan

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China

Correspondence should be addressed to Yuwang Yang; yuwangyang@mail.njust.edu.cn

Received 28 December 2013; Revised 8 February 2014; Accepted 18 February 2014; Published 23 March 2014

Academic Editor: Gelan Yang

This paper proposes a multirate network coding scheme to improve the energy efficiency of wireless sensor networks (WSNs). Firstly, a five-layer network model is introduced, in which a perfect solution can always be achieved. Then, a network topology which belongs to the five-layer network model will be established from the given sensor networks so that the receivers can receive data at a higher rate without introducing excessive additional transmissions, which improves the energy efficiency as a result. The performance of the proposed scheme is evaluated in a simulated instance, and the result shows that 9% of transmission overhead can be saved compared with traditional scheme.

1. Introduction

How to improve energy efficiency is a main issue in the field of wireless sensor networks (WSNs) since sensor nodes are always deployed in a serious environment and the power is supplied by batteries with limited energy. There are many different methods addressing the energy efficiency issue of WSNs, such as topology control [1, 2], coverage control [3, 4], and data fusion [5]. Network coding [6] is a relatively new technology proven to be effective in improving the throughput and energy efficiency. Network coding can be divided into three parts (i.e., encoding at source node, re-encoding at intermediate nodes, and decoding at receivers). Encoding and re-encoding are used to linearly mix original packets or received packets, so any encoded packet may include the information of multiple original packets. For this reason, network coding can be grouped under data fusion.

In the traditional computer communication networks, the principle of network design is to make the intermediate devices perform straightforward function, while the complicated function is performed in the devices working on the border of networks. Therefore, for a long time, it was generally believed that additional operations at the intermediate nodes would not generate any benefits. However, Ahlswede et al. reversed this belief in their study [6] by proving that after network coding, the network could transmit at the rate of multicast capacity (h) equal to the minimal max-flow to different receivers. This rate cannot be always obtained in traditional store-and-forward networks. After that, network coding has attracted many researchers' attention. Li et al. [7, 8] proved that linear network coding is sufficient to achieve the multicast capacity. Ho et al. [9] proposed the Random Linear Network Coding (RLNC) in which the encoding coefficients are randomly selected from a specified finite field. Based on RLNC, Chou et al. [10] proposed a practical network coding scheme which has been widely employed in subsequent studies [11–13]. Moreover, there were some coding schemes based on deterministic algorithm in previous studies [14–16].

Network coding has been proven to be energy efficient in both the wired networks and wireless networks, and many researchers have proposed various network-coding-based routing protocols to improve the energy efficiency of WSNs. We observe that, in most traditional network coding schemes, all receivers receive data at the same rate equal to the multicast capacity. However, when wireless network is adopted to transmit media information, such as figure, sound, and video, the receivers prefer receiving as fast as possible, so

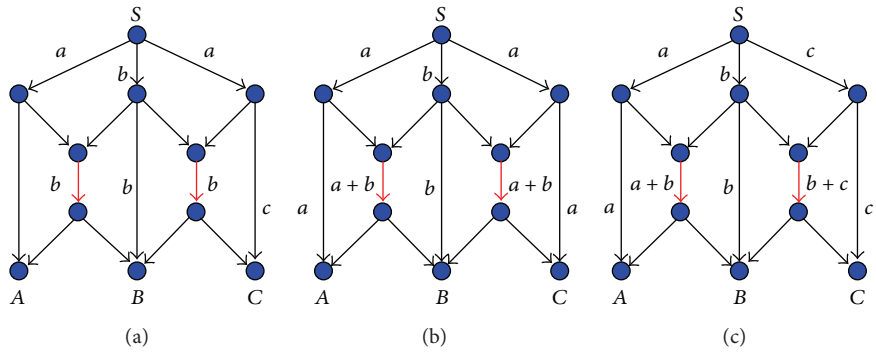


Figure 1: Multicast network with heterogeneous receivers.

it is desirable to transmit at a higher rate, which may even be higher than multicast capacity. Take Figure 1, for example.

In Figure 1, the max-flows of nodes A, B, and C are 2, 3, and 2, respectively, so multicast capacity h of this network is 2. Note that there are two links becoming bottlenecks, which are highlighted in red. If network coding is not allowed, receivers cannot simultaneously achieve multicast capacity due to the bottlenecks, which is shown in Figure 1(a). In Figure 1(b), multicast capacity can be achieved after using network coding. Furthermore, with the coding scheme shown in Figure 1(c), the nodes A and C could receive data at rate of 2 and node B could receive data at the rate of 3, which implies that the total transmission rate has been increased. Such a scheme enables the receivers to receive more data in a transmission period, so the total required transmission times for a given amount of data will be reduced. Based on this idea, this paper proposes a scheme to improve the energy efficiency of WSNs.

Figure 1(c) is a simple example which shows that certain additional throughput can be obtained. In Figure 1, the max-flows for the three receivers are 2, 2, and 3, respectively. With the new encoding scheme, the transmission rate can be increased by 11.6%. In some practical networks, max-flow for each receiver may be different. For example, if the max-flows are 2, 3, and 4, respectively, then the multicast capacity is 2. Therefore, 1/3 of bandwidth is given up of traditional network coding is employed. If we can make all the receivers receive data at the rate of individual max-flows, the average transmission rate will be increased by 50%. In this example, we observe that the network can transmit more data to receivers in a transmission period without involving excessive links. In this sense, the energy efficiency can be increased, especially when the receivers have different max-flows.

Before network coding was proposed, many researchers [17, 18] addressed the multirate multicast networks by employing layered coding, in which the source node encodes data stream into a base layer and several enhancement layers. If a receiver requires the data on layer k, it must obtain the data on layers $1, \ldots, k-1$. In some recent studies [19, 20], the layered multicasting method was combined with network coding to increase the throughput of networks. Figure 2 shows the layered coding scheme.

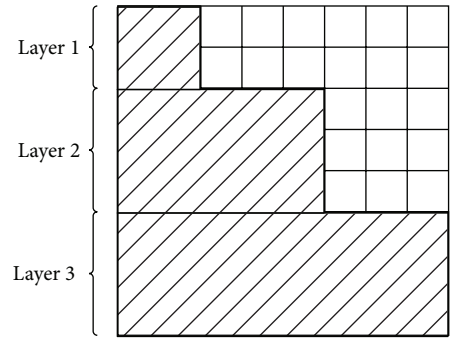


Figure 2: Layered coding scheme.

Figure 2 can be considered as an 8-by-8 matrix, and the area without shadow is filled with zero. After encoding original packets, it is feasible for the receivers to decode the first and second packets with vectors in row 1 and row 2. Similarly, with the packets in layer 1 and layer 2, the receivers can decode the first five packets, and all the eight packets can be obtained with the data in all 3 layers. In this way, the receivers can receive data at different rates.

Moreover, there are also some studies [19–22] on the practical applications based on multi- rate network coding, such as the multimedia communication. Sundaram et al. [21] proposed a polynomial time algorithm to construct multi-rate network codes for multimedia communication network with heterogeneous receivers. Xu et al. [22] also addressed the construction of multirate network codes through linear programming, and they provided two algorithms to improve the throughput of video stream. When network coding is combined with layered multicast, re-encoding can be performed within the same layer or across different layers, which are regarded as inter-layer network coding and intra-layer network coding. In general, inter-layer network coding outperforms intra-layer network coding, so this paper mainly focuses on the former.

This paper proposes a multirate network coding scheme to improve the energy efficiency of WSNs. This scheme can improve energy efficiency on three aspects: firstly, it

can reduce the number of intermediate nodes in which re-encoding operation is required, thus reducing the number of required transmission links; secondly, it can make maximal data fusion with the original packets, which ensures that more data can be sent to receivers in a transmission period; thirdly, this scheme can work on a very small finite field, which indicates that the computation overhead of network coding is low.

The remainder of this paper is organized as follows. In Section 2, some closely related studies are introduced. In Section 3, some preliminaries of linear network coding and employed network model will be described; Section 4 describes the proposed multirate network coding scheme. Section 5 describes some algorithms in the proposed scheme. The result of performance evaluation is provided in Section 6. Finally, the conclusion is made in Section 7.

2. Related Works

Koetter and Médard [23] proved that the perfect solution that can make each receiver receive data at individual max-flow does not always exist. Wu et al. [24] also addressed this issue, and they proved that it is an NP-hard problem to find the multirate solutions that maximize the total transmission rate. Although the perfect solution does not always exist and maximization of the total transmission rate is an NP-hard problem, it is feasible to design algorithms to achieve a good solution.

Kim et al. [25] proposed two prominent heuristic algorithms to construct multirate network codes, called Min-Req and Min-Cut. The two algorithms first determine the max-flows from source node to each node (including both the intermediate nodes and receivers) and then push back the rate requirements from receivers to the source node in an order opposite to the topological order. All the intermediate nodes forward their rate requirements to source node according to Min-Req policy or Min-Cut policy. After the pushback stage, the network code for each link will be assigned according to their rate requirements. Although both methods are efficient and straightforward in implementation, there are some disadvantages. Firstly, all the intermediate nodes and receivers require the knowledge of their max-flows to the source node, so the max-flow algorithm (such as Ford-Fulkerson) needs to be performed many times. Secondly, the decoding at intermediate nodes is required, which will incur higher computation and storage overhead. Thirdly, the link usage rate of these two schemes is very high.

To avoid the decoding operation at intermediate nodes, Widmer et al. [26] also proposed an algorithm. Compared with the two algorithms of Kim et al. [25], the algorithm could achieve a close or even higher throughput. This is a promising scheme in the field of multirate network code construction. Their scheme can be further improved because their flow allocation algorithm may limit the ability of network coding sometimes. When a link has multiple incoming links and is shared by multiple receivers, network coding will be necessary to increase the throughput. However, in their method, the link will be firstly assigned to the receiver with smallest max-flow. After this, when assigning flows for strong receivers, the layer on this link cannot be merged with higher layers or the weak receiver cannot successfully decode. Therefore, their method disqualifies the link for network coding. Theoretically, if network coding is not performed at the intermediate nodes, the potential of network coding will be limited.

Through these algorithms, a good solution can be found for a given network topology. For WSNs, after the sensor nodes are deployed into the environment, the first step is to establish routes such that all nodes are organized into a network. In this paper, we transfer the burden of multirate network code calculation to routing establishment. We will establish a subgraph from the given graph. In the subgraph, we expected that it is straightforward to construct multirate network codes. To this end, we propose a five-layer network model. In this network model, it is easy to construct the multirate network codes. When constructing routing, we ensure that the subgraph must belong to the five-layer network model, and then we can easily assign multirate network codes for WSNs.

In our previous study [27], we also addressed the issue of energy efficiency of WSNs by introducing network coding, and we showed that after combining network coding and opportunistic multipath routing, the energy efficiency could be improved. Through further study on network coding, we observe that the energy efficiency can be further improved. The differences between this paper and the previous one include the following: firstly, a single-rate network coding scheme was employed in the previous study, while a multirate network coding scheme is proposed in this one; secondly, in that paper, it was assumed that all the intermediate nodes were able to reencode, while the number of re-encoding nodes is reduced in this one, which implies that the re-encoding overhead and transmission delay can be reduced; thirdly, in the previous study, the coding scheme was based on a randomized algorithm [10], while the algorithm employed in this paper is a deterministic one proven to be able to perform over a very small finite field. Therefore, the computation overhead of coding is reduced.

Our contribution can be summarized as follows: firstly, we proposed a five-layer network model for multirate network coding and showed that it is easy to assign multirate network codes for the five-layer network model. Moreover, we showed that a perfect solution always exists in the network model. Secondly, we provided a method to establish routing for sensor networks such that the established network belongs to a five-layer network model, and therefore the burden of calculating network codes is reduced. Thirdly, we introduced multirate network coding into WSNs to increase the energy efficiency of transmission.

3. Preliminaries

3.1. Linear Network Coding. Generally, a network can be demonstrated by a directed graph $G = (V, E)$, where V refers to the node set and E refers to the link set. The node set V consists of a source node S, multiple intermediate nodes,

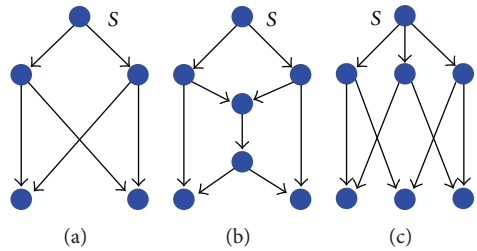


Figure 3: Different kinds of multicast network.

and one or more receivers T. According to previous studies, the network with multiple source nodes can be transformed into the network with one source node, so this paper only addresses the network with one source node. The child node set of node v is denoted by $Child(v)$, and $|Child(v)|$ represents the number of child nodes in the set. Similarly, $Parent(v)$ refers to the parent set of node v. In addition, $\mathrm{MF}(v)$ represents the max-flow from node v to the source node, and r represents the maximal value in $\{\mathrm{MF}(v_i)\}$, $v_i \in T$.

As is known, the "XOR" operation used in Figure 1 is the simplest and most efficient network coding scheme, in which the required field size is 2. In a network with complicated topology, a general method of linear network coding should be employed. When the source node needs to multicast some original packets to a group of receivers, it will firstly select h original packets $P = \{p_1, p_2, \ldots, p_h\}$ and then obtain a $(n \times h)$ matrix G to encode the h original packets. After using a random or deterministic algorithm, the generator matrix G will be obtained, and then new packets $Y = \{y_1, y_2, y_3, \ldots, y_n\}$ can be calculated, as shown in the following:

$$Y = GP = \begin{bmatrix} g_{11} & \cdots & g_{1h} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nh} \end{bmatrix} [p_1, p_2, \ldots, p_h]^T \Longrightarrow y_k \quad (1)$$

$$= \sum_{i=1}^{h} (g_{ki} \times p_i), \quad k \in (1, \ldots, n).$$

After obtaining the new packets, the source node sends the packets Y out along with the corresponding coefficient vectors to its child nodes. If an intermediate node receives one encoded packet, it will forward the packet; if it receives more than one encoded packet, it will reencode the received packets before sending them out. When a receiver gets encoded packets or reencoded packets from its parent nodes, the received coefficient vectors in these packets will form a receiving matrix R according to which the receivers can recover the original packets $P = \{p_1, p_2, \ldots, p_h\}$ through the Gauss eliminating method.

For multirate network coding, the scheme is supposed to make the receivers receive data at different rates, which implies that the receiver could decode part of P from the data it receives, even if it could not decode all the data in P.

3.2. Network Model. Although network coding is proven to be efficient in both the unicast network and multicast network [28], we mainly discuss it in the multicast network, and the unicast network can be considered as a special case of multicast network. In general, the multicast network includes the following several kinds.

Figure 3(b) shows the classical butterfly model, in which there are two receivers, each with two disjoint paths to the source node S, and there is also an intermediate node which needs re-encoding. The network topology shown in Figure 3(a) can also be considered as a multicast network since there are two receivers, and the two receivers are connected to the same group of parent nodes, so they have equal status. For the network topology shown in Figure 3(c), there are three receivers, and each receiver is connected to a different group of parent nodes. In a practical network, the network may be one of these kinds or their combination. In this paper, we try to address the multicast network in generalized networks, so we generalize the multicast network from the following directions.

(i) Generalized Case 1: the multicast capacity should be generalized to $h \geq 2$, and the max-flow for each receiver may be different.

(ii) Generalized Case 2: the number of re-encoding intermediate nodes should be generalized. For example, in Figure 3(b), there is only one intermediate node which needs re-encoding, while in a generalized multicast network, the number is increased.

(iii) Generalized Case 3: the number of the child nodes of the source node $|Child(S)|$ should be greater than the multicast capacity; that is, $|Child(S)| > h$.

In this paper, we construct the multirate network codes for generalized multicast network mentioned above. To this end, we firstly introduce a five-layer network model as shown in Figure 4. The five-layer network model includes Cases 1 and 2, and it can be easily extended to Case 3. Moreover, due to the topology feature of the five-layer network model, it is easy to assign multirate network codes for such kind of network model. In this paper, we will establish routing such that the established network belongs to the five-layer network model, and then the work to assign network codes could be simplified.

In the five-layer network model, r is the maximal value of max-flows from each receiver to the source node. Obviously, r is greater than the multicast capacity (h) or equal to h when all the receivers have the same max-flow. The number of intermediate re-encoding nodes may be greater than 1.

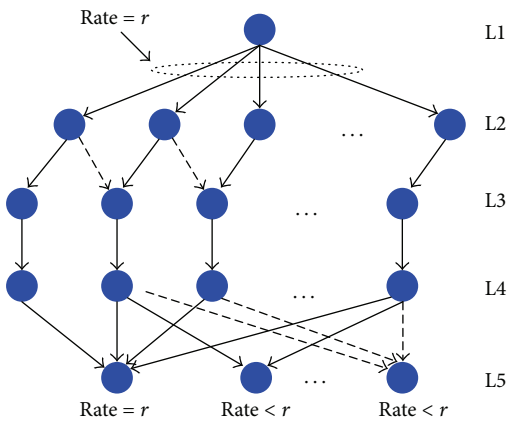


Figure 4: Five-layer network model.

Moreover, the receivers may have different max-flows. It should be noted that the concept of "layer" in the network model is different from that in the layer coding scheme. In the five-layer network model, layer x represents the nodes in $(x-1)$ hops from the source node, while in the layer coding scheme, layer x represents the encoding vectors with which the receivers could decode the original packets.

In this paper, we firstly provide a method to construct the multirate network codes for the five-layer network model, in which Cases 1 and 2 are included. However, Case 3 is not included in the five-layer network model. Then, based on the construction method for the five-layer network model, we propose an edge-eliminating method to construct multirate network codes for the network in Case 3 in Section 4.5.

Notice that, in the five-layer network model, each receiver has several disjoint routes to the source node, and there are four hops from the source node to each receiver. In a practical network, after the routes for each receiver have been established, the number of disjoint routes must be equal to its max-flow, but the hops in each route may be different. Take the network shown in Figure 1 for example; there are some routes to the receivers in two hops. In this case, the network can also be considered as a five-layer network since the network has the features of five-layer network model. On the other hand, if the hops to receivers are greater than 5, it can also be considered as a five-layer network. Therefore, the five-layer network model represents the networks in which MF(i) disjoint routes are established for receiver i, and some links may become the bottlenecks between the third layer and the fourth layer.

4. Proposed Energy-Efficient Network Coding Scheme

Network coding can be divided into three parts: encoding at the source node, re-encoding at the intermediate nodes, and decoding at the receivers, which can be expressed by the following equation:

$$(TG)^{-1}TGP = P. \quad (2)$$

In (2), P is called the data matrix, and it represents the original packets $[p_1, p_2, p_3, \ldots, p_r]$, in which each row vector is an original packet, G is called the generator matrix, and T is called the transfer matrix. After employing network coding, P will experience transformations for three times: firstly, GP corresponds to the encoding at the source node; secondly, multiplying the result of GP by matrix T represents the re-encoding operation at the intermediate nodes; finally, multiplying the result of $(T(GP))$ by matrix $(TG)^{-1}$ represents the decoding operation at the receivers. It should be noted that there are r row vectors in P, and the length of these vectors is arbitrary. Moreover, matrix G and matrix T must be square matrices. In a generalized multicast network, there is more than one receiver. If a network coding scheme makes (2) hold for each receiver, the network coding scheme can be considered valid. It should be noted that different receivers may have different generator matrices and transfer matrices, but they have the same data matrix.

Equation (2) shows the condition of valid network codes, but it holds only in single-rate network coding, where the receivers receive at the same rate h ($h = r$). For each receiver, there must be a generator matrix and transfer matrix satisfying (2) in the unirate network coding scheme. However, in a multirate network ($r > h$), (2) only holds for the receiver with r disjoint routes to the source node, and any other receiver may obtain less than r packets from a subset of the r disjoint routes.

Therefore, the design of multirate network codes can be considered as consisting of the following steps: firstly, we need to obtain a full-rank generator matrix and a transfer matrix for the receiver with r disjoint routes to the source node; secondly, we must ensure that all the other receivers can receive as much data as possible.

4.1. Encoding at the Source Node. In this section, we address how to construct multirate network coding without considering the re-encoding operation at the intermediate nodes; in other words, the transfer matrix in (2) is assumed to be an identity matrix.

For the five-layer network model, the only work of the source node is to obtain a full-rank generator matrix. Generally speaking, many methods can be used to obtain a full-rank matrix. For example, the Vandermonde square-matrix must be a full-rank matrix. By multiplying an upper triangle matrix with a lower triangle matrix, it will definitely generate a full-rank matrix. In a single-rate network coding scheme, all the receivers have the same max-flows, so they must be able to decode simultaneously as long as they receive all the encoded packets. However, in a multirate network coding scheme, the number of disjoint routes for each receiver may be different. The receivers with r disjoint routes can successfully decode r original packets, but the receivers with k ($k < r$) disjoint routes cannot receive r linearly independent packets. In this case, if the receivers want to decode k original packets with k received packets, the receiving matrix R must satisfy that $r-k$ unknown elements are indexed by zero in all the k received vectors.

4.2. Re-Encoding at the Intermediate Nodes. As a matter of fact, in a network-coding-based transmission scheme, the improvement of network performance, such as higher throughput and lower energy consumption, always benefits from the re-encoding operation at the intermediate nodes. Without the re-encoding operation, the scheme cannot be considered as a network coding scheme but only a channel coding scheme, and therefore the benefits of network coding cannot be obtained. As mentioned above, the principle of network design is to make the intermediate nodes perform simple functions. However, network coding employs the re-encoding operation at intermediate nodes, which may increase the complexity. Generally speaking, because the re-encoding operation can significantly improve the network performance, such as the throughput and energy efficiency, it is considered worthwhile to use network coding. Although it is proven that the re-encoding operation does more good than harm, we believe that it is necessary to reduce the number of intermediate re-encoding nodes without sacrificing any performance gain of network coding.

From (2), we know that the receiver can decode the original packets under the condition that both matrices G and T are full-rank matrices. In Section 4.1, we ensure that the generator matrix for each receiver is full-rank, and in this section we describe our method to construct the full-rank transfer matrix T. The transfer matrix T is determined by the intermediate nodes, which implies that the re-encoding operation at any intermediate contributes to the transfer matrix T. In accordance with the theory of matrices, we obtain the following theorem.

Theorem 1. *A sufficient and necessary condition for an invertible square matrix M is that $\exists$ a finite number of elementary matrices $P_1, P_2, \ldots, P_n$ such that $M = P_1 P_2 \cdots P_n$.*

Therefore, we observe that if the transfer matrix T is full-rank, it must be derived from multiple elementary matrices. If we can ensure that each re-encoding operation at the intermediate nodes is equivalent to one or several elementary transformations, the transfer matrix must be full-rank. Our re-encoding method is based on this idea. On the other hand, if a transformation is not equivalent to one or several elementary transformations, the corresponding re-encoding operation must result in a failure at the receiver. In this case, we obtain Lemma 2.

Lemma 2. *A sufficient and necessary condition for a valid re-encoding operation at an intermediate node is that the re-encoding operation must be equivalent to one or several elementary transformations.*

Now, the problem of re-encoding at the intermediate nodes is translated into a pure mathematical problem. Next, we will provide our method and show that re-encoding operation of our method must satisfy the condition in Lemma 2 in the five-layer network model.

First of all, we discuss the re-encoding operation on the butterfly model and then extend the discussion to the five-layer network model. In the butterfly model as shown in Figure 3(b), the following transfer matrices can be obtained:

$$\text{Left Sink } R = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = TG = T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Longrightarrow T = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

$$\text{Right Sink } R = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = TG = T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Longrightarrow T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{3}$$

Obviously, both transfer matrices are elementary matrices. Then, we prove that, in the single rate multicast network model, the re-encoding operation also satisfies the condition in Lemma 2. Before this, let us discuss the number of re-encoding intermediate nodes. According to our assumption, the source node needs to establish h disjoint routes for each receiver, so we consider the upper bound of the number $h-1$, and the lower bound is 0; that is, there are no intermediate nodes required to perform network coding. For each receiver, there are h disjoint routes to the source node. We consider that the difference in the number of hops does not affect the network-coding-based transmission scheme, and at most $(h-1)$ routes will become the bottlenecks. If the number of shared routes is h, the network model will be reduced to the case as shown in Figure 3(a); that is, all the receivers are connected to the same group of parent nodes.

Then, we come back to the discussion of the validity of re-encoding operation at the intermediate nodes in the single-rate multicast network. Note that if the proposition holds when the number of re-encoding nodes is $h-1$, it will definitely hold when the number of re-encoding nodes is less than $h-1$, so we only need to prove the case in which the number of re-encoding nodes is $h-1$. Let the generator matrix be $[\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_h]^T$, in which each element represents a row vector, and after re-encoding at the intermediate node, the received coefficient matrix is $[\alpha_1, \alpha_2 \oplus \alpha_1, \alpha_3 \oplus \alpha_2, \ldots, \alpha_h \oplus \alpha_{h-1}]^T$, which must be a full-rank matrix. Therefore, the condition in Lemma 2 is satisfied. The difference between our method and the pervious deterministic methods is that we ensure the transfer matrix is full-rank when establishing routing, while the previous method assigned network codes after the route establishment. Therefore, our method must ensure that there are no braided routes from layer 2 to layer 3, while the previous methods do not have this restriction.

So far, we provide our method to obtain the full-rank matrices G and T, so the receivers in single-rate network coding must be able to decode. Moreover, the receiver with the maximal max-flow (r) in multirate network coding is able to decode as well. We wonder whether the other receivers in multirate network coding could decode. To this end, we need to address the influence of re-encoding on the decoding of the receivers first.

4.3. Influence of Re-Encoding on Decoding of Receivers. In this section, we address whether the re-encoding operation at intermediate nodes influences the decoding of receivers whose max-flows are less than r in the multirate network coding. In the single-rate network coding, the re-encoding

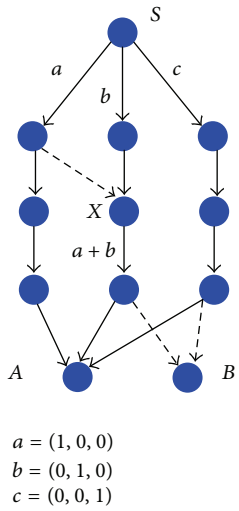


Figure 5: Re-encoding in multirate network coding.

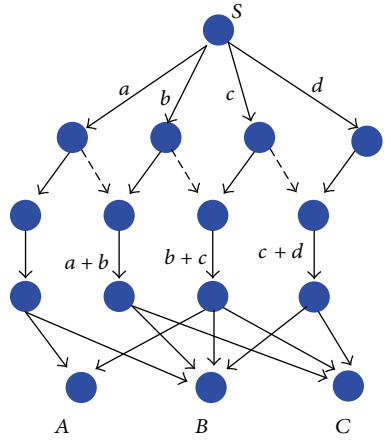


Figure 6: Five-layer network with heterogeneous receivers.

operation will not result in decoding failure at the receivers as long as it is ensured that the re-encoding operation equals one or several elementary transformations. However, it will result in decoding failure in the multirate network coding scheme. In this section, an example will be provided to show the influence of re-encoding at the intermediate nodes on multirate network coding; then, the reason will be addressed; finally, how our method for intermediate nodes can be used to avoid decoding failure will be discussed.

Example 3. Take the network in Figure 5, for example.

As shown in Figure 5, there are three disjoint routes from the source node to receiver A and two disjoint routes to receiver B. After employing re-encoding at the intermediate node X, the receiver A can successfully decode the three original packets. However, the receiver B can only obtain one packet.

Recall that our objective is to avoid the decoding failure induced by the intermediate nodes, and in the meantime the network could multicast at a high rate. Intuitively, we consider that receiver B in Figure 5 can receive data at a high rate, so we need to know what the maximal multicast rate is in our model.

4.4. Maximal Transmission Rate in Five-Layer Network Model. In the five-layer network model, the receivers may have different max-flows. We wonder whether there is a method to make each receiver receive data at the rate of individual max-flow, which suggests that the total transmission rate equals $\sum \mathrm{MF}(t)$, $t \in T$. $\mathrm{MF}(t)$ refers to the max-flow from source node to receiver t in the established subgraph. Take the network in Figure 6, for example.

In Figure 6, the max-flows for nodes A, B, and C are 2, 4, and 3, respectively, and the symbols a, b, c, and d represent the encoding vectors. The network coding scheme must ensure that node A can decode two original packets from $Parent(A)$, B can decode four original packets from $Parent(B)$, and C can decode 3 packets with $Parent(C)$. Then, we address whether there exist network codes which can make all the three receivers receive data at individual max-flows. We believe that, for the five-layer network model, the multirate network codes which make each receiver t ($t \in T$) receive at the rate of $\mathrm{MF}(t)$ always exist. Then, we explain this perspective in the five-layer network, in which r is arbitrary and the receivers have different max-flows. If we consider that valid network codes must exist, what we need to do is to provide an encoding scheme to satisfy the condition, even though there might be many encoding schemes which could satisfy the condition. From the five-layer network model, we know that the receivers are in the fifth layer, and they are connected to the nodes in the fourth layer. Each receiver may be connected to different groups. We observe that if the global encoding vectors for the links from layer 4 to layer 5 are $n_1e_1, n_2e_2, \ldots, n_re_r$ (e_i is an identity vector with the ith element being 1), respectively, each receiver i, $i \in T$ must be able to get original packets, and the number of original packets must be equal to its max-flow. From (2), we know that if the packets from layer 3 to layer 4 are original packets, then there exists generator matrix G such that $GT = I$, where I is an identity matrix. In the last section, we pointed out that the transfer matrix T in the five-layer network model must be a full-rank matrix. Therefore, if the generator matrix G is set to be the inverse of T, then GT must be equal to an identity matrix; thus, all the receivers must be able to receive data at the rate of individual max-flows.

We have shown that it is feasible to multicast at a total rate of $\sum \mathrm{MF}(t)$, $t \in T$, in the five-layer network model. Then, we come back to address the re-encoding operations at the intermediate nodes, and the only work of intermediate nodes is to mix the input packets into new packets, which has been proven to be valid for the five-layer network model. For the source node, it must satisfy two conditions: firstly, the generator matrix must be a full-rank matrix; secondly, each receiver (t) receives data at the rate of $\mathrm{MF}(t)$. We proved that the transfer matrix T must be full-rank in the five-layer network model, and therefore T^{-1} must satisfy condition 1. Moreover, in such encoding scheme, the packets transmitted

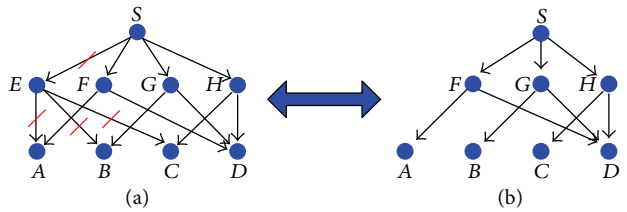


Figure 7: An example of edge-eliminating method.

from layer 3 to layer 4 are original packets, so the second condition can also be satisfied.

4.5. Extend the Method to the Network of Case 3. As mentioned in the last section, the multicast network could be divided into several categories. Moreover, in accordance with Figure 3, we observe that source nodes in Figures 3(a) and 3(b) have two outgoing links, but the source node in Figure 3(c) has three outgoing links even though the multicast capacity for each network is 2. If we divide the network according to the number of outgoing links of the source node, there will be two categories: in one case, $|Child(S)|$ is greater than the multicast capacity (h) and in the other case, $|Child(S)|$ equals h. We observe that the design of multirate network codes for the first category could be transformed to that for the second category. To this end, we propose an edge-eliminating method.

Example 4. Take the network in Figure 7, for example.

In the network shown in Figure 7, the max-flows for nodes A, B, C, and D are 2, 2, 2, and 3, respectively. The disjoint routes to node D are S-F-D, S-G-D, and S-H-D, so we need to eliminate node E and its associated links. Then, the resulting topology is shown in Figure 7(b), and the new max-flows for nodes A, B, C, and D are 1, 1, 1, and 3, respectively. In the new topology, there are no intermediate nodes needed to be reencoded, which implies that the transfer matrix is an identity matrix. Therefore, the network codes generated by the source node will not be affected by any intermediate nodes. The only work of the source node S is to obtain a 3-by-3 generator matrix G' to encode the original packets, which must guarantee that all the receivers could receive at their new max-flows. After obtaining the matrix G', the source node needs to assign a global encoding kernel for $Link(S, E)$, which must guarantee that the nodes in $Child(E)$ will obtain a new packet. Since the encoding vector is 3-dimensional, nodes A, B, and C could receive at rate 1 and node D could receive at rate 3, and the network codes assigned to $Link(S, F)$, $Link(S, G)$, and $Link(S, H)$ must be n_1e_1, n_2e_2, and n_3e_3 ($n_i \neq 0$, e_i is an identity vector with the ith element being 1), respectively. However, we find that we cannot provide a 3-dimensional encoding vector for $Link(S, E)$ such that the nodes A, B, and C can simultaneously achieve their original max-flows.

Then, we consider whether all the receivers can receive data at the rate of their max-flows in the network of Case 3. When we encode the original data at source node, we select a 3-by-3 matrix as the generator matrix for Figure 7(b) since the maximal max-flow to receivers in Figure 7(b) is 3 and the number of outgoing links is 3. Recall that it is impossible to provide a 3-dimensional encoding vector such that the rate of receivers A, B, and C can be increased by 1, so if we provide a 4-dimesional encoding vector n_ie_i ($i = 4$), the receivers could achieve their individual maxi-flows, respectively. Then, the generator matrix G will be extended to a 4-by-4 matrix (e_4, e_1, e_2, e_3).

In the example, $(|Child(S)| - r)$ equals 1, and if $|Child(S)|$ is much greater than r, the multirate network codes can be gradually obtained with the same method. Therefore, we consider that the construction of multirate network codes for the multicast network as shown in Figure 3(c) can be transformed into the construction of multirate network codes for the five-layer network model.

4.6. Disadvantage of the Proposed Scheme. In general, the network code construction algorithms include centralized algorithms and decentralized algorithms. The main difference between these two algorithms is that the former requires overall network topology of the network, while the latter does not. In the last section, we consider that our scheme is not a decentralized one, because the source node in our model requires the knowledge of the network topology, such as all the routes to each receiver. In addition, because the re-encoding operation at intermediate nodes will influence decoding at the receivers in the multirate multicast network, the source node in our scheme requires a network topology from which the transfer matrix can be formulated. Therefore, our method cannot work in a decentralized manner. After the multirate network codes are calculated, a notification will be broadcasted to the nodes in the networks such that every intermediate node gets the information of the GEVs of incoming links and outgoing links, as well as the strategy of re-encoding.

In the literature [25], the authors provided a pushback algorithm to implement the decentralized network coding scheme. In their pushback algorithm, each receiver needs to initialize a message to allow its parent nodes to gather information of the network, and each intermediate node forwards the message to their parent nodes until all messages reach the source node. In our method, each receiver needs to broadcast a routing establishment packet (REP) in which its own ID number is included in the route table. For each intermediate node, if it has not received the packet, it will add its ID number into the route table and forward the packet to its parent nodes. At last, the source node will have the knowledge of the overall network topology, and then it will calculate the max-flow for each receiver. In a practical network, the routing establishment algorithm and code assignment algorithm only perform once before the data dissemination, so we consider the overhead affordable.

5. Algorithms

5.1. Routing Establishment Algorithm. We need to explain why we prefer to establish disjoint routes for each receiver

```
(1) // Step 1. Broadcast routes establishment packet (RDP)
(2) Each receiver broadcasts a RDP
(3) For each non-receiver receiving RDP
(4)     If node id in the routing node list
(5)         Give up the packet.
(6)     Else
(7)         Add its node id in the routing node list of RDP.
(8)     Endif
(9)     Continue broadcasting the received RDP to the network.
(10) Endfor
(11) // The source node calculates the max-flow for each receiver with Ford-Fulkerson method.
(12) For node i = 1 : t                    % t is the number of receivers.
(13)     flow(i) = GetMaxflow(i)
(14) Endfor
(15) Establish flow(i) disjoint routes for each receiver(i).
(16) r = MaxValue(flow(i))                 % Get the big value in flow(i), i ∈ [1, t]
(17) If there is a link braided with the links belonging to receiver with max-flow of r
(18)     Remove the link from the network;
(19) Endif
```

Algorithm 1: Routing establishment algorithm.

```
(1) // Step 1. Obtain the generator matrix.
(2) If n ≥ m                // the number of outgoing links is greater than m
(3)     Use the edge eliminating method to transform the network into five-layer network.
(4) End
(5)   Calculate the transfer matrix T from the matrix A.
(6)   G = T^{-1},            // G and T are r × r matrices.
(7) For i = 1 : 1 : r
(8)       Y_i = Σ_{i=1}^{r} (g_ki × p_i), k ∈ (1, ..., r)      // p_i(i ∈ [1, r]) are original packets.
(9) End
(10) If n ≥ m
(11)     For i = r + 1 : 1 : m
(12)         Add a node V_i and its associated links.
(13)         Assign a global encoding kernel to link(S, V_i) such that the receivers
(14)         connecting to this node V_i receive a new packet.
(15)     End
(16) End
(17) For i = 1 : 1 : m
(18)   The source node sends Y_i to the ith child node.
(19) End
```

Algorithm 2: Encoding scheme at the source node.

before the data dissemination. After deploying the sensor nodes in the environment, some links will be established as long as two nodes are in the transmission radius of each other. The max-flow for each receiver may be obtained with the classical Ford-Fulkerson method, and then some redundant links should be eliminated. Otherwise, more energy will be consumed, and these redundant transmissions will not help improving the performance of network. Most importantly, elimination of the redundant links will not reduce the transmission rate. Therefore, we consider it necessary to establish disjoint routes to the source node for each receiver.

Our scheme must be based on the routing establishment algorithm, and the disadvantage of this idea is that it disables the network to work in a decentralized manner. However, in static WSNs, the routing establishment algorithm only performs once, so we believe that the decision does much more good than harm, and the computation overhead of the algorithm is affordable.

5.2. Encoding Scheme. After calculating the max-flows for each receiver with Algorithm 1, the source node will obtain an adjacency matrix A in which the network topology is stored. Algorithm 2 shows the encoding operation at the source node.

Next, we address the strategies for the intermediate nodes. In our network model, there are two cases of intermediate

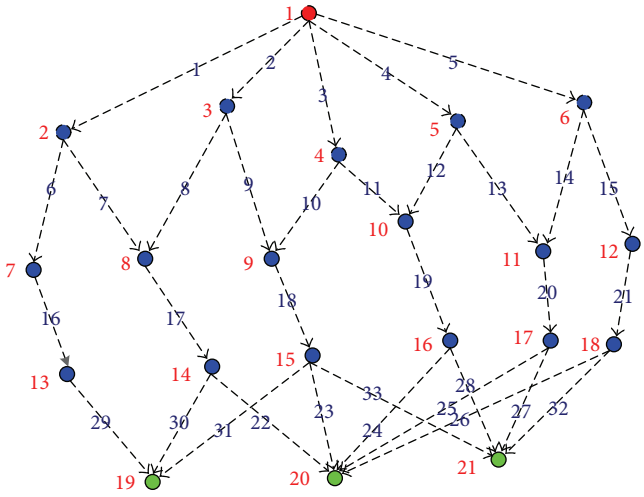


Figure 8: Simulated network topology.

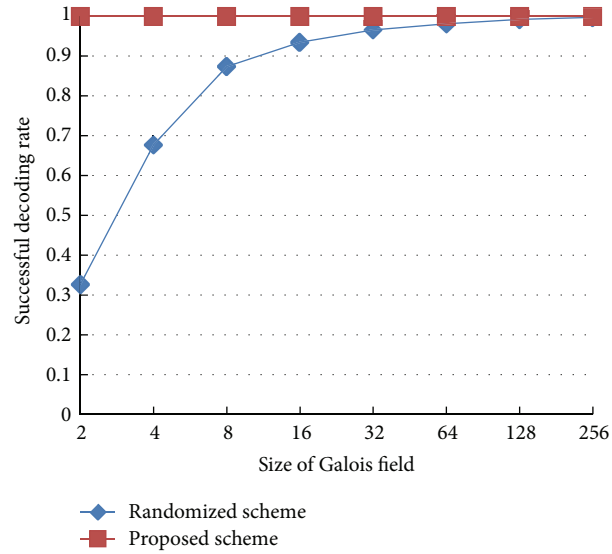


Figure 9: Decoding rates of randomized scheme and the proposed scheme.

nodes: firstly, in the case of $|Parent(V)| > 1$ and $|Child(V)| = 1$, the intermediate node V needs to perform network coding to mix input packets; secondly, in the case of $|Parent(V)| = 1$ and $|Child(V)| \geq 1$, node V needs to forward input packet to all output links without any change. In the network model, the case of $|Parent(V)| > 1$ and $|Child(V)| > 1$ will not appear since we only assign one link for each node in layer 3, which can be considered as the worst kind of bottleneck. Note that the re-encoding operation is only performed at the node whose outgoing link is shared by more than one route. The other intermediate nodes do not need to perform re-encoding. In this sense, the number of re-encoding nodes has been reduced.

For the receivers, they need to receive and decode. The work of decoding is to solve a system of linear equations, and the classical Gauss eliminating method can be employed.

6. Simulation and Analysis

In our simulations, the source node needs to transmit 30 M bits of data to the receivers, and the transmission capacity for each link is 1 M bits per second. The simulated network consists of a source node, three receivers (denoted by A, B, and C, resp.), and 17 intermediate nodes, as shown in Figure 8.

After performing the routing establishment algorithm, the numbers of disjoint routes to the receivers A (node 20), B (node 21), and C (node 19) are 5, 4, and 3, respectively. Therefore, receiver A will first finish receiving all the data. After receiver A has received all the data, the source node needs to send the remaining data to receivers B and C. At that moment, the data required by the two receivers might be different. Take Figure 1 for example; after receiver B has received (a, b, c), receiver A requires packet c, while receiver C requires packet a. If some links are shared by the receivers which have not received all the data, the source node must reencode the remaining packets so that the network could transmit at a higher rate.

Based on the simulated network topology, we evaluated the performance of our scheme in terms of decoding rate, throughput, and energy efficiency.

6.1. Decoding Rate. As mentioned above, the network codes can be obtained with either a randomized algorithm or a deterministic algorithm. For a multirate multicast network, there is also a certain amount of work based on the randomized algorithm [25, 26]. We conducted some experiments and obtained some results of the proposed scheme and the previous randomized algorithms.

In Figure 9, we observe that the proposed scheme can be performed over a very small finite field and all the receivers must be able to decode, since it ensures that the generator matrix and transfer matrix are full-rank matrices, while the randomized algorithm may require a relatively larger finite field as the probability of successful decoding increases with the increase of finite size.

In general, when computing over a Galois field, it is necessary to generate a 2-dimensional multiplication table and a 2-dimensional division table to accelerate the computation, and the size of the tables is proportional to the field size. When the size is large, it will require a long time to construct the two tables and a large space to store the two tables. Accordingly, the computation overhead and storage overhead over a small finite field are less than those over a large finite field.

According to previous studies, the randomized algorithm used to obtain the network codes has its distinct advantages. For example, it can work in a decentralized manner. However, we consider that, in multirate network coding, a deterministic network coding scheme might be a better choice. In single-rate network coding, all the receivers require the same number of independent packets, and the re-encoding operation does not affect the decoding at receivers as long as it equals the elementary transformation. However,

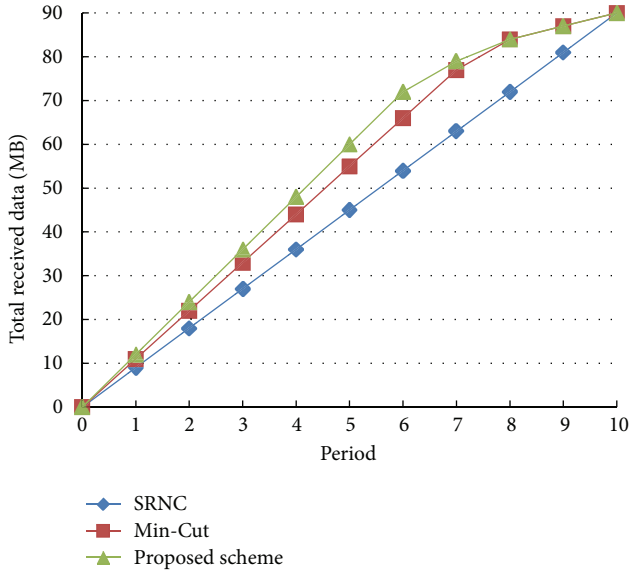


Figure 10: Total throughputs of different schemes.

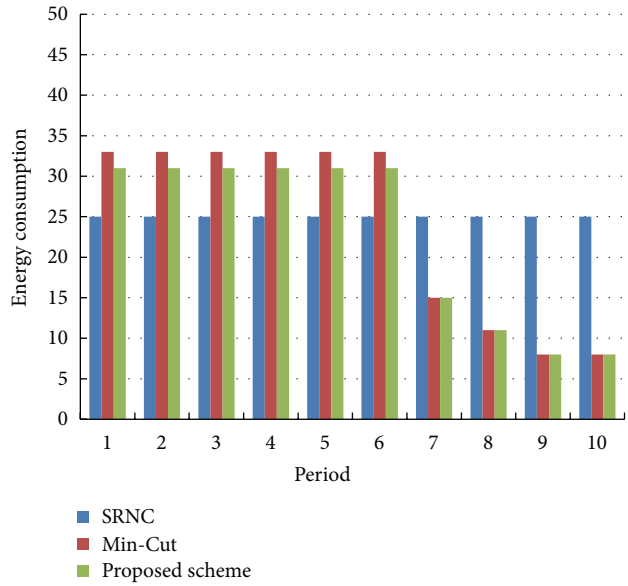


Figure 11: Energy consumptions of different schemes.

in the multirate network coding scheme based on random algorithm, the receivers may fail to decode even if it is ensured that the re-encoding operation is equivalent to the elementary transformations. If the re-encoding operation is performed on the same layer of each receiver, the decoding will not be affected; however, if the re-encoding operation is performed between different layers, the receiver may be unable to decode.

6.2. Improvement of Throughput. In Figure 10, we observe that the total transmission durations of these three schemes are the same due to the bottleneck caused by the receiver with the least max-flow. After applying the proposed scheme, the total transmission rate has been increased because the slope of the proposed scheme is greater than that of SRNC before the receiver with a high transmission rate has received all the data. After that, the source node starts to transmit the remaining data for other receivers, and the receivers which have received all the data do not receive data any more. Therefore, the slope of the proposed scheme becomes smaller than that of SRNC.

Although these methods take the same amount of time, the proposed scheme can significantly improve the performance, because in this scheme, the receivers with r disjoint routes could quickly finish receiving the original data. In this paper, there is only one source node. In a practical network, when a receiver receives all the data, it can be considered as a second source, and if this is the case, the performance of network can be further improved.

Therefore, in this experiment, the performance of Min-Cut is better than that of the traditional SRNC, and the proposed scheme has the best performance. We have to admit that Min-Cut can provide multiresolution multicast, while the proposed scheme cannot. The aim of this paper is to increase the throughput and energy efficiency of WSNs, so the proposed scheme is useful from this perspective.

6.3. Improvement of Energy Efficiency. We know that energy consumption is proportional to the total transmission times, and if a scheme can enable all the receivers to receive all the data with less transmission times, the scheme can be considered as energy-efficient. In the SRNC scheme, since the number of routes to each receiver equals the multicast capacity (h), the receivers with more than h routes to the source node must give up some routes so that the transmission energy will be saved or some redundant data will be received without any performance gain. On the other hand, after a receiver with high max-flows has received all the data, the routes to the receivers will be given up to avoid redundant transmission unless the routes are shared by other receivers. In this way, the comparison between the proposed scheme and the previous schemes is reasonable and fair.

In accordance with Figure 11, the energy consumption in SRNC does not change during all 10 transmission periods. For the proposed scheme, the energy consumption is greater than that of SRNC since more links are required to transmit. After the receiver with high max-flow has received all the data, some links are removed, and therefore the energy consumption is smaller than that of SRNC. Then, we accumulate all the energy consumption during each period, and the result shows that the energy consumption of the proposed scheme is less than that of SRNC. Specifically speaking, 9 percent of transmission overhead can be saved in the simulated instance. Moreover, the energy efficiency of Min-Cut is higher than that of SRNC but lower than that of the proposed scheme.

In our simulated instance, there are only three receivers. We consider that the proposed scheme could provide more performance gain in the network with more heterogeneous receivers.

7. Conclusions

This paper proposes a multirate network coding scheme to improve the energy efficiency of WSNs. The proposed scheme can improve the energy efficiency on three aspects. Firstly, it can reduce the number of re-encoding nodes without compromising the performance of network coding; secondly, the network can transmit more data in a transmission period, and therefore the energy efficiency can be enhanced; thirdly, the scheme can work over a very small finite field.

In the future, some research should be conducted to improve the performance of WSNs and the theory of multirate network coding.

(1) For a generalized network model, we failed to make the source node work in a decentralized manner. In the future work, efforts should be made to improve this problem.

(2) Our routing establishment method may degenerate the theoretical throughput in some networks; how to avoid this problem is worthy of further research.

Conflict of Interests

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

Acknowledgments

The work in this paper is supported by Assembly Research Fund (9140A05010113xxx), Key Technology R&D Program of Jiangsu, China (BE2012386 and BE2011342), the Agricultural Innovation Program of Jiangsu, China (CX(13)3054, CX(10)221, and CX(11)2042), Strategic Emerging Industry Development Program of Shenzhen, China (JCYJ20130331151710105), and the National Natural Science Foundation of China (no. 61301108).

References

[1] P. Berman, G. Calinescu, C. Shah, and A. Zelikovsky, "Efficient energy management in sensor networks," in *Proceedings of the Ad Hoc and Sensor Networks, Series on Wireless Networks and Mobile Computing*, Nova Science Publishers, New York, NY, USA.

[2] N. Li and J. C. Hou, "FLSS: a fault-tolerant topology control algorithm for wireless networks," in *Proceedings of the MobiCom 10th Annual International Conference on Mobile Computing and Networking*, pp. 275–286, New York, NY, USA, 2005.

[3] 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.

[4] 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.

[5] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[6] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[7] S. Y. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[8] S.-Y. R. Li, N. Cai, and R. W. Yeung, "On theory of linear network coding," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '05)*, pp. 273–277, Adelaide, Australia, September 2005.

[9] T. Ho, M. Médard, R. Koetter et al., "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.

[10] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proceedings of the 41st Annual Allerton Conference on Communication Control and Computing*, October 2003.

[11] H. Yomo and P. Popovski, "Opportunistic scheduling for wireless network coding," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 2766–2770, 2009.

[12] C. Fragouli, J. Widmer, and J. Y. le Boudec, "Efficient broadcasting using network coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 2, pp. 450–463, 2008.

[13] Y. W. Yang, C. S. Zhong, Y. M. Sun, and J. Y. Yang, "Network coding based reliable disjoint and braided multipath routing for sensor networks," *Journal of Network and Computer Applications*, vol. 33, no. 4, pp. 422–432, 2010.

[14] N. J. A. Harvey, D. R. Karger, and K. Murota, "Deterministic network coding by matrix completion," in *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '05)*, pp. 489–498, January 2005.

[15] P. Sanders, S. Egner, and L. Tolhuizen, "Polynomial time algorithms for network information flow," in *Proceedings of the 15th Annual ACM Symposium on Parallelism in Algorithms and Architectures*, pp. 286–294, San Diego, Calif, USA, June 2003.

[16] S. Jaggi, P. Sanders, P. A. Chou et al., "Polynomial time algorithms for multicast network code construction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, 2005.

[17] X. Li, S. Paul, and M. H. Ammar, "Multi-session rate control for layered video multicast," in *Proceedings of the Multimedia Computing and Networking*, pp. 175–189, January 1999.

[18] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," in *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 117–130, August 1996.

[19] S. Dumitrescu, M. Shao, and X. Wu, "Layered multicast with inter-layer network coding," in *Proceedings of the 28th IEEE conference on Computer Communications (INFOCOM '09)*, April 2009.

[20] S. Lakshminarayana and A. Eryilmaz, "Multirate multicasting with intra-layer network coding," *IEEE/ACM Transactions on Networking*, vol. 21, no. 4, pp. 1256–1269, 2013.

[21] N. Sundaram, P. Ramanathan, and S. Banerjee, "Multirate media streaming using network coding," in *Proceedings of the 43rd Annual Allerton Conference on Communication, Control, and Computing*, September 2005.

[22] C. Xu, Y. Xu, C. Zhan, R. Wu, and Q. Wang, "On network coding based multirate video streaming in directed networks," in *Proceedings of the 27th IEEE International Performance Computing and Communications Conference (IPCCC '07)*, pp. 332–339, New Orleans, La, USA, April 2007.

[23] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, 2003.

[24] X. Wu, B. Ma, and N. Sarshar, "Rainbow network flow of multiple description codes," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4565–4574, 2008.

[25] M. Kim, D. Lucani, X. Shi, F. Zhao, and M. Médard, "Network coding for multi-resolution multicast," in *Proceedings of the IEEE (INFOCOM '10)*, pp. 1–9, San Diego, Calif, USA, March 2010.

[26] J. Widmer, A. Capalbo, A. F. Anta et al., "Rate allocation for layered multicast streaming with inter-layer network coding," in *Proceedings of the IEEE (INFOCOM '12)*, pp. 2796–2800, 2012.

[27] L. Wang, Y. W. Yang, and W. Zhao, "Network coding-based multipath routing for energy efficiency in wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, article 115, 2012.

[28] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 497–510, 2008.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 159052, 8 pages
http://dx.doi.org/10.1155/2014/159052

Research Article

A Proactive Complex Event Processing Method for Large-Scale Transportation Internet of Things

Yongheng Wang and Kening Cao

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

Correspondence should be addressed to Yongheng Wang; yh.wang.cn@gmail.com

Received 25 December 2013; Revised 23 January 2014; Accepted 18 February 2014; Published 23 March 2014

Academic Editor: Gelan Yang

The Internet of Things (IoT) provides a new way to improve the transportation system. The key issue is how to process the numerous events generated by IoT. In this paper, a proactive complex event processing method is proposed for large-scale transportation IoT. Based on a multilayered adaptive dynamic Bayesian model, a Bayesian network structure learning algorithm using search-and-score is proposed to support accurate predictive analytics. A parallel Markov decision processes model is designed to support proactive event processing. State partitioning and mean field based approximation are used to support large-scale application. The experimental evaluations show that this method can support proactive complex event processing well in large-scale transportation Internet of Things.

1. Introduction

The Internet of Things (IoT) bridges the gap between the physical world and its representation within the digital world. In recent years, with the rapid development of information and communication technologies, bandwidth and storage are no longer restricted in IoT applications. The key issue is how to process the massive events produced by large-scale IoT applications (an event means an atomic occurrence of interest in time).

In IoT applications, event processing engines need to process events that arrive from various kinds of sources such as sensors, RFID readers, and GPS. The events generated by devices directly are called primitive events. The semantic information inside primitive events is quite limited. In real application, users pay more attention to high level information such as business logic and rules. For example, each reading operation of the RFID reader at a garage generates a primitive event, but complex events like "the car leaves the garage" are the kind of events that users are really concerned with. To get complex events, many primitive events need to be combined based on some rules. In IoT application systems, business logic is converted into complex events and business logic is processed based on complex events detection. Complex event processing (CEP) [1] is used to process massive primitive events and get valuable high level information from them. As an example, in logistics industry, CEP is used to track the goods and trigger some actions when exception is found.

In many real-time IoT applications, events are uncertain due to some factors such as measuring inaccuracy, signal disturbance, or privacy protection. Usually, we use probabilities to process such uncertainty. Therefore, it is necessary to develop probabilistic event processing engine.

The traditional type of the event processing is reactive processing which means that actions are triggered by events or by the system states. A proactive event processing system is able to mitigate or eliminate undesired future events or states or to identify and take advantage of future opportunities, by using prediction and automated decision making methods [2]. For example, in a transportation system, we can predict some possible congestion states and take some actions to mitigate or eliminate the congestion states. The proactive event processing systems use predictive analytics (PA) technology which predicts future events or system states through analysis of historical events. The system also uses iterated decision processes (DP) technology which analyzes system states and selects appropriate actions to achieve expected

states. The CEP, PA, and DP technologies have been studied widely, but there are few papers about how to integrate them together to support proactive event-driven system. Proactive event processing in large-scale transportation IoT needs to process massive historical events and analyze complex states iteratively which makes most of the existing algorithms unable to be used directly. Furthermore, proactive event processing systems need high performance since they usually have to take actions in time.

In this paper, we propose a proactive complex event processing (Pro-CEP) method for large-scale transportation Internet of Things. We designed a multilayered adaptive dynamic Bayesian network (mADBN) model for predictive analytics. Based on probabilistic complex event processing, our method uses concurrent actions Markov decision processes (MDP) to integrate CEP and PA.

2. Related Works

2.1. CEP and Probabilistic CEP. Complex event processing detects complex events based on a set/sequence of occurrences of primitive events by monitoring the event flow continuously. Etzion and Niblett introduced the concept, architecture, and methods of complex event processing in their book [3]. Event processing agent (EPA) is a component that applies some rules to generate complex events as output based on a set of input events. Event processing network (EPN) is a network of a collection of EPAs, event producers, and event consumers linked by channels. The network in EPN is used to describe the event processing flow execution. Luckham first introduced the event processing network in the field of modeling [1].

Most of the CEP methods in active databases and RFID applications use fixed data structures such as tree, directed graph or Petri-Net. Those methods cannot extend event query language according to the change of requirement or optimize event query language flexibly. SASE [4] is a high performance complex event processing method which uses nondeterministic finite automaton (NFA). Recently, there is some work about improving the SASE method.

Most probabilistic CEP engines are based on sequential variants of probabilistic graphical models, such as hidden Markov models, dynamic Bayesian Networks, and conditional random fields. Recently, some probabilistic CEP methods based on NFA are proposed. In the work of [5], a data structure called chain instance queues is used to detect complex events with a single scanning of probabilistic event stream. In another paper [6], an optimized method based on SASE is used to calculate the probability of complex events and to obtain the confidence value of the complex patterns given by user against uncertain raw input event stream. Compared with our work, these CEP methods are not integrated with PA and DP to support proactive application.

2.2. Predictive Analytics with Bayesian Networks. Predictive analytics applies some statistical and data mining techniques such as clustering, classification, and regression. In the case of predictive analytics methods for complex event data, some attributes of the monitored system are predicted based on the previously monitored events.

Recently, Bayesian network (BN) has been used in predictive analytics. Castillo et al. used BN to predict the random character of the total mean flow level and the variability of origin-destination pair flows [7]. In the work of Pascale and Nicoli, an adaptive BN was proposed in which the network topology changes according to the nonstationary features of traffic [8]. Gaussian mixture model (GMM) was used to describe the joint probability distribution between the cause nodes and the effect node in a Bayesian network. In the work of Hofleitner et al., dynamic Bayesian network (DBN) is used in PA and a model based on hydrodynamic traffic theory is also proposed to learn the distribution of vehicles on arterial road segments [9]. Compared with our work, these methods are not designed for large-scale IoT applications and cannot process massive event data in proactive event-based system.

2.3. Proactive Event Processing and Markov Decision Processes. Recently, many proactive applications have been developed, for example, proactive security systems, proactive routing in mobile wireless networks, and proactive service level agreement negotiation in service oriented systems. In the work of Engel et al., a proactive event-driven computing framework based on CEP, PA, and Markov decision processes (MDP) is proposed [2, 10]. In this framework, the event processing model is extended and two more types of agent are included: predictive agents which can predict future uncertain events based on the prediction models and proactive agents which can select the best proactive action that should be taken.

MDP has been studied for many years and some variants emerged recently. In the area of large-scale IoT, the main challenge is the combination explosion problem caused by large state space. Mainly, there are two research directions on this problem. The first one tries to simplify the problem by using the information from the application domain, including state aggregation methods, time aggregation methods, and action elimination methods. These methods are related to specific application domain, and it is not easy to find a common solution. In the second direction, approximate methods are developed, including approximate dynamic programming, dynamic programming based on events, and selective approximation. The key issue in these methods is how to approximate the value function during the optimization process. When applied to large-scale IoT, MDP has larger size and some new properties. The model design and large-scale calculation become a new challenge. Compared with our work, current MDP methods are not combined with CEP and cannot process the large state space in large-scale proactive event-based system.

3. Backgrounds

3.1. System Architecture. The system architecture of our work is shown in Figure 1. Raw events are generated by devices such as RFID reader, radar, and GPS. We assume the events can be imprecise because of the limitation of measuring accuracy or signal disturbance. The probabilistic raw events

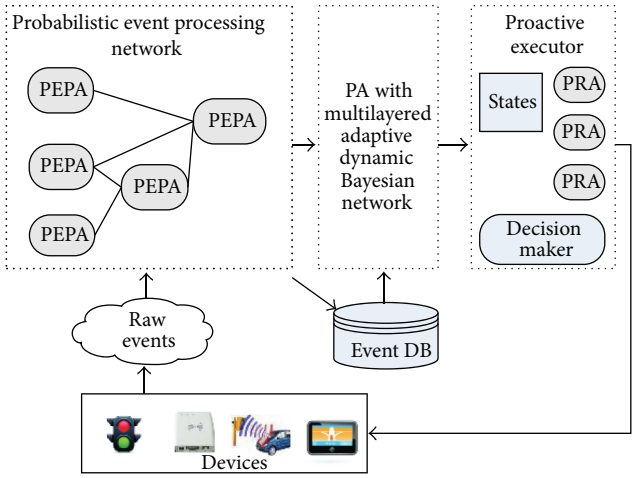


Figure 1: System architecture.

are processed by the probabilistic event processing network (PEPN) which is composed of many interconnected probabilistic event processing agents (PEPA). Different PEPAs can be connected since hierarchical complex event is supported. The PA component can predict the system states from the complex events based on the mADBN model. Complex events are saved to event database and the historical data can be used to train the models in PA components. The decision maker selects appropriate actions according to the predicted states and assigns some proactive agents (PRA) to execute the actions.

3.2. Event Model and Probabilistic CEP

Definition 1 (primitive event). A primitive event in a stream means an atomic occurrence of interest in time. A probabilistic primitive event is represented by $\langle A, T, \Pr\rangle$, where A is the set of attributes and T is the timestamp that the event occurs. Pr is the concrete probability value used to present the occurrence probability of the event.

The probability value represents the possibility that an event is converted accurately from truthful data of nature to digital data used for computing in electronic devices.

Definition 2 (complex event). A complex event is a combination of primitive events or complex events by some rule. A probabilistic complex event is represented by $\langle E, R, \mathrm{Ts}, \Pr\rangle$, where E is the set of elements that compose the complex event, R is the rule of the combination, Ts is the time span of the complex event, and Pr is the probability value.

The main complex event patterns in our work include ALL, ANY, COUNT, and SEQ. In this paper, the COUNT pattern can be used to count the number of vehicles in a specified area during specified time span. The SEQ pattern can be used to represent the moving path of a vehicle. The detailed meaning of the patterns can be found in [3]. Those patterns can be composed to create hierarchical complex patterns.

In a large-scale IoT application, we may need to support distributed event streams processing. In this paper, we assume the primitive events from different event streams are independent. In the same event stream, some primitive events have Markov property which means that the probability of next event is only related to the current event in the sequence but has nothing to do with previous events. Like the work of [6], we use condition probability table (CPT) to save and sort the condition probability.

We extended the SASE method to support probabilistic CEP by adding probability value for each event in the active instance stacks (AIS). Detailed method and algorithm can be found in another paper of the same author [11].

4. Proactive CEP Method

4.1. Predictive Analytics Model. The mADBN model is shown in Figure 2. In this model, there are a state plane and a set of location planes. Each plane is represented by an adaptive dynamic Bayesian network with two dimensions: time and space. In the state plane, nodes denote the states of the system observed at different time instants and/or spatial locations, edges being their probabilistic relations. In Figure 2, the directed edges mean that the state of (i, t) depends on a set of other states before time t. The term "dynamic" in mADBN means that we are modeling a dynamic system. The term "adaptive" means that the graph structure is created based on machine learning from historical data. Each location plane represents the location change of a vehicle (moving path). The directed edges represent the transaction probability from one location to another and the directed edges with solid line represent the real paths.

The graph structure of the state plane can be learned based on analysis of the vehicle location planes. Given a node set V and a constraint set C, the Bayesian network structure learning problem is to find an optimized set of edges and an optimized set of distribution parameters. Mainly, there are two types of methods for Bayesian network structure learning: constraint-based method and search-and-score method. Constraint-based method checks the conditional dependence in data systematically and creates the network structure based on the dependence. Search-and-score method creates reasonable network structure based on a score function. Search-and-score method gets better result but the performance is not good for massive data. In this paper, we combine the two methods together. Global CPT can be created using the historical data in object location planes. Candidates are selected based on global CPT using constraint-based method and then a search-and-score algorithm is used to learn the structure of Bayesian network. In our method, local CPT is first created for each object plane and then all local CPTs are summed up and normalized to get the global CPT. The method was implemented with a MapReduce algorithm to support parallel calculation.

The main idea of the structure learning algorithm using search-and-score is to maximize the score function. Like

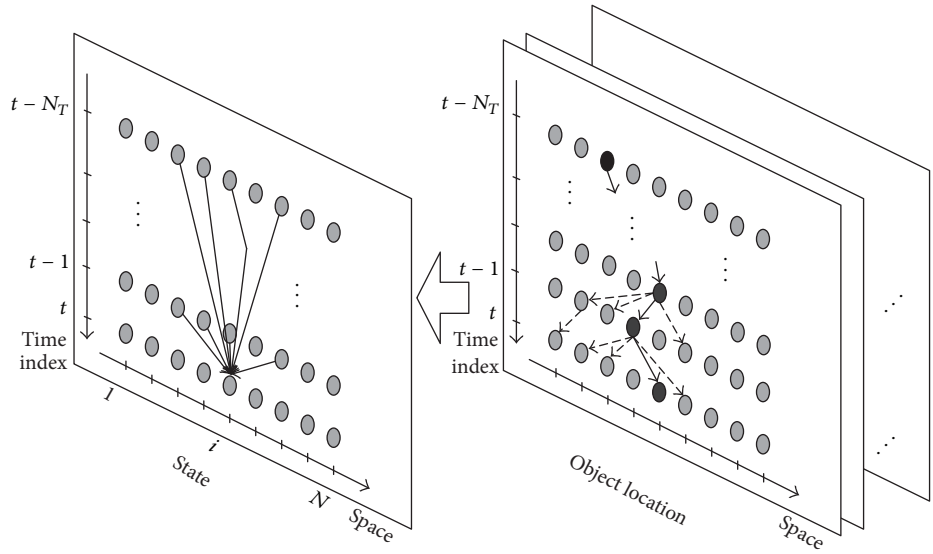


Figure 2: The mADBN model.

the work of [12], we use Bayesian information criterion (BIC) score function which is defined as follows:

$$\mathrm{SBIC}(D, G, \Theta) = \log P\left(D \mid \widehat{\Theta}, G\right) - \frac{d}{2}\log m + o(1), \quad (1)$$

where D is the data set, G is the graph structure, $\widehat{\Theta}$ is the maximum likelihood distribution parameters for D, d is the number of edges, and m is the sample size per vertex. The algorithm has an expansion stage and a contraction stage. During the expansion stage, edges that can increase the score function are added. During the contraction stage, edges are removed if that does not decrease the score function. The candidate parent vertexes for a specific vertex are selected according to the CPT.

4.2. Predictive Analytics Method. In the mADBN model of Figure 2, let $f_{i,t}$ represent the flow state of node (i,t) and let $\mathrm{pa}(i,t)$ represent all the parent nodes of (i,t). N_P is the number of nodes in $\mathrm{pa}(i,t)$. The set of flow states for $\mathrm{pa}(i,t)$ is $F_{\mathrm{pa}(i,t)} = \{f_{j,s} : (j,s) \in \mathrm{pa}(i,t)\}$. The joint distribution of all nodes in the state plane can be expressed as follows:

$$p(F) = \prod_{i,t} p\left(f_{i,t} \mid F_{\mathrm{pa}(i,t)}\right). \quad (2)$$

According to the Bayesian formula, the conditional probability $p(f_{i,t} \mid F_{\mathrm{pa}(i,t)})$ can be calculated by

$$p\left(f_{i,t} F_{\mathrm{pa}(i,t)}\right) = \frac{p\left(f_{i,t}, F_{\mathrm{pa}(i,t)}\right)}{F_{\mathrm{pa}(i,t)}}. \quad (3)$$

The joint distribution $p(f_{i,t}, F_{\mathrm{pa}(i,t)})$ is modeled with GMM as follows:

$$p\left(f_{i,t}, F_{\mathrm{pa}(i,t)}\right) = \sum_{m=1}^{M} \alpha_m g_m\left(f_{i,t}, F_{\mathrm{pa}(i,t)} \mid \mu_m, C_m\right), \quad (4)$$

where M is the number of nodes and $g_m(\cdot \mid \mu_m, C_m)$ is the mth Gaussian distribution which has a $(N_P + 1) \times 1$ vector of mean values μ_m and a $(N_P + 1) \times (N_P + 1)$ covariance matrix C_m. Like the work of [13, 14], we use the EM algorithm to infer the parameters $\{\alpha_m, \mu_m, C_m\}_{m=1}^{M}$ from the historical data. The conditional distribution $p(f_{i,t} \mid F_{\mathrm{pa}(i,t)})$ can be derived from $p(f_{i,t} \mid F_{\mathrm{pa}(i,t)})$ and the estimate $\widehat{f}_{i,t}$ can be calculated from $F_{\mathrm{pa}(i,t)}$ using the minimum mean square error (MMSE) method.

4.3. Decision Making with MDP. According to the properties of proactive complex event processing in IoT application, we extend the traditional MDP as follows.

Definition 3 (parallel MDP for proactive complex event processing). A parallel MDP for proactive complex event processing is represented as $\langle I, S, \vec{A}, P, R, E_c, C, T_c \rangle$, where I denotes the finite set of agents that process actions and S denotes the set of states with a special initial state S_0. $\vec{A} = \times_{i \in I} A_i$ denotes the set of actions, where A_i is the action from the ith agent. $P : S \times A \times S \to [0,1]$ denotes the set of state transformation functions, where $P(s, \alpha, s')$ represents the probability that state s transforms to s' with the execution of action α. $R : S \to \mathrm{Re}$ denotes the set of reward functions (Re means the set of real numbers). Every $(s, \alpha) \in S \times A$ satisfies $\sum_{s' \in S} p(s, \alpha, s') = 1$. E_s is the set of complex events in IoT which affects the system states. C is the context of event and agents act according to different context. $T_s : S \times E_s \times S \to [0,1]$ denotes the transforming probability which represents the affection of complex events to the system states.

The proactive event processing system starts from an original state and selects one or more actions executed by agents according to the policy. The execution of actions

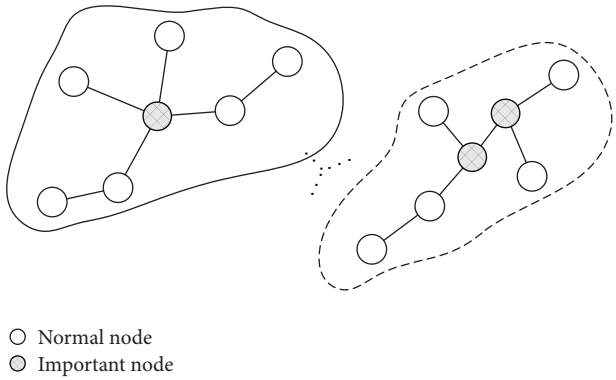


Figure 3: Network partition based on important nodes.

changes the primitive events and the transforming probability values from old states to new states are calculated based on the analysis of the complex events. The key issue of MDP is to find a policy $\pi : S \rightarrow A$ which reflects a state to a set of actions. The best policy maximizes the total rewards. The selection of policy is based on the following equation (value function):

$$v_\pi(s) = r(s, \pi(s)) + \gamma \sum_{s' \in S} p\left(s' \mid s, \pi(s)\right) v_\pi\left(s'\right). \quad (5)$$

The update of policy is based on the following equation:

$$\pi^*(s) = \arg\max_{a \in A}\left(r(s,a) + \gamma \sum_{s' \in S} P\left(s' \mid s, a\right) v_\pi\left(s'\right)\right), \quad (6)$$

where γ is the decay factor. In traditional MDP, actions are executed in order. Recently, some work of parallel MDP is proposed in which actions are partitioned into groups and group of actions executed parallel without affecting each other. In this paper, the MDP model executes actions parallel to the same state and all actions work together to transform the state into an expected one. Our MDP method is based on the following assumptions.

Assumption 4. The vehicles in large-scale IoT can be partitioned based on context and each group is related to a substate of the system. We need to consider the substates of the system but not the state of each vehicle.

An event context is a specification of conditions that groups event instances so that these instances can be processed in a related manner. For example, when processing the traffic congestion problem, the traffic flows at roads or junctions are substates and the system global states are composed of all substates.

Assumption 5. In large-scale transportation IoT, the state of a node depends on the states of its neighbors (the neighbors here mean the nodes that are linked with the current node within some distance).

Assumption 6. In large-scale transportation IoT, nodes can be partitioned into two classes: important and normal. The nodes whose states we want to control proactively are important nodes. The network can be partitioned based on important nodes and the state of important nodes can be changed by changing the state of their neighbors as shown in Figure 3.

According to Assumptions 4, 5, and 6, a new variable $G = (V, E)$ which represents the structure of subgraphs is added to the parallel MDP model, where V is the set of vertexes and E is the set of edges. An edge (i, j) exists which means that the state of node i affects the state of node j.

Definition 7 (neighbor state nodes). For any state node $i \in V$, the neighbor state nodes are $N(i) = \{j \in V \mid d_{i,j} \leq k\}$, where $d_{i,j}$ is the distance of nodes i and j and k is a threshold value.

Definition 8 (state transformation decomposition). Assume there are n substate nodes and n corresponding actions; according to the assumptions, we get

$$p\left(s' \mid s, a\right) = \prod_{i=1}^{n} P_i\left(s_i' \mid s_{N(i)}, a_i\right), \quad (7)$$

where $s_{N(i)} = \{s_j \mid j \in N(i)\}$.

Definition 9 (reward decomposition). Assume there are n substate nodes and n corresponding actions; the total reward can be written as follows:

$$r(s,a) = \sum_{i=1}^{n} r_i\left(s_{N(i)}, a_i\right). \quad (8)$$

A policy π can also be decomposed as $(\pi_1, \pi_2, \ldots, \pi_n)$.

Definition 10 (occupation measure). Let $\mathbf{S}$ be the state space and let $x^0 \in \mathbf{S}$ be the original state of MDP. The occupation measure $P_{x^0,\pi,\gamma} : \mathbf{S} \rightarrow [0,1]$ is defined as follows:

$$P_{x^0,\pi,\gamma}(y) = (1-\gamma)\sum_{t=0}^{\infty} \gamma^t P_\pi\left(S^t = y \mid S^0 = x^0\right), \quad (9)$$

where $y \in \mathbf{S}$. With occupation measure, the value function can be rewritten as follows:

$$V_\pi\left(x^0\right) = \frac{1}{1-\gamma} \cdot \sum_{y \in \mathbf{S}} P_{x^0,\pi,\gamma}(y) \cdot r(y, \pi(y)). \quad (10)$$

The key issue is to find a solution to calculate $P_\pi(S^t = y \mid S^0 = x^0)$ approximately. According to the decomposition we have defined, it can be inferred from $p_i(x_i^t \mid x_{N(i)}^{t-1}, \pi(x_{N(i)}^{t-1}))$. Now, the problem is to find an approximation $\widehat{G}_\pi^{t,i}(x_i^t \mid x_i^{t-1})$ which satisfies

$$\begin{aligned}\widehat{C}_\pi^t\left(y \mid x^0\right) &= \sum_{s^{t-1}} \widehat{G}_\pi^t\left(y \mid x^{t-1}\right) \cdot \widehat{C}_\pi^{t-1}\left(x^{t-1} \mid x^0\right) \\ &= \prod_{i=1}^{n} \sum_{x_i^{t-1}} \widehat{G}_\pi^{t,i}\left(y_i \mid x_i^{t-1}\right) \cdot \widehat{C}_\pi^{t-1,i}\left(x_i^{t-1} \mid x_i^0\right),\end{aligned} \quad (11)$$

where $\widehat{C}$ is the approximation of the conditional probability $P_\pi(S^t = y \mid S^0 = x^0)$ defined by $\widehat{C}_\pi^1(y \mid x^0) = \prod_i \widehat{G}_\pi^{1,i}(y_i \mid x_i^0)$ and $\forall t, \widehat{C}_\pi^t(y \mid x^0) = \prod_{i=1}^n \widehat{C}_\pi^{t,i}(y_i \mid x_i^0)$.

As an effective method for analyzing large and complex stochastic models, mean field theory is widely used in many areas. Using mean field theory, the effect of all other individuals on any given individual is approximated by a single averaged effect. Like the work of Sabbadin et al. [15], we applied mean field approximation which approximates a multidimensional distribution $P(u_1, u_2, \ldots, u_m)$ as the joint distribution G of m independent variables $u_1, \ldots, u_m$. G should be as close to P as possible according to Kullback-Leibler divergence:

$$\mathrm{KL}(GP) = E_Q\left[\log\left(\frac{G}{P}\right)\right]. \tag{12}$$

For state 1 (1 is the first time point and 0 is the start point), we get the following approximation using Kullback-Leibler divergence:

$$\widehat{G}_\pi^1 = \arg\min_{P^1\pi} \mathrm{KL}\left(G_\pi^1\left(S^1 \mid S^0\right) \cdot P^0\left(S^0\right) \mid P_\pi\left(S^1 S^0\right) \cdot P^0\left(S^0\right)\right). \tag{13}$$

After state transformation decomposition, the approximation can be written as

$$\widehat{G}_\pi^{1,i}\left(x_i^1 \mid x_i^0\right) \propto \exp E_{p^0}\left[\log p_i\left(x_i^1 \mid x_i^0, S^0_{N(i)\setminus\{i\}}, \pi_i\left(x_i^0, S^0_{N(i)\setminus\{i\}}\right)\right)\right]. \tag{14}$$

The rest of the iteration is similar to (14). The data processing servers include a master node and N slave nodes. The main iteration is executed at the master node and the calculation of substates is partitioned into slave nodes and executes parallel.

5. Experimental Evaluations

We developed a transportation IoT simulation system based on the road traffic simulation package SUMO [16]. The system supports mobility trace of vehicles by using an OpenStreetMap [17] road map of Beijing. On the roads, we set many virtual induction loops which can detect vehicles that pass corresponding areas. The virtual RFID or GPS readers are implemented using the TraCI interface of SUMO to get the induction loop variables. Each virtual induction loop covers a region. We selected 114 junctions from the map and set 160 thousand vehicles in the system. In order to simulate real traffic system, a series of rules are defined. Each vehicle has a home location and an office location. A vehicle v_i runs between home and office with probability p_i. The vehicles also go to other places such as supermarket and hospital, with corresponding probabilities. Based on this simulation system, we evaluated the precision and performance of Pro-CEP. We used 5 Lenovo ThinkServer RD series servers with 4 GB memory as a cluster and the operating system is Ubuntu 12.

Table 1: Deviation of two methods. The average percent is calculated by ((average deviation)/(average observed vale)) ∗ 100%.

	Deviation	
	ABN	Pro-CEP
Max	286	122
Min	16	12
Average	126.33	71.5
Average percent	16.44%	9.3%

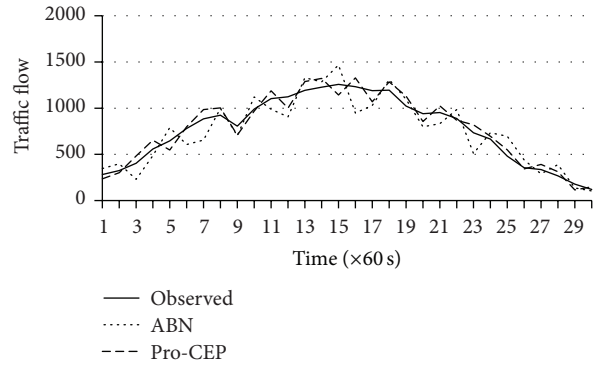


Figure 4: PA accuracy for a typical node.

We first run the simulation for many times to get the historical data of the vehicle paths. In the first experiment, the accuracy of our PA method is compared with ABN [8] which uses Bayesian network but has no vehicle location planes like our model. The result is shown in Figure 4 and Table 1. From the figure, we can see that the accuracy of Pro-CEP is better than traditional method ABN. The reason is that Pro-CEP use the object location plane to create a more precise Bayesian network structure.

In the next experiment, we compared the mean congestion of the system with and without Pro-CEP. Since we have not found other methods that have the same function as Pro-CEP, no other method is compared. The congestion is partitioned into 10 levels, where "0" means no congestion and "9" means the highest congestion. The result is shown in Figure 5. We can see the mean congestion level reduced obviously when Pro-CEP is used. The reason is that Pro-CEP can predict the congestion and take some actions to avoid it. The reduction of congestion level is more obvious when the congestion level becomes high because the system can redirect more vehicles to light loaded nodes in such circumstance.

In the next experiment, we evaluated the performance of Pro-CEP with different server number and data size. The important node number is fixed at 30 and Bayesian model constructing time is not included. The result is shown in Figure 6. As we can see, the running time increases when the vehicle numbers become larger. The reason is that more vehicles need to be rerouted which increases the calculation of the algorithm. The performance for 5 servers is higher than

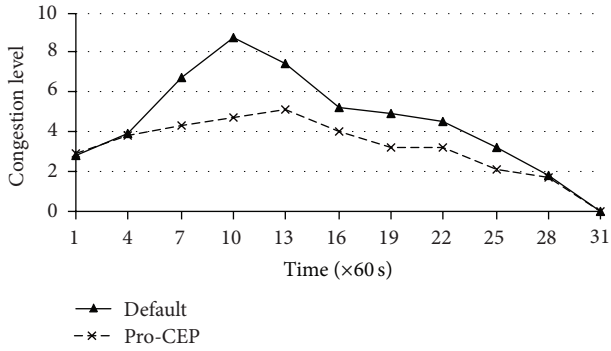


Figure 5: Mean congestion level over time.

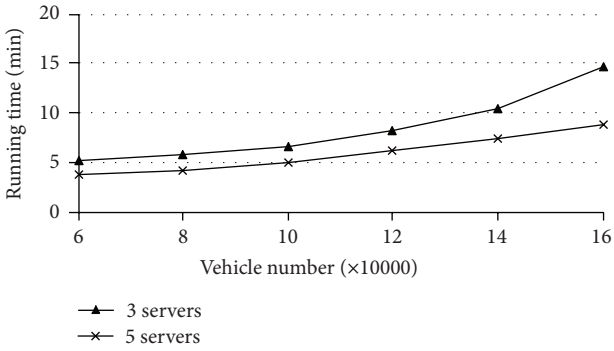


Figure 6: Performance of Pro-CEP with different server number and data size.

3 servers and the running time of the former increases slowly because the parallel method can take advantage of multiple servers.

We also evaluated the performance of Pro-CEP with different server number and important node number. The result is shown in Figure 7. The vehicle number is fixed at 160 thousand. The result shows that the running time increases obviously when the important node number becomes larger. The reason is that more important nodes means more substates in the parallel MDP which increases the calculation of the algorithm.

From all the experiments, we can see that Pro-CEP can support proactive complex event processing in large-scale transportation Internet of Things. The PA method gets high accuracy based on the mADBN model and Bayesian network. The parallel MDP model has obvious effect for congestion control in large transportation IoT. The performance of Pro-CEP decreases when the vehicle number and important node number become large, but we can get better performance with more servers.

6. Discussions and Conclusion

In this paper, we proposed a proactive complex event processing method for large-scale transportation Internet of Things based on a novel mADBN model and parallel MDP. A structure learning algorithm using search-and-score is proposed

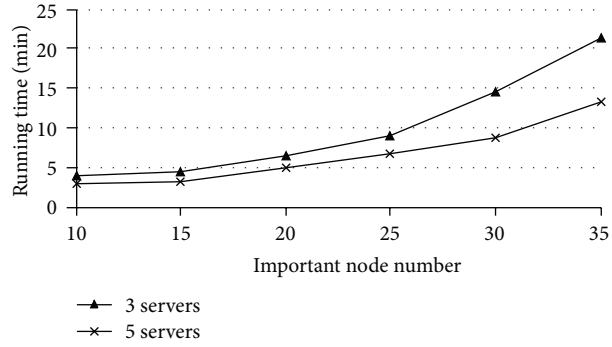


Figure 7: Performance of Pro-CEP with different server number and important node number.

to support accurate PA. A GMM based approximation for Bayesian network, a state partition method, and a mean field based approximation method of parallel MDP are proposed to support large-scale transportation IoT. The experimental evaluations show that this method can support proactive complex event processing well in large-scale transportation Internet of Things.

The performance and scalability of Pro-CEP still need to be improved. In the PA method, the EM algorithm needs to be parallelized to improve scalability. The master node in parallel MDP is the bottleneck and the iteration algorithm still needs to be parallelized further.

Conflict of Interests

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 61371116) and the Hunan Provincial Natural Science Foundation (no. 13JJ3046).

References

[1] D. C. Luckham, *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*, Addison Wesley, Boston, Mass, USA, 2002.

[2] Y. Engel and O. Etzion, "Towards proactive event-driven computing," in *Proceedings of the 5th ACM International Conference on Distributed Event-Based Systems (DEBS '11)*, pp. 125–136, New York, NY, USA, July 2011.

[3] O. Etzion and P. Niblett, *Event Processing in Action*, Manning Publications, Greenwich, Conn, USA, 2010.

[4] E. Wu, Y. Diao, and S. Rizvi, "High-performance complex event processing over streams," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 407–418, Chicago, Il, USA, June 2006.

[5] X. Chuanfei, L. Shukuan, W. Lei, and Q. Jianzhong, "Complex event detection in probabilistic stream," in *Proceedings of the 12th International Asia Pacific Web Conference (APWeb '10)*, pp. 361–363, April 2010.

[6] H. Kawashima, H. Kitagawa, and X. Li, "Complex event processing over uncertain data streams," in *Proceedings of the 5th international conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 521–526, 2010.

[7] E. Castillo, J. M. Menéndez, and S. Sánchez-Cambronero, "Predicting traffic flow using Bayesian networks," *Transportation Research B*, vol. 42, no. 5, pp. 482–509, 2008.

[8] A. Pascale and M. Nicoli, "Adaptive Bayesian network for traffic flow prediction," in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP '11)*, pp. 177–180, June 2011.

[9] A. Hofleitner, R. Herring, and P. Abbeel, "Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network," *IEEE Intelligent Transportation Systems Society*, vol. 13, no. 4, pp. 1679–1693, 2012.

[10] Y. Engel, O. Etzion, and Z. Feldman, "A basic model for proactive event-driven computing," in *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems*, pp. 107–118, Busan, Republic of Korea, 2012.

[11] Y. H. Wang and X. M. Zhang, "Complex event processing over distributed probabilistic event streams," in *Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '12)*, pp. 1489–1493, 2012.

[12] S. Samaranayake, S. Blandin, and A. M. Bayen, "Learning the dependency structure of highway networks for traffic forecast," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 5983–5988, 2011.

[13] J. J. Verbeek, N. Vlassis, and B. Kröse, "Efficient greedy learning of gaussian mixture models," *Neural Computation*, vol. 15, no. 2, pp. 469–485, 2003.

[14] N. Kumar, S. Satoor, and I. Buck, "Fast parallel expectation maximization for gaussian mixture models on GPUs using CUDA," in *Proceedings of the 11th IEEE International Conference on High Performance Computing and Communications (HPCC '09)*, pp. 103–109, June 2009.

[15] R. Sabbadin, N. Peyrard, and N. Forsell, "A framework and a mean-field algorithm for the local control of spatial processes," *International Journal of Approximate Reasoning*, vol. 53, no. 1, pp. 66–86, 2012.

[16] M. Behrisch, L. Bieker, and J. Erdmann, "Sumo—simulation of urban mobility: an overview," in *Proceedings of the 3rd International Conference on Advances in System Simulation*, pp. 63–68, Barcelona, Spain, October 2011.

[17] M. Haklay and P. Weber, "Openstreet map: user-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 760248, 11 pages
http://dx.doi.org/10.1155/2014/760248

Research Article

Fractal Cross-Layer Service with Integration and Interaction in Internet of Things

Bing Jia,[1] Shuai Liu,[1,2] and Yongjian Yang[3]

[1] *College of Computer Science, Inner Mongolia University, Hohhot 010021, China*
[2] *School of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China*
[3] *College of Computer Science and Technology, Jilin University, Jilin 010021, China*

Correspondence should be addressed to Shuai Liu; cs_liushuai@imu.edu.cn

Received 4 December 2013; Revised 10 January 2014; Accepted 20 January 2014; Published 3 March 2014

Academic Editor: Gelan Yang

Under the Internet of Things (IoT), in order to achieve the adaptive and personalized delivery of services, there was an urgent need to build an intelligent cross-layer service platform to support the services with effective integration and interaction. First, this paper focuses on the huge amounts, dynamic, and heterogeneous nature of services in IoT and presents the architecture of cross-layer services platform based on semantics to extend the service source from the common services in registries to the self-organization services in ubiquitous environment, which mainly contains the service management, demand computing, service discovery, and service selection. Then, some of the key foundational models and algorithms are presented, such as IoT service ontology model, semantic-based IoT service description language OWL-S^{iot}, and service management strategy both in registry mode and self-organization by subcluster. Many large-scale, ubiquitous, and heterogeneous services could be integrated in this platform. It could provide multilevel semantic supports for services to meet the common understanding and interoperability.

1. Introduction

Ubiquitous connection, mature service platform, and the scientific service model based on the service logic which can take intelligent action are the most important aspects of services in Internet of Things (IoT) in the future [1]. IoT can achieve ubiquitous connection; so how to use the technology to build a cross-layer platform about the service in Internet of Things (IoT service) which can support the service with effective integration and interoperability is a very important task.

As a sensing-based technology, IoT is a service-oriented network system blending various applications [2]. And its core value is "smart services" [3]. IoT services not only include common services in traditional internet but also include the ubiquitous services in the new network environment (mobile networks, wireless sensor networks, etc.) [4]. IoT services are massive and heterogeneous. They are mostly facing morphological changes, outward expansion, environmental change, business restructuring, and other situational dynamic adaptabilities. IoT services also face different levels of sharing and interoperability issues. These problems also bring challenges to IoT service platform. In order to solve these problems, we are attempting to use Web Services, SOA, SOC, semantics, and so on to achieve the services with effective integration, sharing, discovery, and interoperability. Then we can achieve an efficient service discovery and personalized delivery [5]. We present a semantic-based and cross-layer service platform for IoT service in this paper, including its architecture and key algorithm descriptions, to form a more complete service system.

2. Related Work

Service platform is logic components for supporting services; IoT services need to be provided to appropriate object in the right situations (right time, right place, right environment, etc.) by the right way. Therefore, IoT service platform is bound to be adaptive, using depth and personalized approaches to enhance the experience to deliver services [1].

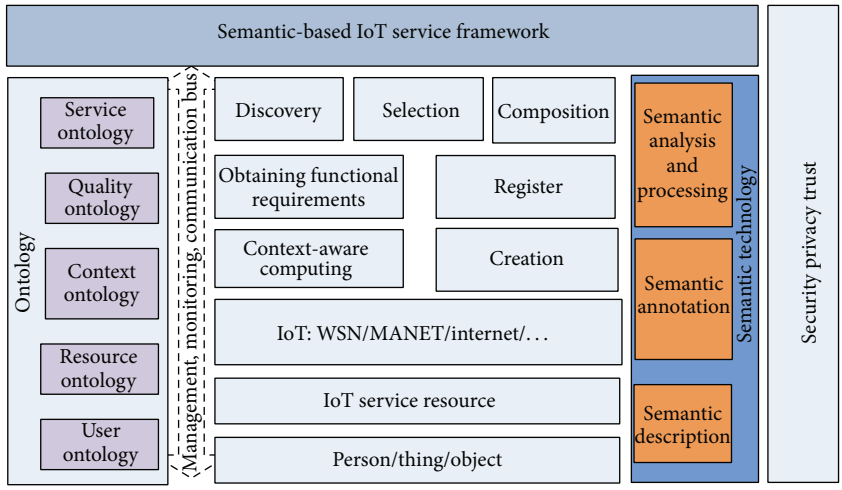


Figure 1: Framework of IoT service based on semantics.

Now the researches about IoT service platform are still in the theory exploratory stage, because the researches of IoT are still in the primary stage. For example, [6] proposed a functional framework of IoT service agent based on enhancing semantics. It uses the adapter agent to handle different types of smart objects' interrelated issues. Li gives the service framework structure based on IoT middleware model, which is future-oriented network, to achieve the transfer of IoT data transfer and the access of services [7]. EU's Seventh Framework Programmer (EU FP7) has organized IoT.est project team dedicated to do researches about the creating and testing of IoT service environment. In its report "Reference Architecture IoT Service Creation and Provision", showed the common reference architecture 5 about IoT service's creation, testing, analysis, and design [8]. They have extended the existing IoT-A architecture by adding a semantic service layer and proposed use-cases that can give an abstract treatment method of heterogeneous objects in IoT [9]. Reference [10] gives semantic-based sensor network services supporting framework, to meet service discovery, composition, compensation, and adaptive functions. Reference [11] proposed a solution about interoperability between devices.

To analyze the service platform from an empirical view, we studied some service platform in mobile communication environment. We mainly selected some more representative platforms such as IMS (IP Multimedia Subsystem), CAMEL (Customized Application for Mobile Enhanced Logic), OSA (Open Service Architecture), WAP (Wireless Application Protocol), and i-mode service platform which represent interaction, information, and integration for Internet [1]. These mobile service platforms are trying to break the traditional mode from end to end, and the service models which are vertical closed two parties (providers/users) are trying to allow the third parties to deploy services and develop application on its platform. Service platform should not be implicitly included within the system anymore. Instead, it should be evolved into a separate part of the display of existing components [12]. The framework and architectures of these platforms have good references for the design of IoT service platform [13].

3. Architecture of IoT Service Platform

3.1. Semantic-Based IoT Service Framework. In this paper, after making a comprehensive analysis on Web service framework and semantic sensor network framework [10] proposed by the EU FP7 [14], we present a semantic-based IoT service framework, as shown in Figure 1.

This framework consists of two parts. One is the service-related concept, both IoT ontology and semantic description, and the other is the corresponding operation of the service. It includes person, thing, and object in the bottom of the framework. And the user ontology is used to map them. IoT service resource needs to be described based on resource ontology. Obtaining functional requirements combines service ontology and quality ontology to achieve semantic description of the user's functional requirements. We can use context ontology to make semantic annotation on context characteristics obtained by context-aware computing. Semantic-based characteristics of the user can help to meet user's adaptive and individual needs of enhanced service. Service ontology is used for service creation and publishing, and it is throughout the whole platform.

This framework includes service discovery, selection, combination and so on, of which the most important are selection and combination. Service composition process can be achieved by the decomposition of requirements. Service discovery is searching and matching process on original advertising services set based on the service requester's service functional requirement. Thereby, it could return the set of service that can meet the functional requirements. The service resource may be organized and managed by registry, or may be decentralized. Service selection is a process to select the personalized service based on the service requester's context.

In particular, it needs to choose service discovery and selection strategy to adapt in different case. Whatever strategy, service ontology, quality ontology, context ontology, and others have played a semantic supporting role. Throughout the outermost of the framework are security, privacy, and trust, which is the basic guarantee for the implementation of the process.

3.2. Architecture of Cross-Layer IoT Service Platform. The Chinese Sensor Network National Standard Working Group has released a report named "Advances in the technology architecture and standards system about Internet of Things." It gives a reference standard of IoT architecture which is mainly divided into three layers, namely, the sensor layer, network layer, and application layer [15]. The sensor layer is the basis for the development and application of IoT. The network layer links the sensor layer and the application layer, including the bearer network and intelligent computing technologies. The application layer can show the business to users. One is a middleware of the business, and the other is a specific application.

It can be seen from IoT's reference architecture that we need a layer of coordinating and handling between the network layer and application layer to achieve the operation model on the data after mixing and cleansing, in order to provide users with more fluid and intelligent services. Then it can better support the industry and the public's application. In addition, since the typical characteristic of IoT is extended to ubiquitous environment like wireless network and ubiquitous sensor network, the service platform should extend the source of IoT service from the traditional common service to the ubiquitous service. In order to meet the extensibility of IoT service, we present the architecture of semantic-based cross-layer IoT service platform according to the standard of IoT architecture and the semantic-based IoT service framework presented in upper section, as shown in Figure 2. It has a special design for open issues-related integrated service produced by the complexity of network.

The platform adopts four-layer structure, adding a data integration and service support layer on the basis of the original IoT architecture. And services are throughout the whole platform. It can meet the requirements of the Times, such as sensor data as a service, network infrastructure as a service, computing platform as a service, and integrated application as a service. Different layers of service have a certain correlation. In general, upper layer service does the job by calling lower layer service. Lower service is the basis of top service. The main functional components of the platform are the various functional modules and associations with the operation between them in the data integration and service support layer. Eventually they can provide support for applications of the "application layer" in specific areas. Platform can also access directly the P2P network formed by the service node in the ubiquitous environment to carry out service discovery in ubiquitous service set.

Sensing layer is mainly used to achieve the real world sensing. The sensing data is also a service. In this layer, we complete collecting a variety of basic data, including user data and environmental data which could be perceived.

The network layer mainly obtains information of network characteristics related to services. Network communication resource performs as a service. This layer also achieves an adaptive mechanism that should transmit the data stably to upper layer and support the transmission of the device control data to the lower layer.

Data integration and service support layer are the core parts of IoT service platform, including service management, demand computing, service discovery, and service selection. Various functional modules and interoperability between them can provide the necessary support for the integration and consolidation in specific different application in application layer. The following describes each module separately.

(i) Service management module: it is the most basic core functionality module in service platform architecture. It is responsible for the standardization of service description and the management of service, respectively, according to the service source's different organizational forms. For service which can be concentrated and registered easily and suitably, we take the management of service registration center model to manage the service's register, update, delete, and so on. But ubiquitous services which has been distributed in the resource pools could not be concentrated and registered easily (e.g., environment-based service, mobile location-based service and highly time-sensitive service, etc.). We can use dynamic autonomous strategies to achieve decentralized management according to the loosely coupled P2P network between them. In particular, for multiple service registration centers which cannot be integrated in a catalogue, each of them could be regarded as a service node. So we can form a loosely coupled P2P service network between them and manage them by a decentralized strategy. In general, decentralized mode is a supplement to registration center model.

(ii) Demand computing module: it achieves interaction with the user and obtains the user's exact needs. One is functional computing, and the other is context-aware computing. Functional requirements are the core primary characteristic of requirements, mainly to know what the specific functions, types, and qualities of service the user wants. Context-aware computing is used to understand the current scene which is the proper way. These are user's enhanced experience needs for services. Context-aware computing is primarily responsible for obtaining the relevant contextual information, making inference and integration to obtain interpretable high-level context (e.g., whether in a meeting, in the hallway, or in the coffee shop, etc.). It could make semantic annotation accordance with the relevant ontology. The demand explanation will be filled into a template to automatically generate a service request description document.

(iii) Service discovery module: it completes the searching and matching of the service. It uses different search and matching algorithms according to the service's management style. It uses a centralized service discovery algorithm for registration center model, focusing on the optimization of semantic matching service. And it uses the P2P service discovery strategy for dispersion mode, because the ubiquitous

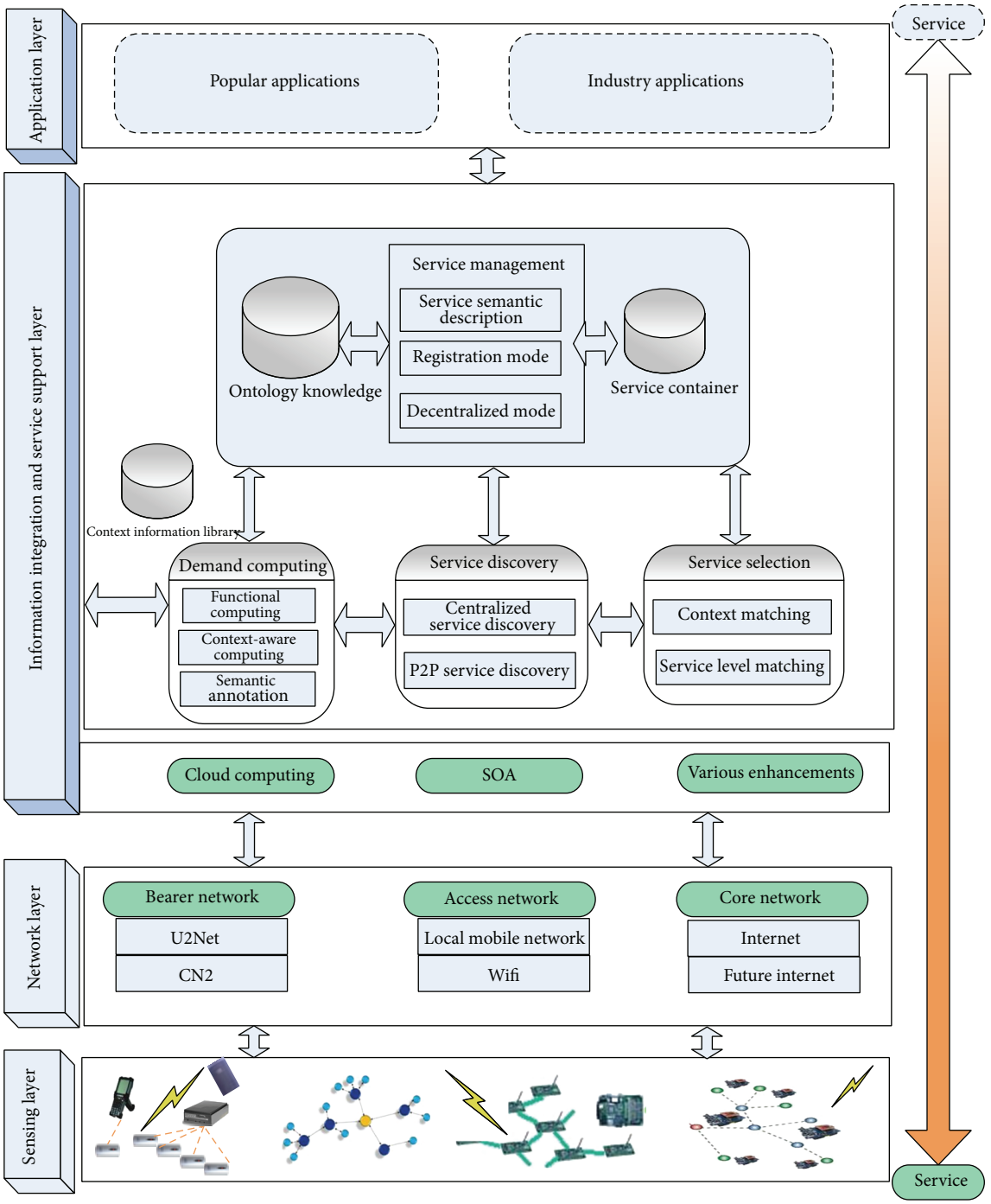


Figure 2: Architecture of IoT service platform.

services could often be formed in a loose P2P network, focusing on the optimization of service search strategy.

(iv) Service selection modules: they select the most adaptive and personalized services from the advertising services, which are obtained from service discovery meeting the functional needs to achieve the personalized pushing on demand. They contain context matching and service level matching.

(v) Service container: it can load services. It is a semantic-based UDDI repository mainly for the registration center model that contains all registered service documentation. The platform can just divide and set the user's region of interest because it cannot predefine the set of services to be discovered for the decentralized mode.

(vi) Ontology knowledge: it is used to store ontology and concepts related to IoT service platform and the relationship

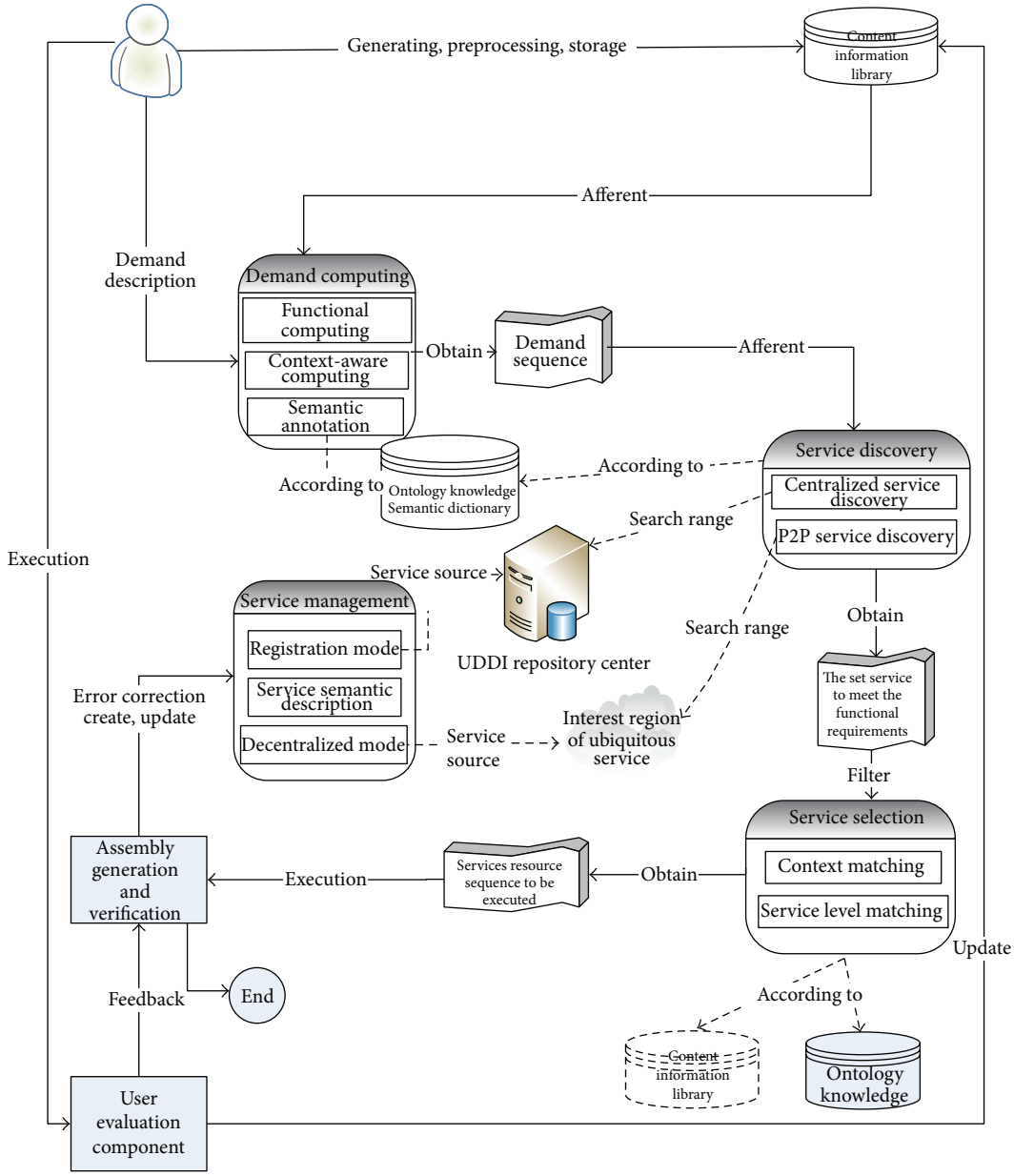


Figure 3: Main flow of service support platform.

between the concepts. It is the core foundation of platform to support semantics.

(vii) Context information library: it is used to store both the original context information and the high-level context after fusion.

Application layer is to show the specific business service in particular areas, including two aspects of the application, both popular applications and industry applications.

3.3. The Main Flow of Each Module of Service Support Layer's Coordination. Figure 3 shows the process of platform processing user's requests. Here the user is the "object" concept instead of the traditional person, which can be any entity in real world.

The key steps in the process are described below.

(i) We can get the user's functional and contextual needs through the interactive between user and demand computing module. Alternatively, we can carry out data mining in the original context information obtained directly through the context-aware computing module to get the implicit functional requirements and context.

(ii) Select a service request description document in turn. First the system should predict in which services source collection could find the suitable service. If it is in a UDDI Services library, it will search and match through a centralized service discovery algorithm. Then it will find a valid service that can meet the functional requirements and constraints. If the searching is in an interest region of

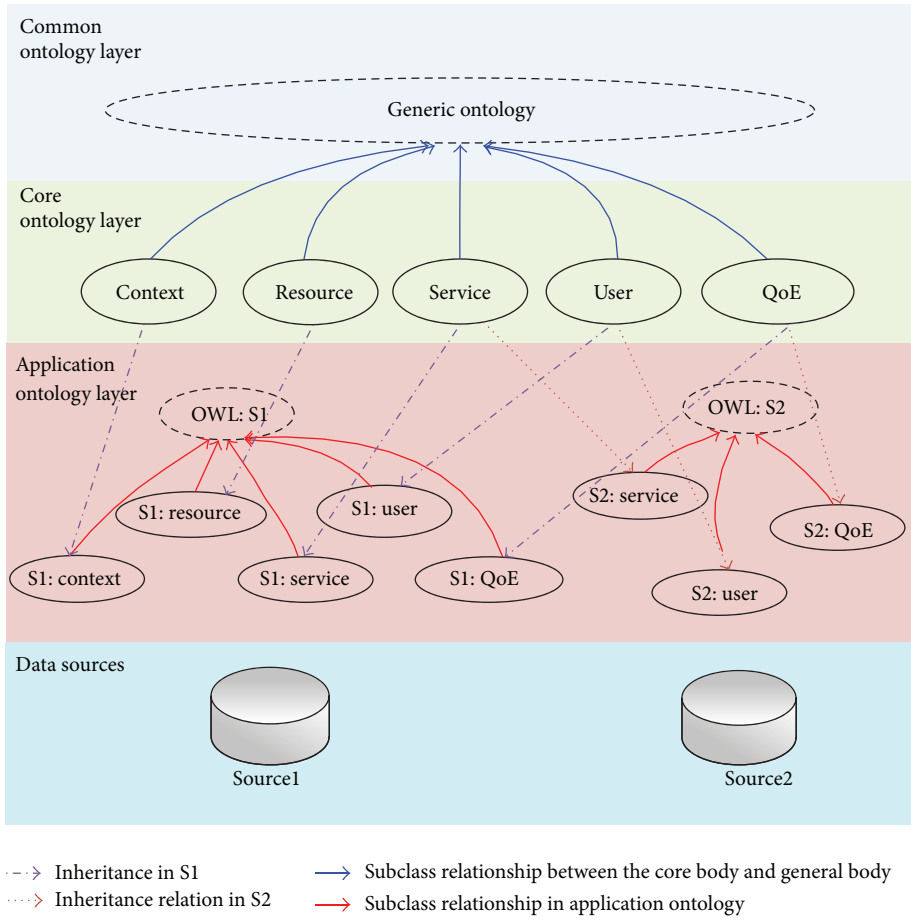


Figure 4: Three-layer ontology model.

ubiquitous service, it needs to call the P2P service discovery algorithms. Platform lets user request client be a Peer by adaptive configuration operation and adds it to a P2P service network formed in the service area which is an interest to the user. The peer can look up and match using P2P service discovery algorithm in the network. In particular, when the user required services cannot be found in the registry, it can also open the P2P service discovery module to continue the search in the ubiquitous service interest areas. Either way, finally we will return a service set sequences that meet the functional requirements and start the next phase of service selection.

(iii) Screen the service sets obtained from the service discovery. It uses more fine-grained personalization selecting based on the user's various context features and service levels. Eventually it returns the most suitable service required sequences adapted to user personalization.

(iv) Take the selected service to users for execution after the assembly verification. The user can choose whether to evaluate after performing the service. Results of the evaluation are returned to service management by the user evaluation component, updating the user's relevant context information.

4. Key Models and Algorithms

4.1. IoT Services Ontology Model Was Constructed. Because of the dynamic of IoT service, it is difficult to construct a single ontology. In order to meet the requirements of the unified sharing and flexible expansion of the related ontologies for IoT service, we propose a three-layer IoT ontology model, as Figure 4 shows.

IoT service ontology includes the generic ontology layer, the core ontology layer, and the application ontology layer from the top to the bottom. In this structure, the lower ontology puts forward the expressing need to the upper ontology. And the upper ontology provides the lower ontology with the design principles. We identified each specific ontology in core ontology layer of the model, namely, the user ontology, service ontology, QoE (quality of experience) ontology, context ontology, and resource ontology. The modeling of each specific ontology in core ontology layer could guide the modeling of service application ontology in the field of various industries. These core ontologies provide the necessary infrastructure to form a moderate coverage for the application ontology for the different information sources. The application ontology layer contains the specific

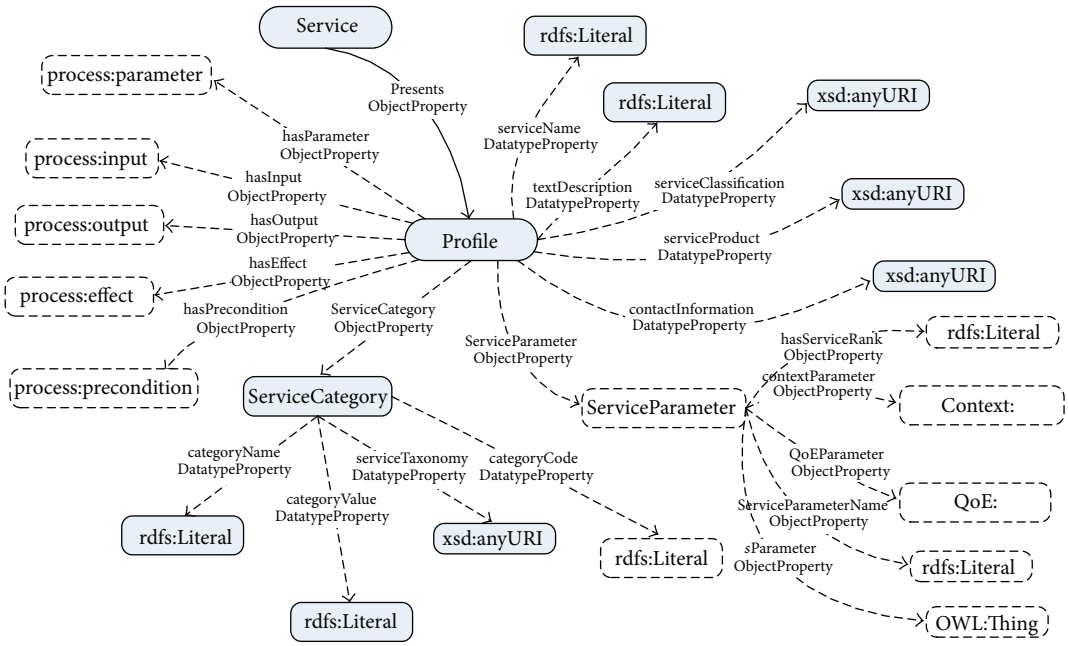


Figure 5: Service profile in OWL-S^{iot}.

application ontology based on the scene-specific information sources, according to the guiding of the core ontology to form a low coverage.

4.2. Semantic-Based IoT Service Description Language OWL-S^{iot}. By contrasting the existing typical Web service semantic description language, OWL-S [16] was selected as a foundation for the evolution of language for its better adaptability and wide application. In order to meet the dynamic flexibility requirements of IoT service description, by some professional processing for OWL-S, a suitable description language for IoT service called OWL-S^{iot} was formed. By Service Parameter in OWL-S, the context ontology and QoE ontology are introduced to form the IoT Service Profile in OWL-S^{iot}, as shown in Figure 5. At the same time, in order to meet the demand of simple cutting, we mark the unnecessary properties of service profile in the resource-constrained environment. These attributes are shown as dashed box in Figure 5.

4.3. Service Management Strategy Both in Registry Mode and Self-Organization by Subcluster. IoT service contained both the traditional common services and the new ubiquitous service. So it needs to be managed, respectively. For the former, we give the service registry architecture based on semantics, as shown in Figure 6.

For the latter, in order to save energy consumption of nodes and to not significantly affect the connectivity of the network, we put forward a self-organization management strategy using sleep scheduling based on "game of life." The main idea and flow of this strategy are described as follows.

(i) Complete the election of service cluster head the election of cluster head takes into account not only random, also the energy consumption of the cluster head [17]. First, the cluster head needs to judge whether they are requested in the last round. If they are, they continue to serve as the cluster head. Otherwise, calculate the node's residual energy threshold in the current round according to (1). Each cluster head node judges whether its current residual energy is greater than the threshold value. And if it is, it continues to be a cluster head. Otherwise, it becomes a common node in the next round. Consider

$$E_{\text{residual}} = E_{\text{initial}} \times \left(1 - \frac{r}{n}\right), \tag{1}$$

where "E_{residual}" is the threshold of residual energy in this round, "E_{initial}" is the average of the initial energy of each node, "r" is the current round, and "n" is the preestimated total rounds.

Then we need to compute a random threshold value in current round to satisfy the randomness of cluster head. The judgment threshold value can be calculated by (2). Then each node generates a random number and judges whether this number is less than the judgment threshold value. If less, then it was elected cluster head, otherwise it becomes a common node. Consider

$$T_{(n)} = \frac{P}{1 - P \times [r \bmod (1/P)]}, \tag{2}$$

where "P" is the percentage of cluster head in all cluster nodes and "r" is the current round.

(ii) The formation of service clusters: cluster head informs all nodes within its communication area that it had become the cluster head by broadcasting. Cluster node will independently judge whether to join the cluster. Because, for a cluster node, it may receive many notification messages from different cluster heads, then it needs to select the nearest cluster head to join to reduce the number of distance-hop

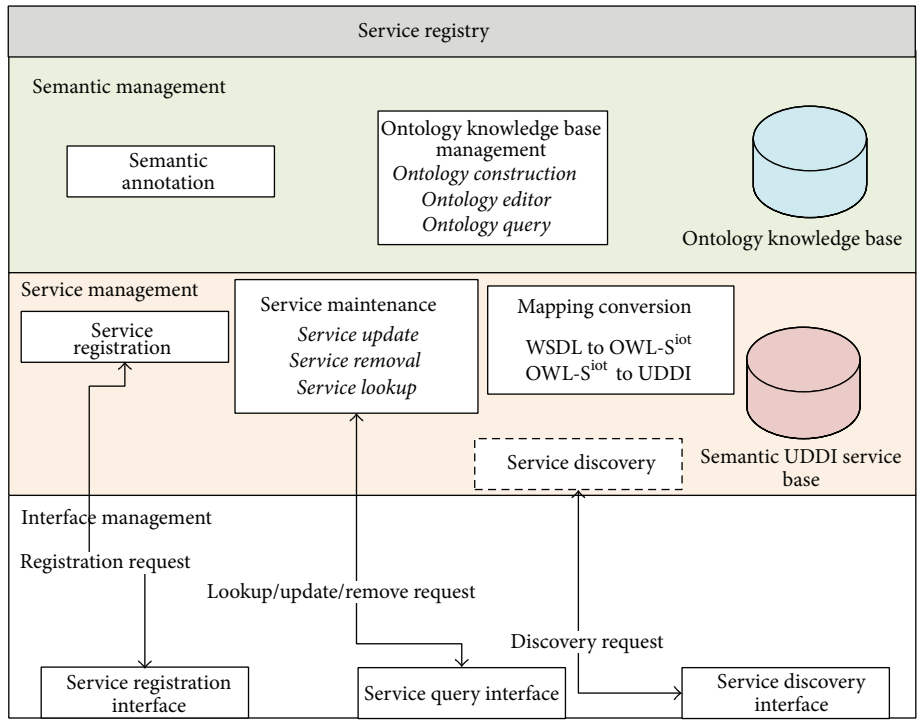


Figure 6: Architecture of service registry.

communication to save energy. Cluster head determines whether the request message is sent to itself after receiving the request message to join its cluster. If so, it allows the node to join the cluster and sends a confirmation message to the requesting node. After a period of interaction, the cluster head could establish a list of the information in this cluster.

(iii) Establishment of a service catalogue: after the formation of the cluster, the cluster node needs to send its service description information to the cluster head. Cluster head finishes collecting service information to create a catalogue of the cluster. Cluster head is responsible for interacting with the client's service request, as well as interacting with other cluster heads.

(iv) Sleep scheduling model based on the "game of life": Initially, the communications functions of all nodes are in the open state. In this period, the cluster node needs to communicate with the cluster head to confirm its identity [18]. If it is the node that needs to execute the customer's service requests, whether the status of the node is working or not in the former round, its state in the next round will be working. And it must immediately communicate and interact with clients to meet the customer's service request. Other cluster nodes without being requested for service by customers will determine their status the next time period based on the "game of life" sleep scheduling rules. The steps are as follows:

(1) determine the survival and birth rule set; they are a set of positive integers;

(2) determine their neighbor nodes "m."

(3) If its last status was working, and $m \leq$ Birth, the status of the node in the next round is working; if $m >$ Birth, then the status of the node in the next round is dormant. If its last status was dormant, and $m \leq$ survival, the status of the node in the next round is working; that is to say, this node will be resurrecting; if $m >$ survival, then the status of the node in the next round is dormant.

After many tests, only when survival value is 1 or 2, birth value is 3; the evolution of the "game of life" sleep scheduling model is relatively stable. And in other situations, the evolution is in shock mode or decay mode. These two modes for this paper do not have research value [17]. So we select the two relatively stable rules "S1B3" and "S2B3."

(v) The interaction process for service request [5]: Figure 7 describes the whole interaction process when a client wants to request the service cluster to service it [19]. We assume that cluster heads have been selected, and the nodes in the cluster have been carried out in accordance with the rules of the sleep schedule.

When a node is in a cluster, it must register its service in the cluster head first. Then it will determine its status in the next period "t" based on sleep scheduling rules. If its status is working, it will listen to the client's request forwarded by cluster head. If its status is dormant, it will send its dormancy request to the cluster head to see if it could go to sleep at time "t". If the current cluster head node does not receive any client requests of the node's service, then allow the node to go to sleep by sending an acknowledgment response. After receiving the dormancy response from the cluster head, the cluster node will be dormant in the following "t" period.

If a client wants to search for a service, it needs to send a request among the cluster heads in the network. The request

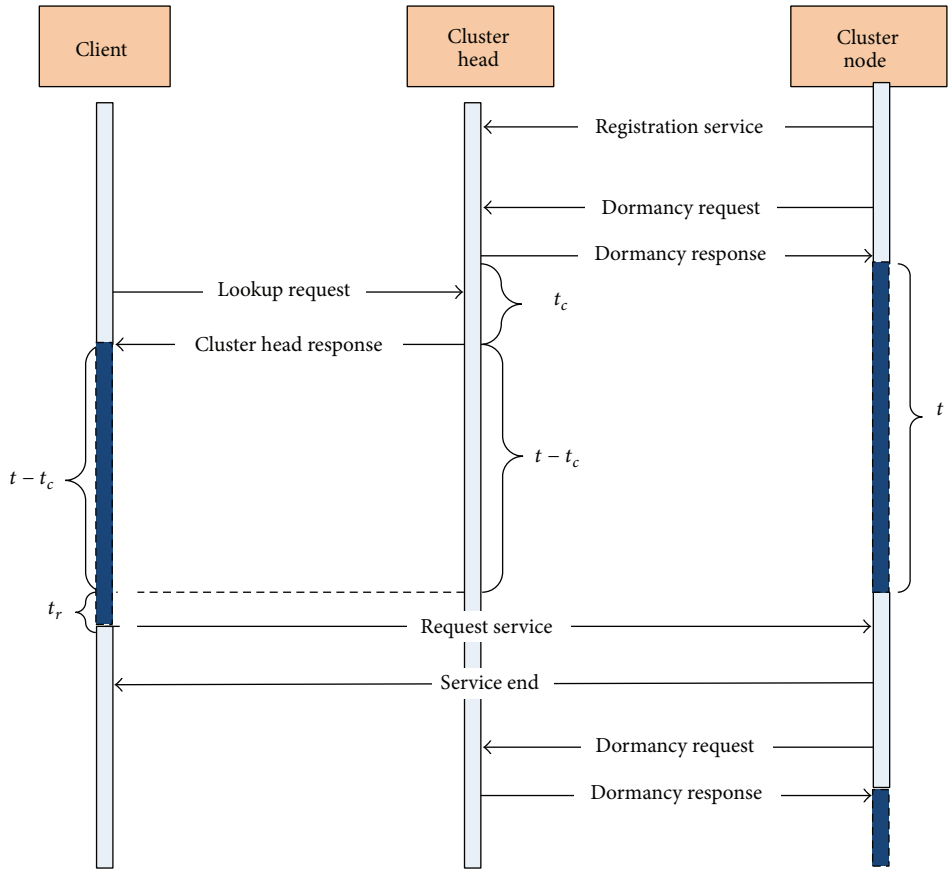


Figure 7: Interaction process for service request.

packet may be transmitted by a hop or multihop between the cluster heads. According to the received service request, each cluster head will find its service catalogue. If the service exists in its service catalogue, then select all the services which meet the requirements to establish a list with the service description. Every service description includes the delayed time ($t - t_c$) that the service could be contacted. The list is sent back to the client. Through this list, the client can choose the most suitable service based on its target. In selecting a service, the client waits for a specified time to contact the service. In this time period, the client can sleep in order to save energy. Before contacting the client service can randomly generate a delay t_r to avoid conflicts with other clients.

(vi) Analysis of the impact of the service interaction delay under this strategy: we use "java" to do a simulation experiment. The amount of service nodes is 100, and we assume that each node provides a service. In the experiment, we ignore the energy management cluster head, periodically and randomly generate cluster heads. We Adopt S1B3 and S2B3 sleep scheduling mechanism for the node to achieve status transition between working and dormant [5]. The time interval of cluster head election is defined as 9 s. The amount number of client requests is a multiple of 10 between 1 and 300. All experiments were repeated 20 times to calculate the average delay time [19]. Each client will wait for a random time interval (selected in [0 s, 9 s]) to broadcast a random service request. If the node that can provide the service is working at this time, it will directly interact without delay. Otherwise, wait for the service node to be awake. As soon as the node is awake, it will form a single request/response interaction with the client. We have designed a service discovery mechanism with no sleep as a comparison method, called nondormancy methods. In order to facilitate the calculation, we ignore the cluster head election delay and state transition delay. Figure 8 shows the S1B3, S2B3, and nondormancy service request interactive delays.

(vii) Nondormancy algorithm has no additional delay besides the base communicating delay, so that it appears as a constant delay. Interaction delay in S1B3 and S2B3 mechanisms is generated when the client waits for the providing service node to wake. For less client requests, the average delay is high. With the increase of the requests by the client, the delay of the interaction is shortened. And the trend is infinitely close to nondormancy algorithm. That is because with the increasing of client's requests, the awake time of the node is longer. In extreme cases, the nodes can be all awake. The connecting delay would be shortened when the service is requested by the client.

5. Conclusion

We propose a semantic-based IoT service framework based on semantic, IoT service features and design principles of IoT

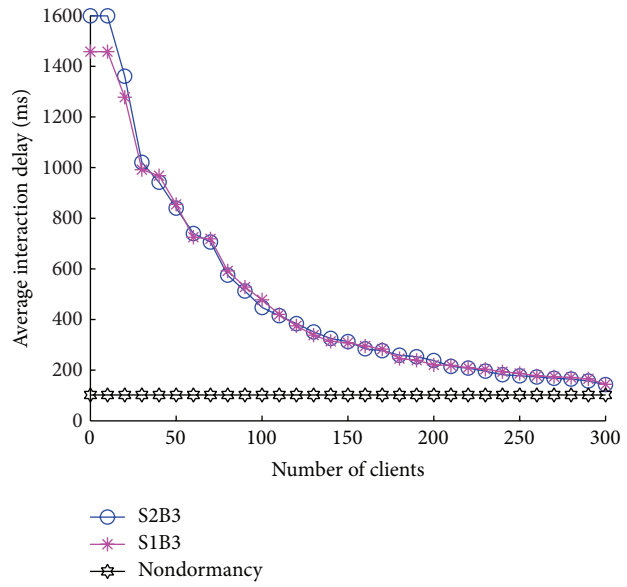


FIGURE 8: The service interaction delay under this strategy.

services framework. The framework is conducted a detailed decomposition referring to the reference standard of IoT architecture. Then we present a semantic-based cross-layer IoT service platform to meet service integration, sharing, and interoperability requirements. And we give the platform's architecture and workflow. Especially, we give the relative fundamental work about semantics, including the model of IoT service ontology and the semantic description method of IoT service. Also we give, respectively, the service registry architecture based on semantics for the traditional common services and self-organization management strategy using sleep scheduling based on "game of life" to save energy consumption of nodes for the new ubiquitous service. The platform extends the range of service source. It could provide the service not only by service providers, but also by normal users.

Conflict of Interests

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

Acknowledgments

This work is supported by Grants from programs of higher-level talents of Inner Mongolia University (nos. 125126, 115117, and 135138), National Natural Science Foundation of China (nos. 61261019, 61262082, and 61272412), scientific projects of High School of Inner Mongolia (no. NJZY13004), Ph.D. Programs Foundation of Ministry of Education of China (no. 20120061110044), Ministry of Education Humanities and Social Sciences Planning Project (no. 13YJAZH130), and Natural Science Foundation of Jilin Province (no. 20130101074JC). The authors wish to thank the anonymous reviewers for their helpful comments in reviewing this paper.

References

[1] H. berndt, *Towards 4G Technologies: Services with Initiative*, John Wiley & Sons, New York, NY, USA, 2008.

[2] Y. H. Liu, *Introduction to Internet of Things*, Science Press, Beijing, China, 2010.

[3] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[4] S. Reddy, V. Samanta, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Mobisense—mobile network services for coordinated participatory sensing," in *Proceedings of the International Symposium on Autonomous Decentralized Systems (ISADS '09)*, pp. 231–236, Athens, Ga, USA, March 2009.

[5] B. Jia, *Research on Semantic-Based Service Architecture and Key Algorithms for the Internet of Things*, Jilin University, 2013, (Chinese).

[6] S. Alam and J. Noll, "A semantic enhanced service proxy framework for internet of things," in *Proceedings of the IEEE/ACM International Conference on Green Computing and Communications & IEEE/ACM International Conference on Cyber, Physical and Social Computing (CPSCom '10)*, pp. 488–495, Hangzhou, China, December 2010.

[7] J. Li, Y. Shvartzshnaider, J. A. Francisco et al., "Enabling internet-of-things services in the mobility first future internet architecture," in *Proceedings of the 13th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '12)*, pp. 1–6, 2012.

[8] W. Wang and J. V. Sesmero, "Report on reference architecture IoT service creation and provision," EC FP7 Project IoT.est, 2012.

[9] J. V. Sesmero and L. L. Muñiz, "Report on use-case scenarios and requirements for the internet of things," EC FP7 Project IoT.est, 2012.

[10] W. Wang, P. Barnaghi, G. Cassar et al., "Semantic sensor service networks," in *Proceedings of the IEEE Sensors Conference*, 2012.

[11] Z. Song, A. A. Cárdenas, and R. Masuoka, "Semantic middleware for the internet of things," in *Proceedings of the 2nd International Internet of Things Conference (IoT '10)*, pp. 1–8, December 2010.

[12] N. Kong, X.-D. Li, W.-M. Luo, and B.-P. Yan, "Model of the resource addressing in the internet of things," *Journal of Software*, vol. 21, no. 7, pp. 1657–1666, 2010 (Chinese).

[13] H. M. Chen, L. Cui, and K. B. Xie, "A comparative study on architectures and implementation methodologies of Internet of things," *Chinese Journal of Computer*, vol. 36, no. 1, pp. 168–188, 2013 (Chinese).

[14] W. Wang, S. De, and A. Lehmann, "Semantic description framework for IoT services," EC FP7 Project IoT.est, 2012.

[15] Chinese Sensor Network National Standard Working Group, Advances in the technology architecture and standards system about Internet of Things [EB/OL], (Chinese), http://wenku.baidu.com/view/4ea7e769a45177232f60a288.html.

[16] OWL-S: Semantic Markup for Web Services[EB/OL], 2004, http://www.w3.org/Submission/OWL-S/.

[17] J. Wang, *Research on Topology Control Algorithm in Delay Tolerant Wireless Sensor Networks*, Jilin University, 2012, (Chinese).

[18] Y. J. Yang, B. Jia, and J. Wang, "An improved algorithm for LEACH protocol in wireless sensor network," *Journal of Beijing University of Posts and Telecommunications*, vol. 36, no. 1, pp. 100–104, 2013 (Chinese).

[19] G. Schiele, C. Becker, and K. Rothermel, "Energy-efficient cluster-based service discovery for ubiquitous computing," in *Proceedings of the 11th Workshop on ACM SIGOPS European Workshop (EW '11)*, Leuven, Belgium, September 2004.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 750601, 15 pages
http://dx.doi.org/10.1155/2014/750601

Research Article

Design Approach Based on EtherCAT Protocol for a Networked Motion Control System

Lei Wang,[1,2] Muguo Li,[2] Junyan Qi,[1] and Qun Zhang[2]

[1] *School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China*
[2] *State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Lei Wang; wang_leiqjy@163.com

Received 28 September 2013; Revised 5 December 2013; Accepted 17 December 2013; Published 13 February 2014

Academic Editor: Gelan Yang

This paper presents a new design approach based on EtherCAT protocol aiming at a special networked motion control system. To evaluate this system, the testing platform of the networked motion control system for ocean wave maker is designed. The paper gives the detailed design for the testing platform including software and hardware. With message access delay and jitter, a set of experiments are done to evaluate the crucial performance for the implemented networked motion control system. By analyzing the experimental results, the systemic performance is verified. And the design approach can be widely extended to other automation application scenarios.

1. Introduction

Motion control, as a subfield of automation, has been finding applications in a broad range of areas such as packaging, printing, textile, semiconductor production, and assembly industries. As the number of the devices of motion control system increases, distributed networked motion control systems are desired in various manufacturing fields. This kind of system must be able to access a mass of device information from any locations in factory floor. In order to solve this problem, various serial communication networks have been designed and implemented to provide reliable and efficient communication paths for data exchange among the system components [1]. In the last two decades, hundreds of proprietary digital network communication protocols named fieldbus have been developed. The construction of networked motion control systems by fieldbus has become a hot topic. In particular, the fieldbuses have been widely used in a great deal of motion control applications for robotics, passenger cars, and aircrafts [1–3]. Nonetheless, because of the relatively low bit rates and (sometimes) changing cycle times, the fieldbuses are difficult to cope with the existing systemic needs for the transmission of a higher quantity of information with the tight timing constraints. Moreover, the software and hardware of multivendor products are incompatible [4]. It is quite difficult to connect the equipment of different manufacturers.

Due to ubiquity, high speed, simplicity, and low cost, Ethernet seems to be the most promising candidate for "the one" network technology to replace fieldbus [5]. Whereas Ethernet is not originally designed for real-time control, the most significant problem is that it is nondeterministic due to its bus arbitration scheme [6]. There are mainly three methods to modify standard Ethernet, including adding an industrial-automation specific application layer on top of TCP/IP, using the priority scheme at the Ethernet MAC layer, and intervening into the scheduling procedure of the MAC layer [7, 8]. The modified Ethernet protocols are called Industry Ethernet protocols which integrated the advantages of both the real-time capability and the merits of Ethernet. So many industrial companies and institutes have shown interests in developing Industry Ethernet protocols. And Industry Ethernet has become the de facto standard for industrial automation networks and widely used in automation and process control domain. In recent years, various Industry Ethernet protocols, such as SERCOS, POWERLINK, ProfiNet/ProfiNet IRT, and EtherCAT, have been proposed to develop networked control systems which have

the control loop closed through a communication network [9–11]. Several researchers also applied the Industry Ethernet technology in motion control [12–14]. Also some analysis and evaluation on the performance of Industry Ethernet protocols have been done [7, 15–18]. At the same time, a considerable number of Industry Ethernet protocols have been introduced into motion control systems to implement networked motion control [19]. However, few researchers investigate the hybrid networked motion control systems with multiaxes servo motors and a great deal of sensors [20]. In view of the special application requirement for high real-time performance and the large amounts of information transmission, we design ocean wave maker networked motion control systems.

The system should accurately simulate ocean waves in the experimental basin and quickly get the experimental data to provide reliable basis for engineering design and scientific research from field sensors. Therefore, the system is constructed with a large number of distributed data acquisition modules and multiservo motors in the field of ocean engineering application. This system should be provided with the capability for transmitting motion control commands and feedback information with lower delay and jitter. Meanwhile, it can help acquire the high efficiency and high real-time large-capacity data more effectively. As a very popular Industry Ethernet protocol, EtherCAT protocol is selected to design the system

The rest of the paper is organized as follows: in Section 2, we briefly introduce the unique features of EtherCAT protocol; and then in Section 3, we describe the overall system architecture and the design of software and hardware. Experimental results are shown in Section 4 for testing and evaluating the system performance. The paper is concluded in Section 5.

2. Major Advantages of EtherCAT Protocol

EtherCAT is a popular real-time Industry Ethernet network defined Beckhoff and supported by the EtherCAT Technology Group (ETG) and currently specified by the IEC 61158, SEMI, ISO [21]. EtherCAT protocol users have explosively grown during the past seven years. As of May 31, 2010, the ETG has 1357 members from 50 countries [21]. It possesses some unique benefits that make it cope with networked motion control needs. Generally speaking, the major advantages of EtherCAT protocol include high data transmission efficiency and speed and high accuracy clock synchronization. Section 2 briefly presents the main merits and functions of the protocol.

2.1. High Transmission Efficiency and Speed

2.1.1. EtherCAT Network Structure. The whole EtherCAT network is made of master station nodes and slave station nodes as shown in Figure 1.

One EtherCAT master (e.g., industrial PC or embedded microcontroller) is connected with a certain number of EtherCAT slaves [13]. All stations are connected together to create a logical ring through standard Ethernet cable. The master station sends the standard Ethernet frames with EtherCAT data around the ring. When the frames pass through each slave station node, they are processed without stopping. At the open end, the frames are transferred backward to the master station.

2.1.2. Structure of EtherCAT Frame. In EtherCAT protocol the basic Ethernet frame structure is not changed. Thus, EtherCAT is compatible to other Ethernet protocols. EtherCAT frame can be distinguished from other Ethernet frames through the Ethernet type field with the special type identifiers 88A4. A standard Ethernet frame consists of five fields: preamble, start of frame delimiter, header, data, and a Frame Check Sequence (FCS), as in Figure 2.

2.1.3. Specific Hardware of Slave Station and Addressing Image. Slave station node is mainly made up of an ASIC chip known as EtherCAT slave controller (ESC) and an application controller. The ESC implements the communication interface function between EtherCAT network and slave application. Each ESC possesses a consistent random access memory (RAM) which is used to exchange data between the EtherCAT master and local slave application controller. The special data link layer hardware units, named Synchronize Managers (SMs) and Fieldbus Memory Management Units (FMMU), carry out the management of the user memory of ESC and the addressing images from EtherCAT PDOs to physical memory units. Figure 3 shows how FMMUs are configured to map physical memory to PDOs images.

Each PDOs image embedded in EtherCAT frame has a unique local address which is assigned by master station. During start-up phase, the master station writes the registers of FMMUs with the values: a local bit start address, a physical memory start address, a bit length, and a type field that specifies the direction of mapping (Input/Output). The application controller of slave station extracts or inserts the data into EtherCAT frame according to the read values stored in the RAM units of ESC from the registers of FMMUs. Both the master station and the slave one need access the RAM; thus ESC must be able to inform both sides of the time when they can access the RAM; otherwise the data consistency and security cannot be guaranteed. To solve the question, the RAM units are managed by a series of SMs. Each SM controls a buffer in RAM and generates interrupt to inform the master and slave station application controller when the access of the other side has finished. SMs consist of a series of registers and they are configured by the master station during start-up phase [22].

2.1.4. Implementation of High Transmission Efficiency and Speed. As shown in Figure 1, when the frames from the master station pass through the slave nodes, they are not encoded or decoded. Only the data are read and written as the bits passing the ESC. After the data are accessed, the frames are automatically forwarded to the next port by the EtherCAT Processing Unit in ESC; accordingly forwarding delay only depends on the receive RAM buffer size and EtherCAT Processing Unit delay. In view of unparalleled

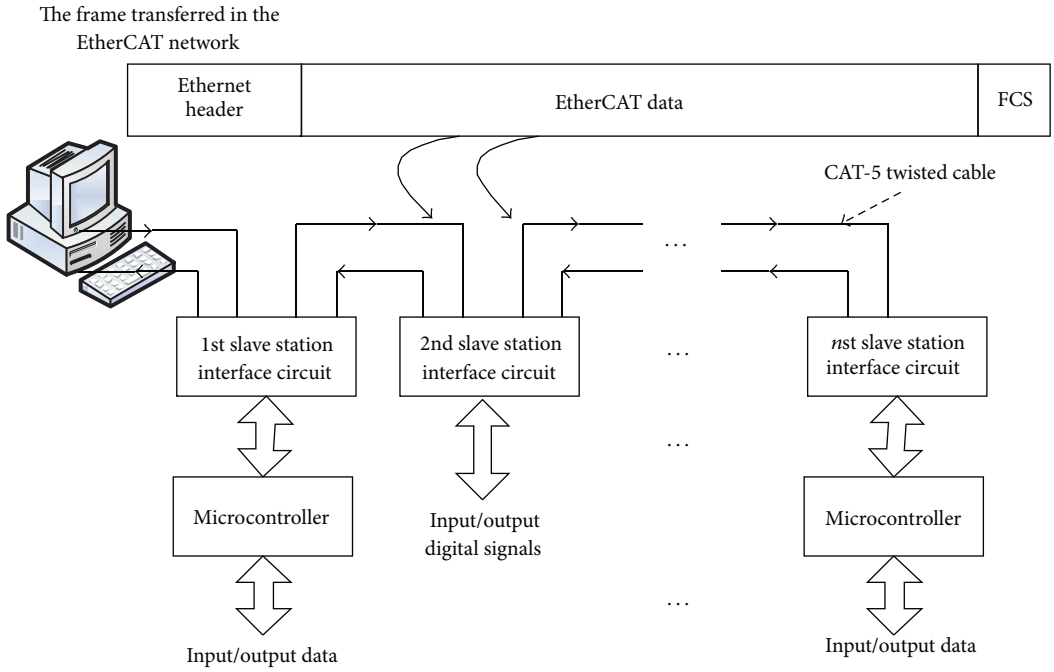


Figure 1: Whole structure of EtherCAT network with two slave stations.

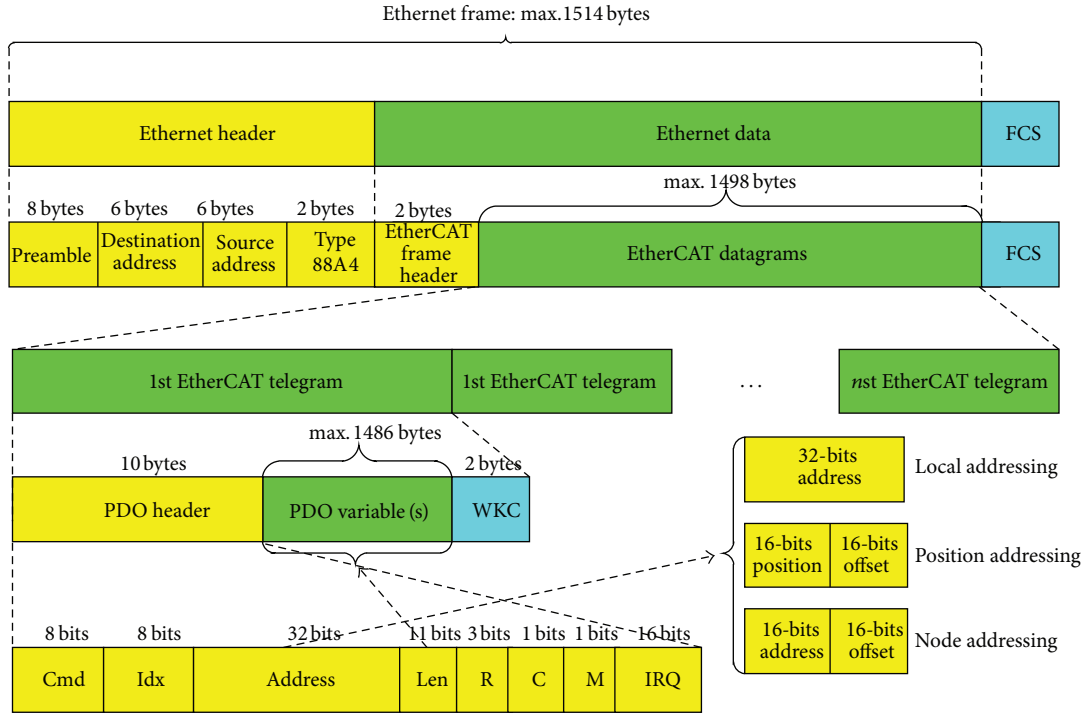


Figure 2: EtherCAT frame structure.

processing technology, the EtherCAT network provides faster communication cycle time than most of Industry Ethernet networks.

Besides, Figure 2 shows that the largest EtherCAT frame size can be up to 1498 bytes; a large amount of Output/Input data to/from multislave station devices are transferred into an EtherCAT frame during one communication period. Only the relevant commands (Cmds: In EtherCAT telegram field) are recognized and executed by either any slave node or several nodes simultaneously. These commands ascertain the master accessing services for the addressable memory section of slave station nodes. Moreover, the frame may comprise the

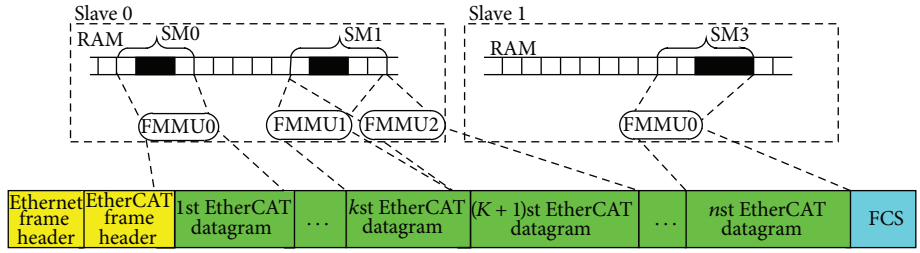


Figure 3: Management of RAM using SMs and FMMUs.

data of many devices both in sending and receiving direction, so the usable data rate increases to over 90%. For example, we can take into account a networked control system with 20 slave stations. When we visit all the stations, for other Industry Ethernet protocols, the master station must send more than 20 different packets. However, for EtherCAT, one long packet that touches all slaves is sent, and the packet contains 20 devices values of data. Besides, if all the slaves need to receive the same data, one short packet is sent, and the slaves all look at the same part of the packet as it is streaming through, optimizing the data transfer speed and bandwidth. So EtherCAT protocol possesses more high transmission efficiency and speed over other Industry Ethernet protocols. It is particularly well adapted to use in the system with a large amount of sensors.

2.2. High Accuracy Clock Synchronization. As we know, the precondition of the high performance of networked motion control is that communication protocol ensures real-time and synchronization of data transmission.

In contrast to many of other Industry Ethernet protocols, EtherCAT adopts accurate clock synchronization scheme referred to as the distributed clock (DC) functionality [23]. This function is realized by using specific hardware module of ESC and provides an identical clock time for both the slave station and master station. To present the principle of distributed clock, a line topology structure with five slave stations is depicted in Figure 4. Three different types of clocks, namely, as master clock, reference clock, and slave clock, are defined in the system [24]. The reference clock is used to synchronize all other slave stations [25]. During the system start-up phase, the master station must set the local time of the reference clock and the other slave clocks to the current reference time. To this end the EtherCAT master sends a special EtherCAT datagram at short intervals (with sufficient frequency that ensures that the slave clocks remain synchronized within the specified limits), in which the EtherCAT slave with the reference clock enters its current time [26]. This information is then read from the same datagram by all other EtherCAT slaves featuring a slave clock. The defined time parameters in Figure 4 are listed in Table 1.

To guarantee accurate clock synchronization, the following operation must be finished beforehand. Firstly, master station passes the propagation delay measurement frame to all slaves and the slave controller records the receive time of the frame. And the master station collects the time stamps afterwards and calculates the propagation between all slaves. Secondly, the local offset time to reference clock of every slave station must be compensated [27]. In order to do that, the time of each slave clock is compared with that of the reference clock, and the difference is stored in each slave to make the slave get the same absolute system time. Finally, the drift between reference clock and slave clock cannot be avoided because each slave station has respective free running oscillator with different parameters. With the parameters information, the master can calculate the delay of each node. The master repeats the calculation for every frame it sends. As the network operates, the enormous sample size means that the master has incredibly accurate data. The inherent ring topology creates an incredibly efficient clock mechanism that increased in accuracy with every message [13, 14].

3. Design of Networked Motion Control System

3.1. Principle of Networked Wave Maker Control System. We design the full closed-loop networked motion control system, as a typical application, for ocean wave maker. Ocean wave maker is a kind of experiment equipment simulating the wave generation according to calculated data under laboratory condition. The basic principle of ocean wave maker is to vibrate the water through the mechanical motion of the wave board to generate waves. The mechanical motion of the wave board should be suitable for the reciprocating smooth variable motion of the time domain wave signals. The servo should respond to changing waves rapidly with high speed and tracking precision. Due to the protocol performance limit, half closed-loop control mode is widely used in existing networked wave maker systems. In the control mode, the data of a full target wave curve are downloaded to slave station memory section, and then the servo motor executes control instructions from servo driver to control push board movement. The actual position values from encoder are feedbacked to slave station application controller. After one wave making period, the data are uploaded to master station and then the master station program modifies the next data values. This kind of control style cannot achieve complicated advanced control algorithm because of the capability limit of slave microcontroller. So the system control accuracy cannot be ensured. To improve the control accuracy of system, it is necessary to construct a full closed-loop between master station and field sensors with less than 1 milliseconds network

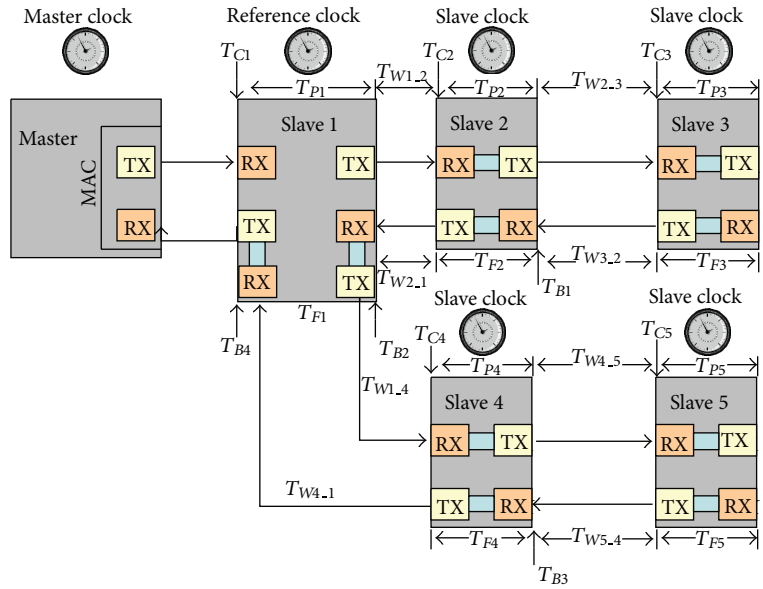


Figure 4: Principle of distributed clock.

Table 1: Parameters for propagation delay calculation.

Parameter	Description
T_{Cx}, T_{By}	Receive time of the first bit of the measurement frame gets to the slave port (x = 1–5, y = 1–4). It will be DC Receive Time 0 register.
T_{Px}	Processing delay of slave x (x = 1–5) when the frame passes through the slave x.
T_{Fx}	Forwarding delay of slave x (x = 1–5) when the frame comes back from the slave x.
T_{Wxy}	Wire propagation delay between slave x and y (x/y = 1–5).

delay. Figure 5 illustrates that the master station sends target position commands to the microcontroller via EtherCAT network. Different from half closed-loop control system, full closed-loop control mode can supply real-time field sensors data to master station before the next commands are sent. The commands are revised by the application program which runs in the master station. Therefore we proposed a full closed loop control structure between master station and field sensors. The commands are able to be revised in one communication period, so actual control curve will be more accurate.

3.2. Detailed Design for Networked Motion Control System of Wave Maker. In this study, we construct the networked motion control system of wave maker including 10 slave stations. Figure 6 shows the whole structure of the system.

The master station is a common PC with a standard Ethernet network card. And each slave station consists of network interface circuit board, 16-bit PIC Microcontroller, A/D and amplifier-filter circuit, servo driver interface circuit, multiplexer circuit for sensors, servo driver, and servo motor. In the following, we will present in detail the system with master station and slave station.

3.2.1. Circuits Principle of Slave Station. Figure 7 shows the designed circuit diagram of system. We will present the circuit principle in terms of the aforementioned four parts.

(A) EtherCAT Network Interface Circuit [14]. In the circuit, the ASIC named ET1100 is regarded as EtherCAT slave controller, which serves as the Data Link Layer. The layer corresponds to Layer 2 in ISO model and provides real-time communication assurance among devices connected via EtherCAT network. Furthermore, the layer carries out the function of the frame check and accomplishes data transmission by extracting data from and/or inserting data into the Ethernet frame, based on the Data Link Layer parameters stored at predefined memory locations. However, the data frames transmission depends on network Physical Layer known as PHY. The layer can receive data bits stream from Data Link Layer and encode the bits into signals. Sequentially, the signals are sent to the transmission medium

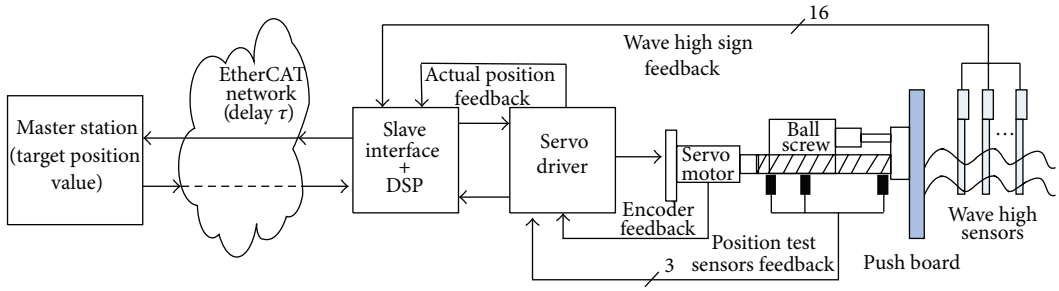


Figure 5: Principle of networked motion control system of wave maker.

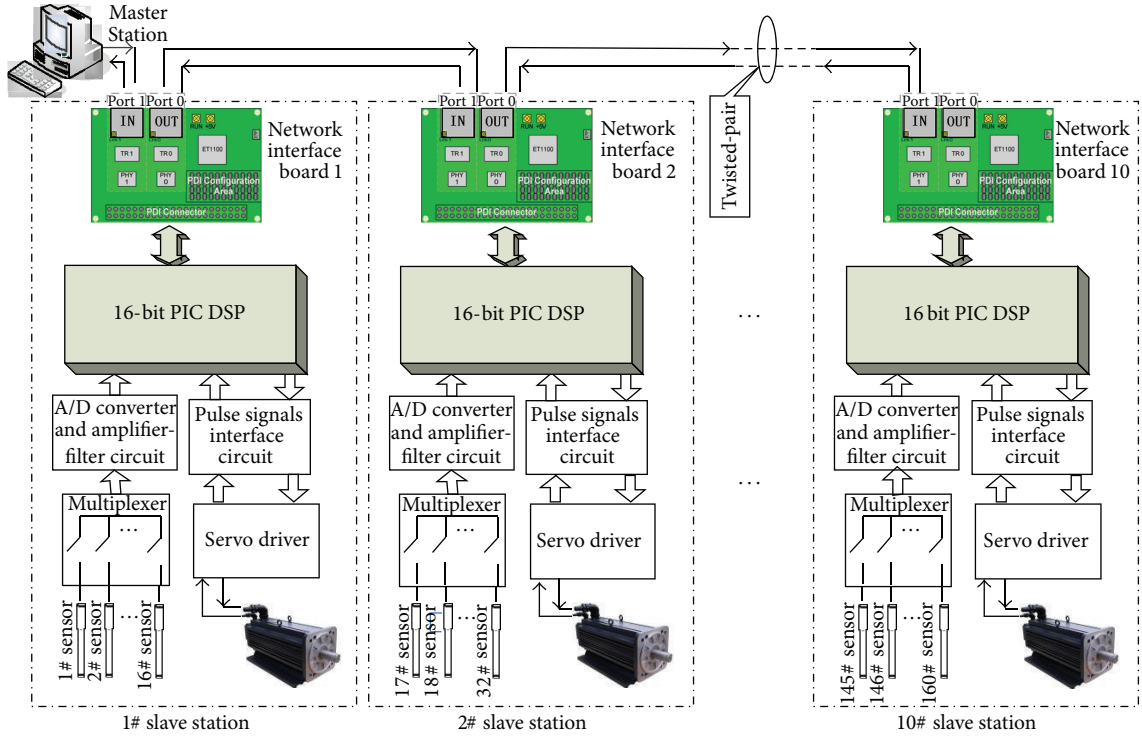


Figure 6: Structure of networked motion control system of wave maker.

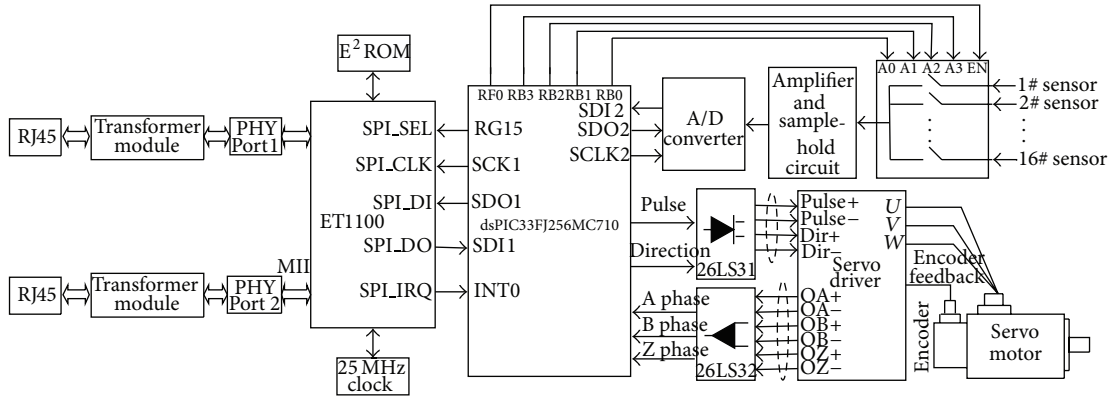


Figure 7: Structure of slave station module circuit.

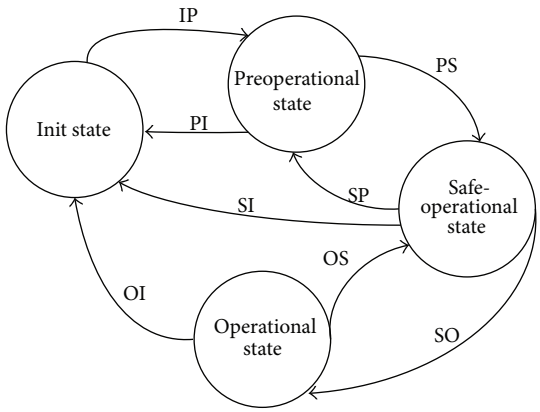


Figure 8: State transferring diagram.

and received by the next Physical Layer interface. After being decoded, the signals are passed on to the Data Link Layer of the next slave controller. To perform the function of Physical Layer, KS8721BL, the Physical Layer interface chip is linked to ET1100 via Medium Independent Interface (MII). At the same time, the chip is connected to medium port connector RJ45 via network transformer HR601680 to improve the signal anti-interference capability. HR601680 is a 1 : 1 transformer with a smaller package and supports 10 M/100 M Ethernet. In addition to the circuit and parts mentioned above, ET1100 provides Serial Peripheral Interface (SPI) to link application layer controller. As an individual network interface board, it provides the 52 pins plug to connect with application controller. Via the plug, the ready data from Data Link Layer will be offered to the application layer microcontroller. The board contains the affiliated circuits, such as clock circuit which generates clock signals by a 25 MHZ crystal oscillator, EEROM circuit used to store device configuration parameters and device description information, reset circuit, and so forth.

(B) Application Controller Circuit. The application controller is the control core, and it takes charge of running the EtherCAT protocol program and application data accessing program. Due to the advanced motor control peripheral features, fast and efficient CPU, and small cost-effective package sizes, we adopt dsPIC Digital Signal Controller (dsPIC33FJ256MC710) as the core controller. The chip has up to 85 programmable digital I/O pins. In addition, several peripheral features are available including four timer/counters and eight Capture/Compare/PWM modules. Especially, there are two SPI modules in the chip, connecting the network interface controller and the network interface chip (ET1100), the network interface controller, and serial ADC, respectively. And the chip provides the control and data lines for all other circuits. It is a bridge linking the EtherCAT network interface circuits and field data acquisition and servo motor control circuits.

(C) Wave Height Data Acquisition Circuit. The data acquisition circuit of each slave station in the system, with the capacity of 16 simple analog channels input, is designed. Sixteen wave height sensors are connected with the channels input ports. Because the output signal from wave sensor is weak, only 0.2 V to 0.8 V, the collected signals must be amplified. To simplify the circuit design, the 16 sensor channels are selected by a multiplexer through the 4-bit binary address lines A0, A1, A2, and A3. These pins receive the address signals from the application microcontroller output pins RB3, RB2, RB1, and RB0 (see Figure 7). The channel number sent by the master station determines the values of the pins. According to the values, the multiplexer selects one of the collected 16-channel signals and forwards the selected input signals into a single outline. The output signal feeds into the amplifier which is capable of implementing gains 5. Hence the signal is amplified up to 1 V to 4 V. Finally, the analogue signals will be converted into digital signals via ADC. However, the ADC must have a stable signal to achieve conversion. Thus, the amplifier out signal is imported into a sample and hold circuitry before it is provided to ADC. The chip AD7276 is chosen in our system. It is 12-bit, high speed, low power, successive approximation ADC, which operates from a single 2.35 V to 3.6 V power supply and feature throughput rates of up to 3 MSPS (Million Samples per Second) and provides a SPI to connect with microcontroller.

(*D*) *Servo Motor Interface and Control Circuit* [15]. The part of circuit includes mainly pulse sending/receiving interface, servo motor encoder feedback interface, servo driver, and servo motor. The microcontroller receives the control commands from the master station and generates the signals of direction and pulse number. The servo motor accordingly rotates in a given angle. Pulses as a square wave are sent to servo driver via interface circuit, the number of pulses determines the angle of rotation, and frequency of square wave determines the speed of rotation. And then the driver controls the motor to move to destination position, which takes both speed and position feedback signals from the encoder of servo motor. The feedback values can be detected and calculated by the microcontroller. In Figure 7, the Minas A4 series driver and servo motor of Panasonic Company are selected and the driver is configured for position control mode. Furthermore, the motor is settled to take 10000 pulses to rotate 360 degrees. 27.8 pulses are required for a rotation of 1 degree. However, the fraction of pulse value can be omitted. So the motor is made to rotate in steps of 1.8 degrees which requires 50 pulses ($(10000 * 1.8)/360 = 50$).

3.2.2. State Machine and Object Dictionary's Definition. State machine is a very important concept in EtherCAT network communication. It is a series state transition responsible for the coordination of master station and slave station application at start-up and during operation. It is achieved both in master station and slave program. All of the state changes are initiated by the master station and the slave stations respond to the changes by executing corresponding program. The requested state by master station is written into the application layer control register and the response resulted by slave station is reflected in the application layer status register. The two registers are located in the chip

ET1100. The detailed state transition diagram is shown in Figure 8.

The four states include [22] the following:

(i) Init: the master station initializes the configuration registers of the chip ET1100 and configures the SMs for mailbox;

(ii) pre-operation: the mailbox communication is activated;

(iii) safe-operation: the mailbox and Process Data input communication is implemented;

(iv) operation: both process data input and process data output are implemented.

Besides, the master will initialize the corresponding resisters of the chip ET1100 during state changes.

Object dictionary is another important concept in EtherCAT network communication. The EtherCAT protocol extends the object dictionary functionality of CAN bus standard. It defines all related data objects of the device in a standardized way. The object dictionary is made up of two sections. The first section includes general device information such as device identification, manufacturer name, and communication parameters. The specific device is functionally described in the second section. All of the accessed information is named as objects. According to the profile definition, a series of index number are used to describe the PDOs. In EtherCAT protocol, there are two types of PDOs including Sync Manager Channel objects and application objects. The indices 0x1C10 to 0x1C2F represent the Sync Manager Channel objects which describe a consistent memory area and manage several PDOs. The application objects are located at indices 0x1600 to 0x16FF for Receive PDOs (RxPDOs) and at indices 0x1A00 to 0x1AFF for Transmit PDOs (TxPDOs). The objects are read through the corresponding entries in the object dictionary. The object dictionary can be described by XML and it is downloaded into the EEROM on the network interface board using configuration tool. In order to access the PDOs rightly, the PDOs mapping relationship is represented in Table 2.

3.3. Master Station Principle. As we can see in Figure 9, the master station comprises three main modules.

The master driver module is the base of implementing EtherCAT communication. It includes a complete protocol stack with the following three types of drivers:

(i) NIC Miniport Driver: this driver program is responsible for sending/receiving frames to/from NIC and provides an interface to upper protocol driver program;

(ii) NDIS Intermediate Driver: the driver between NIC Miniport driver and Protocol Driver can control all traffic being accepted by the NIC. In order to implement EtherCAT protocol, the driver filters all Ethernet data frames and blocks all the frames except EtherCAT frames;

(iii) NDIS Driver: it provides services for application layer clients [16].

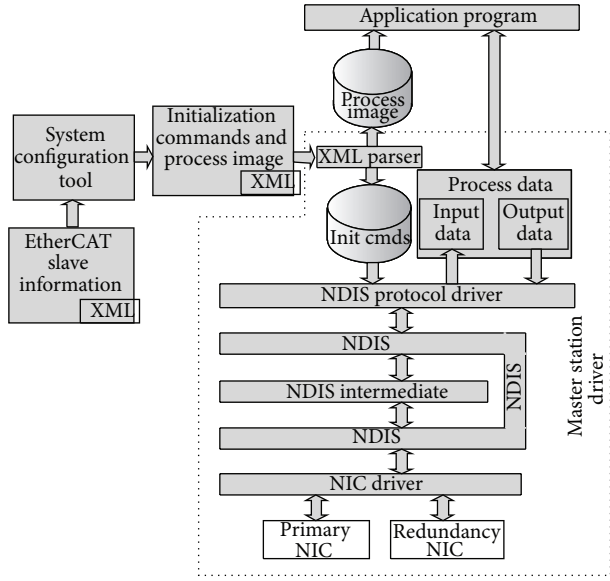


Figure 9: Describing for principle of master station.

Moreover, the master station driver can parse the XML which defines the initialization EtherCAT commands that should be sent during a specific transition according to the EtherCAT state machine and PDOs. And it is loaded through system configuration tool and is read from slave station EEROM during start-up by master station driver. The application program variables build a kind of mapping relationship with PDOs through XML definition. Our aim is to generate the wave height form based on the designed system accurately, the application program can implement the function of calculating and sending revised wave making pulse signals according to the given target wave sequence and the feedback data from field sensors and servo driver encoder. Therefore, a series of array names, such as motor_pulse, motor_direction, channel_number, encoder A, encoder_B, encoder_Z, and fb_sensor, are defined. Each defined array except the array fb_sensor possesses 10 member variables because of the system with 10 slave stations. There are 16 sensors at each slave station, so the total 160 sensors are linked to the system. Therefore, we must define the array fb_sensor including 160 member variables. All of the mentioned member variables are mapped with PDOs in the application program.

The application program sends the pulse signals and rotating direction signals to servo drivers by equal interval time. The target wave curve is converted into a series of discrete data dots. The number of pulse signals represents the data dots size. After the program sends the first group of signals, the feedback signals from servo drivers and wave height sensors are received. Then the next signals which will be sent are modified by the specific algorithm. The system goal is to make precise wave form in the experiment pool, which simulates the target curve wave form. The algorithm is described as "Algorithm 1".

```
  (i) Initialize PDOs image
Call Initialization ();
   {
   int motor_pulse[10] ← 0, motor_direction[10] ← 0;
   int encoder_A[10] ← 0, encoder_B[10] ← 0, encoder_Z[10] ← 0;
   int fb_sensor[160] ← 0, channel_number[160] ← 0;
   }
Execute LoadECATConfiguration ();
   {
   Load the *.xml file;
   /*brief Gets a pointer to the PDOs*/
   INT *ProcessDataPtr (INT imgId, INT inOut, INT offs, INT size);
   /*brief Gets the size of the PDOs*/
   INT ProcessDataSize (INT imgId, INT inOut);
   If (PDOs is output data)
   {
   UNSIGED LONG nData ← ProcessDataSize(0, VG_IN);
   INT *pData ← ProcessDataPtr(0, VG_IN, 0, nData);
   }
   Else if (PDOs is input data)
   {
   UNSIGED LONG nData ← ProcessDataSize(0, VG_OUT);
   INT *pData ← ProcessDataPtr(0, VG_OUT, 0, nData);
   }
   }
(ii) Send PDOs output data
motor_pulse[10] ← the sent pulse values of target curve to every channel;
motor_direction[10] ← the sent direction values of target curve to every channel;
channel_number[160] ← the selected feedback sensor numbers;
For (INT i = 0, i < 10; i++)
{
*pData = motor_pulse[i];
pData++;
*pData = motor_pulse[i]
pData++;
}
*pData ← channel_number[j], where j = 0, . . . , 160;
pData++;
Sendpacket (pData, nData);
(iii) Receive PDOs input data
Receivepacket (pData, nData);
For (INT i = 0, i < 10; i++)
{
encoder_A[i] = *pData;
pData++;
encoder_B[i] = *pData;
pData++;
encoder_Z[i] = *pData;
pData++;
}
fb_sensor[j] ←*pData, where j = 0, . . . , 160;
pData++;
Calculate the next output data according to the received data;
(iv) The program jumps the (ii).
```

Algorithm 1

Table 2: PDOs mapping relationship.

Sync channel 2 indices (process data output)	0X1C12	
RXPDO indices	0X1600	0X1601
Input entry indexes	0X7000, 0X7001, . . ., 0X70FF	0X7010, 0X7010, 0X7012
Input entry names	1# sensor, 2# sensor, . . ., 16# sensor	encoder_A, encoder_B, encoder_Z
Sync channel 3 indices (process data input)	0X1C13	
TXPDO indexes	0X1A00	0X1A01
Output entry indices	0X6000, 0X6001	0X6010
Output entry names	Motor_pulse, motor_direction	Channel_selection

4. Performance Analysis and Experiment Results

Based on the system features, the two most important performance indices including cycle time and delay jitter are analyzed and tested.

4.1. Definition of Cycle Time and Delay Jitter. The cycle time is defined as the necessary time to accomplish an input/output data exchange between the controller and all networked devices [7]. The cycle time T_{Cycle} can be calculated by

$$T_{\text{Cycle}} = T_{\text{SM\_process}} + T_{\text{S\_frame}} + T_{\text{R\_frame}} + T_{\text{SP}_{\text{delay}}} + T_{\text{RP}_{\text{delay}}} + T_{\text{Idle}} + T_{\text{RM}_{\text{process}}}. \quad (1)$$

In (1), $T_{\text{SM\_process}}$ and $T_{\text{RM\_process}}$ are the master sending processing time and master receiving processing time, respectively. They can be described by

$$T_{\text{SM\_process}} = T_{\text{S\_AP\_process}} + T_{\text{S\_Pre\_frame}} + T_{\text{S\_PD\_process}},$$
$$T_{\text{RM\_process}} = T_{\text{R\_AP\_process}} + T_{\text{R\_Pre\_frame}} + T_{\text{R\_PD\_process}}. \quad (2)$$

In (2), $T_{\text{S\_AP\_process}}$ is the time for application program calculating output variables and assigning them to the output PDOs, $T_{\text{S\_Pre\_frame}}$ and $T_{\text{R\_Pre\_frame}}$ are the data frame queuing waiting time for sending and receiving data, respectively, and $T_{\text{S\_PD\_process}}$ and $T_{\text{R\_PD\_process}}$ are the protocol stack program processing time during which the data frame is sent and received.

In (1), $T_{\text{S\_frame}}$ and $T_{\text{R\_frame}}$ are the delay time of sending and receiving data frame, depending on the network communication bit rates; $T_{\text{SP\_delay}}$ and $T_{\text{RP\_delay}}$ represent the propagation delay time of sending and receiving the network data, including the wire propagation delay time, network controller forwarding delay time, slave microcontroller sampling, and processing delay time. T_{Idle} is the waiting time between the two sequentially sent frames. The time defined above can be shown in Figure 10.

Delay jitter in communications refers to the variations between the maximum cycle time and the minimum one [7]. Through this index we can decide whether the network is fast enough to handle the closed-loop servo system, or to use it only to download programs, send commands to the motion controller, and check the status of the controller for diagnostics and remote applications. Some systems can tolerate a little jitter. For example, data acquisition network communication systems that work with transmission rates of 100 milliseconds will not be affected by jitter. Likewise, distributed I/O handles cycle time of 10 milliseconds with little effect from jitter, while motion control systems work best with transmissions that contain less than 1 ms of jitter.

4.2. Cycle Time and Delay Jitter of Different Protocols. The IEC 61784-2 standard defines a set of Performance Indicators (PIs) in order to specify the capabilities of the real-time Ethernet networks [28, 29]. As for the presented system in the paper, we focus on three main PIs: communication cycle data throughput and time synchronization accuracy. To verify the protocol performance superiority, we compare it with two representative Industry Ethernet protocols, namely, POWERLINK and PROFINET. The three protocols are comparable because they are standard Industry Ethernet protocol defined in IEC 61784-2 as a Publicly Available Specification. Moreover, the aforementioned two protocols have been widely used in the field of networked motion control. In particular, they achieve the whole master/slave-model cyclic real-time data exchange. But EtherCAT protocol exchanges the real-time data by using a so-called summation frame [17]. Because the comparison is more meaningful under the identical conditions, we will focus on a specific configuration, namely, the line topological architecture, comprising similar devices to the system designed in the paper. Furthermore, the length of data frame is ascertained as 128 bytes according to the system requirement and the network works at the speed of 100 Mbit/s. So the frame transfer time can be calculated as $T_{\text{frame}} = 10.24\,\mu$s. Moreover, we assume that the length of a copper-based segment between two devices is set to 50 meters, which leads to a medium delay of $T_{\text{medium}} = 227$ ns [17].

4.2.1. Cycle Time. In our analysis, the master process time is set as 200 μs. The communication cycle time can be written by

$$T_{\text{Cycle}} = T_{\text{master\_process}} + T_{\text{S\_frame}} + T_{\text{R\_frame}} + T_{\text{SP\_delay}} + T_{\text{RP\_delay}} + T_{\text{Idle}}, \quad (3)$$

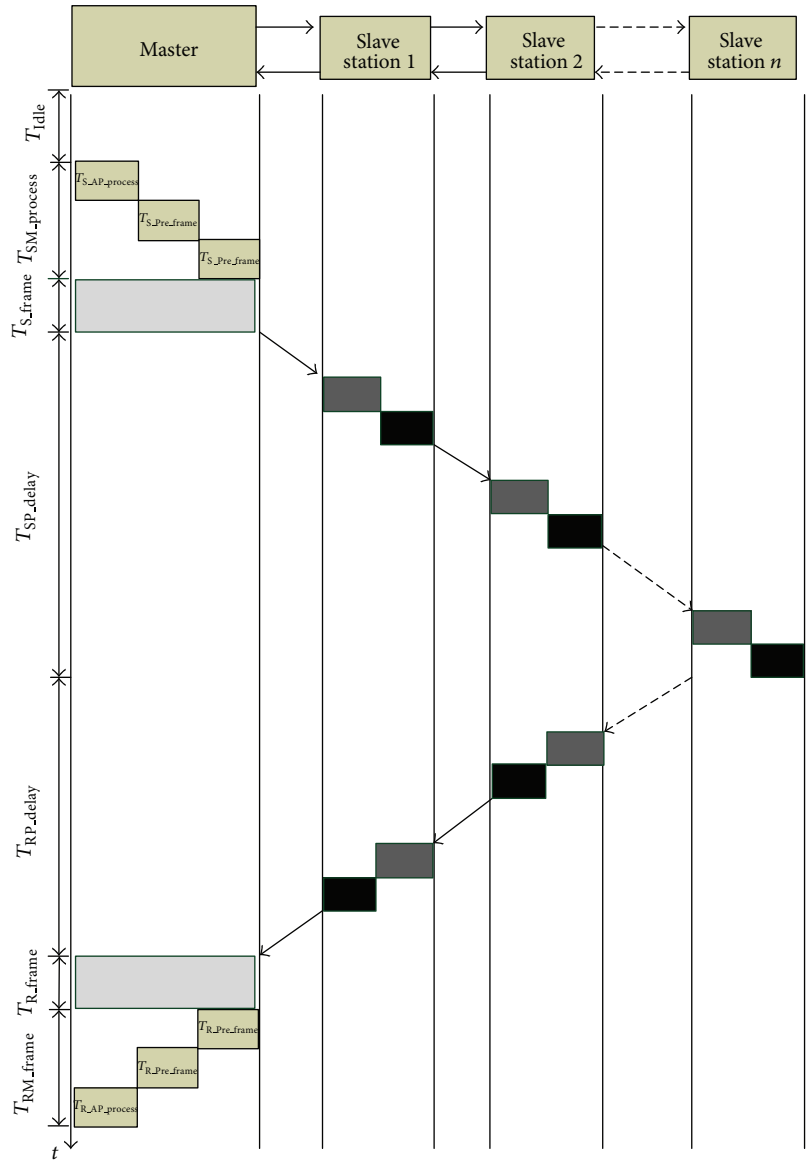


Figure 10: Network communicating cycle time.

where T_{S_frame} is equal to T_{R_frame}, and they can be calculated through the frame length and network speed. T_{SP_delay} is equal to T_{RP_delay}. Consequently, (3) can be defined as follow:

$$T_{Cycle} = T_{master\_process} + T_{tran\_delay} + T_{frame} + T_{Idle}, \quad (4)$$

where

$$\begin{aligned} T_{master\_process} &= T_{SM\_process} + T_{RM\_process}, \\ T_{tran\_delay} &= T_{SP\_delay} + T_{RP\_delay}, \\ T_{frame} &= T_{S\_frame} + T_{R\_frame}. \end{aligned} \quad (5)$$

According to the theoretical analysis provided in [17], T_{tran_delay} can be expressed as

$$\begin{aligned} T_{tran\_delay} &= N_{devices} * \left(T_{ecat\_fwd} + T_{medium}\right), \\ T_{frame} &= T_{data} + T_{ethernet} + T_{ecat\_ov}. \end{aligned} \quad (6)$$

We assume that there is no time delay between two frames ($T_{Idle} = 0$). Consequently, (4) can be defined as

$$\begin{aligned} T_{Cycle} = N_{devices} * \left(T_{ecat\_fwd} + 2 * T_{medium}\right) \\ + T_{frame} + T_{master\_process}. \end{aligned} \quad (7)$$

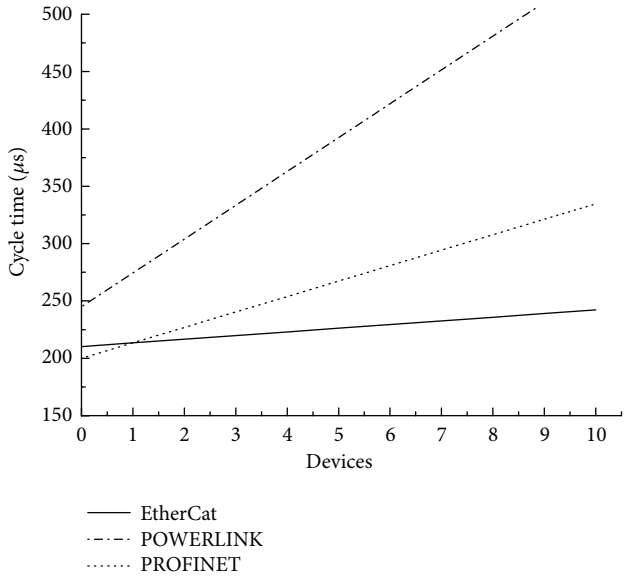


Figure 11: Communication cycle time of Three industry Ethernet protocols.

Similarly, we get (8) from [16, 17]. They represent the communication cycle time for POWERLINK and PROFINET protocol, respectively:

$$T_{\text{Cycle1}} = T_{\text{STA}} + N_{\text{devices}} * \left(T_{\text{SQ}} + T_{\text{QS}} + 2 * T_{\text{frame}}\right) + T_{\text{master\_process}},$$
$$T_{\text{Cycle2}} = N_{\text{devices}} * \left(T_{\text{medium}} + T_{\text{pn\_fwd}} + T_{\text{frame}}\right) + T_{\text{master}_{\text{process}}}. \quad (8)$$

The other typical index values in (7) and (8) are given as follows:

$T_{\text{ecat\_fwd}}$: the forwarding delay time whose value is 2.7 μs;

T_{STA}: the duration of the start period with typical value 45 μs;

T_{SQ}: the time elapsed between the reception of a poll responses frame by the MN and the instant the Poll Requests frame is issued to the next CN and defined as 1 μs [16];

T_{QS}: the time employed by a CN to send a poll responses frame after the arrival of the Poll Requests from the MN with value 8 μs [16];

$T_{\text{pn\_fwd}}$: the forwarding delay time which value is 3 μs [17].

Figure 11 shows a diagram, whose compares the cycle time of EtherCAT, POWERLINK, and PROFINET as a function of the device amount. The diagram presents that the cycle time for EtherCAT protocol is smaller. And as the devices increase, the EtherCAT network cycle time increment is the least. When the devices increase from 1 to 10, the increment for EtherCAT protocol is only 28.75 μs. However, the incremental values are 265.32 μs and 121.2 μs for POWERLINK and PROFINET protocol, respectively.

4.2.2. Data Throughput and Time Synchronization Accuracy. For this kind of network motion control system of ocean wave maker, all of the wave making boards often need to be simultaneously impeled. In other words, it would be better to send all of control commands and data in one frame by master station. And then it is received and executed by all of the slave stations at the same time, which need network protocol with high throughput and time synchronization accuracy. The data exchange method of EtherCAT protocol, using the summation frame, can access a large amount of slave devices over one standard Ethernet frame. So the data throughput can be up to 90% [26]. Considering the minimum cycle time, the reference [28] gives the throughput comparison results between EtherCAT and POWERLINK. It is obvious that the EtherCAT protocol is superior to POWERLINK protocol.

Furthermore, time synchronization accuracy can be up to 15 ns for the EtherCAT. The measured value by oscillograph is given in [30] whereas time synchronization accuracy of the other two protocols reaches only 1 μs [31, 32].

4.3. Experiment Results. We have tested the constructed networked wave maker control system which consists of 10 slave stations in three aspects. We adopt the system making regular sinusoidal wave and irregular wave, respectively, in the experiment shallow water pool. A series of regular sinusoidal data and irregular wave data are sent by the upper master station application program. The regular wave is a sinusoidal periodic signal with period 1 second and 6 cm wave height. And the irregular wave's averaged wave height is 6 cm. The wave height is known as the variance between the maximum value and the minimum one of the wave height curve in wave theory.

Figure 12 depicts the typical wave sensor feedback signals collected by the master station from 1# sensor, 16# sensor, 65# sensor, 145# sensor, and 160# sensor. As we can see in Figure 6, these sensors are linked to the 1# slave station, 5# slave station, and 10# slave station, respectively. Similarly, we make the irregular wave and the results are shown in Figure 13.

The two diagrams of Figure 12 and Figure 13 show that the collected signal curves are smooth enough. Therefore, the system can meet making wave experiment needs.

Next, in order to validate the system communication performance, we have written program code to measure the cycle time. These codes implement the cycle time calculating and are inserted into the main application program. After the data frame with acquired data and servo control commands is sent by the master station, the timer starts until the data frame with the collected sensor data comes back to the master station. In order to illustrate clearly the advantage of EtherCAT, we tested the system with varying number of slave stations. The results are shown in Figure 14.

The results show that the system including a different number of slave stations has almost the same average communication cycle time. But the average communication cycle time concentrates on about 250 μs. The delay time is very small when the data frame passes through the slave station node. The results agree with the theoretical analysis in Section 4.2. The value can be omitted in our system.

Moreover, we analyzed 10 groups of cycle time values and calculated the average values.

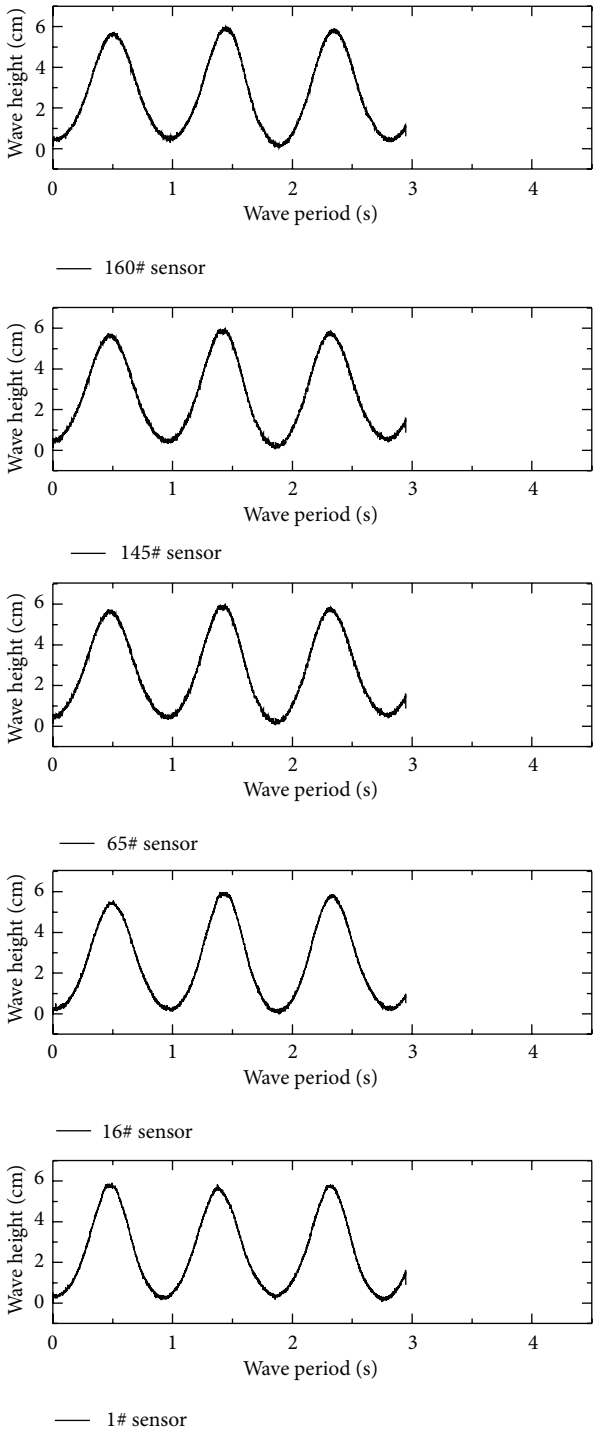


Figure 12: Regular wave height data signals from field sensors.

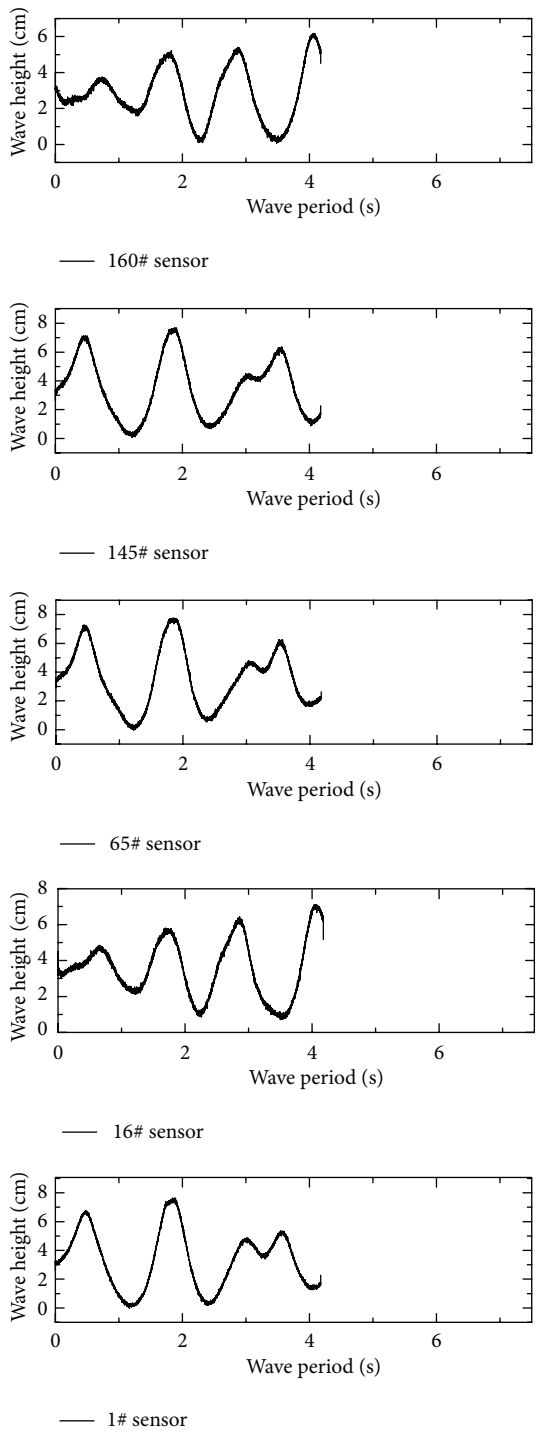


Figure 13: Irregular wave height data signals from field sensors.

These values are depicted in Figure 15, which indicate that the maximum delay jitter is only 3.42 microseconds. The communication system can meet high-speed motion control requirement.

5. Conclusions

The paper introduced the design of a typical networked motion control system based on EtherCAT protocol. This networked motion control system with large amount of measurement sensors asks for not only achieving real-time motion control but also transferring numerous sensor signals by less than 1 millisecond communication cycle time. So far, dozens of Industry Ethernet protocols have been released. Different from other Industry Ethernet protocols, EtherCAT protocol possesses high communication speed and efficiency

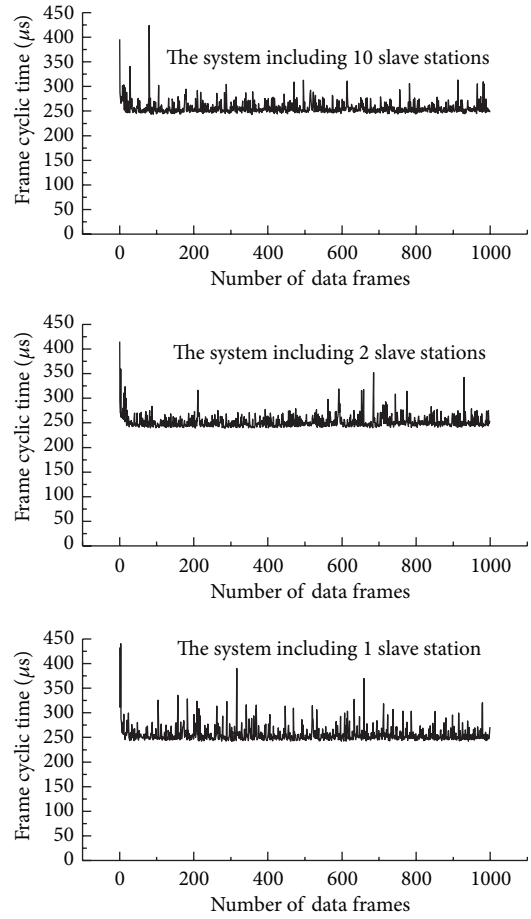


Figure 14: Cycle time of the system connected 1 slave station, 2 slave stations, and 10 slave stations, respectively.

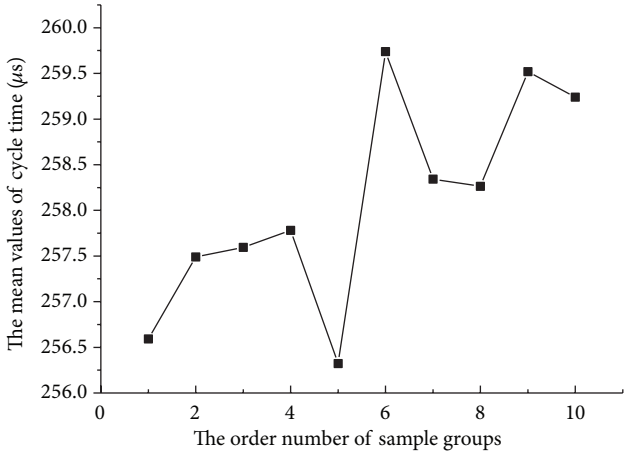


Figure 15: Average cycle time of system with 10 slave stations.

due to the specific communication principle. In particular, unparalleled hardware structure of slave interface controller makes the communication delay up to less than dozens of microseconds. Based on the test by control system platform, the experimental results show that EtherCAT protocol is the most suitable communication protocol for the type of networked motion control system mentioned. The results of this study would be helpful for the design of the networked motion control system with large-capacity data acquisition which requires high efficiency and high real-time.

Conflict of Interests

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

Acknowledgments

The paper is supported by Educational Commission of Henan Province of China (no. 2010A520020), National Natural Science Foundation of China (no. 51174263), Key Science and Technology Project of Henan Province (no. 112102210004), and Ph.D. Programs Foundation of Ministry of Education of China (20124116120004).

References

[1] K. C. Lee, S. Lee, and H. H. Lee, "Implementation and PID tuning of network-based control systems via Profibus polling network," *Computer Standards and Interfaces*, vol. 26, no. 3, pp. 229–240, 2004.

[2] N. P. Mahalik and A. N. Nambiar, "Trends in food packaging and manufacturing systems and technology," *Trends in Food Science and Technology*, vol. 21, no. 3, pp. 117–128, 2010.

[3] I. Erturk, "A new method for transferring CAN messages using wireless ATM," *Journal of Network and Computer Applications*, vol. 28, no. 1, pp. 45–56, 2005.

[4] A. Flammini, P. Ferrari, E. Sisinni, D. Marioli, and A. Taroni, "Sensor interface: from field-bus to ethernet and internet," *Sensors and Actuators A*, vol. 101, no. 1-2, pp. 194–202, 2002.

[5] A. Abarca, M. de la Fuente, J. M. Abril, A. García, and F. Pérez-Ocón, "Intelligent sensor for tracking and monitoring of blood temperature and hemoderivatives used for transfusions," *Sensors and Actuators A*, vol. 152, no. 2, pp. 241–247, 2009.

[6] A. Flammini, P. Ferrari, D. Marioli, E. Sisinni, and A. Taroni, "Wired and wireless sensor networks for industrial applications," *Microelectronics Journal*, vol. 40, no. 9, pp. 1322–1336, 2009.

[7] J. Jasperneite, M. Schumacher, and K. Weber, "Limits of increasing the performance of industrial ethernet protocols," in *Proceedings of the 12th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '07)*, pp. 17–24, Institute of Electrical and Electronics Engineers, Patras, Greece, September 2007.

[8] B. Tao, H. Ding, and Y. L. Xiong, "Design and implementation of an embedded IP sensor for distributed networking sensing," *Sensors and Actuators A*, vol. 119, no. 2, pp. 567–575, 2005.

[9] S. Pal and A. Rakshit, "Development of network capable smart transducer interface for traditional sensors and actuators," *Sensors and Actuators A*, vol. 112, no. 2-3, pp. 381–387, 2004.

[10] L. Bissi, P. Placidi, A. Scorzoni, I. Elmi, and S. Zampolli, "Environmental monitoring system compliant with the IEEE 1451 standard and featuring a simplified transducer interface," *Sensors and Actuators A*, vol. 137, no. 1, pp. 175–184, 2007.

[11] J.-D. Kim, J.-H. Lee, Y.-K. Ham, C.-H. Hong, B.-W. Min, and S.-G. Lee, "Sensor-Ball system based on IEEE 1451 for monitoring the condition of power transmission lines," *Sensors and Actuators A*, vol. 154, no. 1, pp. 157–168, 2009.

[12] R. Dorschner, "Digital servo drives and SERCOS simplify machine installation, optimize performance," *Control Solutions*, vol. 73, no. 10, pp. 50–54, 2000.

[13] "Ethernet-driven motion control goes on a real bender," *Process and Control Engineering*, vol. 59, pp. 21–23, 2006.

[14] H. Makete, "Real-time requirements for discrete time applications in automation systems," *Elektron*, vol. 22, no. 3, pp. 36–40, 2005.

[15] P. Ferrari, A. Flammini, and S. Vitturi, "Performance analysis of PROFINET networks," *Computer Standards and Interfaces*, vol. 28, no. 4, pp. 369–385, 2006.

[16] G. Cena, L. Seno, A. Valenzano, and S. Vitturi, "Performance analysis of Ethernet Powerlink networks for distributed control and automation systems," *Computer Standards and Interfaces*, vol. 31, no. 3, pp. 566–572, 2009.

[17] J. Jasperneite, M. Schumacher, and K. Weber, "Limits of increasing the performance of industrial ethernet protocols," in *Proceedings of the 12th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '07)*, pp. 17–24, September 2007.

[18] L. Seno and S. Vitturi, "A simulation study of Ethernet powerlink networks," in *Proceedings of the 12th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '07)*, pp. 740–743, September 2007.

[19] K. W. Song and G. S. Choi, "Fieldbus based distributed servo control using LonWorks/IP gateway/web servers," *Mechatronics*, vol. 20, no. 3, pp. 415–423, 2010.

[20] L. Wang, P. Orban, A. Cunningham, and S. Lang, "Remote real-time CNC machining for web-based manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 20, no. 6, pp. 563–571, 2004.

[21] M. Rostan, J. E. Stubbs, and D. Dzilno, "EtherCAT enabled advanced control architecture," in *Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference (ASMC '10)*, pp. 39–44, July 2010.

[22] B. A. GmbH, "Hardware Data Sheet ET1100—EtherCAT Slave Controller, Vel'. 1. 8," 2008, http://www.beckhoff.com.

[23] G. Prytz and J. Skaalvik, "Redundant and synchronized EtherCAT network," in *Proceedings of the 5th International Symposium on Industrial Embedded Systems (SIES '10)*, pp. 201–204, IEEE Computer Society, Trento, Italy, July 2010.

[24] M. Rehnman and T. Gentzell, "Synchronization in a force measurement system using EtherCAT," in *Proceedings of the 13th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '08)*, pp. 1023–1030, Institute of Electrical and Electronics Engineers, Hamburg, Germany, September 2008.

[25] G. Cena, I. C. Bertolotti, S. Scanzio, A. Valenzano, and C. Zunino, "On the accuracy of the distributed clock mechanism in EtherCAT," in *Proceedings of the IEEE International Workshop on Factory Communication Systems (WFCS '10)*, pp. 43–52, Institute of Electrical and Electronics Engineers, Nancy, France, May 2010.

[26] G. Cena, S. Scanzio, A. Valenzano, and C. Zunino, "Performance evaluation of the EtherCAT distributed clock algorithm," in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE '10)*, pp. 3398–3403, Institute of Electrical and Electronics Engineers, Bari, Italy, July 2010.

[27] J. C. Lee, S. J. Cho, Y. H. Jeon, and J. W. Jeon, "Dynamic drift compensation for the distributed clock in EtherCAT," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '09)*, pp. 1872–1876, IEEE Computer Society, Guilin, China, December 2009.

[28] L. Seno, S. Vitturi, and C. Zunino, "Real time ethernet networks evaluation using performance indicators," in *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation (ETFA '09)*, September 2009.

[29] L. Winkel, "Real-time ethernet in IEC 61784-2 and IEC 61158 series," in *Proceedings of the IEEE International Conference on Industrial Informatics (INDIN '06)*, pp. 246–250, Institute of Electrical and Electronics Engineers, Singapore, August 2006.

[30] EtherCAT Technology Group, http://www.ethercat.org/.

[31] "Real-Time PROFINET IRT," http://www.profibus.com/.

[32] Ethernet Powerlink Standardization Group, http://www.ethernet-powerlink.org/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 927192, 12 pages
http://dx.doi.org/10.1155/2014/927192

Research Article

Cooperative Transmission Mechanisms in Next Generation WiFi: IEEE 802.11ac

Baofeng Ji,[1] Kang Song,[2] Ying Hu,[2,3] and Hongjun Chen[1]

[1] *College of Information Engineering, Henan University of Science and Technology, Luoyang 471000, China*
[2] *School of Information Science and Engineering, Southeast University, Nanjing 210096, China*
[3] *Institute of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China*

Correspondence should be addressed to Baofeng Ji; baofengji@aim.com

Received 3 September 2013; Accepted 13 December 2013; Published 28 January 2014

Academic Editor: Deguang Le

Very high throughput (VHT) WLAN, known as IEEE 802.11ac, can provide compelling performance and has attracted extensive attention for achieving transmission rate over 1 G bps for 5 GHz band owing to involving the MU-MIMO and maximum 160 MHz bandwidth transmission. Despite extensive studies and dramatic performance, the conventional carrier sensing mechanism emerges some drawbacks especially in the overlapping BSS scenario responsible for employment of MU-MIMO and bandwidth expansion. In order to address the issue of conventional carrier sensing mechanism, namely, redundant or useless network allocation vector (NAV) setting, this paper proposes an enhanced NAV transmission mechanism, which not only needs little modifications to current standard draft but also can achieve performance improvement significantly. On this basis, in order to solve the SINR inaccuracy calculated based on null data packet (NDP) in the actual MU-MIMO transmission, the SINR feedback mechanism is proposed and the frame structure modifications are displayed in detail. Furthermore, theoretical analysis is also performed on the proposed mechanism and the formulas of the achieved gains are also derived. Numerical results confirm the validation of our theoretical analysis and further substantiate that the proposed scheme obtains obvious throughput gain over the conventional mechanism.

1. Introduction

With the development of modern broadband communication systems, there is growing evidence that the timeless and convenience of the information acquisition become more and more important to user experience. It has gained popularity recently due to its easy and quick deployment with low cost. In wireless local area network (WLANs), the wireless channel is shared by all the nodes and hence a medium access control (MAC) protocol is needed to coordinate their transmissions to reduce the collision. Although it was initially standardized for WLANs, IEEE 802.11DCF (Distributed Coordination Function) was known as carrier sense multiple access with collision avoidance (CSMA/CA), with an optional use of RTS/CTS [1]. However, the employment of multiple user multiple input multiple output (MU-MIMO) and bandwidth expansion mechanism makes the existing mechanism of IEEE 802.11 emerge some drawbacks. For instance, when multiple stations are simultaneously active and have different packet lengths to send, the network allocation vector (NAV) should be set as the longest duration to avoid interference; "NAV" is the time that the neighboring node will finish its ongoing transmission. But for the station in the neighboring BSS which only interferes with part of the active users, it is very possible that its NAV setting is redundant; that is, thus, even those transmissions that will not interfere with the ongoing one are still blocked. Therefore, how to address these issues and improve the performance of VHT WLANs is of particular importance. Motivated by this, the paper aims to find solutions to these problems in the framework of TXOP sharing mechanism adopted in the current standard [2]. The proposed scheme could solve these drawbacks with some subtle frame structure modifications to WLAN IEEE 802.11ac draft.

It is worth noting that the fundamental access method of the IEEE 802.11 MAC is a DCF known as carrier sense multiple access with collision avoidance (CSMA/CA) with an option of RTS/CTS, which is a little similar to carrier

sense multiple access with collision detection (CSMA/CD). The four-way handshake procedure (RTS/CTS/DATA/ACK), which is used to deal with the hidden terminal problem, is as follows. Before a node begins to transmit, it should first sense the channel to determine whether there is any ongoing transmission. If the channel is busy, the node will defer until the channel is sensed idle for a period of DIFS. Then the node randomly chooses a backoff period according to the contention window and starts a back-off timer to backoff. The back-off timer decreases by 1 after the channel is idle for the duration of a slot. If the channel is sensed busy during any slot in the back-off interval, the back-off timer will be suspended. It can be resumed only after the channel is idle for a period of DIFS again. After the back-off timer reduces to 0, the sender transmits a RTS omnidirectionally. This transmission ends after the receiver correctly receives the data and responses with an ACK. All four kinds of frames contain an estimated duration of the rest time of the transmission. Other nodes that receive these frames update their NAVs (network allocation vectors) with the duration. Every NAV decreases by 1 after a time slot. Those nodes are only allowed to transmit after they sense the channel idle for a period of DIFS after their NAVs expire [3].

Another kind of access method of IEEE 802.11 MAC is a EDCA (enhanced distributed channel acess). With EDCA, high-priority traffic has a higher chance of being sent than low-priority traffic: a station with high-priority traffic waits a little less before it sends its packet, on average, than a station with low-priority traffic. This is accomplished by using a shorter contention window (CW) and shorter arbitration interframe space (AIFS) for higher priority packets. The exact values depend on the physical layer that is used to transmit the data. In addition, EDCA provides contention-free access to the channel for a period called a transmit opportunity (TXOP). A TXOP is a bounded time interval during which a station can send as many frames as possible (as long as the duration of the transmissions does not extend beyond the maximum duration of the TXOP). If a frame is too large to be transmitted in a single TXOP, it should be fragmented into smaller frames. The use of TXOPs reduces the problem of low rate stations gaining an inordinate amount of channel time in the legacy 802.11 DCF MAC. A TXOP time interval of 0 means it is limited to a single MAC service data unit (MSDU) or MAC management protocol data unit (MMPDU) [2].

Though WLAN has been extensively studied, some parts of them still focus on the conventional DCF mechanism of IEEE 802.11. The performance analysis of DCF mechanism of 802.11 was firstly researched by Bianchi in [4] on JSAC 2000, which described the access protocol of CSMA using bidimensional chains of Markov and analyzed the throughput performance of IEEE 802.11a. The theoretical upper bound of IEEE 802.11a capacity was derived by Cali et al. in [5]. Reference [6] investigated the MU-MIMO transmission schemes based on WLAN 802.11n, which considered an opportunistic channel-aware scheduling policy to achieve simultaneous downlink transmission to multiple users and one of the drawbacks of this scheme proposed was random beamforming technique at the physical layer. Reference [7] gave some research about the MU-MIMO transmission under

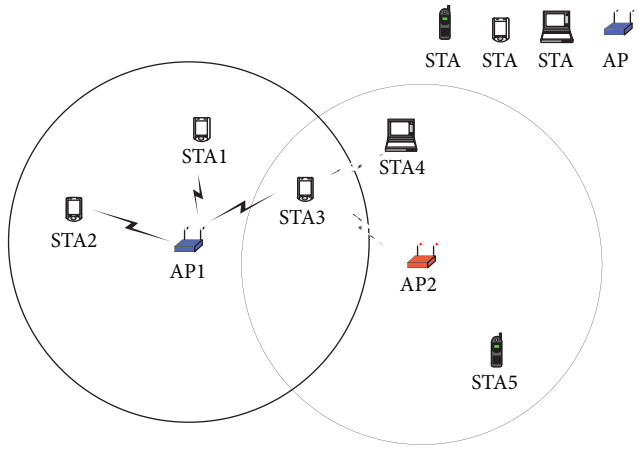


Figure 1: IEEE 802.11ac overlapping BSS scenario.

IEEE 802.11n. The contention window of WLAN IEEE 802.11 was also studied based on DCF to maximize throughput by Weng et al. [8] and Babich and Comisso [9]. References [10–12] gave a network allocation vector analysis for IEEE 802.11 based on DCF mechanism.

On the other hand, there are also some papers focusing on EDCA (enhanced distributed channel access) mechanism, which is an enhanced mechanism for DCF. Reference [13] presented the throughput performance of EDCA and compared the case of full NAV setting and packet by packet NAV setting with and without NAV clear mechanism. The results showed that EDCA in the case of packet by packet NAV setting outperforms that in the case of full NAV setting and EDCA with NAV clear mechanism outperforms that without NAV clear mechanism. Reference [14] presented a NAV scheme to improve the system throughput, the proposed scheme displayed considerable throughput gains. The author in [15] proposed and investigated a scheme that modified network allocation vector update process of voice terminals based on EDCA. Mario and coauthors proposed a priority based CSMA/CA mechanism to support deadline aware scheduling in home automation applications using IEEE 802.15.4 [16]. The authors in [17] also presented an adaptive energy-management framework for sensor nodes with constrained energy scavenging profiles; the scheme proposed boosts the system throughput to some extent.

Up to now, to the best of our knowledge, the research on the performance analysis of TXOP sharing with MU-MIMO and solving problems as similar as Figure 1 is still an open problem. For example, as shown in Figure 1, when multiple stations are simultaneously active and have different packet lengths to send, the NAV should be set as the longest duration to avoid interference; "NAV" is the time that the neighboring node will finish its ongoing transmission. But for the station in the neighboring BSS which only interferes with part of the active users, it is very possible that its NAV settings is redundant; that is, the duration in the NAV is longer than the actual interfering duration. More seriously, if the station in the overlapping area responds to the MU-MIMO polling but finally fails to obtain the transmission opportunity, the overlapping nonserving BSS will be set NAV mistakenly. This

will result in some performance loss. Motivated by this, we will propose a two-level NAV mechanism to address this issue after elucidating that the problem is how to form and further provide theoretical analysis on its achieved performance.

2. Problem Formulation

It is known that the available channel in IEEE 802.11ac could only support up to two 160 MHz or five 80 MHz bandwidths simultaneously. Furthermore, the introduction of MU-MIMO in IEEE 802.11ac will result in inefficient use of the limited frequency resource with the conventional enhanced distributed channel access (EDCA) mechanism, becoming of redundant or useless NAV setting. The problem will be elaborated as follows.

As shown in Figure 1, the two-BSS overlapping network considered in this paper is as follows: the two BSSs involve different stations showed in Figure 1, respectively. It is worth noting that the STA3 locates in the overlapping BSS (it is the so-called OBSS scenario) position and associates with AP1. According to IEEE 802.11ac standard, AP1 may transmit RTS to poll STA1, STA2, and STA3 in order to realize MU-MIMO and decide how wide the bandwidth to use. It is critical that the STA3 may be polled by AP1 and respond to the polling with CTS sending to AP1, the action of which will lead to STA4 and AP2 set NAV to avoid interference. However, the bandwidth and duration in NAV are determined such that each STA in the following MU-MIMO transmission is protected according to the TXOP mechanism in EDCA. It is likely that this makes the NAV setting for STA4 and AP2 become redundant or useless. There are two typical scenarios for the issue of mistaken or redundant NAV setting for the overlapping BSS.

For the first case, two neighboring BSSs have different primary channels; namely, the primary channel of BSS2 overlaps the secondary channel of BSS1, which is a frequent scenario for several BSSs usually. In this case, STA3 will suffer interference from different BSSs due to its location in the overlapping area (the more the BSS number increases, the stronger interference OBSS station received); it is most likely that it cannot satisfy the condition required by AP1 and will be excluded in the list of the following MU-MIMO transmission such as the unequal bandwidth or other algorithms for achieving an objective. If this happens, the NAV setting for STA4 and AP2 becomes mistaken due to the polling response of STA3 and cannot be cleared based on conventional mechanism, resulting in noticeable throughput loss of the VHT WLANs.

For the second case, the scenario of adjacent BSS has the same primary channel; namely, the primary channel of BSS2 overlaps the primary channel of BSS1, which occurs regularly in practical generally. Nevertheless, different STAs may have different lengths of packets to transmit in practical, then AP1 can clear the redundant TXOP duration through CF-END frame when STA3 finishes the transmission; however, STA4 and AP2 cannot receive the CF-END frame (CF-END is a no response frame), resulting in that the BSS2 is out of work in the redundant TXOP duration and decreasing the throughput of VHT WLANs inevitably.

Furthermore, the accuracy of the SINR feedback in Downlink MU-MIMO of IEEE 802.11ac plays a key role in improving the performance of MU-MIMO transmission. The SINR calculated based on NDP is usually far from accuracy, since at that moment (in the sounding process) the possible interuser interference in the actual MU-MIMO transmission cannot be estimated effectively. Generally speaking, there are several continuous data frames for each STA in a TXOP; it is therefore possible to correct the SINR estimation during the process of data transmission via BA frame.

Focusing on these cases above, we will propose solutions to solve the problems above in the next section.

3. Proposed Solutions

3.1. Enhanced NAV Mechanism. Here we would like to elaborate a two-level NAV mechanism to solve the above-mentioned two redundant NAV setting cases. Though the description of the proposed schemes may seem to only address the special cases shown in Figure 1, the extension to a more general case is straightforward.

Case 1. Firstly, AP1 obtains the available channel use for MU-MIMO implementation and may transmit RTS/data to poll the stations to be served. STA3 may respond with CTS/BA; then AP2 and STA4 in BSS2 will set NAV to enter power saving state. AP1 can decide transmit bandwidth after receiving all the response frame and AP1 also needs to calculate which bandwidth can maximize the throughput (different algorithms may be adopted and this is not the focus in this paper). If AP1 finally decides that STA3 will not be served for unequal bandwidth or other reasons, then AP2 and STA4 may still be mistakenly in NAV setting. We propose the following two steps to address this issue.

Secondly, AP2 and STA4 in BSS2 need to wait a timeout duration (such as DIFS, SIFS) to make sure that they are in this mistaken NAV duration case.

Thirdly, to clear the mistaken NAV setting while guaranteed no interference on others, AP2 and STA4 need to check two conditions to clear NAV setting after receiving the final frame announcement. One is the timeout; the other is as shown in Figure 2; the NAV duration to be cleared should be in the range between the longest duration and the second longest duration.

Case 2. Firstly, similar to the first case, AP2 and STA4 in BSS2 will set NAV after overhearing CTS/BA frame. If the packet length of STA3 is much shorter than the TXOP duration, at the end of STA3's packet transmission, the redundant NAV duration for AP2 and STA4 actually can be used without causing any interference. But according to conventional mechanism, AP2 and STA4 have to wait until the TXOP duration is ended. We propose the following solution to address this issue. In our solution the final BAR frame needs to carry the information of the last frame, and the response BA frame of STA3 also needs to carry information to announce AP2 and STA4. The last frame announcement can be obtained through two methods. One is via "more data" field indication; namely, "more data" equal to 1 represents

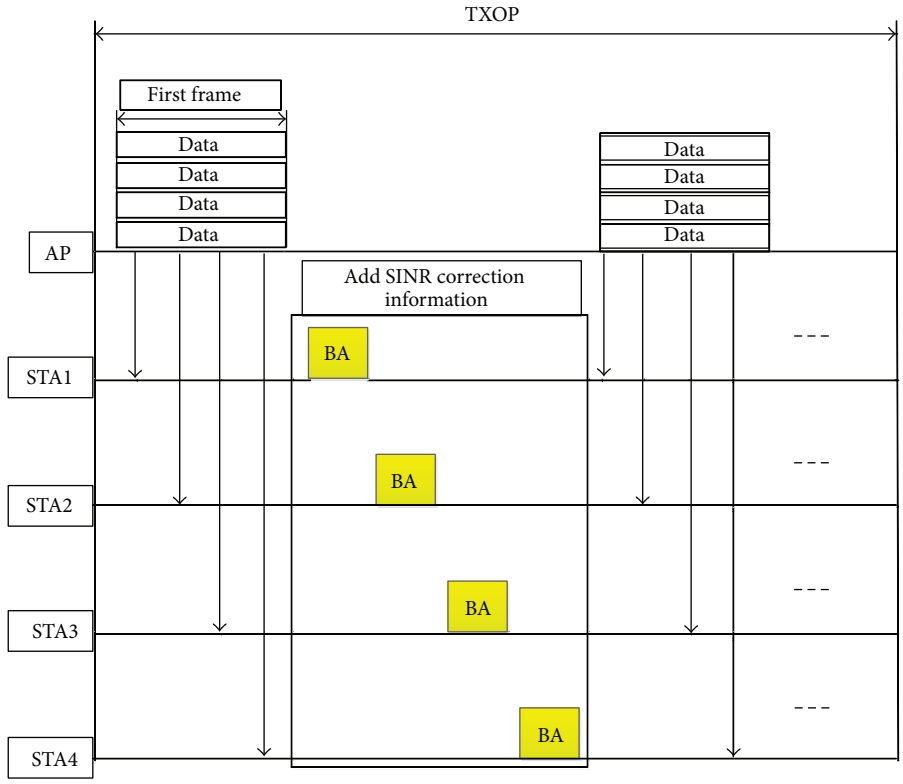


Figure 2: BA frame correction process.

the last frame announcement, otherwise not the last frame; another can be realized through BAR control field, which has 9 reserved bits to be used and any of these bits available equal to 1 indicates the last frame announcement, otherwise not the last frame.

Secondly, STA3 needs to inform the last frame announcement to its surroundings so as to clear useless NAV settings, which can be implemented through reserved bits in control field of BA frame; namely, a reserved bit in BA control field of STA3 equal to 1 represents the last frame announcement, otherwise not the last frame.

Thirdly, AP2 and STA4 check two conditions to see whether their NAV setting can be cleared or not, which is the same as that in the first case. Based on that, AP2 and STA4 can clear the possible useless NAV after waiting SIFS without causing interference to other STAs.

3.2. SINR Correctness Feedback Mechanism. The accuracy of the SINR feedback in downlink MU-MIMO of IEEE 802.11ac plays a key role in improving the performance of MU-MIMO. The SINR calculated based on NDP is usually far from accuracy, since at that moment (in the sounding process) the possible interuser interference in the actual MU-MIMO transmission cannot be estimated. In general there are several continuous data frames for each STA in a TXOP; it is therefore possible to correct the SINR estimation during the process of data transmission via BA frame, as illustrated in Figure 2.

Therefore, we propose the SINR correctness feedback mechanism in order to satisfy the channel state information (CSI) accuracy requirement for MU-MIMO. Namely, the transmitter can update the precoding matrix adaptively through the SINR feedback. Especially, the secondary access category (AC) users can join the MU-MIMO transmission when the TXOP is not expired; then, the CSI of the secondary AC users are not accurate enough, which can have great effect on the MU-MIMO transmission. So we can give a SINR correctness feedback mechanism to solve the problem above-mentioned and the steps of correcting the estimated SINR during TXOP can be seen as follows.

(i) AP transmits the first data frame in an MU-MIMO TXOP using an MCS lower than that indicated by the feedback SINR during sounding.

(ii) STA reestimates the SINR based on the first data frame, in which the interuser interference can be incorporated.

(iii) Based on the SINR reestimation, STA sends the SINR correction information (usually defined as a difference between the originally estimated and the reestimated SINR) to AP through the subsequent BA frame, as illustrated in Figures 3 and 4.

(iv) On receiving the SINR correction information, AP transmits the remaining data frames in this TXOP using a corrected MCS.

In order to trigger a STA to reestimate the SINR in the process of data transmission in a TXOP, an indication is needed in the data frame. There are two ways:

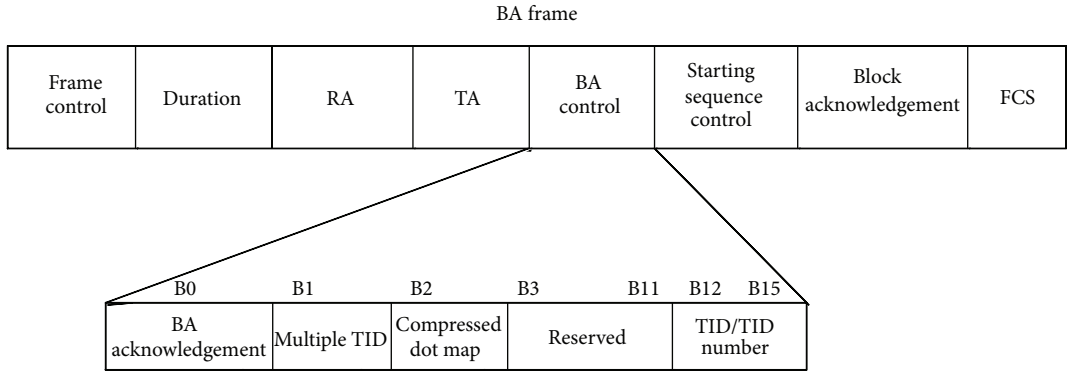


Figure 3: BA frame modification Scheme 1.

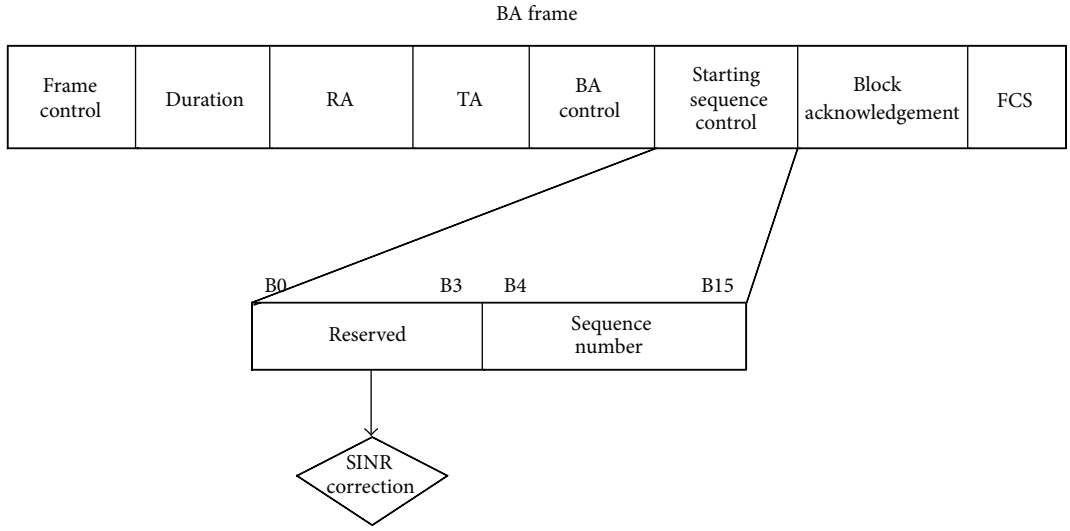


Figure 4: BA frame modification Scheme 2.

(i) using one reserved bit in the VHT-SIGA field to explicitly indicate this;

(ii) alternatively, using scrambling to implicitly indicate this, that is, using a particular scrambling initialization in the service field to trigger the SNR reestimation.

Then STA reobtain the SINR per space stream per subcarrier and can calculate the average value; namely, STA sums the SINR per stream per subcarrier and then subtracts the stream number to obtain the total average SINR, which can be compared with the former feedback average SINR to obtain the SINR correction. The SINR correction can be quantized to four bits and feedback to the transmitter through BA frame. The four bits (B3 to B6) can represent this correction. Due to each STA at most can receive four space streams, so the group can feedback four average SINR correction and needs eight bits (B3 to B10) to indicate.

B3 and B4 indicate the first space stream, B5 and B6 indicate the second space stream, B7 and B8 indicate the third space stream, and B9 and B10 indicate the fourth space stream. The main process can be seen in Figure 2. Firstly, the transmitter can send the first MU-MIMO frame in TXOP duration with low modulation and coding scheme (MCS); secondly, the STA can recalculate the SINR after receiving the first MU-MIMO frame; thirdly, the STA feedback the SINR correction to the transmitter through BA frame; fourthly, the transmitter can update the precoding matrix adaptively and give better power allocation after overhearing the BA frame with SINR correction to maximize the system throughput.

The advantages of this scheme mainly conclude that the system does not need the control package frame; this not only reduces the complexity; but also does not increase frame exchange, which makes the transmitter obtain the more accurate SINR and thus improves the MU-MIMO performance. Certainly, the scheme proposed occupies four reserved bits in order to implement the enhancement of the MU-MIMO transmission.

4. Performance Analysis

4.1. The Performance of Enhanced NAV Mechanism. In this section, we analyze the performance of the proposed enhanced NAV mechanism.

It is assumed that the service time X of one frame transmission completed in a TXOP follows exponential

distribution with parameter λ; then m frames transmission completion are independent identically distributed and can be obtained $Y = X_1 + \cdots + X_m$ which follows m-Erlang distribution [18]; namely,

$$f_y(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{m-1}}{(m-1)!}, \quad y \geq 0. \tag{1}$$

It is assumed that the business arrival Z follows the Poisson distribution with parameter $\widetilde{\lambda}$ as (4):

$$\text{Pois}\left(\widetilde{\lambda}\right) = \frac{\widetilde{\lambda}^k}{k!}e^{-\widetilde{\lambda}}, \quad k \in 0, 1, 2, 3, \ldots. \tag{2}$$

Based on the above model, then, the average length of the packet payload can be calculated through computing the expectation of distribution as follows:

$$E[y] = \int_0^\infty y f(y)\, dy = \frac{m}{\lambda}. \tag{3}$$

Let P_u represent the user's packet payload; then the average packet length EP_u is constant and can be expressed as $EP_u = \lambda_s = m_u/\lambda$; let P_u^{Possion} note the uth user business arrival probability, which follows Possion distribution. TXOP is called transmit opportunity, which is a duration and stations can be won through contention; however, different stations have different data volumes and then different stations have different redundant times $T_{\text{redundant}_t}$ in a TXOP duration.

The redundant time $T_{\text{redundant}_t}$ in a TXOP duration when the transmission mode is SU-MIMO and frame numbers are m can be written as

$$T_{\text{redundant}_t} = \text{TXOP} - \Delta - Y, \tag{4}$$

where the $\Delta = \text{RTS}/s_{\text{Data}} + \text{SIFS} + \delta + \text{CTS/ACK} + \text{SIFS} + \delta$ because every transmission starting needs TXOP initialization and needs RTS/CTS or short data polling. δ is transmission delay time.

Then we give the throughput gains G_a expression as

$$G_a = \frac{\sum_{u=1}^{\overline{\mathcal{N}}} E[P_u] P_u^{\text{Possion}}}{T_{\text{redundant}_t}}, \tag{5}$$

where $\overline{\mathcal{N}}$ is user number, which refers to users released when OBSS stations finish frame transmission as the proposed solution procedure and can be calculated using campel theorem as [19, 20].

We have elaborated the redundant time $T_{\text{redundant}_t}$ in a TXOP duration when transmission mode is SU-MIMO and frame number is m, however, MU-MIMO transmission mode has some difference compared with SU-MIMO mode. It is assumed that a station service time of m frames interaction in SU-MIMO mode follows m-Erlang distribution as the network model, nevertheless, MU-MIMO transmission should need to wait some extra times to complete interaction when each adding a station, the extra time can be expressed as extra_time = SIFS + δ + BAR + δ + SIFS + δ + BA + δ, we can assume that the MU-MIMO transmission has G stations, then the redundant time in MU-MIMO transmission mode can be expressed as $T_{\text{MU\_redun}} = \text{TXOP} - \Delta - \text{extra\_time} \cdot \sum_{u=2}^{G} \lambda_u - X$, where X expresses all the transmission frames in MU-MIMO mode.

Then we will give some propositions to evaluate the performance of the proposed scheme.

Firstly, we can calculate the distribution of the redundant time as Proposition 1.

Proposition 1. *The distribution about the redundant time $T_{redundant_t}$ in a TXOP duration when the transmission mode is SU-MIMO and frame number is m can be obtained in the following:*

$$f_{SU\_redundant}(x) = \frac{\lambda e^{-\lambda(t-\Delta x)}[\lambda(t-\Delta-x)]^{m-1}}{(m-1)!}. \tag{6}$$

Then on the basis of Proposition 1, we can obtain the distribution of SU-MIMO throughput gains using the proposed enhanced NAV scheme as Proposition 2.

Proposition 2. *The distribution of the throughput gains using proposed enhanced NAV scheme can be obtained as follows when the transmission mode is SU-MIMO:*

$$\begin{aligned} f_{SU\_gains}(x) = {} & \sum_{k=0}^{\lfloor x(t-\delta)\rfloor} \frac{\lambda e^{-\lambda(t-\Delta)}}{(m-1)!} e^{-\lambda\overline{\mathcal{N}}} \\ & \times \frac{\lambda_{\overline{\mathcal{N}}}^k}{(k-1)!}\frac{k}{x} e^{(\lambda\lambda_s k)/x} \\ & \times \left[\lambda\left(t-\Delta-\frac{\lambda_s k}{x}\right)\right]^{m-1}, \end{aligned} \tag{7}$$

where $\lambda_{\overline{\mathcal{N}}} = \sum_{u=1}^{\overline{\mathcal{N}}} \lambda_u$.

The proof of Proposition 2 can be seen in Appendix A.

Furthermore, we can give the expectation of throughput gains in SU-MIMO mode as Corollary 1.

Corollary 1. *The expectation of throughput gains in SU mode can be obtained as*

$$\begin{aligned} & E_{SU\_gains} \\ & = \frac{\lambda_s \lambda_{\overline{\mathcal{N}}}}{(m-1)!} \sum_{k=1}^{\infty} \left(\frac{1}{t-\Delta}\right)^k \left(\frac{1}{\lambda}\right)^{k-1} \gamma(m+k+1, \lambda(t-\Delta)), \end{aligned} \tag{8}$$

where $\gamma(\cdot,\cdot)$ is an incomplete gamma function [21].

The proof of Corollary 1 can be seen in Appendix B.

Based on this section's beginning, the distribution of the redundant time in MU-MIMO mode can be obtained as Proposition 3.

Proposition 3. *The distribution of the redundant time T_{MU_redun} in a TXOP duration when transmission mode is*

MU-MIMO and frame number is m can be obtained in the following:

$$f_{MU\_redun}(x) = e^{-\lambda_G}\frac{\lambda^m}{(m-1)!}\sum_{k=0}^{\lfloor(t-\Delta-x)\rfloor}\frac{\lambda_G^k}{k!} \times e^{-\lambda(t-\Delta-x-k)} \times (t-\Delta-x-k)^{m-1}, \tag{9}$$

where $\lambda_G = con\_t\sum_{u=2}^{G}\lambda_u$.

The proof of Proposition 3 can be seen in Appendix C.

Similar to the SU-MIMO mode, we can also calculate the distribution of throughput gains using the proposed scheme in MU-MIMO mode as Proposition 4.

Proposition 4. *The distribution of throughput gains using the proposed scheme can be obtained as follows when transmission mode is MU-MIMO:*

$$f_{MU\_gains}(x) = e^{-\lambda_{\overline{\mathcal{M}}}}e^{-\lambda_G}\frac{\lambda^m}{(m-1)!} \times \sum_{l=0}^{\lfloor x(t-\delta)\rfloor}\sum_{k=0}^{\lfloor(t-\delta-l/x)\rfloor}\frac{\lambda_{\overline{\mathcal{M}}}^l}{l!} \times \frac{l}{x}\frac{\lambda_G^k}{k!}e^{-\lambda(t-\Delta-(l\lambda_s/x)-k)} \times \left(t-\Delta-\frac{l\lambda_s}{x}-k\right)^{m-1}. \tag{10}$$

The calculating method of Proposition 4 is similar to Proposition 3.

Furthermore, the expectation of the throughput gains in MU-MIMO can be obtained as Corollary 2.

Corollary 2. *The expectation of the throughput gains in MU-MIMO transmission mode can be calculated as follows:*

$$E_{MU\_gains} = e^{-\lambda_G}\frac{\lambda^m}{(m-1)!}\sum_{k=0}^{\lfloor t-\Delta\rfloor}\frac{\lambda_G^k}{k!}(I_1 - I_2), \tag{11}$$

where $I_1 = (-1)^{m-2}k^{m-1}e^{\lambda k}Ei(-\lambda k) + \sum_{i=1}^{m-1}(i-1)!(-k)^{m-1-i}\lambda^{(-i)}$ *and* $I_2 = e^{\lambda k}\sum_{j=0}^{m-1}C_{m-1}^{j}(t-\delta-k)^j(t-\Delta)^{m-1-j}\Gamma(m-j)\Gamma(-(m-1-j), \lambda(t-\Delta))$.

The proof of Corollary 2 can be seen in Appendix D.

It is noted that the redundant time $T_{\text{redundant}_t}$ may be the whole TXOP duration when the primary channel of one BSS overlaps the secondary channel of another BSS, namely, Case 1.

4.2. The Performance Analysis of the MU-MIMO in IEEE 802.11ac. In this section, we will analyze the MU-MIMO performance in IEEE 802.11ac. Assume that $\overline{X}$ is the average transmission frame length and $\overline{T}$ is the average transmission time. If P_{AP} denotes the probability of AP contending the channel right use, then the throughput can be expressed as

$$S(m, N, r_\gamma) = P_{\text{AP}}\cdot\frac{\overline{X}(m, N, r_\gamma)}{\overline{T}_{m,N,r_\gamma}}, \tag{12}$$

where $\overline{X}(m, N, r_\gamma)$ is the average transmission frame length for N users under m time slots with transmission threshold r_γ and $\overline{T}(m, N, r_\gamma)$ is the average transmission time for N users under m time slots with transmission threshold r_γ, which can be expressed as

$$\overline{X}(m, N, r_\gamma) = \sum_{i=1}^{n_t} i\cdot l\cdot\left(\sum_{\omega=\gamma}^{R}P_f(i, m, N, r_\gamma, r_\omega)\right),$$
$$\overline{T}(m, N, r_\gamma) = T(i=0, m)\cdot P_f(i=0, m, N, r_\gamma) + \sum_{i=0}^{n_t}\sum_{\omega}^{R}T(i, m, r_\omega)\cdot P_f(i, m, N, r_\gamma, r_\omega), \tag{13}$$

where index i represents the different frame types; $i = 0$ represents the null frame without transmission; $i = 1$ represents the frame involving one beam and length is l bits; $i = 2$ represents the frame involving two beams and length is $2\cdot l$; $i = 3$ represents the frame involving three beams and length is $3\cdot l$; $i = 4$ represents the frame involving four beams and length is $4\cdot l$. Then the average frame length can be calculated through different frame types with probability. Let $P_f(i, m, N, r_\gamma, r_\omega)$ denote the transmission probability of i frame for N users under m time slots with threshold r_γ and transmission rate r_ω.

If the probability P_{AP} of AP contending the channel right use can be expressed as [22]

$$P_{\text{AP}} = \frac{m!n!}{m^n}\sum_{j=1}^{\min(m,n)}\frac{(-1)^{j+1}(m-j)^{n-j}}{(j-1)!(m-j)!(n-j)!}. \tag{14}$$

Due to the users in MU-MIMO group needing not to contend the channel, then one user can be located in different groups; therefore, the AP needs to select users in downlink MU-MIMO. If the SINR threshold is r_γ, then the probability for selecting n users from N and SNR $> r_\gamma$ can be given by

$$P_{\text{select}}(n, r_\gamma) = C_N^n(1-F(\gamma))^n(F(\gamma))^{N-n}. \tag{15}$$

Then the $P_f(i, m, N, r_\gamma, r_\omega)$ can be written as

$$P_f(i, m, N, r_\gamma, r_\omega) = \sum_{n=1}^{N}\left(P_{\text{select}}(n, r_\gamma)\cdot P_{ri}(r_\omega, i)\right), \tag{16}$$

where $P_{ri}(r_{\omega,i})$ is the probability for users with i beams with available rate r_ω, which is owing to all the beams and where

one user can have only one MCS in IEEE 802.11ac and can be given by

$$\begin{aligned} & P_{ri}(r_\omega, i) \\ & = P_r\{(\text{at least 1 beam with } r_\omega) \cap (\text{no beam with } r > r_\omega)\} \\ & = P_r\{(\text{at least 1 beam with } r_\omega) \mid (\text{no beam with } r > r_\omega)\} \\ & \quad \cdot P_r\{\text{no beam with } r > r_\omega\} \\ & = (1 - P_r\{1 - P_r\{\text{all beams with } r < r_\omega\}\}) \\ & \quad \cdot P_r\{\text{all beams with } r < r_\omega\} \\ & = \left(\left(1 - \left[\frac{\sum_{v=\gamma}^{\omega-1} P_r(r_v)}{\sum_{v=\gamma}^{\omega} P_r(r_v)}\right]^i\right)\left[\frac{\sum_{v=\gamma}^{\omega} P_r(r_v)}{\sum_{v=\gamma}^{R} P_r(r_v)}\right]^i\right), \end{aligned} \tag{17}$$

where $P_r(r_v)$ is the probability with transmission rate less than r_ω.

Then, we can obtain the $\overline{X}(m, N, r_\gamma)$ and the $T(i, m, r_\omega)$ can be calculated as

$$T(i, m, r_\omega) = T_{\text{data}}(i, r_\omega) + T_{\text{overhead}}(i, m), \tag{18}$$

where $T_{\text{data}(i,r_\omega)}$ denotes the transmission time for i frame with transmission rate r_ω and $T_{\text{overhead}}(i, m)$ is the frame control overhead (such as RTS, CTS) for i frame under m time slots.

5. Simulation

In this section, we will verify the performance gains and analysis of the scheme proposed through simulations.

We first show the probability density function (PDF) of the redundant time in SU mode compared with simulation and analysis in different service frame numbers in Figure 5; then it can be observed that the measurement shows good agreement between the theoretical analysis and experimental results. Meanwhile, it is also illustrated that the redundant time in TXOP reduces gradually with the transmission frames increasing.

We then compare the theoretical analysis and experimental results of the redundant time PDF in MU mode with different MU-MIMO transmission station number and service frame number in Figure 6; it can be observed that the redundant time in TXOP will reduce with the service frame number increasing as well as SU mode. Furthermore, the station number in group also has greater effect on the redundant time in TXOP. The reason is that the more the station number in group, the longer the transmission time; then the redundant time in TXOP may decrease more obviously.

As we can see from the theoretical and numerical results of throughput gains in SU mode in Figure 7, we can know that the measurement shows good agreement between the theoretical analysis and experimental results with different released station number and frame number variation. The results further illustrate that the throughput gains will boost considerably especially with the released station number increasing and no matter how many antennas, the calculation

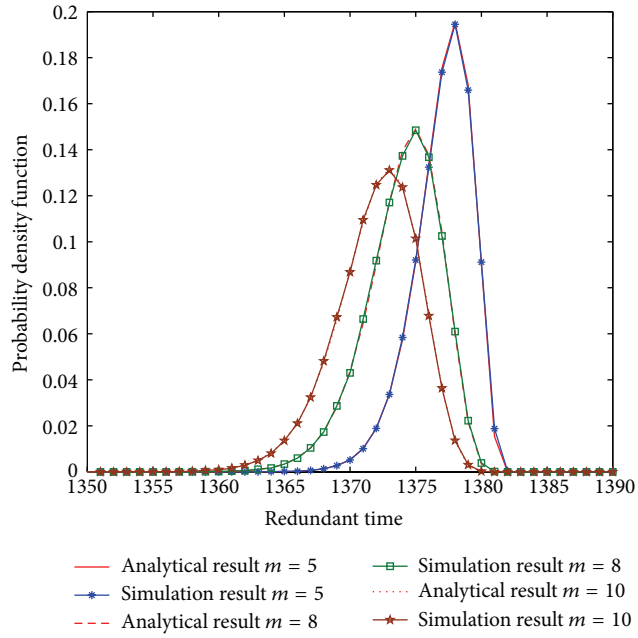


Figure 5: The PDF of the redundant time in SU mode.

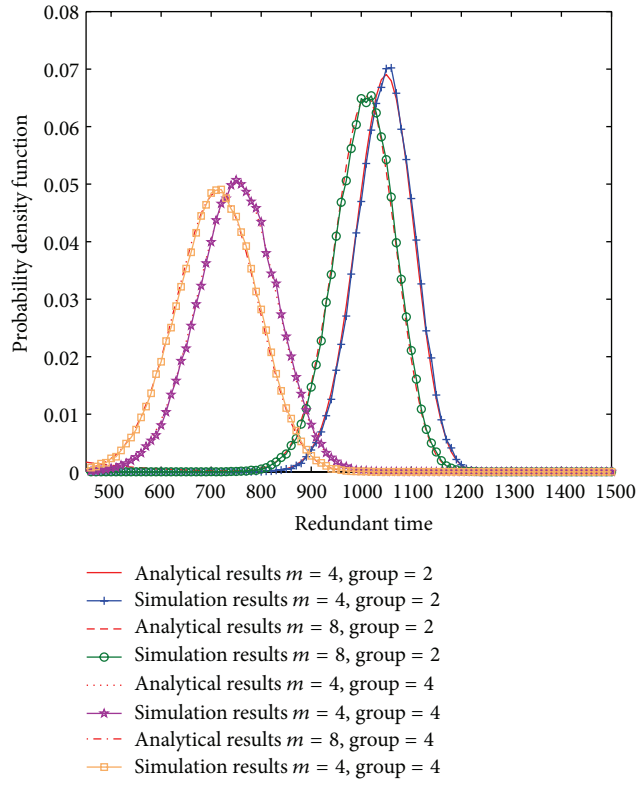


Figure 6: The PDF of the redundant time in MU mode.

of the released station number can be carried out by Campel theorem [19].

Figure 8 demonstrates that the proposed two-level network allocation vector scheme can be vastly superior to the conventional mechanism with the redundant time increasing, which verifies the superiority and correctness of the proposed scheme. The growing gap in throughput outcomes

Table 1: The threshold r_γ and available rates r_ω.

Index	Available rates (Mbps)	Threshold γ (dB)
1	0	$\gamma \le 8$
2	6	$8 < \gamma \le 12.5$
3	9	$12.5 < \gamma \le 14$
4	12	$14 < \gamma \le 16.5$
5	18	$16.5 < \gamma \le 19$
6	24	$19 < \gamma \le 22.5$
7	36	$22.5 < \gamma \le 26$
8	48	$26 < \gamma \le 28$
9	54	$\gamma > 28$

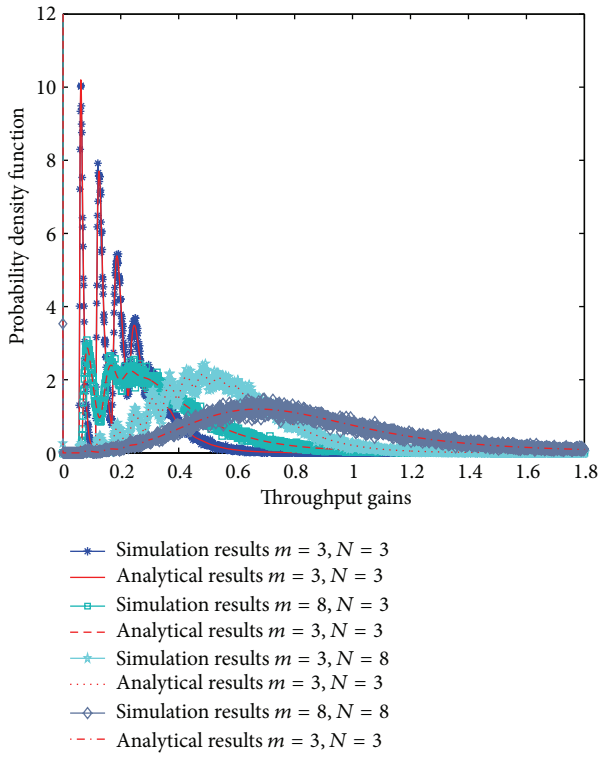


Figure 7: The PDF of the throughput gains in SU mode.

is due to more stations which can win over more time for data transmission; then the throughput can be enhanced greatly with the redundant time in TXOP increasing.

Figure 9 gives the performance analysis of MU-MIMO in IEEE 802.11ac with parameters as Table 1. From this figure, we can observe that the smaller the threshold is, the larger the throughput will be; the reason is that the user number may reduce gradually if the threshold increases. Furthermore, as expected, the correctness of theory analysis is verified through simulation results.

System goodput is a critical assessment consideration for wireless networks and can be generalized to evaluate the performance of the practical network; it is defined as the amount of network data it receives correctly and can be expressed as (1 − Per) ∗ throughput, where the Per is the packet error ratio. As can be seen from Figure 10, we can observe that the proposed scheme is superior to the conventional mechanism; the reason is that the performance

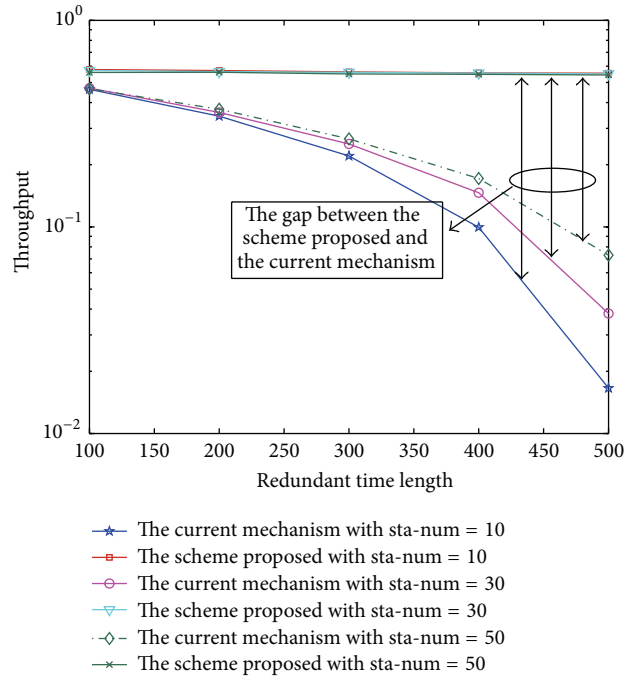


Figure 8: The tendency comparison with the redundant time increasing.

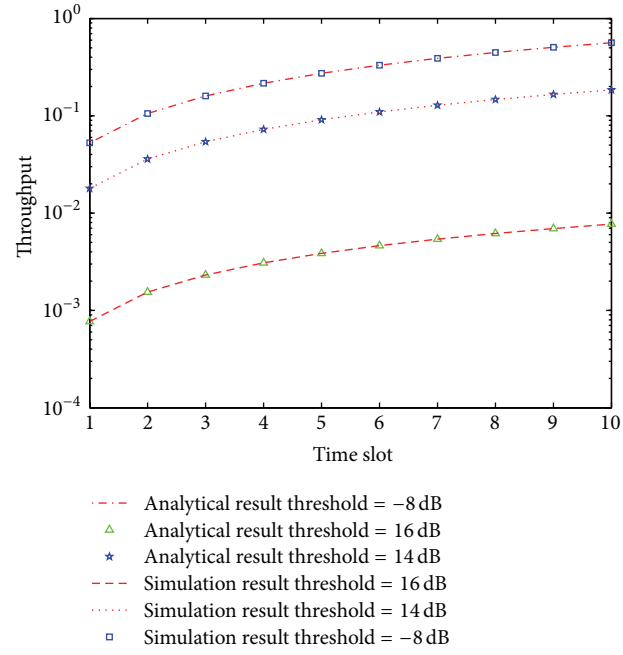


Figure 9: Performance analysis of the MU-MIMO in IEEE 802.11ac.

gains come from the limited feedback, which can bring AP more accurate channel state information. Furthermore, the BA frame in traditional mechanism also needs to transmit; what we have done is to take advantage of the reserved bits effectively.

In Figure 11, the throughput performance of the secondary AC users polling is compared to the traditional IEEE 802.11ac draft. It can be seen that the throughput

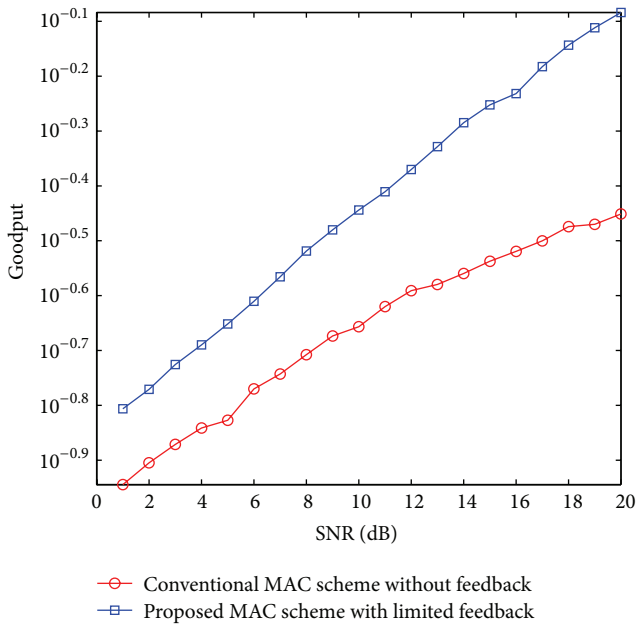


Figure 10: The performance comparison of the proposed feedback mechanism with conventional scheme.

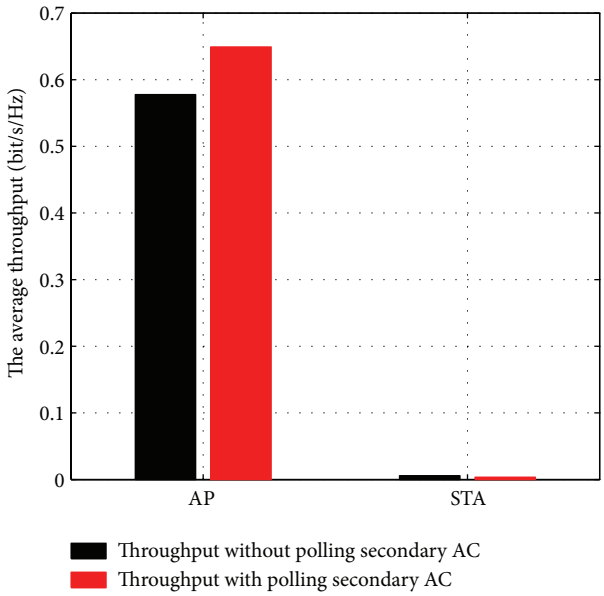


Figure 11: Secondary AC user polling based on TXOP fairness.

of AP is larger than the traditional mechanism, which is resulting from that the polling of the secondary AC users can extend the downlink transmission time of AP. On the contrary, the AP just polls the primary AC user in traditional mechanism may result in unfairness for other STAs in group. Meanwhile, the proposed scheme has little effect on the downlink transmission time of STAs in group because MU-MIMO transmission is only executed by AP in IEEE 802.11ac. Therefore, we can observe that the proposed secondary AC users polling in TXOP can not only increase the system throughput but also enhance the success probability of TXOP initialization.

6. Conclusion

In this paper, we propose the enhanced NAV mechanism based on IEEE 802.11ac draft, which only needs subtle frame structure modifications and can bring large improvement of system throughput. Both performance analysis and numerical simulations are also provided for the proposed enhanced NAV scheme. The results showed that the proposed enhanced NAV scheme can obtain noticeable performance gains than the conventional mechanism. From the results, we can know that the performance of conventional mechanism varies with what parameters and also obtain how to adjust parameters related to the proposed scheme to improve the system throughput. Using the scheme we proposed, one can solve the drawbacks of conventional mechanism easily and conveniently and also can evaluate the performance of VHT WLANs fast and obviously. On this basis, in order to solve the SINR inaccuracy calculated based on null data packet in the actual MU-MIMO transmission, the SINR feedback mechanism is proposed and the frame structure modifications are displayed in detail. Furthermore, theoretical analysis is also performed on the proposed mechanism and the formulas of the achieved gains are also derived. Numerical results confirm the validation of our theoretical analysis and further substantiate that the proposed scheme obtains obvious throughput gain over the conventional mechanism. Additionally, the throughput performance of the secondary AC users polling is compared to the traditional IEEE 802.11ac draft. It can be concluded that the throughput of AP is larger than the traditional mechanism, which is resulting from the fact that the polling of the secondary AC users can extend the downlink transmission time of AP. Finally, we can observe that the proposed secondary AC users polling in TXOP can not only increase the system throughput but also enhance the success probability of TXOP initialization.

Appendices

A. The Proof of Proposition 2

Here we will give the proof of Proposition 2. Due to the throughput gains definition and $E[P_u] = \lambda_s$, then we can rewrite the definition as

$$\text{Throughput\_gains} = \eta = \frac{\lambda_s \sum_{u=1}^{\overline{\mathcal{N}}} \lambda_u}{T_{\text{redundant}_t}} = \frac{x}{y}. \tag{A.1}$$

Then we can know the distribution of throughput gains that can be calculated as follows:

$$\begin{aligned} f(\eta) &= \int_0^{t-\Delta} f(y) f(\eta y) y dy \\ &= \int_0^{t-\Delta} \frac{\lambda e^{-\lambda(t-\Delta-y)} [\lambda(t-\Delta-y)]^{m-1}}{(m-1)!} \frac{\lambda^{\eta y}_{\overline{\mathcal{N}}}}{(\eta y)!} \\ &\quad \times e^{-\lambda_{\overline{\mathcal{N}}}} \delta(\eta y - k) y dy \\ &= \sum_{k=0}^{\lfloor \eta(t-\Delta) \rfloor} \frac{\lambda e^{\lambda(t-\Delta)}}{(m-1)!} e^{-\lambda_{\overline{\mathcal{N}}}} \frac{\lambda^k_{\overline{\mathcal{N}}}}{k!} \frac{k}{\eta} e^{\lambda k/\eta} \left[\lambda\left(t-\Delta-\frac{k}{\eta}\right)\right]^{m-1}. \end{aligned} \tag{A.2}$$

B. The Proof of Corollary 1

Due to $E_{\text{SU\_gains}} = E[x/y] = E[x]E[1/y] = \lambda_s \lambda_{\overline{\mathcal{M}}} E[1/y]$, then

$$\begin{aligned} E\left[\frac{1}{y}\right] &= \int_{1/(t-\Delta)}^{\infty} \frac{\lambda^m}{(m-1)!} y \frac{1}{y^2} e^{-\lambda(t-\Delta-(1/y))} \left(t-\Delta-\frac{1}{y}\right)^{(m-1)} dy \\ &\xrightarrow{t-\Delta=a} \frac{\lambda^m}{(m-1)!} \int_{1/a}^{\infty} \frac{1}{y} e^{-\lambda(a-(1/y))} \left(a-\frac{1}{y}\right)^{(m-1)} dy \\ &\overset{s=1/y}{=} \frac{\lambda^m}{(m-1)!} \int_0^a \frac{1}{s} e^{-\lambda(a-s)} (a-s)^{(m-1)} ds \\ &= \frac{1}{(m-1)!} \sum_{k=1}^{\infty} \left(\frac{1}{a}\right)^k \left(\frac{1}{\lambda}\right)^{k-1} \gamma(m+k-1, \lambda a). \end{aligned} \tag{B.1}$$

Then the Corollary 1 can be obtained.

C. The Proof of Proposition 3

Due to the redundant time of MU mode being $T_{\text{MU\_redun}} = t-\Delta-\text{con\_}t \sum_{u=2}^{G} \lambda_u - X$, then the distribution of the $T_{\text{MU\_redun}}$ contains the sum of Poisson variable x and m-Erlang variable y. We firstly obtain the $z = x + y$ distribution as follows:

$$\begin{aligned} f(z) &= \int_0^z f(x) f(\eta - x) dx \\ &= \int_0^z \frac{\lambda_G^x}{(x)!} e^{-\lambda_G} \delta(x-k) \frac{\lambda e^{-\lambda(z-x)} [\lambda(z-x)]^{m-1}}{(m-1)!} dx \\ &= e^{-\lambda_G} \frac{\lambda^m}{(m-1)!} \sum_{k=0}^{\lfloor z \rfloor} \frac{\lambda_G^k}{(k)!} e^{-\lambda(z-k)} (z-k)^{m-1}. \end{aligned} \tag{C.1}$$

Then the distribution of the $T_{\text{MU\_redun}}$ can be obtained as follows:

$$\begin{aligned} f_{\text{MU\_redun}}(x) = e^{-\lambda_G} \frac{\lambda^m}{(m-1)!} \sum_{k=0}^{\lfloor t-\Delta-x \rfloor} \frac{\lambda_G^k}{k!} e^{-\lambda(t-\Delta-x-k)} \\ \times (t-\Delta-x-k)^m. \end{aligned} \tag{C.2}$$

D. The Proof of Corollary 2

Due to $E_{\text{MU\_gains}} = E[x/y] = E[x]E[1/y] = \lambda_s \lambda_{\overline{\mathcal{M}}} E[1/y]$, then

$$\begin{aligned} E\left[\frac{1}{y}\right] &= \int_{1/(t-\Delta)}^{\infty} \frac{e^{-\lambda_G} \lambda^m}{(m-1)!} \frac{1}{y} \sum_{k=0}^{\lfloor 1/y \rfloor} \frac{\lambda_G^k}{k!} e^{-\lambda((1/y)-k)} \left(\frac{1}{y}-k\right)^{m-1} dy \\ &\xrightarrow[x=1/y]{t-\Delta=a} \frac{e^{-\lambda_G} \lambda^m}{(m-1)!} \int_0^a \sum_{k=0}^{\lfloor 1/y \rfloor} \frac{\lambda_G^k}{k!} e^{-\lambda(x-k)} (x-k)^{m-1} \frac{1}{x} dx \\ &= \frac{e^{-\lambda_G} \lambda^m}{(m-1)!} \sum_{k=0}^{\lfloor 1/y \rfloor} \frac{\lambda_G^k}{k!} \int_k^a e^{-\lambda(x-k)} (x-k)^{m-1} \frac{1}{x} dx. \end{aligned} \tag{D.1}$$

The integral term can be calculated as follows:

$$\begin{aligned} \text{Integral} &= \int_k^a e^{-\lambda(x-k)} (x-k)^{m-1} \frac{1}{x} dx \\ &= \int_0^{a-k} e^{-\lambda t} t^{m-1} \frac{1}{t+k} dx \\ &\xrightarrow{u=a-k} \int_0^u e^{-\lambda t} t^{m-1} \frac{1}{t+k} dx \\ &= \int_0^{\infty} e^{-\lambda t} \frac{t^{m-1}}{t+k} dx - \int_u^{\infty} e^{-\lambda t} \frac{t^{m-1}}{t+k} dx = I_1 - I_2, \end{aligned} \tag{D.2}$$

where I_1 can be calculated using [23, 3.353.5]

$$\begin{aligned} I_1 &= \int_0^{\infty} e^{-\lambda t} t^{m-1} \frac{1}{t+k} dx \\ &= (-1)^{m-2} k^{m-1} e^{\lambda k} Ei(-\lambda k) \\ &\quad + \sum_{i=1}^{m-1} (i-1)! (-k)^{m-1-i} \lambda^{-i}, \\ I_2 &= \int_u^{\infty} e^{-\lambda t} t^{m-1} \frac{1}{t+k} dx \\ &= \int_{u+k}^{\infty} e^{-\lambda(s-k)} (s-k)^{m-1} \frac{1}{s} ds \\ &= \int_b^{\infty} e^{\lambda k} e^{-\lambda s} (s-b+b-k)^{m-1} \frac{1}{s} ds \\ &= e^{\lambda k} \sum_{j=0}^{m-1} C_{m-1}^j (b-k)^j \int_b^{\infty} e^{-\lambda s} \frac{(s-b)^{m-1-j}}{s} ds, \end{aligned} \tag{D.3}$$

where $\int_b^{\infty} e^{-\lambda s} (s-b)^{m-1-j} (1/s) ds$ can be obtained through [23, 3.383.9]. Then the results can be seen as follows:

$$\begin{aligned} I_2 = e^{\lambda k} \sum_{j=0}^{m-1} C_{m-1}^j (t-\delta-k)^j (t-\Delta)^{m-1-j} \\ \times \Gamma(m-j) \Gamma(-(m-1-j), \lambda(t-\Delta)). \end{aligned} \tag{D.4}$$

Conflict of Interests

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

References

[1] "Wireless LAN medium access control(MAC) and physical layer(PHY) specifications," IEEE Standards Working Group, 1999.

[2] IEEE 802.11ac/D2.0, "Draft STANDARD for Information Technology—Telecommunications and information exchange between systems: local and metropolitan area networks-Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, Amendment 5: Enhancements for Very High Throughput for Operation in Bands below 6 GHz," 2012.

[3] P. Li, X. Geng, and Y. Fang, "An adaptive power controlled MAC protocol for wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 226–233, 2009.

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

[5] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 wireless LAN: capacity analysis and protocol enhancement," in *Proceedings of the 17th Annual IEEE Conference on Computer Communications (INFOCOM '98)*, vol. 1, pp. 142–149, April 1998.

[6] E. Kartsakli, N. Zorba, L. Alonso, and C. V. Verikoukis, "A threshold-selective multiuser downlink mac scheme for 802.11n wireless networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 857–867, 2011.

[7] Y.-J. Choi, N.-H. Lee, and S. Bahk, "Exploiting multiuser MIMO in the IEEE 802.11 wireless LAN systems," *Wireless Personal Communications*, vol. 54, no. 3, pp. 385–396, 2010.

[8] C.-E. Weng, C.-Y. Chen, C.-H. Chen, and C.-H. Chen, "Optimal performance study of IEEE 802.11 DCF with contention window," in *Proceedings of the 6th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA '011)*, pp. 505–508, October 2011.

[9] F. Babich and M. Comisso, "Optimum contention window for 802.11 networks adopting directional communications," *Electronics Letters*, vol. 44, no. 16, pp. 994–995, 2008.

[10] M. Najimi, Y. Darmani, M. S. Haghighi, and K. Mohamedpour, "An adaptive network allocation vector for IEEE 802.11-based multi-hop networks," in *Proceedings of the 2nd International Conference on Future Generation Communication and Networking (FGCN '08)*, vol. 1, pp. 279–282, December 2008.

[11] S. W. Kim, B.-S. Kim, and Y. Fang, "Network allocation vector (NAV)-based opportunistic prescanning process for WLANs," *Electronics Letters*, vol. 46, no. 24, pp. 1630–1632, 2010.

[12] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "Influence of channel BER on IEEE 802.11 DCF," *Electronics Letters*, vol. 39, no. 23, pp. 1687–1689, 2003.

[13] J. Yan, M. Zhang, J. Li, and X. Xu, "Performance comparison of IEEE 802.11s EDCA based on different NAV settings," in *Proceedings of the 12th IEEE International Conference on Communication Technology (ICCT '2010)*, pp. 755–758, November 2010.

[14] B. F. Ji, K. Song, C. G. Li et al., "Throughput enhancement for VHT WLANs based on two-level allocation vector," in *Proceedings of the IEEE 50th Global Telecommunications Conference*, pp. 881–885, 2012.

[15] Y. Yu, Y. Bai, and L. Chen, "NAV (Network Allocation Vector)-based differentiation mechanism for VoWLAN enhancement," in *Proceedings of the 67th IEEE Vehicular Technology Conference-Spring (VTC '08)*, pp. 2198–2202, May 2008.

[16] C. Mario, S. Gianfranco, and P. Giovanni, "A priority based CSMA/CA mechanism to support deadline aware scheduling in home automation applications using IEEE 802.15.4," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 139804, 12 pages, 2013.

[17] R. S. Agnelo, M. Y. Liu, and M. Mahta, "An adaptive energy management framework for sensor nodes with constrained energy scavenging profiles," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 272849, 33 pages, 2013.

[18] M. Chen, *Stochastic Process in Information and Communication Engineering*, Science Press, Beijing, China, 3rd edition, 2009.

[19] B. François and B. Bartłomiej, *Stochastic Geometry and Wireless Networks. Volume I. Theory*, NoW Publishers, 2009.

[20] B. François and B. Bartłomiej, *Stochastic Geometry and Wireless Networks. Volume II. Application*, NoW Publishers, 2009.

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic, San Diego, Calif, USA, 6th edition, 2000.

[22] W. Szpankowski, "Analysis and stability considerations in a reservation multiaccess system," *IEEE Transactions on Communications*, vol. 31, no. 5, pp. 684–692, 1983.

[23] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series*, vol. 3, Gordon and Breach Science, 1986.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 792814, 10 pages
http://dx.doi.org/10.1155/2014/792814

Research Article

An Improved Multipoint Relaying Scheme for Message Propagation in Distributed Peer-to-Peer System

Zhiping Liao,[1] Song Liu,[2] and Shengfeng Xi[1]

[1] *Information Science and Engineering College, Hunan City University, Yiyang 413000, China*
[2] *Commercial College, Hunan City University, Yiyang 413000, China*

Correspondence should be addressed to Zhiping Liao; 729627398@qq.com

Received 18 August 2013; Accepted 17 December 2013; Published 22 January 2014

Academic Editor: Deguang Le

Message propagation is a primary means to locate resources in current unstructured peer-to-peer (P2P) systems and how to reduce effectively the redundant messages of propagation while keeping its high network coverage is what many researchers go in for in this field. In this paper, we propose the algorithm of LMPR, a scheme of improved MPR for message propagation in distributed unstructured P2P system. Based on analyzing the process and message redundancy of MPR, this paper proposes to add to such scheme a mechanism of specified relay-list to reduce further the number of its retransmitting of message. By attaching to each relay-peer a relay-list with peers it will relay message to, the scheme can assure that in the view of each decision-peer a broadcast message can be dispatched and only be dispatched once to every neighbor peer within 2 hops of it. Simulation results show that the scheme of LMPR has better performance than that of MPR in message redundancy, network overhead, and fluctuation brought to the system.

1. Introduction

Peer-to-peer (P2P) systems become immensely popular in recent years and they are used in a variety of contexts, from file sharing applications, distributed storage, to content streaming, and so forth [1, 2]. Among them, fully distributed unstructured P2P systems are the most common ones, with the collaboration of all participants and by virtue of the huge number of peers, objects can be widely replicated, providing the opportunity for high availability and scalability. In such systems, there is no centralized coordinator, each peer only maintains a small amount of neighbor links and the locations of resources are unrelated to the topology of network [3]. Therefore in current fully distributed unstructured P2P systems the locating of resources mostly relies on message propagation, also known as message broadcast, which is to send message from a source peer to all other peers in the network [4, 5].

Many protocols have been proposed for message propagation and flooding-based broadcast is the basic idea of them [2, 6–9]. Although this classic technique can diffuse a message reliably and robustly in the network, it consumes a large amount of bandwidth as its cost. As we know, in P2P systems, bandwidth capacity is one of the most precious resources and the traffic should be restricted as much as possible particularly for those of wireless and mobile participants [10, 11]. If this constraint of communication traffic is not considered while designing an algorithm, the system may suffer greatly from performance degradations due to the high overloads or congestion of message flooding.

In broadcasting, message needs a mechanism to reach the far away nodes. And the concept of intermediate peers which serve as relays to pass the messages between the source and the destination is one of the solutions. By reducing the number of redundant intermediate peers, the scheme of multipoint relaying (MPR) [10] can achieve equal good results to that of pure flooding with much less control traffic. However, there are still many redundant messages re-transmitted under MPR scheme. To further improve the performance of message broadcasting, we propose the algorithm of LMPR where a specified relay-list is assigned to each relay-peer on the base of MPR. Our claim is that if we can reduce the redundant retransmissions of MPR while

maintaining approximate message coverage with MPR, the scheme can favor the message propagation and improve the performance of MPR.

The remaining of this paper is organized as follows. Section 2 briefly presents the basic idea of scheme of MPR. Section 3 analyzes the message redundancy of MPR and introduces our inspiration for improving the MPR scheme. And Section 4 describes our proposed scheme of LMPR, multipoint relaying with specified relay-list. In Section 5 simulation methodology and the simulated results are reported. Finally, we conclude this paper and make suggestions for further research in Section 6.

2. Algorithm of Multipoint Relaying

Pure flooding is the simplest broadcast scheme in distributed P2P overlay networks. It starts with the source peer S sending a message package to all its neighbor peers. On receipt of the message first time, a neighbor peer forwards the message to each of its own neighbor peers except for the sender. Thus, a message floods in the network. When a peer receives a repeat message, it drops it and does not retransmit it. The pure flooding scheme is easy to implement and has a high probability to disseminate a message to all other peers in the network. However it implies a huge network overhead due to its characteristic of message explosion.

Multipoint relaying (MPR) scheme can be seen as developed to reduce the number of duplicate retransmissions of pure flooding. In this protocol, the number of message re-transmitters is restricted to a small subset, not all the sender's neighbor peers that receive the message first time, like pure flooding scheme. This subset of retransmitters is called multipoint relays of a given network in scheme of MPR. These multipoint relays can cover the same network region as all the re-transmitters of flooding algorithm. Each node decides the set of multipoint relays in the view of its own neighbor peers within 2 hops, completely independent of other nodes' selection of their MPRs. To select the multipoint relays for a node x, let the set of 1-hop neighbors of node x be $N(x)$, and the set of its 2-hop neighbors $N2(x)$. Let the selected multipoint relay set of node x be $MPR(x)$. And the heuristic selecting of multipoint relays proposed by MPR is presented as follows [10].

(1) Start with an empty multipoint relay set $MPR(x)$, $MPR(x) = \varphi$.

(2) First select every 1-hop neighbor peer y in $N(x)$ as a multipoint relay which is the only neighbor peer of some peer in $N2(x)$; that is, $N(y) \cap N(x) = \varphi$ and add the 1-hop neighbor peer y to the multipoint relay set $MPR(x)$.

(3) While there still exists some peer in $N2(x)$ which is not covered by the multipoint relay set $MPR(x)$: (a) For each peer y in $N(x)$ which is not in $MPR(x)$, compute the number of peers that it covers among the uncovered peers in the set $N2(x)$. (b) Add that peer of $N(x)$ in $MPR(x)$ for which this number is maximum.

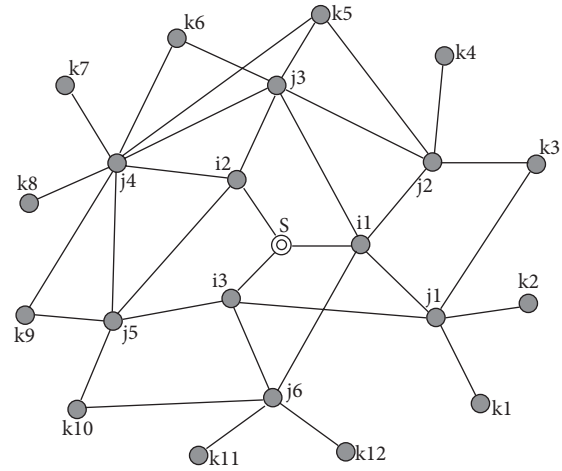


Figure 1: Network topology.

3. Analysis of Message Redundancy in MPR and Inspiration for Improving the MPR Scheme

3.1. Network Model. To analyze intuitively the message redundancy of MPR in broadcasting, we consider a scenario of message propagation in the miniature of a typical unstructured P2P network. With this network model the message redundancy of pure flooding and MPR is analyzed. And the network topology is given in Figure 1 where each node represents a peer and each edge is the neighbor link between the two end peers. Here, the source peer S wants to disseminate a message to all the other peers in the network. According to the hops away from S, these peers are divided into three layers and named separately with the prefix of "I," "j" and "k." More specifically, as is shown in Figure 1 the names of 1-hop peers from S are prefixed with character "I," named i1, i2, and i3 in sequence. Similarly, the peers named from j1 to j6 are all 2 hops away from S and peers with the distance of 3 hops are named k1, k2,..., k12 in sequence.

In order to focus on the problem of message redundancy in MPR when broadcasting, in this paper we assume a simplified network model where there is no message failure, where messages are delivered instantly and where each peer has a unique identity assigned by the system. And we also assume that the less hops a peer away from S the earlier it receives the broadcasting message. For those peers who both have the same hops away from S and receive the broadcasting message transmitted by the same source or relay-peer, the smaller the serial number of a peer, the earlier it receives message from the transmitter. For example, the source peer S transmits the message to its neighbors in order of i1, i2, and i3, and the peer j1 transmits its relay message in order of k1, k2, and k3. For convenience of presenting, we divide peers into two groups: the peers who relay the broadcasting message to their neighbors after receiving it are called relay-peers, and the others are called reception-peers who do not forward the broadcasting message further after receiving it.

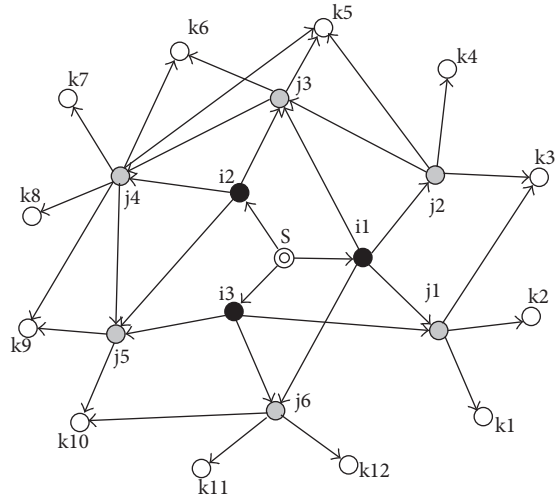


Figure 2: Effect of message propagation.

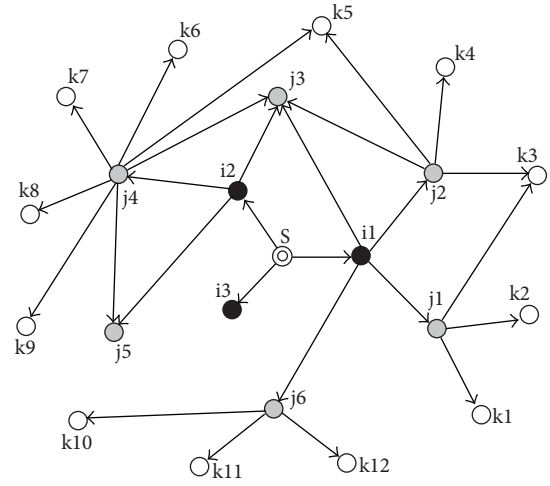


Figure 3: Effect of message propagation.

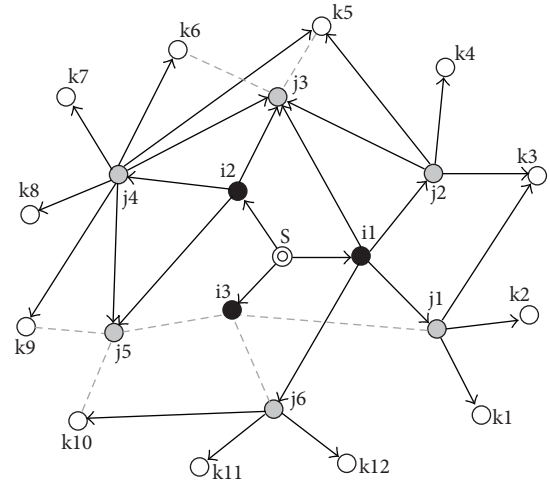


Figure 4: Effect of message propagation.

3.2. Analysis of Message Redundancy

3.2.1. Pure Flooding. In the scheme of pure flooding, all the peers except for the source are relay-peers, and every relay-peer always transmits the message to all its neighbors except for the senders. Therefore, each link in network corresponds to a message transmission. As for the network scenario with topology given in Figure 1 the flooding effect is shown in Figure 2 where each arrow represents a message from the tail node to the head one and the number of messages is 34 in 3 hops.

3.2.2. Multipoint Relaying. The main objective of MPR scheme is to obtain a minimized set of relay-peers to cover all peers in network. According to the MPR scheme, in the network scenario given above, firstly the source peer S decides to choose i1 and i2 as relay-peers, then peer i1 chooses j1, j2, j6 as relay-peers, and peer i2 chooses j4 as relay-peer to cover all the 2-hop neighbors in their view. As a result, within 3 hops from S the message reaches 21 peers: 6 relay-peers (i1, i2, j1, j2, j6, and j4) and 15 reception-peers (i3, j3, j5, and k1~k12). And the major improvement is that the three peers (i3, j3, and j5) simply act as reception-peers in MPR while they are relay-peers in pure flooding. The propagation effect of MPR is shown in Figure 3. Totally the number of messages is 27 in 3 hops. And 7 duplicate retransmissions are reduced in comparison with the flooding scheme. For convenience of comparison these reduced redundant messages are plotted using dash lines in Figure 4 instead of solid lines in Figure 2.

In MPR scheme, although the number of relay-peers is minimized on the base of pure flooding, each of the selected relay-peers still transmits a message to all its 1-hop neighbors except for the sender. Thus if a reception-peer has direct links to many relay-peers at one time, it will surely receive separately a message from each of these neighboring relay-peers. Similarly, if one relay-peer is neighboring to another relay-peer there always exits a message from the one who is the earlier receiver to the other. However, for each peer one message is theoretically enough, no matter it is a relay-peer or not, all the other messages it received are redundant. As for the network scenario we discussed, there are 21 peers within 3 hops from S, so 21 messages are enough for an optimal scheme to broadcast the message to every peer. Nevertheless, the MPR scheme expends 27 messages to realize the broadcasting; although this number is 7 less than that of pure flooding, there is still 6 redundant messages to optimize compared with the optimal scheme. Therefore, if the relay-peers selected by MPR scheme have many common neighbors or have many links between themselves, the number of redundant messages is surely nontrivial. In fact, in the current P2P networks which are generally of power-law distribution characteristics, this situation is always there [12–15].

3.3. Inspiration for Improving the MPR Scheme. As can be seen from the above analysis of message redundancy that although the MPR technique reduces greatly the number of retransmissions of pure flooding, the scheme still brings

heavy extra message load to the P2P applications which are generally built on the complex network with characters of power-law distribution. The reason is that relay-peers in MPR scheme do not share the information of their relay targets, which leads to the fact that after one relay-peer has already sent a message to a neighboring peer there are still many other relay-peers transmitting message to the same target peer. In other words, many relay-peers transmit repeatedly the message to common target peers because of their blindness of one another's relaying information. Aiming at improving this situation, this paper proposes the scheme of LMPR which is to add to the MPR scheme a technique of specified relay-list. By attaching to each relay-peer a relay-list with peers it will relay message to, a decision-peer assigns for each of its 2-hop neighbors a responsible relay-peer, so as to assure that multiple relay-peers will not transmit message to one common peer in the 2-hop view of the decision-peer.

4. Multipoint Relaying with Specified Relay-List (LMPR)

4.1. The Basic Idea of LMPR. The scheme of LMPR works in a distributed manner and starts from the source peer of message propagation. According to the role in relaying process we divided the working states of a peer into two kinds: decision-peer and normal-peer. When a source peer begins to broadcast message to its 1-hop neighbors or when a peer receives the broadcasting message, it turns into the state of a decision-peer who decides how to forward the message to peers within 2 hops of its view. A decision-peer decides independently the set of relay-peers in its own view and assigns for each of these relay-peers a specified list containing the subset of all 2-hop neighbors of the decision-peer. (Here, the specified list is used to tell the corresponding relay-peer which 2-hop neighbors of the decision-peer it should relay message to.) And then the decision-peer sends to each neighbor a message accompanied by the specified list. After transmitting the message accompanied by relay-list to every neighbor, the peer turns into the state of normal-peer in which the peer updates its neighboring information, manages the information used in LMPR, or does other things which are required in P2P application but beyond the discussion of basic idea of LMPR. Therefore the basic idea of LMPR is generally implemented by the actions of decision-peer, which includes mainly three parts: multipoint relaying with relay-list, generating of N2Set(x), and generating of relay-lists.

4.1.1. Multipoint Relaying with Relay-List. First, Similar to MPR, we assume that each peer in the system knows all its neighbors within 2 hops and their neighboring relations between each other. And a peer can obtain this information by exchanging the 1-hop neighboring information with all its 1-hop neighbors. Based on this information and the relay-list named list0 which is received from its upstream decision-peer, each decision-peer x calculates its own set of N2Set(x) which includes all those 2-hop neighbors need to be covered by the message originated from the current decision-peer itself. Then a decision-peer x must also assign for each of its 1-hop neighbors included in list0 a relay-list that specifies which 2-hop neighbors in N2Set(x) the neighbor should relay message to. If a 1-hop neighbor is relay-peer then the list contains IDs of those peers it will relay message to; otherwise the list contains nothing. Finally a decision-peer sends the message with corresponding specified relay-list to each of those 1-hop neighbors given in list0. Thus, through the message relaying of these relay-peers a broadcast message can be dispatched and only be dispatched once to every peer within 2 hops away from a decision-peer in its view. And the details of generating N2Set(x) and generating relay-lists are presented separately in the following two subsections.

4.1.2. Generating of Set N2Set(x). Since a relay-peer only transmits the message to the part of its 1-hop neighbors specified by list0 when it turns into a decision-peer, therefore only those 2-hop neighbors covered by the 1-hop neighbors given in list0 are needed to be broadcast to. And the remaining 2-hop neighbors of it can be broadcasted message to by the relay-peers not included in list0 and such transmitting assignment is considered by the upstream decision-peer according to the principle of LMPR. Otherwise, this redundancy problem will arise when the message is relayed to 3-hop neighbors of the decision-peer. Thus, to extend the optimizing mechanism of relay-list to the peers farther than 2 hops, the set of N2Set(x) must be defined as different from that of MPR in which the set includes all the 2-hop neighbors of a decision-peer. In scheme of LMPR, when a peer y receives the message accompanied by a relay-list named list0 from x, y becomes a new decision-peer. If list0 is not empty, then peer y picks out one by one a peer from the list and adds all those 1-hop neighbors of this peer into the N2Set(x) when they are still not exists in the set. This process continues until all peers in list0 are gone through and thus the N2Set(x) of the new decision-peer is formed.

4.1.3. Generating of Relay-Lists. To avoid that several relay-peers transmit message to a common neighbor in its view, a decision-peer assigns for each of its relay-peers a specified list containing the IDs of those 2-hop neighbors it will relay message to. And the mechanism of relay-list makes ensure that each 2-hop peer must appear and only appears once in one of the lists. To generate such relay-lists, decision-peer x can initially add all the peers in list0 into a set named RSet(x). Here, if the decision-peer x is a source peer of broadcasting, then list0 is composed of all its 1-hop neighbors; otherwise list0 is the relay-list sent by the upstream decision-peer. Then x picks out one by one a relay-peer y from RSet(x), each time x generates a specified list named relay-list$_y$ which contains all IDs of those peers that are the 1-hop neighbors of y as well as the elements of N2Set(x). If there is no such peer meeting this condition, then the relay-peer y is actually a reception-peer and relay-list$_y$ is an empty list. As a relay-list has been generated the decision-peer x removes separately the picked relay-peer y from RSet(x) and its responsible 2-hop neighbors of x from N2Set(x). This process continues until RSet(x) becomes empty. Thus the decision-peer x can generate for each of its relay-peers a tuple consisting of the relay-peer's

identity and specified relay-list; that is, $\langle ID_y, \text{relay-list}_y \rangle$. Finally, x transmits the broadcast message attached by relay-list_y to the neighbor with identity of ID_y. And y becomes a decision-peer after it receives the broadcast message and x turns into the state of normal-peer when it finishes all its message sending.

4.2. Pseudocode Description of LMPR. The working procedures of a decision-peer x in scheme of LMPR can be as follows.

Step 1. If it is a source peer, x generates a unique identity named MessageID for the message to be broadcast, sets list0 $\leftarrow$ {all 1-hop neighbors of x}, N2Set(x) $\leftarrow$ {all 2-hop neighbors of x}, and then skips to Step 4. Otherwise x starts from Step 2.

Step 2. x Accepts the message package including $\langle$MessageID, Message, list$\rangle$ from a upstream decision-peer. Based on the data of MessageID stored locally, x decides if the message has been broadcast before. If the message is a repeated one, then x skips to Step 7; otherwise it turns to Step 3.

Step 3. x Stores the MessageID, and sets: list0 $\leftarrow$ list, N2Set(x) $\leftarrow$ {N2Set(x) = φ; for each $y_i \in$ list0, adds into N2Set(x) all the peers that are y_i's 1-hop neighbors as well as the 2-hop neighbors of x while they do not exist in the set.}

Step 4. x Sorts the list0 according to the MPR selection algorithm for the sake of efficiency.

Step 5. For each $y_i \in$ list0, x initializes a new empty relay-list_{yi} firstly, then extracts from N2Set(x) the peers that are 1-hop neighbors of y_i, and puts them into relay-list_{yi}. If there is no such satisfied peer for some y_i, then its relay-list_{yi} keeps empty. Thus x generates for each relay-peer y_i specified by list0 an item of relaying information; that is, $\langle ID_{yi}, \text{relay-list}_{yi} \rangle$.

Step 6. For each item of $\langle ID_{yi}, \text{relay-list}_{yi} \rangle$, x sends to peer y_i with identity of ID_{yi} the message package with $\langle$MessageID, Message, list$\rangle$ where the list is relay-list_{yi}.

Step 7. x Turns into the state of normal-peer.

4.3. Message Propagation of LMPR in the Given Network Scenario. To better illustrate the scheme of LMPR and show its improvement over MPR, we still take the network scenario given in Section 3.1 as an example, presenting the process of message propagation in scheme of LMPR and analyzing its message redundancy in this subsection.

As for the network shown in Figure 1, the scheme of LMPR starts its message broadcast from the source peer S. Firstly, according to the mechanism of LMPR, S generates for all its three 1-hop neighbors items of relaying information: $\langle$i1, {j1, j2, j3, j6}$\rangle$, $\langle$i2, {j4, j5}$\rangle$, and $\langle$i3, {NULL}$\rangle$. It means that decision-peer S only chooses in its view peers i1 and i2 as relay-peers and regards i3 as a reception-peer. And through the mechanism of relay-list peer S allocates {j1, j2, j3, j6} of all its six 2-hop neighbors to relay-peer i1 for relaying message to, and the remaining two of {j4, j5} are assigned to i2 for message transmitting. Thus all the six 2-hop neighbors of S can receive and only receive one broadcast message in the view of S. Then after message sending of S, peers i1 and i2 become decision-peers successively. Similarly, i1 calculates its four items of relaying information as $\langle$j1, {k1, k2, k3}$\rangle$, $\langle$j2, {k4, k5}$\rangle$, $\langle$j6, {k10, k11, k12}$\rangle$ and $\langle$j3, {NULL}$\rangle$; and i2 generates its two items of relaying information as: $\langle$j4, {k6, k7, k8, k9}$\rangle$ and $\langle$j5, {NULL}$\rangle$. Certainly, it can be seen from these relaying items that the four peers of j1, j2, j6, and j4 become decision-peers in the next round of propagation and each decision-peer has its own relaying target peers assigned by specified relay-list, and so on. The propagation effect of LMPR can be illustrated as Figure 5.

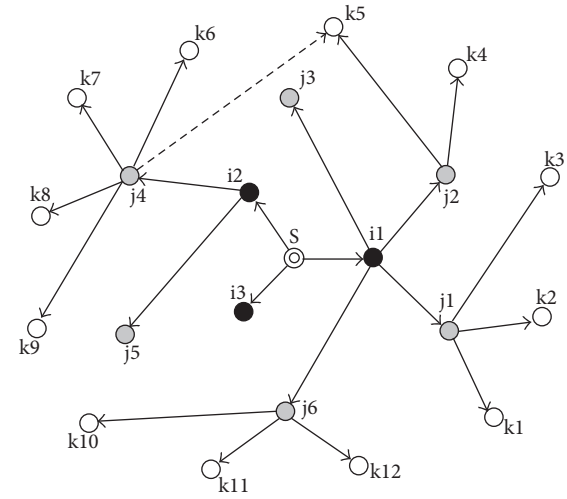


Figure 5: Effect of message propagation.

From the process of message propagation presented above, we can see obviously that a relay-peer does not forward a message to all its neighbors as it does in MPR. However it only transmits the message to those assigned by the specified relay-list, so as to reduce the number of messages while the message coverage keeps unchangeable. Although there is still a redundant message from peer j4 to k5 shown as a dash line in Figure 5 because of some blindness of relay-list mechanism beyond 3 hops of decision-peer, the total number of messages spread in LMPR is 22 which is 5 less than that of MPR. To show intuitively its improvement of the message redundancy over that of MPR, we illustrate further the propagation effect of LMPR as Figure 6 where the reduced redundant messages are plotted using dash lines instead of solid lines in Figure 3. And all the 5 redundant messages reduced on the base of MPR can be seen apparently.

5. Simulations

The performance of MPR scheme and its advantage over the pure flooding has been studied by [10]. In this section, we evaluate the performance of the proposed LMPR scheme

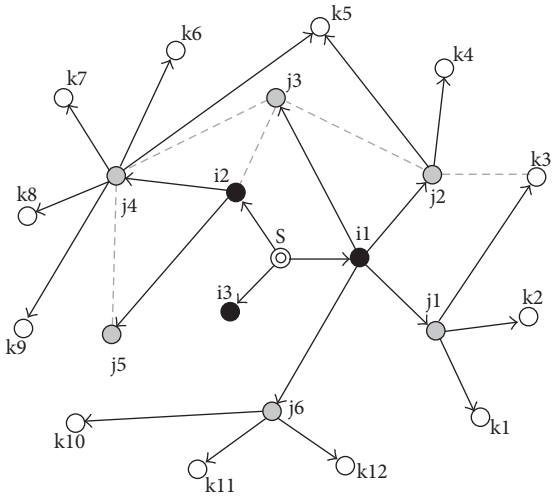


Figure 6: Effect of message propagation.

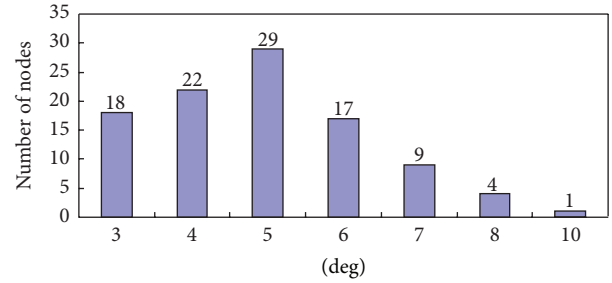


Figure 7: Degree distribution of all nodes in the simulated network.

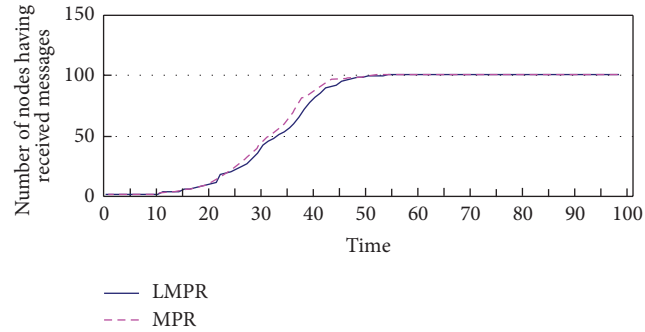


Figure 8: Number of nodes having received the broadcast message.

and compare it with the protocol of MPR by analyzing their message overhead, propagation convergence, and network coverage of broadcasting.

We divide the simulations into two subsections: the performance in a classic stable P2P network and the stability of performance under varying network size and packet loss probability. The first subsection is to evaluate the main performance of the two schemes of LMPR and MPR in a classic stable P2P network. And the second subsection shows their performances under P2P networks with different sizes and packet loss probabilities, which is to testify the stability and representative of performances observed by the first subsection.

5.1. Simulation Methodology. To observe the performance of LMPR and MPR which are basically two algorithms of message broadcasting in P2P network, the process of message propagation between peers following these two schemes should be simulated. The simulation work in this paper is carried on the Java based PeerSim [16] which is one of the most known simulators among P2P researchers. The PeerSim engine has two simulation models: cycle-based model and event-based model. We have chosen the event-based model where message sending is modeled in more detail than the cycle-based one. To simulate and observe the performance of the two schemes of LMPR and MPR, we have designed and programmed separately these two protocols on the platform of PeerSim. As far as programming technology is concerned, simulation works of these two protocols are similar. Therefore, for simplicity we only present the implementation of LMPR protocol which includes mainly six Java classes whose function descriptions are given in Table 1.

5.2. Performance in a Classic Stable P2P Network. This part shows the values of main performance characters of the two observed algorithms and their changing with time under the simulation performed in a classic static undirected P2P network. And the main configuration parameters of the network in simulation are listed in Table 2.

In our simulation experiments, the network topology is generated by having each node that represents a peer in the network randomly select two other nodes and build edges with the latter ones. These edges represent the neighboring relationships between peers. Thus the topology of generated network is in the form of an undirected random graph [17]. As for the network generated in this part of simulation its average degree of all network nodes is 4.94, the minimum degree is 3 and the maximum degree is 10. And the degree distribution of all nodes in the simulated network is shown in Figure 7.

As a scheme of message propagation, the primary function of algorithm LMPR is to transmit the broadcast message to entire network as soon as possible. To observe this performance, we have measured the number of nodes having received the broadcast message on schemes of both LMPR and MPR in the simulated network and the result is plotted in Figure 8. As is shown in the figure, the number increases apparently from the 11th time unit and when it comes to the 55th time unit all 100 nodes in the network have received the message. It can be seen that the increase of node coverage rate of broadcast message is similar between the two algorithms of LMPR and MPR. However, the increasing curve of LMPR is lower and smoother than that of MPR which means that the scheme of LMPR runs a little slower but steadier than MPR and the end time of message propagation of both algorithms is almost the same. The reason is that the specified relay-lists of scheme LMPR only reduce the redundant messages in broadcasting and the pruning of messages does not influence the rate of its network coverage.

Table 1: Function descriptions of main classes designed for simulating LMPR algorithm.

Class name	Function description
LmprED	LmprED, the core class of the algorithm LMPR, implements the interfaces of both CDProtocol and EDProtocol. The former interface is mainly used to simulate the action of initiating message broadcasting from source peer. And the later interface is mainly used to simulate the action of message transferring by general peers. Both two actions have the function of generating set of N2Set(x) and corresponding relay-lists.
RelayItem	RelayItem, an auxiliary class of LmprED, is in nature a data structure that helps to store the relay-information which includes relay-peers and relay-lists.
LmprIdleProtocol	LmprIdleProtocol is an auxiliary class of protocol LmprED. It stores the broadcast message to and from the current peer, which includes message identity, sender, receiver, response state, total number of messages, and number of redundant messages. And LmprIdleProtocol also has the function of deciding if the current peer is in active at any given time.
DecisionMakingState	DecisionMakingState, an auxiliary class of LmprIdleProtocol, is a data structure used to store the number of sent messages, the number of responded messages and the time when each message is sent.
LmprMessage	LmprMessage is an auxiliary class of protocol LmprED and it is a data structure used to store the message package transmitted between peers. This data structure includes mainly a message's identity, content, sender, and relay-list.
LmprObserver	LmprObserver is in charge of outputting the simulation results of algorithm LMPR. Based on the parameters set in configuration file, it counts and outputs regularly the observed simulation results. Simulation results related to the scheme of LMPR in this paper are all output by the function of this class.

Table 2: Main configuration parameters of the network in simulation.

Parameter	Value	Description
SIZE	100	Number of nodes or peers in the simulated network.
DROP	0	Network loss ratio, the value of zero means no packet loss in the network.
MINDELAY	10	A parameter used to generate the minimum delay of transmitting after receiving a message, takes one percent of a cycle time as a unit. The minimum delay of transmitting is (CYCLE ∗ MINDELAY)/100.
MAXDELAY	100	A parameter used to generate the maximum delay of transmitting after receiving a message, takes one percent of a cycle time as a unit. The maximum delay of transmitting is (CYCLE ∗ MAXDELAY)/100.
STEP	CYCLE/20	The time interval between two observations on simulation parameters takes the value of CYCLE/20 as a time unit when outputting simulation results.
CYCLE	SIZE ∗ 10000	A parameter used to simulate the changing of time in milliseconds.
MessageSize	20	The size of broadcast message is supposed to be 20 bytes. And this parameter is determined by the application, none of relay-list or relay-peer.

The amount of messages resulted by a proposed algorithm is an important index of its performance. Figure 9 shows the total number of messages disseminated in the network over time. From the 25th time unit the amount of messages generated by scheme LMPR is less than that of MPR notably. Till the 70th time unit, the overall numbers of messages of both algorithms arrive at their highest value where the total number of messages transmitted in LMPR is 267 and that number in MPR is 325. Although the total number of messages generated by both algorithms is much more than the ideal value of 100 when each node only receives one message, the performance of LMPR is still better than MPR. And it demonstrates that the mechanism of relay-list cuts down apparently the amount of messages in the network on the base of algorithm MPR.

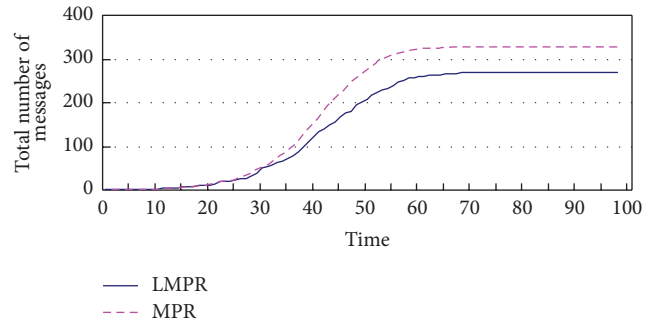


Figure 9: Total number of messages disseminated in the network.

As is known from Figure 8 that all nodes have received the broadcast message at time unit 55, however, Figure 9 shows that the amounts of messages resulted by both algorithms are still increasing after the time unit 55. It is because that some nodes are still in active state at the moment and these active nodes continue to transmit the message to their neighbors. The number of messages stops increasing till all nodes become inactive. In Figure 10 the number of active nodes in the network is measured over time. From the 11th time unit on, the number of active nodes of both algorithms

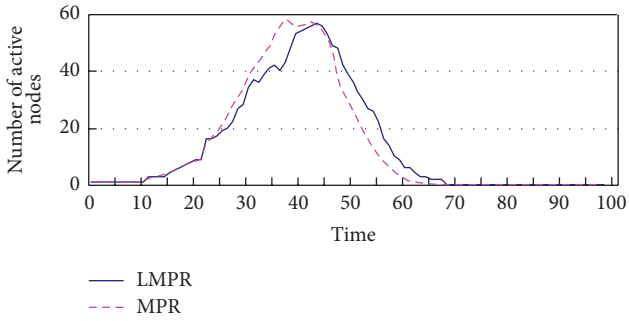


Figure 10: Number of active nodes in the network.

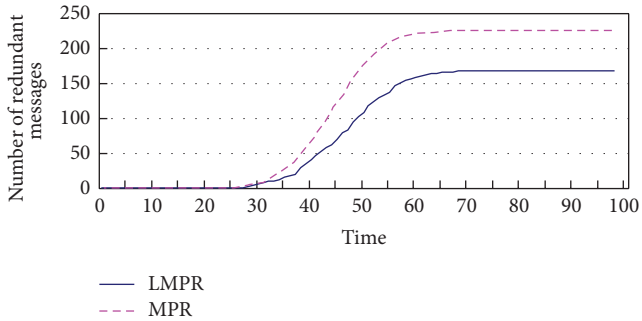


Figure 11: Number of redundant messages in the network.

increases notably and the value becomes the highest at about the 40th time unit, and then the number begins to decrease and drops to zero at about the 70th time unit. The figure shows that the varying of the number of active nodes in LMPR is more slow and flatter than in scheme of MPR. And we can see that the scheme of LMPR brings a smaller fluctuation to the system than that of MPR.

The main objective of the proposed algorithm of LMPR in this paper is to reduce the amount of redundant messages in message propagation on the base of MPR while keeping its rate of network coverage almost unchangeable. Therefore, the number of redundant messages in the network is observed in our simulation and the results are shown in Figure 11. It can be seen from the figure that although the redundant messages in the two observed schemes both start to emerge at about the time unit 30 and then increase rapidly, the quantity resulted by algorithm of LMPR is much less than that of MPR.

Although the number of redundant messages generated by LMPR is less than that of MPR, the size of each message package in the former algorithm is larger than that of the latter because of the additional overhead brought by the assigned relay-lists described in Section 4.1. Therefore, the network overhead resulted by the two schemes becomes another big concern in the simulation. If the size of the message broadcast is supposed to be 20 bytes in our simulation, then the network overhead resulted by the two observed algorithms can be plotted as in Figure 12. From the figure we can see that the network overhead brought by LMPR is higher than the other algorithm. However, since the size of relay-lists is invariable at a given network topology, with the increase of the size of broadcast message the influence of network overhead brought by assigned relay-lists will surely decrease. In the simulation environment assumed by this paper, if the size of the broadcast message is larger than 48 bytes, then the network overhead resulted by the algorithm of LMPR becomes lower than that of MPR. The varying of network overhead with the size of broadcast message is shown in Figure 13.

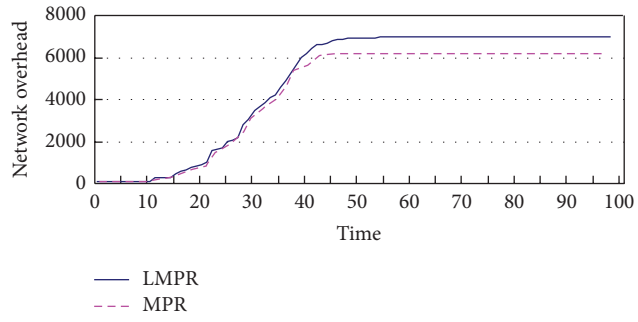


Figure 12: Network overhead resulted by message broadcasting.

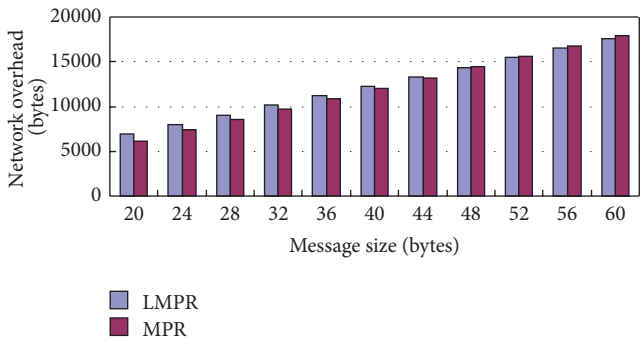


Figure 13: Varying of network overhead with the size of broadcast message.

5.3. Stability of Performance under Varying Network Sizes and Packet Loss Probabilities. The simulation results presented by the previous part mainly show the performance of the two algorithms of LMPR and MPR in a classic stable P2P network with size of 100 nodes. To check if those performance characteristics observed can still hold on when the two algorithms work under P2P networks with different sizes and packet loss probabilities is the main task of the simulation described in this subsection. Figure 14 presents the average number of redundant messages per node in networks with different sizes. It is shown by the figure that the quantity of average redundant messages increases for some extent when the size of network adds from 100 to 500, and then the value keeps mostly unchangeable after the size is larger than 500. In general, the average number of redundant messages per node in the scheme of MPR is about 2.1 while that number in LMPR is about 1.8, which is of better performance than the former algorithm.

Finally, the number of nodes having received the broadcast message is measured when the network size is set to 1000

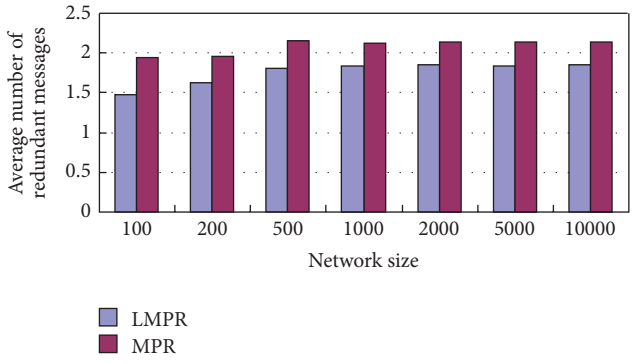


Figure 14: Average number of redundant messages per node.

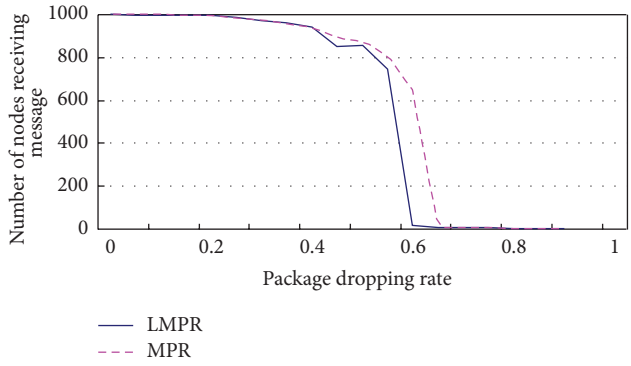


Figure 15: Number of nodes having received message varying with the packet loss probability.

and the packet loss probability varies from 0 to 1 with a step of 0.1. The simulation results are plotted in Figure 15 which shows that both algorithms have high message coverage when the packet loss probability is lower than 0.2. And with the increase of probability, message coverage begins to decrease in both schemes. Then as the packet loss probability increases to about 0.6, the number of nodes having received the broadcast message in LMPR goes down sharply and turns to zero soon after. And this happens when the packet loss probability increases to about 0.7 in MPR. It is because that there are more redundant messages contributing to the reliability of the system in the scheme of MPR. Anyhow the simulation result demonstrates that the scheme of LMPR also performs well in tolerating the packet loss probability.

6. Conclusions

Message propagation is used widely in unstructured P2P systems for resource searching and distributing. In this paper we have presented LMPR, a scheme of improved MPR for message propagation in distributed P2P system. Aiming at reducing the message redundancy of scheme of MPR while keeping its message coverage almost unchangeable, this paper proposes to add to the MPR scheme a technique of specified relay-list. By attaching to each relay-peer a relay-list with peers it will relay message to, the scheme assures that a broadcast message can be dispatched and only be dispatched once to every peer within 2 hops away from a decision-peer in its view.

Simulation experiments prove that the number of redundant messages resulted by algorithm of LMPR is much less than that of MPR. And if the size of the broadcast message is larger than some constant value such as 48-bytes of our experiment, then the network overhead resulted by the algorithm of LMPR becomes lower than that of MPR. Furthermore, the simulation results also show that the scheme of LMPR brings a smaller fluctuation to the system than MPR and the main performance characteristics can still hold on under P2P networks with different sizes and packet loss probabilities.

However, owing to the openness and anonymity of P2P systems, there always exists invalid relay-peers in message propagation such as free-rider or failure of peer [15, 18, 19]. Therefore, to improve the performance of LMPR, further research is also encouraged to apply the reputation mechanism to the selecting of relay-peers and generating of relay-lists.

Conflict of Interests

There is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work is supported by the Research Plan of Science and Technology of City Yiyang of China under Grant No. 2011JZ29 and the Colleges Oriented Science Research Project of Hunan Province of China under Grant Nos. 11C0248, 12C0571.

References

[1] S. Androutsellis-Theotokis and D. Spinellis, "A survey of peer-to-peer content distribution technologies," *ACM Computing Surveys*, vol. 36, no. 4, pp. 335–371, 2004.

[2] E. Meshkova, J. Riihijärvi, M. Petrova, and P. Mähönen, "A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks," *Computer Networks*, vol. 52, no. 11, pp. 2097–2128, 2008.

[3] B. F. Cooper and H. Garcia-Molina, "Ad hoc, self-supervising peer-to-peer search networks," *ACM Transactions on Information Systems*, vol. 23, no. 2, pp. 169–200, 2005.

[4] M. Yang and Z. Fei, "A novel approach to improving search efficiency in unstructured peer-to-peer networks," *Journal of Parallel and Distributed Computing*, vol. 69, no. 11, pp. 877–884, 2009.

[5] J. J. Jung, "Trustworthy knowledge diffusion model based on risk discovery on peer-to-peer networks," *Expert Systems with Applications*, vol. 36, no. 3, pp. 7123–7128, 2009.

[6] S. Jiang, L. Guo, and X. Zhang, "LightFlood: an efficient flooding scheme for file search in unstructured peer-to-peer systems," in *Proceedings of International Conference on Parallel Processing*, 2003.

[7] T. Yeferny and K. Arour, "LearningPeerSelection: a query routing approach for information retrieval in P2P systems,"

in *Proceedings of the 5th International Conference on Internet and Web Applications and Services (ICIW '10)*, pp. 235–241, Barcelona, Spain, May 2010.

[8] A. Rosenfeld, C. V. Goldman, G. A. Kaminka, and S. Kraus, "PHIRST: a distributed architecture for P2P information retrieval," *Information Systems*, vol. 34, no. 2, pp. 290–303, 2009.

[9] F. Javier Ortega, J. A. Troyano, F. L. Cruz, C. G. Vallejo, and F. Enríquez, "Propagation of trust and distrust for the detection of trolls in a social network," *Computer Networks*, vol. 56, pp. 2884–2895, 2012.

[10] A. Qayyum, L. Viennot, and A. Laouiti, "Multipoint relaying for flooding broadcast messages in mobile wireless networks," in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pp. 3866–3875, 2002.

[11] B. Yang and H. Garica Molina, "Improving search in peer-to-peer networks," in *Proceeding of the 22nd IEEE International Conference on Distributed Computing Systems (ICDCS '02)*, pp. 1–10, 2002.

[12] L. A. Adamic, R. M. Lukouse, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks," *Physical Review E*, vol. 64, Article ID 046135, 8 pages, 2001.

[13] B. Wu and A. D. Kshemkalyani, "Modeling message propagation in random graph networks," *Computer Communications*, vol. 31, no. 17, pp. 4138–4148, 2008.

[14] I. Keidar, R. Melamed, and A. Orda, "EquiCast: scalable multicast with selfish users," *Computer Networks*, vol. 53, no. 13, pp. 2373–2386, 2009.

[15] C. Li, B. Yu, and K. Sycara, "An incentive mechanism for message relaying in unstructured peer-to-peer systems," *Electronic Commerce Research and Applications*, vol. 8, no. 6, pp. 315–326, 2009.

[16] A. Montresor and M. Jelasity, "PeerSim: a scalable P2P simulator," in *Proceedings of the IEEE 9th International Conference on Peer-to-Peer Computing (P2P '09)*, pp. 99–100, Seattle, Wash, USA, September 2009.

[17] A. Steger and N. C. Wormald, "Generating random regular graphs quickly," *Combinatorics Probability and Computing*, vol. 8, no. 4, pp. 377–396, 1999.

[18] X. Li, P. Zhao, and L. Li, "Resilience and reliability analysis of P2P network systems," *Operations Research Letters*, vol. 38, no. 1, pp. 20–26, 2010.

[19] F. G. Mármol and G. M. Pérez, "Security threats scenarios in trust and reputation models for distributed systems," *Computers & Security*, vol. 28, pp. 545–556, 2009.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2014, Article ID 596983, 9 pages
http://dx.doi.org/10.1155/2014/596983

Research Article

Optimal Deployment and Scheduling with Directional Sensors for Energy-Efficient Barrier Coverage

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

[1] *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China*
[2] *College of Electronics and Information Engineering, Nanjing University of Technology, Nanjing, Jiangsu 211816, China*
[3] *School of Business, Hunan University, Changsha, Hunan 410082, China*

Correspondence should be addressed to Guangwei Bai; bai@njut.edu.cn and Yanhui Jiang; huihui0733@163.com

Received 5 September 2013; Revised 25 November 2013; Accepted 12 December 2013; Published 2 January 2014

Academic Editor: Yong Jin

In recent years, barrier coverage problem in directional sensor networks has been an interesting research issue. Most of the existing solutions to this problem aim to find as many barrier sets as possible to enhance coverage for the target area, which did not consider the power conservation. In this paper, we address the efficient sensor deployment (ESD) problem and energy-efficient barrier coverage (EEBC) problem for directional sensor networks. First, we describe a deployment model for the distribution of sensor locations to analyze whether a target area can be barrier covered. By this model, we examine the relationship between the probability of barrier coverage and network deployment parameters. Moreover, we model the EEBC as an optimization problem. An efficient scheduling algorithm is proposed to prolong the network lifetime when the target area is barrier covered. Simulation results are presented to demonstrate the performance of this algorithm.

1. Introduction

Directional sensors, such as image sensors [1], video sensors [2] and infrared sensors [3], have been widely used to improve the performance of wireless sensor networks. Directional sensor has a limited angle of sensing range due to the technical constraints or cost considerations, which is different from omnidirectional sensor. Directional sensors may have several working directions and adjust their sensing directions during their operation.

Power conservation is another important issue for directional sensor, just like omnidirectional sensor. Power consumption of sensor nodes has a great impact on the lifetime of sensor networks. Most directional sensors have limited power sources which are supplied by batteries. The batteries of sensors are not rechargeable due to the hostile or inaccessible environments in many scenarios. Due to the small size of existing batteries, sensor nodes cannot last as long as desired.

Barrier coverage is an efficient way for many applications in directional sensor networks (DSNs), such as intrusion detection and border surveillance [4, 5]. There are two kinds of barrier coverage, that is, weak and strong [6], which are illustrated in Figure 1. Weak barrier coverage guarantees detections of intruders moving along congruent crossing paths, but it does not guarantee the detection of intruders moving along arbitrary crossing paths. Strong barrier coverage guarantees that no intruders can cross the region undetected no matter what crossing paths they choose. Constructing a strong barrier with directional sensors for a target region is a challenging problem. In this paper, we will focus on strong barrier coverage for DSNs. For convenience, this paper will use barrier coverage to refer to strong barrier coverage.

In this work, we separate barrier coverage problem for DSNs into two subproblems, that is, (1) whether a target area can be barrier covered by the directional sensors and (2) how long a target area can be barrier covered.

On the one hand, whether a target area can be barrier covered is affected by sensor deployment methods. In certain applications, for example, indoor application, sensors are placed with desired locations. In other cases, such as remote or inhospitable environments, sensors are deployed

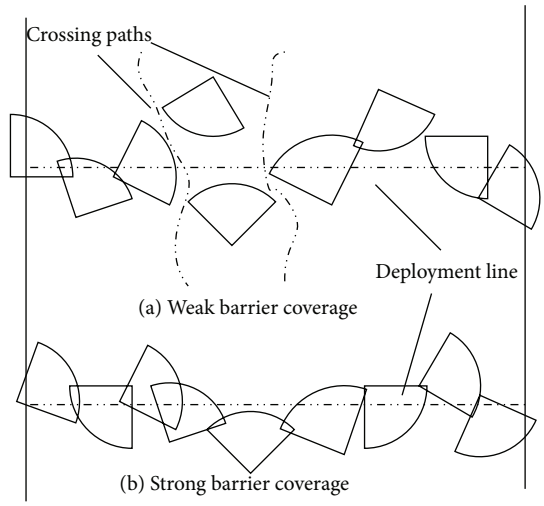


Figure 1: Strong and weak barrier coverage for DSNs.

randomly by airplanes. In determined placement, we need to minimize the number of sensors by optimizing their locations to form a barrier. In random placement, barrier coverage depends on the distribution of sensor locations. Besides, network parameters, such as the number of nodes, the sensing radius of each sensor, and the number of directions per sensor, have important effects on the probability of barrier coverage.

On the other hand, how long a target can be barrier covered is affected by scheduling algorithm with directional sensors. We call a subset of directions of the sensors in which the directions achieve a barrier for the target area as a barrier set. Directional sensor has multiple working directions, but different directions of the same sensor cannot work at the same time. So there is no more than one direction from the same sensor in a barrier set. We may find multiple barrier sets for a directional sensor network. The work time of each barrier set is limited to the remaining work time of sensor in the barrier set. Directional sensor will die when its power is exhausted. To conserve energy, we make different barrier sets work at different times. We leave the necessary sensors which are in a working barrier set in the active state and put other sensors into sleep state. So the target area can be barrier covered continuously by the barrier sets. The network lifetime is the total work time of the barrier sets.

In this paper, we study the energy-efficient barrier coverage problem for directional sensor networks. We consider the following scenario. The target area is a two-dimensional Euclidean plane. A number of directional sensors are deployed side by side along straight line across the region by airplane. After they have been deployed, they will never move. We first describe the sensing model of each sensor and then define ESD and EEBC problems we need to solve. Our contributions are as follows. First, we describe a deployment model for the distribution of sensor locations to solve the ESD problem. We analyze the overlap conditions of the neighboring sensors, and get a function to compute the probability of barrier coverage with some parameters. Then the relationships between the probability of the barrier coverage and network parameters are studied in different deployment scenarios. Our results investigate that the probability of barrier coverage is more sensitive to a number of sensors and sensing radius of each sensor. Second, we propose an energy-efficient scheduling algorithm to prolong the network lifetime when the target area is barrier covered. We model the EEBC as an optimization problem. We introduce a directed graph to model the overlap of directional sensors. Based on this graph, we describe an integer linear programming formulation to find out barrier sets in the network. By solving the optimization problem, we get the optimal solutions, the barrier sets, and their work times. These barrier sets work continuously with their work times to maximize the network lifetime.

The remainder of this paper is organized as follows. In Section 2, we briefly survey the related work on barrier coverage of sensor networks. In Section 3, we define the ESD and EEBC problems. In Section 4, we propose and evaluate a solution to the ESD. In Section 5, we present an energy-efficient algorithm to solve the EEBC and give the simulation results for the EEBC. Finally, we conclude this paper in Section 6.

2. Related Work

In the past years, the barrier coverage problem in wireless sensor networks has been fairly well studied in the literature.

Sensor deployment strategies have direct impact on the barrier coverage of wireless sensor networks. Different deployment strategies can result in significantly different barrier coverage. Kumar et al. [7] defined the notion of barrier coverage of a belt region using wireless sensors. They proposed efficient algorithm to determine whether a region is barrier covered or not after deployment. Then they established the optimal deployment pattern for achieving barrier coverage. In [8], Saipulla et al. established a tight lower bound for the existence of barrier coverage under line-based deployment. Their results provided important guidelines to the deployment and performance of wireless sensor networks for barrier coverage. Balister et al. [9] computed the probability that the nodes provide barrier coverage when they are deployed by the Poisson point process.

Most of the barrier coverage literatures assumed that every sensor node is stationary in the sensor networks. Zhang et al. [10] studied a strong barrier coverage problem for DSNs. They presented an integer linear programming formulation for the problem. Then several efficient centralized algorithms and a distributed algorithm were proposed to solve the problem. Wang and Cao [11] considered the problem of constructing camera barrier in both random and deterministic deployment. They proposed a novel method to select camera sensors to form a camera barrier, which is essentially a connected zone across the monitored field such that every point within this zone is full view covered. In [12], Chen et al. introduced the concept of local barrier coverage and showed that local barrier coverage almost always provides global barrier coverage for thin belt regions.

They also developed a novel sleep-wake-up algorithm to maximize network lifetime. Liu et al. [6] believed quality of barrier coverage is not binary. They proposed a metric for measuring the quality of k-barrier coverage and established theorems that can be used to precisely measure the quality using the proposed metric. In [13], Chen et al. provided theoretical foundations for the construction of strong barriers in a sensor network. They obtained the critical conditions for strong barrier coverage in a strip sensor network. Based on this result, they further proposed an efficient distributed algorithm.

There are some literatures which focus on barrier coverage for sensor networks with mobile sensors. Saipulla et al. [14] studied the barrier coverage with mobile sensors of limited mobility. They first explored the fundamental limits of sensor mobility on barrier coverage and presented a sensor mobility scheme that constructs the maximum number of barriers with minimum sensor moving distance. They further devised an algorithm that computes the existence of barrier coverage under the limited sensor mobility constraint and constructed a barrier if it exits. In [15], He et al. considered the barrier coverage problem where m sensors are needed to guarantee full barrier coverage and there are only n mobile sensors available ($n < m$). They first modeled the arrival of intruders at a specific location as a renew process. Then they proposed two sensor patrolling algorithms to solve the problem: periodical monitoring scheduling (PMS) and coordinated sensor patrolling (CSP). In [16], Keung et al. demonstrated the intrusion detection problem as a classical kinetic theory of gas molecules in physics. By examining the correlations and sensitivity from the system parameters, they derived the minimum number of mobile sensors that needs to be deployed in order to maintain the k-barrier coverage for a mobile sensor network.

Most of existing solutions to barrier coverage problem aim to find as many barrier sets as possible to enhance coverage for the target area. However, power conservation is still an important issue in directional sensor networks. Therefore, in this paper, we will study the energy-efficient barrier coverage problem with directional sensors.

3. Problem Statement

In this section, we give the assumptions and formally define the problems we need to solve, followed by a simple example of the EEBC to briefly describe this problem.

3.1. Assumptions and Problem Definition. In this paper, we consider the following scenario. The target region is a belt region, which is like a two-dimensional rectangular area. Sensors are deployed in the area along the deployment line (as shown in Figure 1) by airplane. After sensors have been deployed, they will never move. We assume that all the sensors have the same sensing radius and angle of view. The size of the angle of view may theoretically change from 0 to 2*pi. So each sensor has multiple working directions, and sensing region of each working direction is a sector. Different directions of the same sensor do not overlap. No more than one direction of each sensor can be active at the same time.

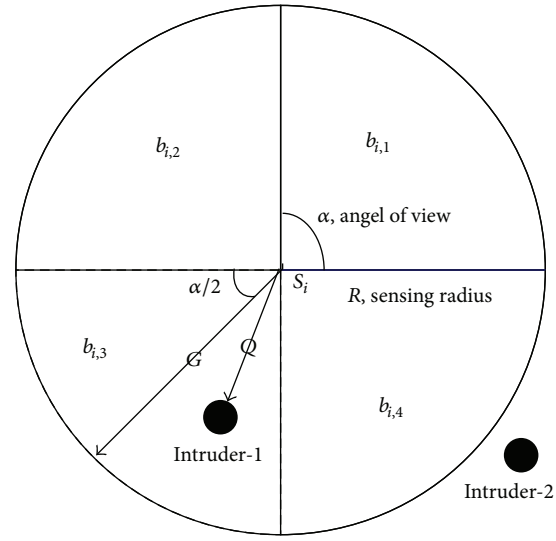


Figure 2: Sensing model of a directional sensor.

We adopt some notations throughout the paper. Let L and W be the length and the width of the target area, respectively. S_i denotes the ith sensor in the network. R denotes the sensing radius of each sensor. We define M as the number of directions of each sensor. $b_{i,j}$ ($i = 1, \ldots, N$, $j = 1, \ldots, M$) denotes the jth direction of ith sensor. Each sensor initially has an equal lifetime, which is defined as T.

Figure 2 is an example to illustrate the sensing model of directional sensor in this paper. The angel of view of sensor S_i is pi/2. So S_i have four directions: $b_{i,1}$, $b_{i,2}$, $b_{i,3}$, and $b_{i,4}$ (as shown in Figure 2). Intruder-1 can be detected by the direction $b_{i,3}$ if it satisfies the two following conditions:

$$\left|\vec{Q}\right| \leq R, \tag{1}$$

$$\vec{G} \cdot \vec{Q} \geq \left|\vec{Q}\right| \cos \frac{\alpha}{2}. \tag{2}$$

Conditions (1) and (2) mean that the intruder is detected by a sensor if it is located anywhere within the active working direction of the sensor. As shown in Figure 2, intruder-1 can be detected by $b_{i,3}$. However, intruder-2 cannot be detected by any direction of S_i.

We provide the following problem definitions.

Definition 1 (ESD (efficient sensor deploy) problem). For a given target area (length L, width W), find out the relationships between the probability of barrier coverage and the deployment parameters: number of nodes (N), sensing radius of each sensor (R), and the number of directions of each sensor (M).

Definition 2 (EEBC (energy-efficient barrier coverage) problem). Find a family of K barrier sets $B_1, B_2, B_3, \ldots, B_k$, with nonnegative work time $t_1, t_2, \ldots, t_k$, such that $t_1 + t_2 + \cdots + t_k$ is maximized, and for each sensor, the total work time is less than the initial lifetime (T).

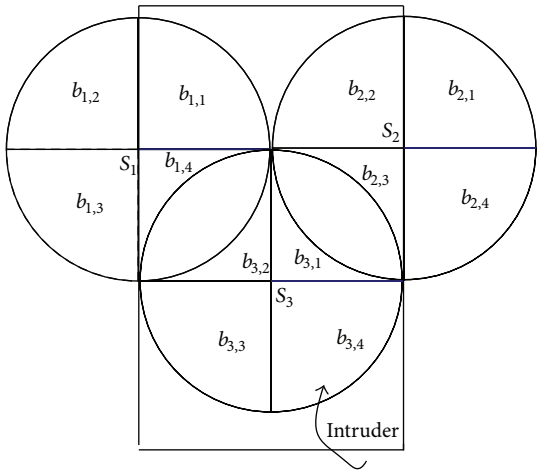


Figure 3: Example of EEBC.

3.2. Example for EEBC. Figure 3 shows a directional sensor network with three sensors (S_1, S_2, and S_3) deployed to detect the intruders who want to cross the belt region. Each sensor has an initial lifetime of 1 (time unit). Each sensor has four working directions (illustrated in Figure 3). For example, $\{b_{1,4}, b_{2,2}\}$ is a barrier set for the network. We can find out multiple barrier sets for this network. If we set any barrier set with the work time of 1, the network lifetime is 1. On the other hand, we can get the following barrier sets: $B_1 = \{b_{1,1}, b_{2,2}\}$ with the work time of 0.5, $B_2 = \{b_{1,4}, b_{3,1}\}$ with 0.5, and $B_3 = \{b_{3,2}, b_{2,3}\}$ with 0.5. The total work time of each sensor is no more than 1. Finally, the network lifetime is maximized to 1.5. This implies that efficient scheduling can prolong the lifetime of the network.

4. Deployment Model and Analysis

In this section, we will solve the ESD problem.

4.1. Deployment Model for DSNs. In this section, we introduce a deployment model for directional sensor networks in this work.

Suppose that N sensors are deployed by the deployment line evenly from the left to the right. We get the location of each sensor. The coordinates (x_i, y_i) of the ith sensor S_i are:

$$x_i = i * \frac{L}{(N+1)} + \gamma_i, \quad i = 1, 2, \ldots, N, \tag{3}$$

$$y_i = \tau_i, \quad i = 1, 2, \ldots, N. \tag{4}$$

Because of mechanical inaccuracy, terrain constraints, wind, and other environmental factors, the sensors cannot be deployed with the location we desired. Therefore, γ_i and τ_i denote the offset of the ith sensor in the horizontal and vertical directions, respectively. And in this model, we assume that the random offsets γ_i and τ_i are independently and identically distributed with a normal distribution of zero mean and δ variance.

In this paper, each directional sensor has different working directions and can detect any intruder within its active direction. And a barrier set is formed by overlapped directions that intersect both the left and right boundaries of the target area. Such a barrier set can guarantee that no intruders can cross the target area without being detected. We consider the following cases to find out the overlap conditions of neighboring sensors, that is, S_1 and S_2 (as shown in Figure 4). Each sensor in this figure has four directions. We define d ($d = |S_2 - S_1|$) as the distance between the two neighboring sensors, S_1 and S_2.

Case 1. Figure 4(a) shows the distance between S_1 and S_2 that is no more than R. In this condition, we can see that $b_{2,3}$ can overlap any direction of S_1 and also $b_{1,1}$ can overlap any direction of S_2. This means that, no matter which direction of S_1 is active, S_2 can find a direction to be active to overlap S_1.

Case 2. Figure 4(b) shows the distance between S_1 and S_2 that is more than R but less than $2R$. In this condition, there are some directions of these two sensors that can overlap each other, like $b_{1,1}$ and $b_{2,2}$. However, there exist some directions, which none directions of the other sensor can overlap. For example, if $b_{1,2}$ is active, none of directions of S_2 can overlap it.

Case 3. Figure 4(c) shows the distance between S_1 and S_2 that is more than $2R$. In this condition, we cannot find any direction from S_1 to overlap the direction from S_2.

Theorem 3. *When N directional sensors are deployed from the left to the right in a target area with the coordinates (x_i, y_i) and the sensing radius R, the probability of the target area can be barrier covered by these sensors given by*

$$P(\textit{barrier}) \geq P(d_0 < R)\, P(d_N < R) \prod_{i=1}^{N-1} P(d_i < R), \tag{5}$$

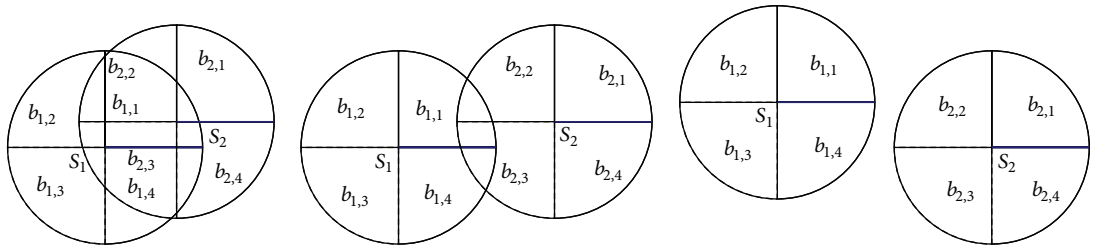
where d_0 is the distance between the sensor S_1 and the left boundary. d_N is the distance between the sensor S_N and the right boundary. d_i is the distance between S_{i+1} and S_i.

Proof. There are N sensors with the following conditions. (1) $d_0 < R$, which means the left boundary can be covered by the sensor. (2) $d_i < R$, which means the neighboring sensors could overlap each other. (3) $d_N < R$, which means the right boundary can be barrier covered. When they satisfy these three conditions, the sensors can form a barrier for the target area. These three conditions are a sufficient condition for barrier coverage with directional sensors. Hence, the theorem holds.

We can get

$$d_i = |S_{i+1} - S_i| = \frac{L}{N+1} + (\gamma_{i+1} - \gamma_i) + \tau_{i+1} - \tau_i. \tag{6}$$

Formulation (6) implies that d_i follows a Ricean distribution [14]. Also, we can easily verify that $P(d_0 < R) > P(d_i < R)$



(a) $d \le R$ (b) $R < d \le 2R$ (c) $d > 2R$

FIGURE 4: Overlap conditions of neighboring sensors.

and $P(d_N < R) > P(d_i < R)$. Therefore, we denote d as the distance between neighboring sensors. We get

$$\begin{aligned} &P\,(\text{barrier}) \\ &\quad \ge P(d < R)^{N+1} \\ &\quad = \left(1 - Q_1\left(\frac{L}{(2N+1)\sqrt{2}\delta}, \frac{R}{\sqrt{2}\delta}\right)\right)^{N+1}, \end{aligned} \tag{7}$$

where Q_1 is Marcum's Q-function of the first order. By formulation (7), we can make a rough calculation of the barrier coverage probability for a directional sensor network with given parameters. □

4.2. Model Analysis and Simulation Results. In this section, we compare our model analysis described in Section 4.1 and simulation results in different scenarios. We study the relationships between the barrier coverage probability and the network parameters, such as the number of sensors (N), sensing radius of each sensor (R), and the number of directions of each sensor (M).

Figure 5 plots the relationship between the probability of barrier coverage and the number of sensors (N) when $L = 2000$ m, $R = 60$ m, and $M = 4$. Figure 6 plots the relationship between the probability of barrier coverage and sensing radius (R) when $L = 2000$ m, $N = 40$, and $M = 4$. In these two figures, we consider three different variances $\delta = 8$, 6, and 4. Figure 7 plots the probability of barrier coverage when M values 3, 4, 6, and 12. In this figure, $L = 2000$ m, $N = 40$, $R = 65$ m, and $\delta = 4$.

As shown in Figures 5, 6, and 7, there is a good match between our model analysis and simulation results. The match improves as the variance δ decreases. Also we can verify that our model analysis is indeed a lower bound than the simulation results.

In Figures 5 and 6, we can see that the probability of barrier coverage increases monotonically as the number of sensors (N) and the sensing radius (R) increase. The reason is that when the number of sensors increases, the distance between neighboring sensors will decrease, which leads to the more probability of barrier coverage. When the sensing radius increases, each direction can overlap more other directions. Therefore, the network has more probability to form a barrier. However, Figure 7 implies that the number of directions of each sensor has an effect on the probability of barrier coverage both in our analysis and simulation results.

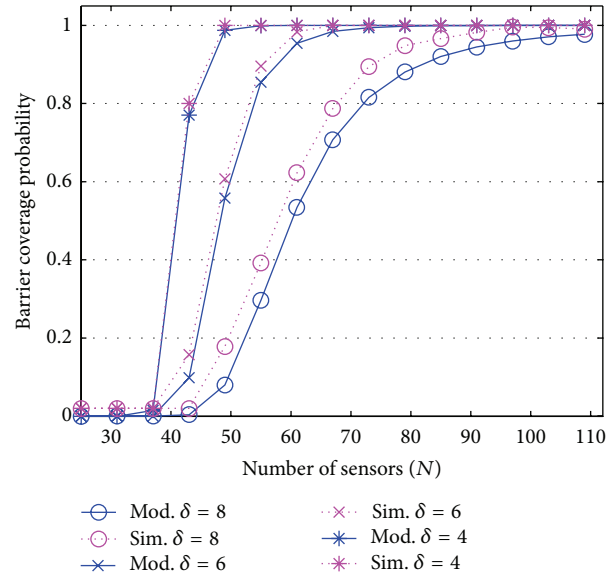


FIGURE 5: Barrier coverage probability versus N, when $L = 2000$ m, $R = 60$ m, $M = 4$.

We can observe from the results above that increasing the number of nodes or sensing radius is an efficient way to improve the probability of barrier coverage. And our analysis can guide the deployment methods and parameters in the realistic application.

5. Energy-Efficient Barrier Coverage

This section focuses on the EEBC problem. By the deployment model and analysis in Section 4, we get several methods to improve the probability of barrier coverage. Next, we discuss how to schedule directional sensors in the network to form energy-efficient barriers.

5.1. Optimization Formulation for EEBC. In this section, we model optimization formulation for EEBC problem. In the EEBC problem, we wish to compute the value L which is

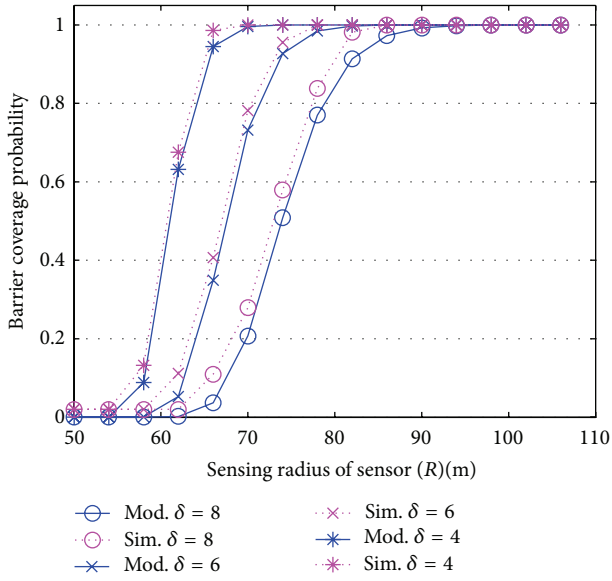


FIGURE 6: Barrier coverage probability versus R, when $L = 2000$ m, $N = 40$, and $M = 4$.

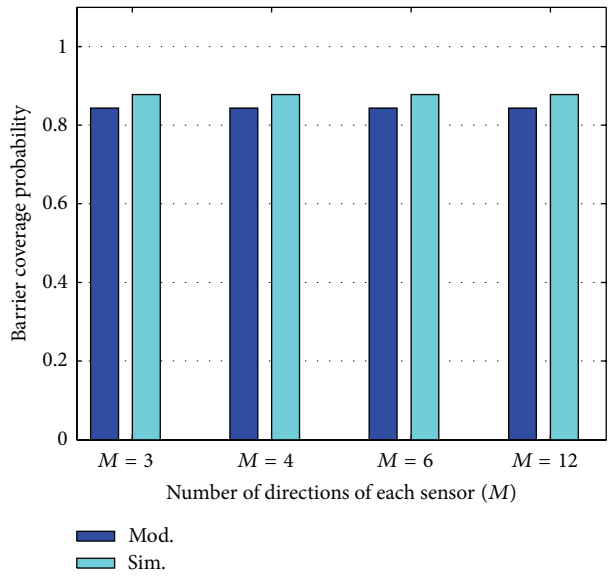


FIGURE 7: Barrier coverage probability versus M, when $L = 2000$ m, $N = 40$, $R = 65$ m, and $\delta = 4$.

the longest lifetime of the network, where the network lifetime is defined as the time duration when the target area is barrier covered by a set of active directions of different sensors. The lifetime of a network is limited by the remaining energy of single sensor which works in the network. When one of the working sensors runs out of energy, the barrier set which it belongs to cannot work.

We assume that N sensors are deployed in a target area to form barriers. The network can find K barrier sets by some algorithm. The kth barrier set is denoted by B_k, which is formed of $b_{i,j}$ $(i = 1, \ldots, N, j = 1, \ldots, M)$. t_k denotes the work time of B_k. We set a Boolean variable $C_{i,j,k}$ as

$$C_{i,j,k} = \begin{cases} 1, & b_{i,j} \in B_k \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Thus, we obtain the following optimization formulation to compute the maximum lifetime of the network:

$$\text{maximize} \quad L = t_1 + t_2 + t_3 + \cdots + t_k \tag{9}$$

subject to

$$\sum_{j=1}^{M} \sum_{k=1}^{K} C_{i,j,k} t_k \le T, \quad i = 1, 2, \ldots, N \tag{10}$$

$$\sum_{j=1}^{M} C_{i,j,k} \le 1, \quad i = 1, 2, \ldots, N, \ k = 1, 2, \ldots, K \tag{11}$$

$$C_{i,j,k} = \{0, 1\}, \quad t_k \ge 0. \tag{12}$$

The objective function (9) maximizes the lifetime of the network. The maximum lifetime is the work time of the network, which meets the constraints (10)–(12) and maximizes the lifetime value. It is the total work time of K barrier sets.

The constraint (10) indicates that each barrier set cannot contain different directions which belonging to the same sensor. No more than one direction of the same sensor can work at the same time. The constraint (11) shows the lifetime constraint for each sensor. The total work time of the M directions belong to the same sensor is no more than the initial lifetime of the sensor. The constraint (12) shows the restrictions on the variables. The variable $C_{i,j,k}$ can be either 1 or 0, which means the direction $b_{i,j}$ works either in the kth barrier or not. The work time t_k cannot be negative.

This optimization formulation is a mixed integer programming problem. If we find the barrier sets of the network, it can be solved as a linear problem.

5.2. Energy-Efficient Scheduling Algorithm. In this section, we consider that N directional sensors are deployed in a target area and the probability of barrier coverage is 1. As described in Section 5.1, we need to find multiple barrier sets and set a reasonable time for each one to prolong the lifetime of the network.

First, we model a directed graph $G(V, E)$ for the DSNs. Vertex v $(v \in V)$ represents directions of each sensor, there is a directed edge between u $(u \in V)$ and v $(v \in V)$ if u and v overlaps and they belong to different sensors. We distinguish two nodes for G, a source s and a sink t. There is an edge directed from s to v if the direction which v is corresponding to covers the left boundary. Also there is an edge directed form v to t if the direction which v is corresponding to covers the right boundary.

The importance of $G(V, E)$ which we modeled above is to reduce the barrier problem to the path problem. We call the path (from s to t) in which the corresponding directions can

form a barrier for network as a barrier path. There are not directions which belong to the same sensor in each barrier path. However, the directions belongs to the same sensor can participate in different barrier paths, because different barrier sets do not work at the same time but continuously to prolong the network lifetime.

In order to find the maximum number of barrier paths, we denote $f(u, v)$ as the flow leaving u to v or each edge (u, v) in G. The initial value of $f(u, v)$ is 0. We describe the following mixed integer linear program formulation to find the barrier paths in the network:

$$\text{maximize} \quad \sum_{v \in V} f(s, v) \tag{13}$$

subject to

$$\sum_{v \in V} f(s, v) = \sum_{v \in V} f(v, t), \tag{14}$$

$$\sum_{u \in V} f(u, v) = \sum_{u \in V} f(v, u), \quad \forall v \in V, \tag{15}$$

$$f(u, v) = \{0, 1\}, \quad v \in V, \ u \in V. \tag{16}$$

The objective function (13) is the maximum flow of G. Based on the computed flow allocation, we can get the maximum number of barrier paths in DSNs. Constraints (14) and (15) indicate the flow-conservation property, which is often informally referred to as "flow in equals flow out." These make sure that each path from s to t we get can form a barrier for the target area. Constraint (16) shows the variable $f(u, v)$ values 0 or 1, which means the corresponding vertex can be either in a barrier path or not.

The Ford-Fulkerson algorithm [15] is used to solve the problem (13)–(16). By this algorithm we can identify the multiple barrier paths in G. DSNs can form a strong barrier by the corresponding directions in the paths to detect intruders. It means we get multiple barrier sets for DSNs. Then we can ease the optimization formulation (described in Section 5.1) for EEBC problem as a linear program problem.

This linear program problem has K variables. Because there are N sensors in the network, it has $N + K$ constrains. The simplex algorithm [15] is the classical method for solving linear programs. In contrast to most of other algorithms, its running time is not polynomial in the worst case. It does yield insight into linear programs, however, and is often remarkably fast in practice. We use the simplex algorithm to solve this linear program. It returns an optimal solution to the linear program. That is K-vector $t = (t_1, \ldots, t_k)$. Finally, we get K barrier sets and the work time of each one. These barrier sets work continuously with their work times to maximize the network lifetime.

5.3. Simulation Results. In this section, we evaluate the performance of our algorithms in simulation methodology. In the simulation, we assume that our algorithm is computed in the sink node. Before the network starts to monitor the target area, the information of all the barrier sets is broadcast to the sensor nodes.

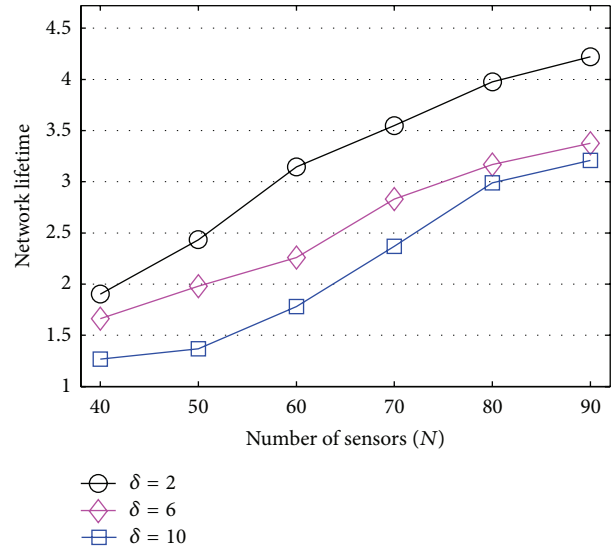


Figure 8: Network lifetime versus N when $L = 1000$ m, $R = 100$ m, $M = 4$.

For simplicity, we assume the initial lifetime of each sensor is unit 1 ($T = 1$). The energy consumed when a sensor switches from one direction to another is omitted in this paper. Therefore, the time duration when the network works is affected by the following parameters: the number of sensors (N), the sensing radius of each sensor (R), and the number of possible directions per sensor (M). We evaluate their impacts by setting them to different values in different scenarios.

Figure 8 shows the network lifetime of our algorithm when N varies from 40 to 90, $L = 1000$ m, $R = 100$ m, and $M = 4$. Figure 9 shows the network lifetime when R varies from 60 to 100, $L = 500$ m, $N = 50$, and $M = 4$. Different curves in these two figures from up to down correspond to the chosen variance, $\delta = 2$, 6, and 10. Each data of simulation results in Figures 8 and 9 is an average over multiple experiments.

As shown in Figures 8 and 9, the network lifetime is sensitive as the number of nodes and sensing radius. Also the variance has small effect on the lifetime. The lifetime increases as δ decreases. In Figure 8, as the number of nodes increases, the total number of directions in the network increases which can form more barrier sets to prolong the lifetime. In Figure 9, the sensing radius increases, and the total number of directions does not increase. But each direction can overlap more other directions; that is, a barrier set can be formed of a less number of directions. So the number of barrier sets increases in the network which prolongs the network lifetime.

We can observe from Figures 8 and 9 that, generally, increasing the number of sensors (N) or sensing radius (R) can prolong the network lifetime. However, Figure 10 shows a special case of the growth of network lifetime when we increase N or R, respectively. In this simulation, the initial parameters for a given network are $L = 500$ m, $N = 40$, $R = 40$ m, $M = 4$ and, $\delta = 2$.

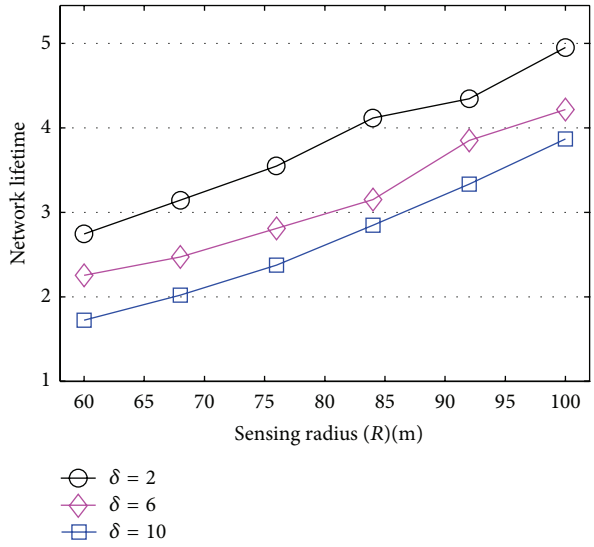


Figure 9: Network lifetime versus R when $L = 500$ m, $N = 50$, $M = 4$.

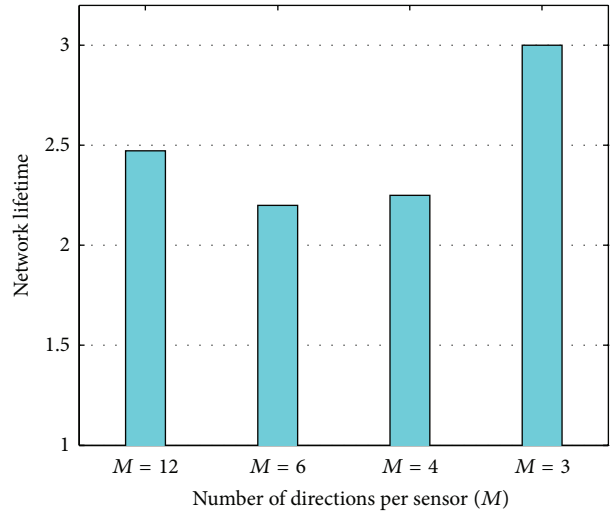


Figure 11: Network lifetime versus M, when $L = 500$ m, $R = 50$ m, $N = 50$, $\delta = 2$.

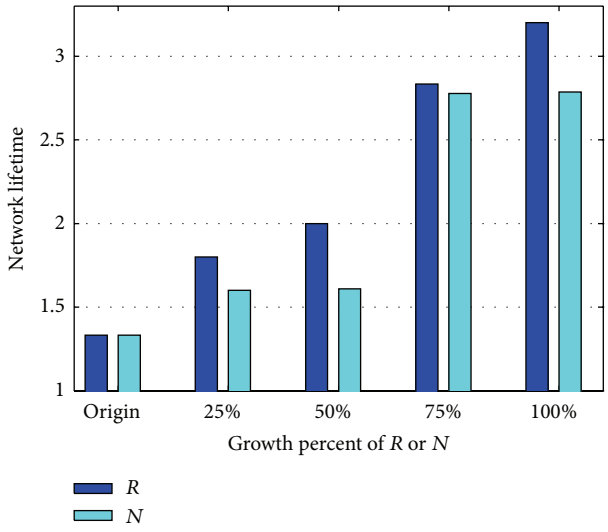


Figure 10: Network lifetime versus R or N for a given network.

In Figure 10, we can see that when we increase the sensing radius, the network lifetime maintains the momentum of steady growth. The reason is that extending the length of sensing radius for a given network means the target area can be barrier covered by a less number of sensors. The number of barrier sets increases, which leads to the network lifetime increases. However, when we increase the number of sensors, the growth ratio of network lifetime is nonsignificant (like 25 percent and 50 percent). When we deploy more sensors for a given network, the location of each sensor is randomly distributed. The characteristic of randomness may lead to the fact that the number of barrier sets cannot be increased with the sensors we added. So the network lifetime could not be extended under such circumstances.

Figure 11 shows the network lifetime when M varies, $L = 500$ m, $R = 50$ m, $N = 50$, and $\delta = 2$. We find out that the network lifetime does not monotonically increase or decrease with the number of directions per sensor. On one hand, the number of directions increases meaning more different directions can overlap, which leads to more barrier sets in the network. This results in the increase of the network lifetime. On the other hand, the number of directions increases meaning the sensing region is smaller than before. The total number of directions each direction can overlap becomes less. More sensors are needed to form barriers, leading to the decrease of the network lifetime.

We can observe from the results above that each data of simulation results is more than 1, which means our proposed algorithms could prolong the network lifetime. Moreover, network lifetime has proportional relationship to the number of sensors and sensing radius, respectively. The number of directions per sensor also has impact on the network lifetime, but it is not proportional to it.

6. Conclusions and Future Work

Sensor deployment provides important effects in wireless directional sensor networks for barrier coverage. And efficient scheduling algorithm can prolong the network lifetime. In this paper, we have studied the problem of what the probability of barrier coverage depends on, the ESD, and the problem of how to schedule the sensors in the network to prolong the lifetime, the EEBC. We have proposed a deployment model to solve the ESD. An optimization formation was described to solve the EEBC. In the future works, we will consider efficient scheduling for other coverage issues, such as area coverage and target coverage.

Conflict of Interests

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant nos. 60673185 and 61073197, the Natural Science Foundation of Jiangsu Province under Grant nos. BK2010548, Scientific and Technological Support Project (Industry) of Jiangsu Province under nos. BE2011186, National Science Fund for Distinguished Young Scholars nos. 70825006, and National Science Fund for Young Scholars nos. 71201052.

References

[1] M. Rahimi, R. Baer, O. I. Iroezi et al., "Cyclops: in situ image sensing and interpretation in wireless sensor networks," in *Proceedings of ACM SenSys*, pp. 192–204, San Diego, Calif, USA, November 2005.

[2] W. Feng, E. Kaiser, W. C. Feng, and M. L. Baillif, "Panoptes: scalable low-power video sensor networking technologies," *ACM Transactions on Multimedia Computing Communication and Applications*, vol. 1, no. 2, pp. 151–167, 2005.

[3] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys '04)*, pp. 214–226, Baltimore, Md, USA, November 2004.

[4] M. Amac Guvensan and A. Gokhan Yavuz, "On coverage issues in directional sensor networks: a survey," *Ad Hoc Networks*, vol. 9, no. 7, pp. 1238–1255, 2011.

[5] D. G. Costa and L. A. Guedes, "The coverage problem in video-based wireless sensor networks: a survey," *Sensors*, vol. 10, no. 9, pp. 8215–8247, 2010.

[6] B. Liu, O. Dousse, J. Wang, and A. Saipulla, "Strong barrier coverage of wireless sensor networks," in *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '08)*, pp. 411–419, Hong Kong, China, May 2008.

[7] S. Kumar, T. H. Lai, and A. Arora, "Barrier coverage with wireless sensors," in *Proceedings of ACM MobiCom*, pp. 284–298, Cologne, Germany, September 2005.

[8] A. Saipulla, C. Westphal, B. Liu, and J. Wang, "Barrier coverage of line-based deployed wireless sensor networks," in *Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM '09)*, pp. 127–135, Rio de Janeiro, Brazil, April 2009.

[9] P. Balister, B. Bollobas, A. Sarkar, and S. Kumar, "Reliable density estimates for coverage and connectivity in thin strips of finite length," in *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking (MobiCom' 07)*, pp. 75–86, Montreal, Canada, September 2007.

[10] L. Zhang, J. Tang, and W. Zhang, "Strong coverage with directional sensors," in *Proceedings of IEEE GLOBECOM*, pp. 1816–1821, Hilton Hawaiian Village, USA, December 2009.

[11] Y. Wang and G. Cao, "Barrier coverage in camera sensor networks," in *Proceedings of the 12th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '11)*, no. 12, Paris, France, May 2011.

[12] A. Chen, S. Kumar, and T. H. Lai, "Designing localized algorithms for Barrier coverage," in *Proceedings of ACM MobiCom*, pp. 63–74, Montreal, Canada, September 2007.

[13] A. Chen, T. H. Lai, and D. Xuan, "Measuring and guaranteeing quality of barrier coverage for general belts with wireless sensors," *ACM Transactions on Sensor Networks*, vol. 6, no. 1, article 2, 2009.

[14] A. Saipulla, B. Liu, G. Xing, X. Fu, and J. Wang, "Barrier coverage with sensors of limited mobility," in *Proceedings of the 11th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '10)*, pp. 201–210, Chicago, Ill, USA, September 2010.

[15] S. He, J. Chen, X. Li, X. Shen, and Y. Sun, "Cost-effective barrier coverage by mobile sensor networks," in *Proceedings of IEEE INFOCOM*, pp. 819–827, Orlando, Fla, USA, March 2012.

[16] G. Y. Keung, B. Li, and Q. Zhang, "The intrusion detection in mobile sensor network," *IEEE/ACM Transactions on Networking*, vol. 20, no. 4, pp. 1152–1161, 2012.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 390936, 9 pages
http://dx.doi.org/10.1155/2013/390936

Research Article

Optimal Placement of Wireless Sensor Nodes for Bridge Dynamic Monitoring Based on Improved Particle Swarm Algorithm

Ting You,[1] Huixia Jin,[2] and Peijiang Li[3]

[1] *College of Electrical and Information Engineering, Quzhou University, Zhejiang 324000, China*
[2] *College of Communication and Electronic Engineering, Hunan City University, Hunan 413000, China*
[3] *School of Information Engineering, Quzhou College of Technology, Zhejiang 324000, China*

Correspondence should be addressed to Ting You; youtingmail@shu.edu.cn

Received 6 September 2013; Revised 17 November 2013; Accepted 20 November 2013

Academic Editor: Gelan Yang

The issue of optimal placement of wireless sensor networks (WSNs) is one of the major challenges for dynamic monitoring of bridges. It should be solved based on combining the effective monitoring of the dynamic performance of a bridge with the energy consumption of WSNs. Thus, the relationship between the bridge modal and the energy consumption of wireless networks is derived. Optimizing sensor location is achieved by using the improved wavelet particle swarm algorithm, which overcomes the disadvantage of the conventional particle swarm algorithm by applying mutation operator to the selected particles with certain small probability. Based on the comparison between the ANSYS simulation results and the measured bridge data, as well as the energy distribution diagram of WSNs, the proposed method is shown to reflect the dynamic performance of the test bridge well. Moreover, the method effectively controls the energy consumption of WSNs and achieves reasonable optimization effects. Therefore, this method can be used for dynamic monitoring of wireless sensors in various bridge types.

1. Introduction

With the rapid development of bridge construction, the safety of the bridges is attracting more and more attention. Once completed and used, the structure or constructional elements of a bridge suffer from continuous degradation, aging of material performance, and external factors. Vehicles, wind, earthquakes, fatigue, overload, and human behaviors are several external factors that damage bridges. Therefore, monitoring and evaluating the physical conditions of bridges to determine their health information are essential. The dynamic monitoring of a bridge structure has become the primary technological means for maintaining the safety of such structures [1–3].

The dynamic monitoring of a bridge structure is based on sensor nodes which collect data on the bridge, such as current structural strains, stresses, and loads. These data are then transmitted to the monitoring center by using a wireless sensor network (WSN). In general, if more sensor nodes are used in a monitoring system, then the structural characteristics of a bridge can be described more accurately. However, because of limits on sensor quantity, economic cost, installation difficulty, and other aspects, the rationality in the number and positioning of sensor node layout is regarded as the premise for guaranteeing quality of structural monitoring [4, 5].

The characteristics of the WSN should also be considered in the actual layout process of its nodes. Excessive sensor nodes will result in oversized transmission data from the WSN. Consequently, an overload in network energy consumption will lead to delayed search speed and poor network stability [6]. Therefore, the optimal layout of sensor nodes is considered as a multiobjective optimization issue based on the aforementioned contents.

Scholars and experts worldwide have studied the optimization issue of sensor layout in bridges. Huang and Zhu considered the maximum nondiagonal elements in a MAC matrix as the objective function and optimized sensor layout

according to the contribution of the optional positions of sensors to the MAC matrix [7]. Shan et al. introduced an improved adaptive genetic algorithm for the optimal layout of sensors [8]. Sun et al. adopted classical optimization criteria and the improved genetic algorithm in the form of fitness to obtain optimal layout [9]. Zhang et al. adopted the Pareto genetic algorithm and designed corresponding genetic operators and coding scheme for solving the optimal position of sensors under double criteria [10].

Based on previous researches, sensor layout optimization mainly focuses on a single optimization objective that reflects the static or dynamic characteristics of bridges. Therefore, this paper presents optimal layout conditions of sensor nodes and the optimal conditions of energy consumption of WSNs. This work also established an improved sensor node optimal layout algorithm based on the improved wavelet particle swarm optimization (PSO). The result of the ANSYS simulation analysis was compared with the measured data of the bridge structure. The energy distribution of the WSN was obtained. Lastly, the validity of the optimal layout was verified.

2. Analysis of the Sensor Layout Model for Bridge Monitoring

2.1. Bridge Modal Analysis. Bridge modal analysis is effective in the dynamic monitoring and overall evaluation of the structural state of a bridge [11].

The dynamic equilibrium equation of a bridge can be simplified as

$$EI\frac{\partial^4 u}{\partial x^4} + m\ddot{y} = p(x,t). \tag{1}$$

This dynamic equilibrium equation belongs to nonlinear dynamic systems, wherein u represents the dynamic deflection of the beam, EI is a constant that represents the flexural rigidity of the beam, m represents beam quality per unit length, and $p(x,t)$ represents the dynamic load. Suppose that the corresponding vibration model function of the beam is $\varphi_k(x)$, the generalized coordinate $q_k(t)$ is introduced, and the vibration mode superposition method is applied. Then, the dynamic deflection $u(x,t)$ of the beam under forced vibration can be expressed in the form of a vibration mode series. This vibration mode series can be described as follows:

$$y(x,t) = \sum_{k=1}^{\infty} \varphi_k(x)\, q_k(t). \tag{2}$$

By substituting formula (2) into formula (1) and simplifying the resulting equation, the dynamic response of the beam can be obtained as follows:

$$u(x,t) = \frac{2F}{ml}\sum_{k=1}^{\infty}\frac{1}{\omega_k^2 - \Omega_k^2}\left(\sin\Omega_k t - \frac{\Omega_k}{\omega_k}\sin\omega_k t\right)\sin\frac{k\pi x}{l}, \tag{3}$$

where $\omega_k = (k\pi/l)^2\sqrt{EI/m}$ represents the natural vibration frequency of the bridge, $\Omega_k = k\pi v/l$ represents the generalized disturbance frequency, v represents the average passing speed of vehicles, l represents the beam span, and F represents the applied force of vehicles on the bridge.

The measurement of the dynamic property of a bridge primarily depends on numerous strain gauges. These strain gauges are distributed throughout the bridge space and connected with WSNs. Suppose that $\{\varepsilon\}$ represents the strain of the position to be measured, n represents the quantity of the strain sensor nodes, k represents the modal number, ψ_j represents the strain modal vector at the jth order, $\Psi \in R^{n\times j}$ represents the strain modal vector matrix, q_j represents the strain modal coordinate at the jth order, and $\{q\} \in R^{k\times j}$. Then, the strain of the bridge can be expressed as follows:

$$\{\varepsilon\} = \sum_{l=1}^{m}\psi_l q_l = \Psi\{q\}. \tag{4}$$

The least square solution of formula (5) is given as

$$\{\overline{q}\} = \left[\Psi^T\Psi\right]^{-1}\Psi^T(\varepsilon). \tag{5}$$

Suppose that $\{\varepsilon\} = \Psi\{q\} + \{\gamma\}$, where $\{\gamma\}$ represents the Gaussian white noise with the variance of σ^2. Then, the covariance of $\{\overline{q}\}$ and $\{q\}$ can be expressed as follows:

$$\begin{aligned}[W] &= E\left[(\{q\} - \{\overline{q}\})(\{q\} - \{\overline{q}\})^T\right] = \left[\sigma^2\Psi^T\Psi\right]^{-1} \\ &= \frac{1}{\sigma^2}[Q]^{-1}.\end{aligned} \tag{6}$$

Suppose that $\|Q\| = \|\Psi^T\Psi\|$. When 2-norm Q reaches the maximum value, the covariance $[W]$ becomes the minimum. Then, good estimated results of the strain modal coordinate q_j can be obtained [12–14], that is, good optimal layout results of sensor nodes.

As shown in formulas (4) to (6), the applied force F on the bridge in formula (3) can only be obtained when the strain modal coordinate q_j and the number of strain sensor nodes n are determined. The dynamic response of the bridge can then be obtained.

The estimate of the optimal strain modal coordinate q_j, or the required and reasonable optimal layout for the sensor nodes of the bridge structure, can be determined via bridge modal analysis. However, the node layout is not only about single-objective optimization in the actual layout process of WSN nodes. Node layout also considers the characteristics of WSN nodes. The number of actual sensor nodes in the monitoring system is limited. Excessive sensor nodes may result in oversized transmission data from the WSN. Consequently, an overload in network energy consumption may result in delayed search speed and poor network stability. Therefore, establishing an optimized and energy-efficient WSN structure is extremely important.

2.2. Energy Optimization Analysis of the WSN. In this research, the dynamic monitoring system of the structural state of the bridge for collecting the dynamic signals of the bridge is based on WSN nodes. Considering the numerous and large-scale network nodes, the entire network is divided

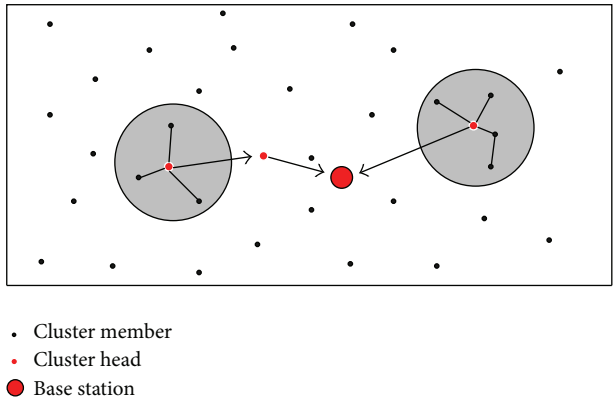


Figure 1: Transmission and retransmission of signals.

into small subnetworks called clusters. The manager of each cluster is called the cluster head. To select the cluster head with proper quantity and optimal position [15, 16], the entire WSN adopts the hierarchical topology establishment method with optimal global energy consumption based on location information. In this manner, the WSN will have a cluster hierarchical topological structure (Figure 1), which can reduce the overall energy consumption of the network and extend its life cycle [17].

The following is the description for the computation of the overall energy consumption.

Suppose that L represents the length of the bridge and M represents the width of the bridge. The WSN nodes are randomly distributed in the $L \times M$ area. $(x_{\mathrm{CH}}, y_{\mathrm{CH}})$ represents the coordinate of the cluster head node, wherein all cluster member nodes CP are in the Voronoi diagram, and all nodes are uniformly distributed [18]. Suppose that the quantity of cluster heads in the $L \times M$ region is n_{CH}; then the average coverage region of each cluster head node can be given as $\sqrt{(L \times M)/n_{\mathrm{CH}}} \times \sqrt{(L \times M)/n_{\mathrm{CH}}}$.

Similarly, suppose that $(x_{\mathrm{CP}}, y_{\mathrm{CP}})$ represent the coordinates of the cluster member node CP which are also randomly distributed. In the bridge deck region of $L \times M$, the mathematical expectations of the cluster head CH and the cluster points CP are given as follows:

$$E\left[x_{\mathrm{CH}}\right] = E\left[y_{\mathrm{CH}}\right] = E\left[x_{\mathrm{CP}}\right] = E\left[y_{\mathrm{CP}}\right] = \frac{1}{2} \times \sqrt{\frac{L \times M}{n_{\mathrm{CP}}}}, \tag{7}$$

$$E\left[x_{\mathrm{CH}}^2\right] = E\left[y_{\mathrm{CH}}^2\right] = E\left[x_{\mathrm{CP}}^2\right] = E\left[y_{\mathrm{CP}}^2\right] = \frac{L \times M}{3n_{\mathrm{CH}}}. \tag{8}$$

The location of the gateway node is set at (0, 0). Then, the mathematical expectations of the cluster head CH and the gateway node BS at a particular distance are given as follows:

$$E\left[d_{\mathrm{avg(CH,BS)}}^2\right] = E\left[\left(x_{\mathrm{BS}} - x_{\mathrm{CH}}\right)^2 + \left(y_{\mathrm{BS}} - y_{\mathrm{CH}}\right)^2\right] = \frac{2}{3} \times L \times M. \tag{9}$$

Suppose that the cluster distribution function is as follows:

$$\rho\left(x, y\right) = \frac{n_{\mathrm{CH}}}{L \times M}. \tag{10}$$

Then, the mathematical expectations of cluster member nodes CP and cluster head CH on a squared distance are given as follows:

$$E\left[d_{\mathrm{avg(CP,CH)}}^2\right] = \iint \left(x^2 + y^2\right) \rho\left(x, y\right) dx\, dy = \frac{L \times M}{2\pi n_{\mathrm{CH}}}. \tag{11}$$

The consumed energy E_{CP} of cluster member nodes CP can be expressed as follows:

$$\begin{aligned} E_{\mathrm{CP}} &= n_{\mathrm{CP}} \times \left(E_{\mathrm{elec}} + E_{\mathrm{fs}} \times d^2\right) \\ &= n_{\mathrm{CP}} \times \left(E_{\mathrm{elec}} + E\left[d_{\mathrm{avg(CP,CH)}}^2\right]\right) \\ &= \left(n - n_{\mathrm{CH}}\right) \times \left(E_{\mathrm{elec}} + E_{\mathrm{fs}} \times \frac{L \times M}{2\pi n_{\mathrm{CH}}}\right). \end{aligned} \tag{12}$$

E_{CH} has three parts: E_{RX} (the energy that receives data from cluster member nodes), E_{pro} (the energy that processes data), and E_{TX} (the energy that sends data to the gateway BS), as shown in the following:

$$\begin{aligned} E_{\mathrm{CH}} &= E_{RX} + nE_{\mathrm{pro}} + E_{TX} \\ &= \left(n - n_{\mathrm{CH}}\right) \mathrm{E}_{\mathrm{elec}} + nE_{\mathrm{pro}} \\ &\quad + \alpha n_{\mathrm{CH}} \left(E_{\mathrm{elec}} + E_{mp} \times \frac{4}{9} \times L^2 \times M^2\right). \end{aligned} \tag{13}$$

The total energy EP consumed during the topology formation process of the WSN is given in formula (14). EP primarily consists of two parts: E_{CP} (the consumed energy of the cluster points) and E_{CH} (the consumed energy of the cluster head). Consider the following:

$$\begin{aligned} EP = E_{\mathrm{CP}} + E_{\mathrm{CH}} &= \left(n - n_{\mathrm{CH}}\right) \times \left(E_{\mathrm{elec}} + E_{\mathrm{fs}} \times \frac{L \times M}{2\pi n_{\mathrm{CH}}}\right) \\ &\quad + \left(n - n_{\mathrm{CH}}\right) E_{\mathrm{elec}} + nE_{\mathrm{pro}} \\ &\quad + \alpha n_{\mathrm{CH}} \left(E_{\mathrm{elec}} + E_{mp} \times \frac{4}{9} \times L^2 \times M^2\right). \end{aligned} \tag{14}$$

As shown in formula (14), the number of the optimal cluster head nodes n_{CH} and their sets can be obtained when the EP is at its minimum. The consumed energy of the sensor network is also at its minimum during this moment.

As stated earlier, the optimal layout of sensors in bridge monitoring is essentially a multiobjective optimization issue. The number and locations of sensors do not only influence the test accuracy of several-order modals but also affect the energy and data transmission of the WSN. The present study adopts a linear weighting method to design the evaluation function. This weighting method can be described as follows:

$$y = \lambda_1 \left[W\right] + \lambda_2 \cdot EP, \tag{15}$$

where $\lambda_1 + \lambda_2 = 1$ (λ_1 and λ_2 represent the weighting coefficients). Suppose that $\lambda_1 = 0.65$ and $\lambda_2 = 0.35$. Then, focus is put on the accuracy degree of the modal in the optimization. The optimal solution method is used to obtain the number of sensors and their coordinate positions. In this manner, the multiobjective optimization issue of sensor nodes can be solved.

3. Optimal Solution of the Layout Based on Wavelet Particle Swarm Algorithm

The wavelet particle swarm algorithm belongs to an intelligent optimization algorithm based on population search which has been developed in recent years. Derived from the simulation of bird population behaviors, this algorithm is simple and easy to perform; moreover, no gradient information is required and fewer parameters are needed. The present work used the improved particle optimization algorithm [19] to obtain optimal solutions.

Suppose that a population consisting of N particles exists in an objective search space of D dimension, where particle i represents the vector of D dimension, as shown in the following:

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{iD}), \quad i = 1, 2, \ldots, N. \tag{16}$$

The flying speed of particle i is expressed as follows:

$$V_i = (v_{i1}, v_{i2}, \ldots, v_{iD}), \quad i = 1, 2, \ldots, N. \tag{17}$$

The optimal position of particle i that has been searched thus far is described as follows:

$$p_{\text{best}} = (p_{i1}, p_{i2}, \ldots, p_{iD}), \quad i = 1, 2, \ldots, N. \tag{18}$$

The optimal position of the entire particle swarm that has been searched thus far is the global extremum, as shown in the following:

$$g_{\text{best}} = (p_{g1}, p_{g2}, \ldots, p_{gD}). \tag{19}$$

Particles update their speeds and positions according to the following:

$$V_{id}^{k+1} = \omega \cdot V_{id}^{k} + c_1 \cdot r_1 \cdot \left(P_{id}^{k} - X_{id}^{k}\right) + c_2 \cdot r_2 \cdot \left(P_{gd}^{k} - X_{id}^{k}\right), \tag{20}$$

$$x_{id} = x_{id} + v_{id}, \tag{21}$$

where ω represents acceleration factors c_1 and c_2 and r_1 and r_2 represent the uniform random numbers within the scope of $[0, 1]$.

To improve the performance of the basic PSO algorithm and to avoid premature convergence of the algorithm, the mutation operation in genetic algorithms is added to the former. Hence, an improved PSO algorithm is obtained. Suppose that the mutation probability of each particle is $p \in [0, 1]$, wherein value is determined by the dimensions of the particles. Suppose that $x_i^k = (x_{i1}^k, x_{i2}^k, \ldots, x_{ij}^k)$ represents the selected ith mutation particle x_{ij}^k at the kth time iteration. Then, $px_{\max i}$ in the jth dimension of the mutation particle represents the upper limit of the search space. Similarly, $px_{\min i}$ is the lower limit of the mutation particle in the search space. The mutation formula is given as follows:

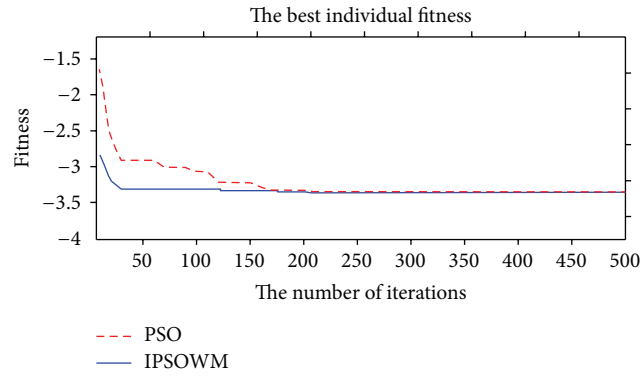


Figure 2: The change in fitness values during evolution.

$$\text{mut}\left(x_{ij}^k\right) = \begin{cases} x_{ij}^k + \sigma \cdot \left(px_{\max i} - x_{ij}^k\right), & \text{if } \sigma > 0, \\ x_{ij}^k + \sigma \cdot \left(x_{ij}^k - px_{\min i}\right), & \text{if } \sigma \le 0, \end{cases} \tag{22}$$

where σ represents the value of the wavelet function. Then, the Morlet wavelet function is selected and its calculation formula is given as follows:

$$\sigma = \frac{1}{\sqrt{a}} \exp\left(-\frac{\Phi^2}{2a^2}\right) \cos\left(5\left(\frac{\Phi}{2}\right)\right), \tag{23}$$

where the value range of Φ falls between $[-2.5a, 2.5a]$. The calculation formula for a is given as follows:

$$a = \exp\left(-\ln(g) \times \left(\frac{k}{\text{num}}\right)^{\xi\omega m}\right) + \ln(g), \tag{24}$$

where $\xi\omega m$ represents the shape parameter, k represents the number of iterations, g represents the upper limit of a, and num represents the maximum number of iterations. The wavelet mutation formula, that is, formula (22), is improved to obtain the following:

$$\text{mut}\left(x_{iD}^k\right) = \begin{cases} gb_D^k + \sigma \times \left(px_{\max i} - gb_D^k\right), & \text{if } \sigma > 0, \\ gb_D^k + \sigma \times \left(gb_D^k - px_{\min i}\right), & \text{if } \sigma \le 0, \end{cases} \tag{25}$$

where gb_D^k represents the optimal solution of the population at the kth iteration. The difference between formulas (22) and (25) is obvious. In formula (22), one particle is selected in a certain probability for wavelet mutation and this particle may fail to jump out of the local optimal node after mutation. As a result, the global optimal point will not be found. Meanwhile, formula (25) performs wavelet mutation operation on the optimal solutions of the current population, and then transmits the value after mutation to the current particle selected

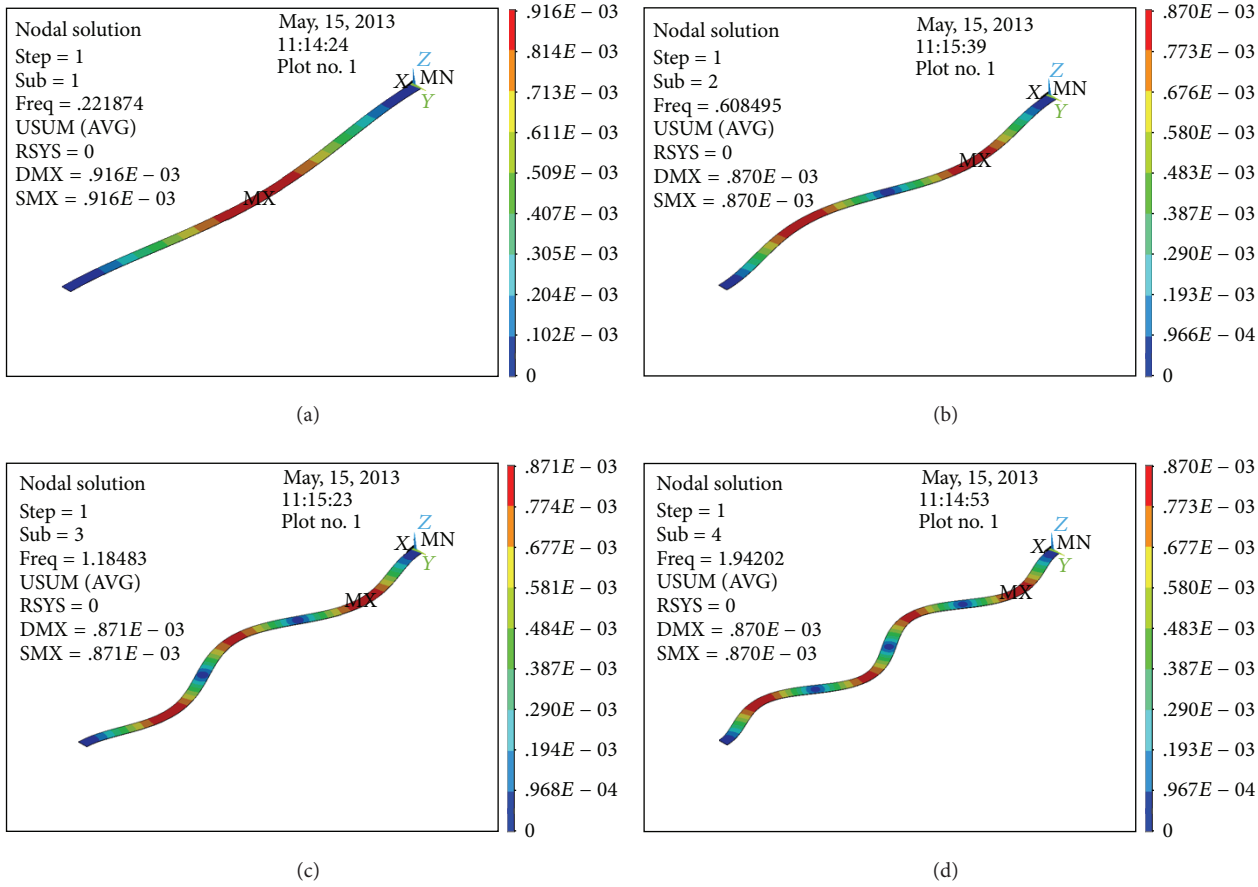


Figure 3: The calculation result of dynamic characteristic of bridge: (a) first order, (b) second order, (c) third order, and (d) fourth order.

in a certain probability. As such, formula (25) can conduct an effective search around the current optimal solution. This formula can also guide the particle swarm to move forward the global optimal point and to reduce the local optimum possibility of the particles in each iteration.

Based on the preceding analysis, the fitness function is shown in formula (15), $y = \lambda_1[W] + \lambda_2 \cdot EP$. Supposing that $\lambda_1 = 0.65$, $\lambda_2 = 0.35$, and the acceleration factor in formula (20) is $c_1 = c_2 = 1.4945$; then the number of population particles is 80, and the evolution iteration times of the algorithm are 500. Changes in the fitness values during the evolution process are shown in Figure 2. The red dotted line in the figure represents the curve from the fundamental form of the PSO algorithm. The blue solid line represents the curve from the improved wavelet mutation PSO algorithm. Based on the experimental results, the improved algorithm is superior to the fundamental form of PSO in terms of convergence rate.

Based on the relationship between cluster head nodes and member nodes of the aforementioned fitness function, the number of cluster head nodes is confirmed to be 12 and that of sensor nodes is 130. All nodes are discretely distributed throughout an 800 M bridge deck. Several sensor coordinates are shown in Table 1.

Table 1: The coordinates of several sensors.

No.	X-coordinate (m)	Y-coordinate (m)
1	−367.210	−3.61
2	−310.031	1.691
3	−227.368	4.712
4	136.811	−1.067
5	53.161	2.343
6	0.123	−0.214
7	55.324	−0.061
8	144.377	−2.347
9	231.331	−1.920
10	371.019	−0.341

4. Simulation and Experimental Analysis

In actual bridge dynamic tests, installing any number of sensor nodes to acquire the dynamic parameters of the entire bridge is not feasible. The acquisition of dynamic parameters is subject to limitation under various conditions. The finite element analysis method can be used to obtain dynamic situations under certain conditions. This research used the

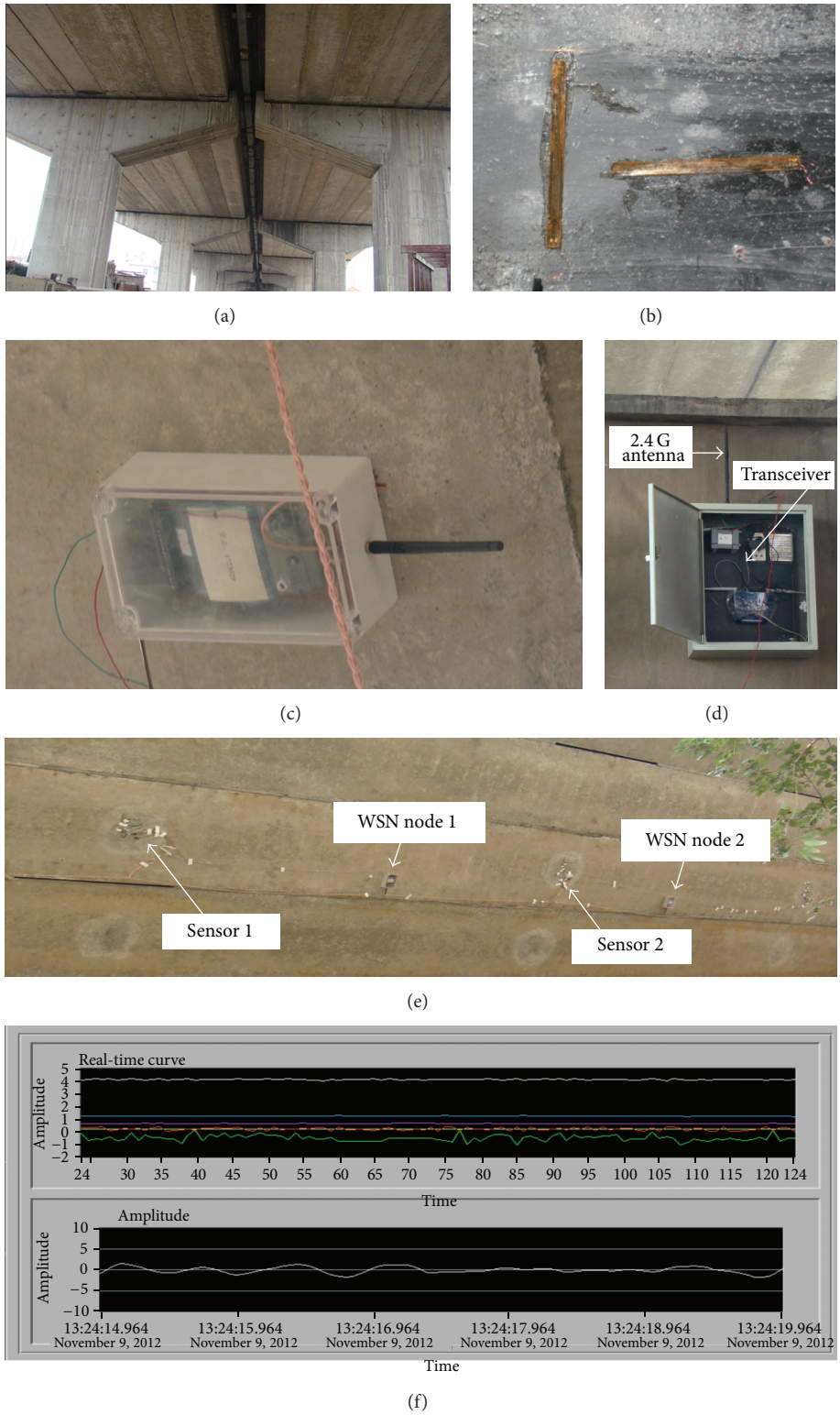


Figure 4: (a) The underbridge condition of the Xiazhang Highway Bridge, (b) strain gauges, (c) WSN nodes, (d) the collection node box, (e) installed network nodes, and (f) the interface of data collection.

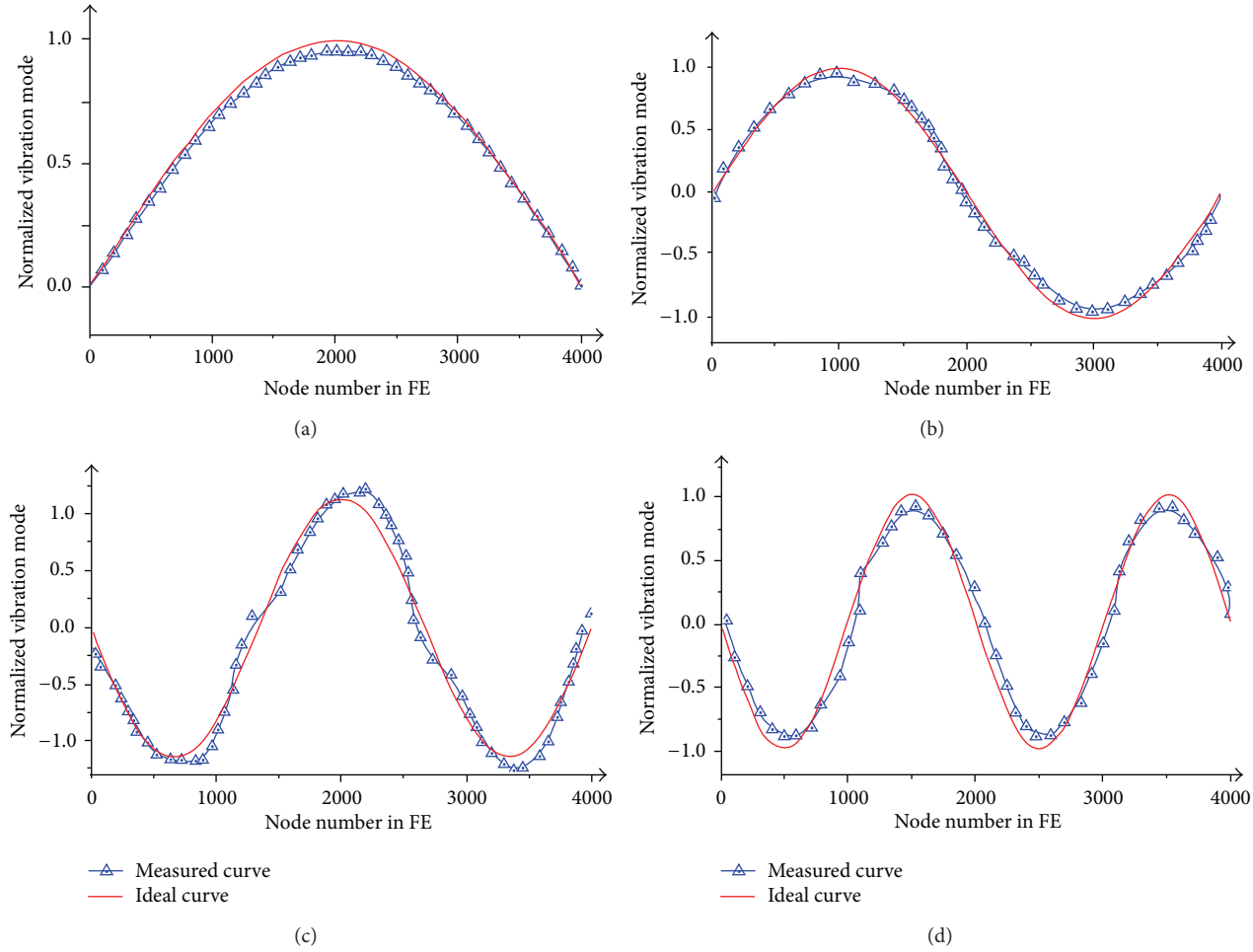


Figure 5: Comparison of ANSYS analytical curves with measured experimental data. (a) The first-order mode. (b) The second-order mode. (c) The third-order mode. (d) The fourth-order mode.

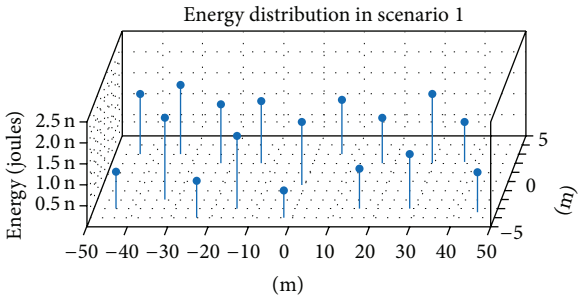


Figure 6: Energy distribution in scenario 1.

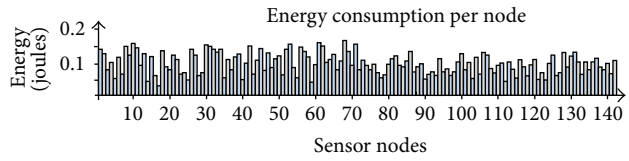


Figure 7: Energy consumption per node.

finite element software ANSYS to analyze the intrinsic mode of a bridge measuring 800 M long and 10 M wide. The SOLIDE185 was also adopted to establish the model. The elasticity modulus was $E = 30$ GPa and Poisson's ratio was $\nu = 0.2$. The model was divided into 4,000 units and 8,822 finite element nodes, which focused on the in-plane vertical bending vibration of the bridge deck. Therefore, the concerned target was a vertical vibration modal. The initial four-order vertical vibration data were selected as the to-be-detected modal vibration data. The analysis results are shown in Figure 3. The obtained vibration data can be compared with the actual measured values.

The test object in this paper is the Xiazhang Highway Bridge in Quzhou. The condition of its underbridge is shown in Figure 4(a). Based on the position relationship between cluster head nodes and cluster member nodes, the positions of 130 sensor nodes and 12 cluster head nodes discretely distributed throughout the 800 M long bridge deck were determined in this study. Free-scale MMA7361 micromechanical acceleration sensors were used to detect vibration data on the aforementioned positions (surface of the underbridge).

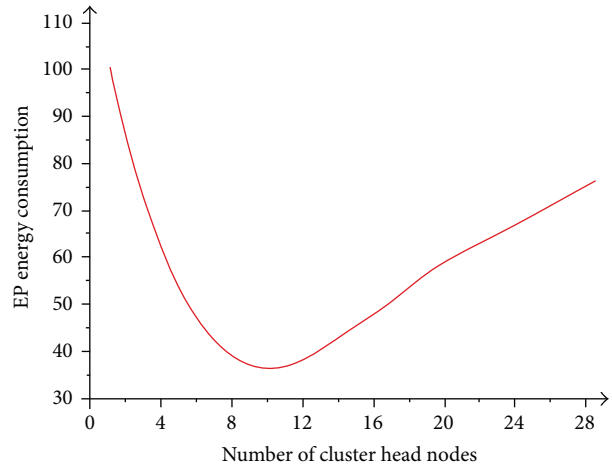


Figure 8: Relationship between energy consumption of sensor network and number of cluster head nodes.

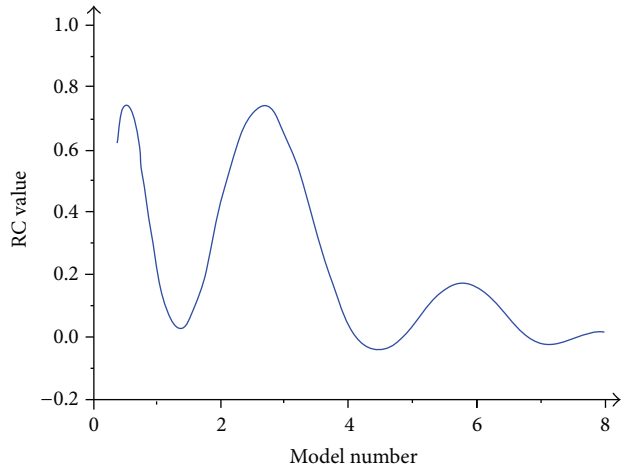


Figure 9: Relationship between RC value and modal number.

BX120-100AA strain gauges were placed on these points for the strain test. The specific bounded strain gauge condition is shown in Figure 4(b). The WSN nodes are shown in Figure 4(c). As shown in Figure 4(d), the collection nodes have two functions. One function is used to receive data from the nodes in the network, while another function is used to transmit data to the remote monitoring center by 3G communication network. Figure 4(e) shows the installed network nodes and Figure 4(f) illustrates the interface of data collection. The tested data implemented normalization processing. The cubic-spline method *rounding* was used for the fitted curve. After comparison with the ANSYS analysis results, the initial four-order modal shape diagram was obtained (Figure 5).

The goodness of fit between the ANSYS analysis results and the test results is relatively high, thus indicating that the distribution points have relatively high accuracy. Among the vibration modes, the first two-order vibration modes can accurately reflect the actual vibration mode despite the few measuring points. Orders 3 and 4 vibration modes decrease as the number of measuring points decreases and their goodness of fit becomes poorer. The accuracy test for the higher-order vibration modes requires an increase in the number of sensor nodes.

In the regular operation of each sensor node (with a 250 kbps transmission rate), 1 byte needs 4 μs and the transmission power of the node is 0.0004 W. When the two values are multiplied, 1.6 nJ is obtained. In this manner, the consumed energy of each sensor node per second can be measured. Figure 6 shows the energy consumption diagram of each node. Figure 7 shows the energy distribution diagram.

The relationship between energy consumption and cluster nodes is calculated by formula (15) (Figure 8). According to the fitness function, the number of cluster nodes, that is, 12, is close to the minimum value of the energy consumption curve. This result illustrates the rationalization of the proposed optimization.

Based on formula (6), the change rate of the first $i+1$ and i orders can be expressed as follows:

$$\mathrm{RC} = \frac{\|Q_{i+1}\|_2 - \|Q_i\|_2}{\|Q_i\|_2}, \quad 1 \le i \le n-1, \tag{26}$$

where i represents the number of computing modals and n represents the number of all modals. The curve is shown in Figure 9. Formula (26) is significant because the first i order modal information can represent itself when the increment of the two norms of the Q matrix in the first $i+1$ order modal becomes zero or a relatively small value. The first i order modal basically contains the modal information of the entire bridge information, thus providing relatively complete modal information of the optimal layout of the sensors. The modal becomes flat after seven orders, thus demonstrating that most modal information of the bridge is contained within seven orders (Figure 9). Therefore, the initial four-order modal that was selected can reveal the major dynamic parameters of the bridge.

5. Conclusion

In this paper, an optimization procedure was put forward and was applied on the dynamic monitoring project of Xiazhang Highway Bridge in Quzhou. The experimental results indicated that the accuracy of the bridge modal measurement could be improved by using a multiobjective optimal layout of sensor nodes. The entire network energy could also be reduced. On the other hand, the improved wavelet particle swarm algorithm improved the convergence rate of the optimization process. However, bridge modal information in an ideal situation was acquired by means of ANSYS simulation. The actual modal data of the bridge must have some deflection with the simulation data. The proposed method has practical value and theory value. It can be analogized to the monitoring issues of WSNs for other types of bridges.

Conflict of Interests

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

Acknowledgments

This study is financially supported by the Public Welfare Project of the Science and Technology Department of Zhejiang Province (2012C31013), the Teacher Development Project of the Visiting Scholars in Colleges and Universities of the Education Department of Zhejiang Province (FX2012090), and the Faculty Development Fund of Quzhou University (XNZQN201111).

References

[1] S. W. Doebling, C. R. Farrar, and M. B. Prime, "A summary review of vibration-based damage identification methods," *Shock and Vibration Digest*, vol. 30, no. 2, pp. 91–105, 1998.

[2] F. N. Catbas, D. L. Brown, and A. E. Aktan, "Use of modal flexibility for damage detection and condition assessment: case studies and demonstrations on large structures," *Journal of Structural Engineering*, vol. 132, no. 11, pp. 1699–1712, 2006.

[3] Z. Li, A. Li, and X. Han, "Operational modal identification of suspension bridge based on structural health monitoring system," *Journal of Southeast University*, vol. 25, no. 1, pp. 104–107, 2009.

[4] O. S. Salawu, "Detection of structural damage through changes in frequency: a review," *Engineering Structures*, vol. 19, no. 9, pp. 718–723, 1997.

[5] G. Movva, Y. Wan, S. Fu et al., "Optimal sensor placement for structural health monitoring: a comparative study between the control engineering and civil engineering approaches," in *SPIE Smart Structures and Materials and Nondestructive Evaluation and Health Monitoring*, International Society for Optics and Photonics, 2013.

[6] J. Wu, S. Yuan, X. Zhao, Y. Yin, and W. Ye, "A wireless sensor network node designed for exploring a structural health monitoring application," *Smart Materials and Structures*, vol. 16, no. 5, pp. 1898–1906, 2007.

[7] M.-S. Huang and H.-P. Zhu, "Serial method for optimal placement of sensors and its application in modal tests of bridge structure," *Bridge Construction*, no. 5, pp. 80–83, 2007.

[8] D.-S. Shan, Z.-H. Wang, and Q. Li, "Sensors optimal placement for long-span steel truss railway cable-stayed bridges," *Journal of Civil, Architectural, Environmental Engineering*, no. 1, pp. 18–21, 2011.

[9] X.-D. Sun, H.-W. Li, and J.-P. Ou, "Application of Improved genetic algorithm in optimal sensor placement for dynamic test of Harbin Sifangtai Bridge," *Journal of Xi'an University of Architecture and Technology*, vol. 38, no. 5, pp. 624–628, 2006.

[10] L.-Z. Zhang, Q. Huang, and C.-H. Wang, "Optimal sensor placement based on multi-object genetic algorithm," *Engineering Mechanics*, vol. 24, no. 4, pp. 168–172, 2007.

[11] F. Magalhães, Á. Cunha, and E. Caetano, "Structural health monitoring based on automated operational modal analysis: application to an arch bridge," in *Earthquakes and Health Monitoring of Civil Structures*, pp. 241–268, Springer Netherlands, 2013.

[12] Y. Q. Li, Z. H. Xiang, M. S. Zhou, and Z. Z. Cen, "Optimal sensor placement for mode-shape based damage identification on bridges," *Journal of Tsinghua University*, vol. 50, no. 2, pp. 312–315, 2010.

[13] Z.-Y. Shuai, "Research on Optimization of Sensor Placement for High-Speed Railway Bridge Health Monitoring System," Southwest Jiaotong University, Chengdu, China, 2008.

[14] R. He, "Modal Identification and Its Error Analysis and Damage Detection in Structural Monitoring Systems of Long-Span Bridges," Tsinghua University, Beijing, China, 2009.

[15] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, 2002.

[16] Z. Liu, "Research on Topology Construction Algorithm and Application Technology of Wireless Sensor Network," Huazhong University of Science and Technology, Wuhan, China, 2011.

[17] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[18] J. M. Ko and Y. Q. Ni, "Technology developments in structural health monitoring of large-scale bridges," *Engineering Structures*, vol. 27, no. 12, pp. 1715–1725, 2005.

[19] D.-H. Gao, P. P. Dong, Y.-B. Tian, and H.-T. Zhou, "An improved particle swarm optimization algorithm with wavelet mutation," *Computer Engineering*, vol. 38, no. 21, pp. 145–147, 2012.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 691042, 9 pages
http://dx.doi.org/10.1155/2013/691042

Research Article

Random Walk Based Location Prediction in Wireless Sensor Networks

Zhaoyan Jin, Dianxi Shi, Quanyuan Wu, and Huining Yan

College of Computer, National University of Defense Technology, Changsha 410073, China

Correspondence should be addressed to Zhaoyan Jin; jinzhaoyan@163.com

Received 3 September 2013; Revised 5 November 2013; Accepted 19 November 2013

Academic Editor: Yong Jin

With the development of wireless sensor network (WSN) technologies, WSNs have been applied in many areas. In all WSN technologies, localization is a crucial problem. Traditional localization approaches in WSNs mainly focus on calculating the current location of sensor nodes or mobile objects. In this paper, we study the problem of future location prediction in WSNs. We assume the location histories of mobile objects as a rating matrix and then use a random walk based social recommender algorithm to predict the future locations of mobile objects. Experiments show that the proposed algorithm has better prediction accuracy and can solve the rating matrix sparsity problem more effectively than related works.

1. Introduction

Recently, WSNs have been applied in many areas, such as environmental monitoring [1], target tracking [2], and intrusion detection [3]. However, in all of these areas, location is very important, for data without location in WSNs will be useless. Research of localization is a hot topic in WSN, and it includes two kinds, that is, localization of sensor nodes in the WSN itself and localization of mobile objects using a WSN. For localization of sensor nodes in the WSN itself, several anchor nodes with known location are selected, and locations of other sensor nodes can be calculated with predefined anchor nodes [4]. For localization of mobile objects using a WSN, all sensor nodes are assumed to be anchor nodes, and the location of mobile objects can be calculated with observed anchor nodes [5]. The most popular localization system of this kind is the global position system (GPS).

Given a WSN, where all sensor nodes are anchor nodes, the locations of mobile objects can be calculated with the WSN. Furthermore, if we save all location records of mobile objects, then we can predict future locations of mobile objects with the observed location histories and the WSN. In this paper, we study the problem of future location prediction in WSNs. Here, we assume a WSN as a connected network, where each sensor node has connections with all its neighbors. The reason is that if a mobile object visits location A, it will probably visit A's neighborhood. In a WSN, each node records the number of visits of all mobile objects to itself. Hence, we have a network of sensor nodes and a rating matrix of sensor nodes on all mobile objects. The purpose of this paper is to predict which mobile object will probably visit the desired locations or sensor nodes in the future.

2. Related Works

If we assume each sensor node as a user and each mobile object as an item or a commodity, then the number of visits from a mobile object to a sensor can be assumed to a $|\text{User}| \times |\text{Item}|$ rate matrix and the sensor network becomes a $|\text{User}| \times |\text{User}|$ graph, and thus the location prediction problem can be transformed into a social recommender problem. In this section, we review related works on recommender systems.

2.1. Traditional Recommender Systems. As mentioned by Huang et al. [6], collaborative filtering (CF) method is one of the most commonly used and successfully deployed methods in recommender systems. The CF recommender method can be classified into memory based and model based [3, 4].

The memory-based CF method is heuristic, and it aims to find a neighborhood of similar users [7, 8] or items [9–11] for recommendation. The user-based method predicts ratings

based on ratings of neighborhood users, and the item-based method predicts ratings based on ratings of neighborhood items. Here, the similarity between users or items could be Pearson correlation coefficient [12], vector space similarity [13], cosine similarity [14], and so on.

In model-based CF method, one can build a model by the priori |User| × |Item| matrix and predict unknown User-Item ratings with the model [15, 16]. The clustering models divide the matrix into several smaller clusters and use a neighborhood cluster to predict unknown user-item ratings [17, 18]. If the user-item ratings belong to different categories, then classification models can be applied [19]. Here, the priori |User| × |Item| matrix, where each element belongs to a category, is assumed to be training data. With the training data, classification model can be used to predict unknown ratings.

Recently, due to its efficiency in dealing with large datasets, several matrix factorization methods [20–24] have been proposed. The matrix factorization methods factor the |User|×|Item| matrix into two or three low-rank matrixes and employ these low-rank matrixes to make further predictions. The purpose of matrix factorization is to minimize the sum-squared errors of all elements in the priori |User| × |Item| matrix. Given a $n \times m$ rating matrix $R_{n\times m}$, the Singular Value Decomposition (SVD) [20, 21] is to factor the rating matrix $R_{n\times m}$ into three low-rank matrixes, that is, $R_{n\times m} \simeq A_{n\times k}\boldsymbol{\lambda}_{k\times k}B_{k\times m}$, where $k \ll n, m$. First, one can factor the rating matrix $R_{n\times m}$, such that $R_{n\times m} = A_{n\times n}\boldsymbol{\lambda}_{n\times m}B_{m\times m}$, where $\boldsymbol{\lambda}_{n\times m}$ is a positive semidefinite diagonal matrix and $\lambda_{i\times i} \geq \lambda_{j\times j} \geq 0$ for all $0 \leq i \leq j \leq \min(m, n)$, and then truncate $A_{n\times n}$, $\boldsymbol{\lambda}_{n\times m}$, and $B_{m\times m}$ into $A_{n\times k}$, $\boldsymbol{\lambda}_{k\times k}$, and $B_{k\times m}$, respectively, while keeping the k biggest singular values. Another matrix factorization (MF) method [22–24] is factoring the rating matrix $R_{n\times m}$, such that $R_{n\times m} = U_{n\times k}V_{k\times m}$ $(k \ll n, m)$.

2.2. Trust-Based Recommender Systems. Traditional recommender systems have been well studied and developed in both academia and industry, but they all assume that users are independent and identically distributed and ignore the relationships between users. However, in practice, recommendations from different users should have different weights. For example, a preference or advice from a close friend should be more trustable than that of a stranger. Based on this idea, many researchers have recently begun to analyze trust-based recommender systems [25–30]. Trust-based recommender systems use both the rating matrix and the network of users to predict unknown ratings.

TidalTrust [25] applies the social network to find neighbors and aggregates their ratings weighted by the trusts between the source user and his neighbor. The searching of neighbors is based on breadth first search, which finds the nearest raters in the social network, and the trust between the source user and a rater is aggregating the trust value between source user's direct neighbors and the rater, weighted by the trust values between the source user and its direct neighbors. MoleTrust [26] is similar to TidalTrust, except that the computation of trust between the source user and a rater is based on direct neighbors of the raters.

Golbeck and Hendler [27] studied how trust information can be mined and integrated into social-rating systems. They defined trust as "trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome," presented two algorithms for inferring trust relationships between individuals that are not directly connected in the network, and proposed a prototype email client—TrustMail, which could filter spam for an email system.

TrustWalker [28] is a combination of item-based CF and trust-based recommendation, and it considers both ratings of similar items by trustable friends and ratings of the exact item by further neighbors in the social network. The random walk model of TrustWalker is as follows. Starting from the source user u, at step k and at node v, if v has rated i, then it returns to the rating $r_{v,i}$; with the probability $\Phi_{v,i,k}$, the random walk stops, randomly chooses an item j rated by v, and returns to $r_{v,j}$; and with the probability $1 - \Phi_{v,i,k}$, the random walk goes to a direct neighbor of v. If $\Phi_{v,i,k} = 0$, TrustWalker is a pure trust-based recommender system; if $\Phi_{v,i,k} = 1$, the random walk never starts, and thus TrustWalker becomes a traditional item-based CF recommender.

The works in [25–28] are all heuristic or memory based. However, model-based recommender methods, such as SocialTrustEnsemble [29] and SocialMF [30], can also be used in social recommender systems. Both SocialTrustEnsemble and SocialMF apply the matrix factorization method, and they factor both the |User| × |User| matrix and the |User| × |Item| matrix into two low-rank matrixes. While choosing the neighborhood users, although they both use direct neighbors of the source user, they have different weighting methods. SocialTrustEnsemble gives different weights to ratings from different neighborhood users, whereas SocialMF assumes that the weight of a rating recursively depends on the neighbor's neighbors.

2.3. Other Types of Recommendation. Besides recommending items for users, the social recommender systems can also recommend users [31–33], communities [34, 35], and tags [36, 37], as well as all kinds of recommendations for a group of users [38–40]. For more details about social recommender systems, readers can refer to [5].

3. SocialSlopeOne Location Prediction Algorithm

3.1. Problem Definition. In this paper, we consider each sensor node as a user, each mobile object as an item, and the number of visits from a mobile object to a sensor as the rating of user (sensor node) on that item (mobile object), and then we have a |User| × |User| graph and a |User| × |Item| rating matrix.

Assuming a WSN with n sensor nodes and m mobile objects, the WSN network $G = (U, E)$ can be represented as an $n \times n$ adjacency matrix M_G,

$$m_{G,ij} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

Table 1: Notations.

Notation	Description
n	Number of sensor nodes
m	Number of mobile objects
G	WSN graph
M_G	Adjacency matrix of G
P_G	Transition probability matrix of G
R	Rating matrix
$r_{i,j}$	The rating of user i on item j
$\widehat{r}_{i,j}$	The predicted rating of user i on item j
M_R	Unweighted adjacency matrix of R
P_R	Transition probability matrix of R
No_i	Neighbor nodes of outgoing edges for i in R
Ni_i	Neighbor nodes of ingoing edges for i in R
π_i	Stable distribution vector starting from i
$s_{i,j}$	Similarity between user nodes i and j by the random walk model

or a transition probability matrix P_G,

$$p_{G,ij} = \frac{m_{G,ij}}{|\mathrm{No}_i|}, \tag{2}$$

where No_i is the set of nodes linked by outgoing edges of node i and $|\mathrm{No}_i| = \sum_{j=1}^{n} m_{G,ij}$ is the cardinality.

Moreover, the rating matrix R is $n \times m$, and

$$r_{i,j} = \begin{cases} r_{i,j}, & \text{if user } i \text{ has rated item } j \\ \text{null}, & \text{otherwise.} \end{cases} \tag{3}$$

If we consider R as an unweighted graph, then the adjacency matrix of R is M_R,

$$m_{R,ij} = \begin{cases} 1, & \text{if } r_{ij} \neq \text{null} \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

and the transition probability matrix of R is P_R,

$$p_{R,ij} = \frac{m_{R,ij}}{|\mathrm{No}_i|}, \tag{5}$$

where $|\mathrm{No}_i|$ is the same as (2).

Given a WSN with a network G and a rating matrix R, the purpose of the paper is to predict unknown ratings $\widehat{r}_{i,j} = f(i, j)$ in R for $r_{i,j} = \text{null}$, where f is built with G and R. Table 1 lists all notations used in the paper.

3.2. Overall Recommender System. While predicting an unknown rating of user i on j, our overall recommender system works as follows.

(1) It constructs a social-rating graph with the WSN graph and the rating matrix. Here, we merge the common nodes of users in the two graphs. Figure 1 is an example of a social-rating graph, and Table 2 is its rating matrix, where u_i $(1 \leq i \leq 7)$ are users and i_i $(1 \leq i \leq 5)$ are items.

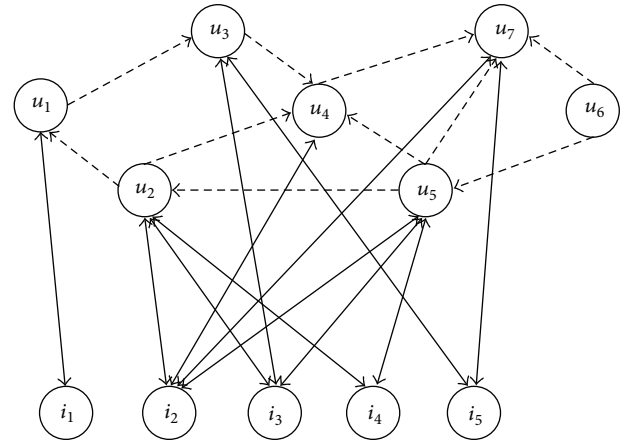


Figure 1: An example of social-rating graph.

Table 2: The rating matrix.

Users	Items				
	i_1	i_2	i_3	i_4	i_5
u_1	2				?
u_2		4	3	5	
u_3			4		5
u_4		5			
u_5		5	3	4	
u_6				?	
u_7		4	?		4

(2) While predicting the rating $\widehat{r}_{i,j}$ of user i on j, we compute the random walk based similarity, defined in Section 3.3, between user i and the other users.

(3) Once we get the similarities between i and other users, we can weigh the ratings of each user based on its similarity to i. With the rating matrix, we design a social slope one algorithm to predict the rating $\widehat{r}_{i,j}$, and this is described in Section 3.4.

3.3. Random Walk Based Similarity. While predicting the rating $\widehat{r}_{i,j}$ of user i on j, we can compute the similarities between user i and all other users, that is, $U - \{i\}$, with a random walk with restart model. The main idea of the model is as follows:

(i) starting from the source node i;

(ii) at each step k and at user node u:

(a) with probability ε, return to i;

(b) with probability $1-\varepsilon$, continue the random walk:

(1) firstly, randomly walk to a direct neighbor user u' of u in G;

(2) secondly, randomly walk to a direct neighbor item i of u' in R;

(3) thirdly, randomly walk to a direct neighbor user u'' of i in R.

The fingerprint of a random walk is a chain of nodes, which is $\langle \text{user}, \underbrace{\langle \text{user, item, user} \rangle}_{1}, \underbrace{\langle \text{user, item, user} \rangle}_{2}, \ldots, \underbrace{\langle \text{user, item, user} \rangle}_{k} \rangle$. After the kth steps, the random walk would stop at a user node. For example, $\langle u_1, \langle u_3, i_3, u_5 \rangle, \langle u_4, i_2, u_2 \rangle \rangle$ is a fingerprint on Figure 1 with $k = 2$. We use the stable distributions, which are the probabilities of stopping at all user nodes, to represent the similarities between the source user i and others and call them random walk based similarities. For a source user i, the random walk based similarity vector π_i is

$$\pi_i = (1 - \varepsilon) P_R P_R^T P_G^T \pi_i + \varepsilon \delta_i, \tag{6}$$

where the superscript T is the transpose of a matrix, and δ_i is a personalized vector defined as follows:

$$\delta_i (j) = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \tag{7}$$

3.4. The Social Slope One Algorithm. As the slope one algorithm is an effective CF algorithm in recommender systems, we apply the idea of slope one for predicting unknown ratings. The slope one algorithm tries to find slope one suggestions in the rating matrix and uses slope one to predict unknown ratings. However, in a social recommender system, the rating matrix is sparse, and the slope one algorithm cannot predict ratings for users who do not have any slope one suggestion.

In order to solve the sparsity of the rating matrix, we propose a social slope one algorithm. While predicting a rating for user u on i, our proposed algorithm works as follows.

3.4.1. Predicting Ratings for New Users. In social recommender systems, a new user is a person that has not rated any item yet, that is, $\text{No}_u = \emptyset$. In this situation, we can take suggestions for the specified item from other users. However, the suggestions from different users should have different weights. Here, we use the average of ratings, which other users have rated for the specified item, weighted by their similarities to the active user,

$$\widehat{r}_{u,i} = \sum_{v \in \text{Ni}_i} s_{u,v} \cdot r_{v,i}, \tag{8}$$

where $s_{u,v}$ is the normalized similarity between u and v,

$$s_{u,v} = \frac{\pi_u (v)}{\sum_{w \in \text{Ni}_i} \pi_u (w)}. \tag{9}$$

Taking Table 2, for example, u_6 is a new user and $\widehat{r}_{6,4}$ is the average of $r_{2,4}$ and $r_{5,4}$ weighted by $s_{6,2}$ and $s_{6,5}$, respectively, that is, $\widehat{r}_{6,4} = s_{6,2} \cdot r_{2,4} + s_{6,5} \cdot r_{5,4}$.

3.4.2. Predicting Ratings for Non-Cold-Start Users. While predicting the rating $\widehat{r}_{u,i}$ of user u on i, if user u is not a new user and the item i is not a new item, then we try to construct slope one suggestions from the rating matrix, so that the slope one algorithm can be used to predict unknown ratings.

Input: Ni_i and Ni_j
Output: $\langle \text{diff}_r, \text{dis}_r \rangle$ for $1 \leq r \leq \min(|\text{Ni}_i|, |\text{Ni}_j|)$
(1) For $r = 1$ to $\min(|\text{Ni}_i|, |\text{Ni}_j|)$ do
(2) Find $p \in \text{Ni}_i, q \in \text{Ni}_j$, such that min dis$(p, q)$;
(3) $\text{diff}_r = r_{p,i} - r_{q,j}$, $\text{dis}_r = \frac{s_{u,p} + s_{u,q}}{2}$;
(4) Let $\text{Ni}_i = \text{Ni}_i - \{p\}$, $\text{Ni}_j = \text{Ni}_j - \{q\}$;
(5) End for;
(6) Return $\langle \text{diff}_r, \text{dis}_r \rangle$ for $1 \leq r \leq \min(|\text{Ni}_i|, |\text{Ni}_j|)$;

Algorithm 1: SocialDiff(Ni_i, Ni_j).

If all items that u has rated have been rated only by u itself, that is, $\text{Ni}_j = \{u\}$ for all $j \in \text{No}_u$, then we can treat $\widehat{r}_{u,i}$ as a new user problem, and use (8) to predict the rating. Taking user u_1 and item i_5 in Table 1, for example, $\text{No}_1 = \{i_1\}$ and $\text{Ni}_1 = \{u_1\}$, so $\widehat{r}_{1,5} = s_{1,3} \cdot r_{3,5} + s_{1,7} \cdot r_{7,5}$.

Otherwise, if $\text{Ni}_j \neq \{u\}$ for at least one $j \in \text{No}_u$, then we can construct slope one suggestions from the rating matrix, so that the slope one algorithm can be used. For each item $j \in \text{No}_u$, which has been rated by others besides the user u, we construct slope one suggestions using Ni_i and Ni_j. While constructing the rating differences between items i and j, at each time, we select one user p from Ni_i and one user q from Ni_j and minimize the following distance metric:

$$\text{dis}(p, q) = \frac{|s_{u,p} - s_{u,q}|}{s_{u,p} \cdot s_{u,q}}. \tag{10}$$

Equation (10) means that it is better to choose two ratings for items i and j from users similar to each other, and moreover, both of them are similar to the active user. If users p and q are the same user, then $\text{dis}(p, q) = 0$ is the smallest, and thus the slope one suggestion between i and j is $r_{p,i} - r_{p,j}$ and the weight is $(s_{u,p} + s_{u,q})/2 = s_{u,p}$. If users p and q are not the same user, then the slope one suggestion is $r_{p,i} - r_{q,j}$, and the weight is $(s_{u,p} + s_{u,q})/2$. For Ni_i and Ni_j, we can construct $\min(|\text{Ni}_i|, |\text{Ni}_j|)$ slope one suggestions and add them together weighted by the normalized distance, so the predicted rating $r_{u,i}^j$ using the slope one algorithm is

$$r_{u,i}^j = \sum_{r=1}^{\min(|\text{Ni}_i|, |\text{Ni}_j|)} \frac{\text{dis}_r}{\sum_{s=1}^{\min(|\text{Ni}_i|, |\text{Ni}_j|)} \text{dis}_s} \cdot \text{diff}_r + r_{u,j}, \tag{11}$$

where the difference diff_r and its weight dis_r ($1 \leq r \leq \min(|\text{Ni}_i|, |\text{Ni}_j|)$) are calculated by Algorithm 1.

Then, for all $j \in \text{No}_u$ such that $\text{Ni}_j \neq \{u\}$, we add the predicted rating $r_{u,i}^j$ together weighted by $\min(|\text{Ni}_i|, |\text{Ni}_j|)$, that is, $w_{u,i}^j = \min(|\text{Ni}_i|, |\text{Ni}_j|)$, and then $\widehat{r}_{u,i}$ can be calculated by the following equation:

$$\widehat{r}_{u,i} = \sum_{j \in \text{No}_u \wedge \text{Ni}_j \neq \{u\}} \frac{w_{u,i}^j}{\sum_{k \in \text{No}_u \wedge \text{Ni}_k \neq \{u\}} w_{u,i}^j} r_{u,i}^j. \tag{12}$$

Table 3: Statistics of datasets.

Dataset	No. of users	No. of edges	No. of items	No. of ratings	No. of new users
Epinions	49290	487183	139738	664824	7739
Flixster	786936	7058819	48794	8196077	36360

In Table 2, if we want to predict the rating $\widehat{r}_{7,3}$, then $\mathrm{No}_7 = \{i_2, i_5\}$, $\mathrm{Ni}_2 = \{u_2, u_4, u_5, u_7\}$, and $\mathrm{Ni}_5 = \{u_3, u_7\}$. If we assume that $s_{7,k}$ $(1 \le k \le 6)$ are equal, then $r^2_{7,3} = 4+((3-4)+(4-5)+(3-5))/3 = 8/3$, $w^2_{7,3} = 3$, $r^5_{7,3} = 4 + ((4-5)/1) = 3$, $w^5_{7,3} = 1$, and finally, we can have $\widehat{r}_{7,3} = (3/(1+3))\cdot(8/3)+(1/(1+3))\cdot 3 = 2.75$.

3.5. Complexity Analysis. In the problem of location prediction, our approach include two steps, that is, computing nodes' similarities with the random walk model and computing recommender results with the social slope one algorithm.

Theorem 1. *The computational complexity of the random walk based similarity, defined by formula* (6)*, is* $k(n^2 + 2mn)$*, where* k *is the number of iterations.*

Proof. The computation of formula (6) is the process of iterations, and the number of iterations decides the precision of the results. Here, we assume the number of iterations is k. At each iteration, there are three matrix-vector multiply operations, and their computational complexities are $n \times m$, $m \times n$, and $n \times n$, respectively. So, the computational complexity of formula (6) is $k(n^2 + 2mn)$. □

Theorem 2. *The computational complexity of the social slope one algorithm is* $(1/3)\cdot(r/n)\cdot(r/m)^4$*, where* r *is the number of nonnull elements in the rating matrix* R*.*

Proof. We assume that the rating matrix R contains r nonnull elements, and then each item is rated by about r/m users and each user rates about r/n items. While constructing slope one suggestions from one item, the SocialDiff$(r/m, r/m)$ algorithm iterates r/m times (in line 1). At each iteration, the algorithm needs i^2 $(1 \le i \le r/m)$ time (from line 2 to line 4), that is, $(r/m)\cdot((r/m)+1)\cdot(2(r/m)+1)/6$. So, the computational complexity of the social slope one algorithm is $(1/3)\cdot(r/n)\cdot(r/m)^4$. □

From Theorems 1 and 2, we can have the following corollary.

Corollary 3. *The computational complexity of our random walk based social slope one algorithm is* $k(n^2 + 2mn) + (1/3)\cdot(r/n)\cdot(r/m)^4$.

4. Experiments

In this section, we perform experiments on two datasets, report the experimental results, and compare the results with existing CF algorithms. As the acquirement of sensor network data is very difficult, we use two social network data instead. However, both of them have similar characteristics with WSN data.

4.1. Datasets. The datasets in our experiments are Epinions [26] and Flixster [30], and both of them contain a social network and a rating matrix. The Flixster dataset is a social network service, where users can rate movies, the relationships between users are bilateral, and the rating values are 10 discrete numbers in the range [0.5, 5] with step size 0.5. The Epinions dataset is also a social network service containing a social network and a rating matrix. However, the relationships in the Epinions dataset are unilateral, and the rating values are integers from 1 to 5. Basic statistics of the two datasets are in Table 3.

4.2. Metrics. In the experiments, we treat the social network as known data and make 90% of the rating data as the training set and the remaining 10% as the test data. Our evaluation metrics are RMSE (root mean square error), precision, and coverage.

The RMSE of a model prediction with respect to the estimated variable $r_{u,i}$ is defined as the square root of the mean squared error:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{(u,i)\in R_{\mathrm{test}}} \left(r_{u,i} - \widehat{r}_{u,i}\right)^2}{\left|R_{\mathrm{test}}\right|}}, \tag{13}$$

where $|R_{\mathrm{test}}|$ is the cardinality of test set R_{test}.

Precision (also called positive predictive value) is the fraction of retrieved instances that are relevant:

$$\mathrm{Precision@}k = \frac{\#\mathrm{hit}}{k}, \tag{14}$$

where k is the number of the highest rating items that we recommend to the active user and #hit is the number of recommended items which are real in the top-k list. In our experiments, we choose $k = 5$, that is, Precision@5.

Coverage is the ratio of successful recommendations in the whole test set:

$$\mathrm{Coverage} = \frac{\#\mathrm{success}}{\left|R_{\mathrm{test}}\right|}, \tag{15}$$

where #success is the number of test data which we can make a recommendation successfully.

4.3. Baseline Algorithms. In our experiments, we compare the predicting results of different algorithms. Following is

Table 4: Experimental results for all users.

Algorithm	RMSE		Precision		Coverage	
	Epinions	Flixster	Epinions	Flixster	Epinions	Flixster
PearsonCF	1.235	0.973	0.721	0.797	0.673	0.739
SlopeOneCF	1.203	0.847	0.805	0.836	0.704	0.729
TrustWalker	1.134	0.904	0.853	0.861	0.952	0.692
SocialMF	1.197	0.896	0.826	0.903	0.718	0.736
SocialSlopeOne	1.096	0.838	0.893	0.897	0. 998	0.882

the description of labels that we use to denote each of these algorithms.

PearsonCF. It is user based CF algorithm with Pearson correlation coefficient as similarity measure [12].

SlopeOneCF. It is the slope one CF algorithm proposed in [16]. However, the algorithm does not take the social network into consideration and cannot deal with sparse matrix effectively.

TrustWalker [28]. This algorithm takes the social network into consideration and considers both ratings of similar items by trustable friends and ratings of the exact item by further distance users.

SocialMF [30]. It is model based matrix factorization algorithm considering both the social network and the rating matrix.

4.4. Results. In this subsection, we present the results of our experiments, first for all users and then for cold start nodes, and finally evaluate the impact of number of neighbors on the results of the proposed SocialSlopeOne algorithm.

4.4.1. Performance on All Users. We evaluate the performances (RMSE, precision, and coverage) of all algorithms on the whole test set, and the results are listed in Table 4. For better illustrating the results, we compare the results in Table 4 with Figures 2 and 3.

As shown in Figure 2, for the Flixster dataset, SocialSlopeOne has similar prediction error to SlopeOneCF, and both of them have lower prediction error than other algorithms; for the Epinions dataset, SocialSlopeOne has the lowest prediction error. In Figure 3, for precision, SocialMF is the best in the Flixster dataset, whereas SocialSlopeOne is the best in the Epinions dataset; for coverage, SocialSlopeOne is the best in both datasets. Moreover, SocialSlopeOne is the best for the sum of coverage and precision for all datasets. Hence, we can have that SocialSlopeOne has the best prediction accuracy and coverage in both datasets.

4.4.2. Performance on Cold Start Nodes. We also evaluate the performances (RMSE, precision, and coverage) of all algorithms on cold start nodes, and the result is listed in Table 5. In our experiments, we define cold start nodes as ratings for new items and new users. For better illustrating the results, we compare the results in Table 5 with Figures 4 and 5.

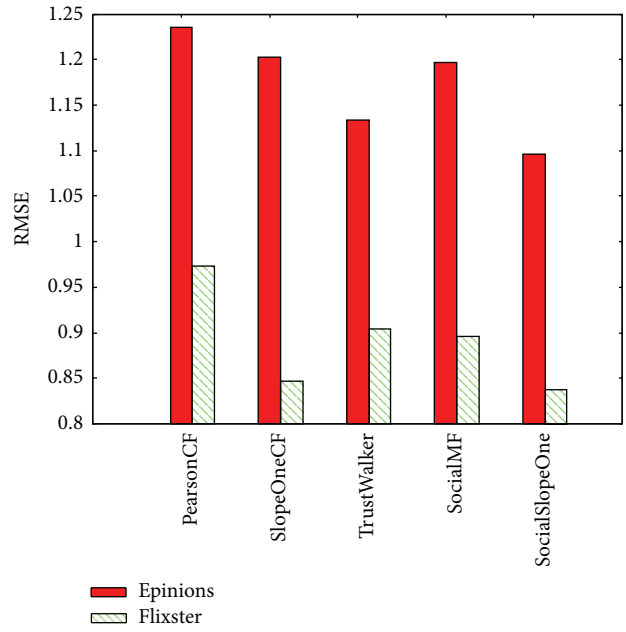


Figure 2: Comparison of RMSE for all nodes.

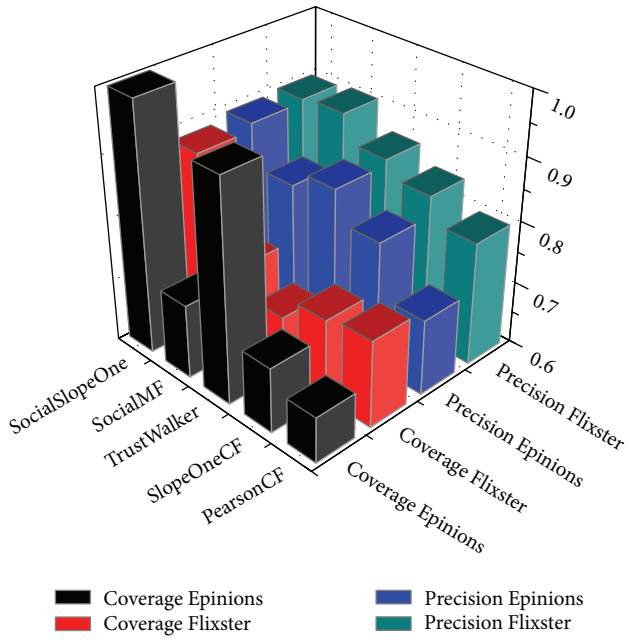


Figure 3: Comparison of precision and coverage for all nodes.

Table 5: Experimental results for clod start nodes.

Algorithm	RMSE		Precision		Coverage	
	Epinions	Flixster	Epinions	Flixster	Epinions	Flixster
PearsonCF	1.382	1.239	0.218	0.453	0.369	0.417
SlopeOneCF	null	null	null	null	null	null
TrustWalker	1.218	1.205	0.734	0.793	0.735	0.665
SocialMF	1.289	1.182	0.609	0.729	0.701	0.682
SocialSlopeOne	1.125	1.026	0.807	0.795	0.897	0.793

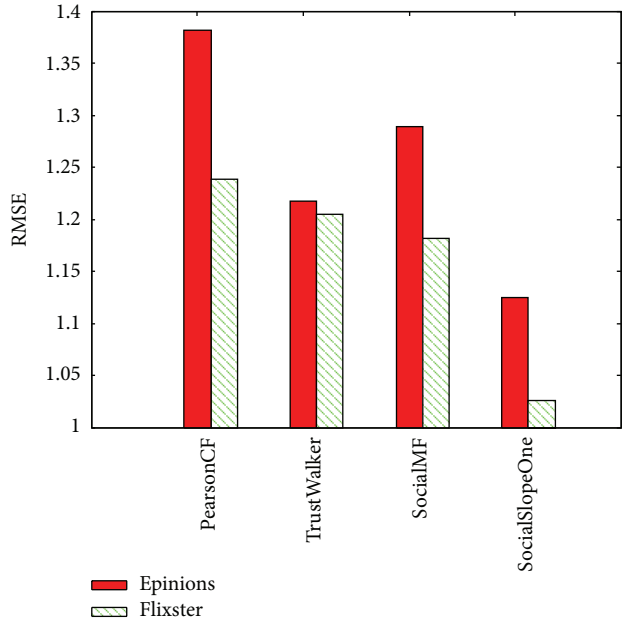


Figure 4: Comparison of RMSE for cold start nodes.

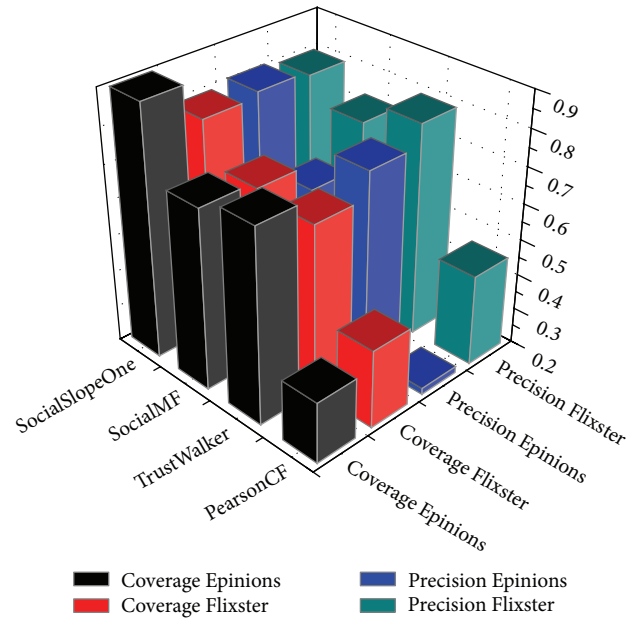


Figure 5: Comparison of precision and coverage for cold start nodes.

As shown in Figure 4, SocialSlopeOne has the lowest prediction error in both datasets for cold start nodes. In Figure 5, for precision, SocialSlopeOne and TrustWalker have similar precision, and they are better than others; for coverage, SocialSlopeOne is also the best in both datasets. Moreover, SocialSlopeOne is the best for the sum of coverage and precision for all datasets. Hence, we can conclude that SocialSlopeOne has the best prediction accuracy and coverage for cold start nodes in both datasets.

4.4.3. Impact of k on the Results. In the SocialSlopeOne algorithm, we need to construct $\min(|\mathrm{Ni}_i|, |\mathrm{Ni}_j|)$ slope one suggestions and distances from Ni_i and Ni_j. In large Ni_i and Ni_j, constructing all slope one suggestions will take a long time. However, small number of these suggestions, especially those from the same user, can approximate the results effectively. We set k as the top list with small distances defined in (10) and use the top-k slope one suggestions to approximate the results.

This top-k method cannot be applied in new items, so we observe the RMSE along with k for new users and all nodes. From Figure 6 we can see that, for both all nodes and new users, the RMSEs decrease with the rising of k. For small k, RMSEs decrease quickly, but as the increasing of k, RMSEs decrease slower and slower, which means that we can use small k to approximate the results for large Ni_i and Ni_j while taking suggestions from other users.

5. Conclusions

Traditional localization technologies focus on calculating the location of sensor nodes or mobile objects using WSNs. However, this paper studies the problem of future location prediction for mobile objects in WSNs. In addition to the WSN, we assume the location histories of all mobile objects as a rating matrix and transfer future location prediction problem to a social recommender problem. In social recommender systems, although current recommender algorithms could give recommender results, they have either low prediction accuracy or high computational complexity and cannot deal with sparse data efficiently. In this paper, we propose a random walk based similarity metric to find similar users and use slope one suggestions from similar users to construct recommender results. With our defined similarity metric, we can find more similar users, and this solves the matrix

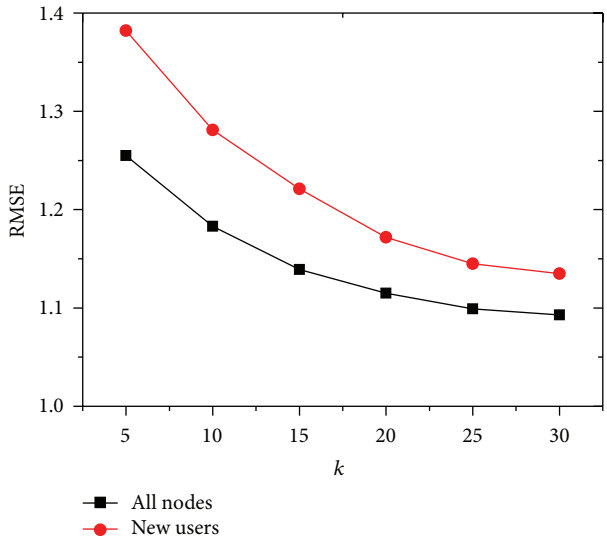


Figure 6: RMSE VS neighbors.

sparsity problem. In addition, we propose the social slope one recommender method, which has higher prediction accuracy and lower computational complexity. Experiments show that the proposed algorithm has better prediction accuracy and can solve the rating matrix sparsity problem more effectively than related works.

Conflict of Interests

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

Acknowledgments

This work was supported by the National High Technology Research and Development Program ("863" Program) of China (no. 2012AA011201).

References

[1] C. Alippi, R. Camplani, C. Galperti, and M. Roveri, "A robust, adaptive, solar-powered WSN framework for aquatic environmental monitoring," *IEEE Sensors Journal*, vol. 11, no. 1, pp. 45–55, 2011.

[2] S. Bhatti and J. Xu, "Survey of target tracking protocols using wireless sensor network," in *Proceedings of the 5th International Conference on Wireless and Mobile Communications (ICWMC '09)*, pp. 110–115, IEEE, August 2009.

[3] A. P. R. Da Silva, A. A. F. Loureiro, M. H. T. Martins, L. B. Ruiz, B. P. S. Rocha, and H. C. Wong, "Decentralized intrusion detection in wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Quality of Service and Security in Wireless and Mobile Networks (Q2SWinet '05)*, pp. 16–23, ACM, October 2005.

[4] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, no. 10, pp. 2529–2553, 2007.

[5] A. Pal, "Localization algorithms in wireless sensor networks: current approaches and future challenges," *Network Protocols and Algorithms*, vol. 2, no. 1, pp. 45–73, 2010.

[6] Z. Huang, H. Chen, and D. Zeng, "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 116–142, 2004.

[7] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 43–52, 1998.

[8] R. Jin, J. Y. Chai, and L. Si, "An automatic weighting scheme for collaborative filtering," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 337–344, July 2004.

[9] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 143–177, 2004.

[10] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.

[11] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web*, pp. 285–295, 2001.

[12] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008.

[13] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[14] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *10th Asian Conference on Computer Vision (ACCV '10)*, pp. 709–720, Springer, 2010.

[15] N. N. Liu and Q. Yang, "EigenRank: a ranking-oriented approach to collaborative filtering," in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR '08)*, pp. 83–90, July 2008.

[16] D. Lemire and A. Maclachlan, "Slope one predictors for online rating-based collaborative filtering," *Society for Industrial Mathematics*, vol. 5, pp. 471–480, 2005.

[17] C.-F. Tsai and C. Hung, "Cluster ensembles in collaborative filtering recommendation," *Applied Soft Computing Journal*, vol. 12, no. 4, pp. 1417–1425, 2012.

[18] S. K. Shinde and U. Kulkarni, "Hybrid personalized recommender system using centering-bunching based clustering algorithm," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1381–1387, 2012.

[19] K. Balog, H. Ramampiaro, N. Takhirov, and K. Nørvåg, "Multi-step classification approaches to cumulative citation recommendation," in *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, pp. 121–128, 2013.

[20] C. Cobos, O. Rodriguez, J. Rivera et al., "A hybrid system of pedagogical pattern recommendations based on singular value decomposition and variable data attributes," *Information Processing & Management*, vol. 49, no. 3, pp. 607–625, 2013.

[21] A. B. Barragáns-Martínez, E. Costa-Montenegro, J. C. Burguillo, M. Rey-López, F. A. Mikic-Fonte, and A. Peleteiro, "A hybrid content-based and item-based collaborative filtering approach

to recommend TV programs enhanced with singular value decomposition," *Information Sciences*, vol. 180, no. 22, pp. 4290–4311, 2010.

[22] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[23] N. Guan, D. Tao, Z. Luo, and B. Yuan, "NeNMF: an optimal gradient method for nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, 2012.

[24] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[25] J. A. Golbeck, *Computing and applying trust in web-based social networks [Ph.D. thesis]*, University of Maryland at College Park, 2005.

[26] P. Massa and P. Avesani, "Trust-aware recommender systems," in *Proceedings of the 1st ACM Conference on Recommender Systems (RecSys '07)*, pp. 17–24, October 2007.

[27] J. Golbeck and J. Hendler, "Inferring binary trust relationships in Web-based social networks," *ACM Transactions on Internet Technology*, vol. 6, no. 4, pp. 497–529, 2006.

[28] M. Jamali and M. Ester, "TrustWalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 397–405, July 2009.

[29] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*, pp. 203–210, July 2009.

[30] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 135–142, September 2010.

[31] N. L. Muscanell and R. E. Guadagno, "Make new friends or keep the old: gender and personality differences in social networking use," *Computers in Human Behavior*, vol. 28, no. 1, pp. 107–112, 2012.

[32] E. M. Daly, W. Geyer, and D. R. Millen, "The network effects of recommending social connections," in *Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 301–304, September 2010.

[33] J. Hannon, M. Bennett, and B. Smyth, "Recommending Twitter users to follow using content and collaborative filtering approaches," in *Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 199–206, September 2010.

[34] W. Chen, J. Chu, J. Luan et al., "Collaborative filtering for orkut communities: discovery of user latent behavior," in *Proceedings of the 18th International Conference on World Wide Web*, pp. 681–690, 2009.

[35] A. Akther, H. Kim, M. Rawashdeh, and A. El Saddik, *Applying Latent Semantic Analysis To Tag-Based Community Recommendations,*, Advances in Artificial Intelligence, Springer, 2012.

[36] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction tensor factorization for personalized tag recommendation," in *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM '10)*, pp. 81–90, February 2010.

[37] S. Tuarob, L. C. Pouchard, and C. L. Giles, "Automatic tag recommendation for metadata annotation using probabilistic topic modeling," in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2013.

[38] H. Kim and A. El Saddik, "Exploring social tagging for personalized community recommendations," *User Modeling and User-Adapted Interaction*, vol. 23, no. 2-3, pp. 249–285, 2013.

[39] S. Berkovsky and J. Freyne, "Group-based recipe recommendations: analysis of data aggregation strategies," in *Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 111–118, September 2010.

[40] L. Baltrunas, T. Makcinskas, and F. Ricci, "Group recommendations with rank aggregation and collaborative filtering," in *Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 119–126, September 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 723260, 10 pages
http://dx.doi.org/10.1155/2013/723260

Research Article

Predictive Analytics by Using Bayesian Model Averaging for Large-Scale Internet of Things

Xinghui Zhu,[1] Fang Kui,[1] and Yongheng Wang[2]

[1] *College of Information Science & Technology, Hunan Agricultural University, Changsha 410128, China*
[2] *College of Information Science and Engineering, Hunan University, Changsha 410082, China*

Correspondence should be addressed to Xinghui Zhu; 152062475@qq.com

Received 7 August 2013; Accepted 18 September 2013

Academic Editor: Gelan Yang

Massive events can be produced today because of the rapid development of the Internet of Things (IoT). Complex event processing, which can be used to extract high-level patterns from raw data, has become an essential part of the IoT middleware. Prediction analytics is an important technology in supporting proactive complex event processing. In this paper, we propose the use of dynamic Bayesian model averaging to develop a high-accuracy prediction analytic method for large-scale IoT application. This method, which is based on a new multilayered adaptive dynamic Bayesian network model, uses Gaussian mixture models and expectation-maximization inference for basic Bayesian prediction. Bayesian model averaging is implemented by using Markov chain Monte Carlo approximation, and a novel dynamic Bayesian model averaging method is proposed based on event context clustering. Simulation experiments show that the proposed prediction analytic method has better accuracy compared to traditional methods. Moreover, the proposed method exhibits acceptable performance when implemented in large-scale IoT applications.

1. Introduction

The Internet of Things (IoT) is a novel paradigm that aims to bridge the gap between the physical world and its representation in the digital world. IoT is expected to become an integrated part of the Internet in the future. IoT is defined as a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols. Bandwidth and storage are no longer considered restricting factors in IoT applications because of the rapid development of novel information and communication technologies. The main issue is how to process the massive events produced by IoT applications, such as the possibility of processing incomplete data streams and historical data from various data sources.

In a large-scale IoT application, the system must process events that arrive from various sources. Such sources include sensors, which constitute the wireless sensor networks, radio-frequency identification (RFID) readers, global positioning systems (GPS), and social media. The events that are directly generated by RFID readers or sensors are primitive events. The semantic information inside primitive events is quite limited. Thus, only simple information can be obtained from primitive events. In real-life applications, people give more attention to higher-level information, such as business logic and rules. For example, each reading operation of an RFID reader at a garage generates a primitive event. However, the user is actually concerned with a complex event such as "the car leaves the garage." Numerous primitive events have to be combined according to certain rules to obtain such complex event. An IoT application system converts business logic into complex events and then detects business logic on the basis of the detected complex events. Complex event processing (CEP) [1] is used to process huge primitive events to obtain valuable information. As an example, in logistics industry, CEP is used to track the goods and trigger some actions when an exception is found.

CEP in active databases has been studied extensively and has recently become a popular research area because of the rapid development of IoT. Most CEP methods assume that the data are deterministic. However, events are imprecise in several real-time IoT applications because of a number of factors,

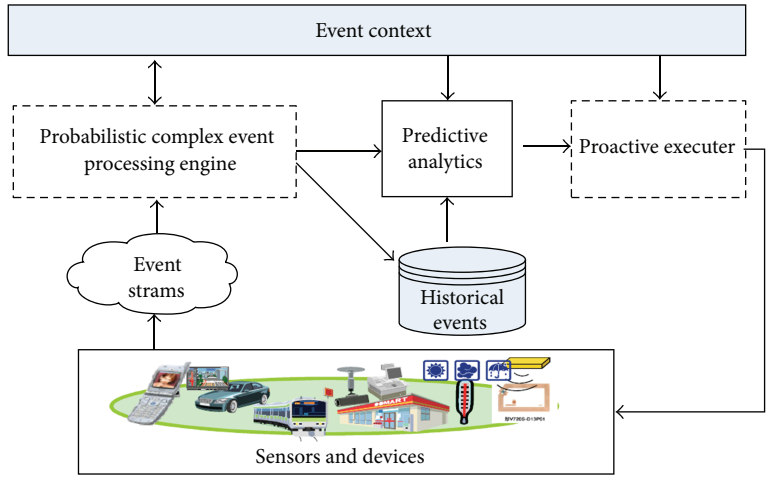


Figure 1: Proactive event processing in IoT.

such as limitation of measuring accuracy, signal disturbance, or privacy protection. The uncertainty is usually treated as probabilities. Therefore, the event processing engine must be capable of supporting probability.

In certain IoT applications, actions can be executed to change the state of the system. Most of these event processing methods are reactive, which indicates that the action is triggered by the state change of the system. A proactive event processing system has the ability to mitigate or eliminate undesired future events or to identify and take advantage of future opportunities by applying prediction and automated decision-making technologies [2]. For example, in a transportation IoT, we can predict a number of congestion states and then perform a number of actions to avoid the congestion states. Figure 1 shows a proactive event processing system in the IoT. The most important part of proactive event processing is predictive analytics (PA), which analyzes historical data to generate predictions regarding a future event.

PA applies several statistical and data mining techniques, such as clustering, classification, and regression. The Bayesian network (BN) [3] and its variations, such as dynamic Bayesian network (DBN) [4] and adaptive Bayesian network (ABN) [5], are extensively used in PA. Bayesian model averaging (BMA) [6] technology, which can average a great number of different competing models to help account for the uncertainty inherent in the model selection process, is also used in PA. When processing massive events in large-scale IoT applications, the performance of PA with BN and BMA is restricted by the challenges posed by big data. In 2012, "Big data" was recognized as one of the three leading edge technology trends that a CEO cannot overlook [7]. Big data can be characterized by volume, velocity, variety, and veracity ("data in doubt"). IoT is one of the most important areas in which big data must be processed. However, most of the existing PA algorithms cannot directly process the big data from IoT. Furthermore, IoT applications, particularly proactive event processing systems, require high-performance PA because such applications have to perform an action at an exact time. To date, few papers have reported on how to integrate CEP and PA to support proactive event-driven systems.

In this paper, we propose a predictive analytics method based on Bayesian model averaging (PABMA). This method can support proactive event processing in large-scale IoT applications. DBN can be created by employing probabilistic CEP. Consequently, basic predictive analytic method is proposed. A model comparison method for BMA is proposed to address the model uncertainty issue. A novel context clustering-based data partitioning and model selecting method is also proposed. The remainder of this paper is organized as follows. In Section 2, we provide a general review of related works on CEP, PA, and BMA technologies. In Section 3, we describe the event context model of our work. The PABMA method is described in detail in Section 4. In Section 5, we present the experiments we conducted to test PABMA. Finally, the discussion and conclusion are presented in Section 6.

2. Related Work

2.1. Complex Event Processing. CEP detects complex events on the basis of a set/sequence of occurrences of single events by continuously monitoring the event stream and then reacting to detected situations. Etzion and Niblett defined the basic concept and architecture of CEP [8] in their book. Event processing agent (EPA) is a component that applies logic on a set of input events to generate a set of complex events as output. Event processing network (EPN) is a network of a collection of EPAs, event producers, and event consumers linked by channels. The network is used to describe the event processing flow execution. Luckman first introduced EPN in the field of modeling [1]. The main idea of complex event detection consists of four steps: (1) primitive events are extracted from large volume data; (2) event correlation or event aggregation is detected to create a business event with event operators according to specific rules; (3) primitive or composite events are processed to extract their time, causal, hierarchical, and other semantic relationships; and

(4) response is sent to the actionable business information because of the guaranteed delivery of events to the subscribers.

The CEP engine must process streams of events with timestamps. Thus, numerous event pattern recognition methods are based on sequential variants of probabilistic graphical models, such as hidden Markov models [9], DBNs [4], and conditional random fields [10]. Methods for detecting complex events in probabilistic event streams based on nondeterministic finite automaton (NFA) have been recently proposed. Xu et al. proposed a data structure called chain instance queues to detect complex events that satisfy query requirements with single scanning probabilistic stream [11]. Conditional probability indexing tree is defined to store conditional probabilities of BN to improve its performance. Kawashima et al. proposed an optimized method for calculating the probability of outputs of compound events and for obtaining the value of confidence of the complex pattern given by the user against uncertain raw input data stream generated by distrustful network devices [12]. The present work is based on CEP and probabilistic CEP.

Proactive applications have been continuously developed in recent years. A number of examples include proactive security systems [13], proactive routing in mobile ad-hoc wireless networks [14], and proactive service level agreement negotiation in service-oriented systems [15]. Engel et al. proposed a proactive event-driven computing framework based on CEP, PA, and Markov decision processes [2, 16]. Engel at al. extended the event processing agent model to include two more types of agents, namely, predictive agents, which may derive future uncertain events based on prediction models, and proactive agents, which compute the best proactive action that should be performed. Proactive CEP requires the prediction of future events or states; such prediction is the main focus of this paper.

2.2. Predictive Analytics with Bayesian Networks. For PA methods based on complex event data, certain attributes of the monitored system can be predicted according to previously monitored events. Such prediction process can be divided into four steps: (1) collect and preprocess raw data; (2) transform preprocessed data into a form that can be easily handled by the (selected) machine learning method; (3) use the transformed data to create the learning model (training); and (4) use the previously created learning model to report predictions to the user. Thus, future events can be predicted by using the recent data on the basis of the learning model trained for previously monitored events.

Bayesian methods are becoming increasingly popular as frameworks for model selection and forecasting tools. Castillo et al. used BN, which considers the random character of the level of the total mean flow and the variability of origin-destination pair flows [17]. Pascale and Nicoli proposed an adaptive BN in which the network topology changes according to the nonstationary characteristics of traffic [18]. In this study, two major stationary areas were recognized as principal phases of traffic flows. Sun et al. modeled traffic flows within adjacent road links in a transportation network as a BN. The joint probability distribution between the cause nodes and the effect node in a constructed BN is described as a Gaussian mixture model (GMM) [19]. Hofleitner et al. used DBN and introduced a model based on hydrodynamic traffic theory to study the density of vehicles on arterial road segments and to illustrate the distribution of delay within a road segment [20]. In contrast to our work, these methods use single BN models and do not consider the massive data from IoT applications.

2.3. Bayesian Model Averaging for Predictive Analytics. Standard data analysts ignore the uncertainty in model selection, resulting in overconfident inferences and decisions that are riskier than they actually seem. The uncertainty inherent in the model selection process is often neglected by traditional statistical analyses. BMA is a technique designed to help account for this uncertainty. By averaging over several different competing models, BMA incorporates the model uncertainty into the conclusions regarding parameters and prediction. The traditional method for BMA is the mixture of experts model [21] proposed by Jacobs et al. In this model, the predictive distribution of various submodels is composed with weight to obtain the predictive distribution of the composed model. The main issue is how to compare and select the models to determine the coefficient of the mixture of expert models.

Model comparison and selection have been recently proposed. Zhou et al. constructed posterior probabilistic properties and model parameters on the basis of sequential Monte Carlo sampling and used these properties to compare different models [22]. With regard to the appropriateness of different genic models to different biological systems, Milias-Argeitis et al. compared and selected Bayesian genic models according to numerous methods, such as annealed importance sampling and approximate Bayesian computation [23]. Karabatsos and Walker proposed a mixed multinomial logit model and presented a Markov chain Monte Carlo (MCMC) algorithm to sample and estimate the posterior distribution of the model parameters [24]. Tawara et al. investigated the effect of the differences of the optimization methods for the multiscale GMM. MCMC-based method was compared with variational Bayesian method in a speaker clustering experiment [25]. Compared with our work, these methods were not optimized for big data from IoT.

3. Event and Context Model for IoT

We present an example of a transporting system based on the IoT to illustrate the event model and the function of the system. The system can obtain information, such as ID, location, and speed, from vehicles by using RFID, radar, GPS, and camera. Other information such as temperature and brightness can be obtained through a wireless sensor network. Through the use of complex events, the system can predict a number of future states (e.g., the congestion state of roads) and then perform actions to support proactive event processing.

Definition 1 (probabilistic primitive event). A primitive event in a stream indicates an atomic occurrence of interest in time.

A probabilistic primitive event is represented by $\langle A, T, Pr\rangle$, where A is the set of attributes, T is the timestamp when the event occurs, and Pr is the concrete probability value used to represent the occurrence probability of the event. The probability value represents the possibility that an event is converted accurately from truthful data to digital data used for computing in electronic devices.

In the transporting system example, each read operation of the devices generates a primitive event. Sometimes, the primitive event is not certain; for example, two RFID readers may find the same object at the same time. The primitive event may also be uncertain when a car accident is detected through a camera. A probability value is used to represent such uncertainty.

Definition 2 (probabilistic complex event). A complex event is a combination of primitive events or complex events according to certain rules. A probabilistic complex event is represented by $\langle E, R, Ts, Pr\rangle$, where E represents the elements that compose the complex event, R represents the rule of the combination, Ts represents the time span of the complex event, and Pr is the probability value.

Definition 3 (event type). The event type is a specification for a set of event objects that have the same semantic intent and the same structure. Every event object is considered an instance of an event type. An event type can represent either primitive events derived from a producer or complex events produced by an event processing agent.

The main complex event patterns in our work include ALL, ANY, COUNT, and SEQ. In this paper, the COUNT event can be used to represent the number of objects in a specified area during a specified time span. The SEQ event can be used to represent the moving path of an object. The detailed meanings of the patterns can be found in [8]. Such patterns can be employed to create hierarchical complex patterns.

Definition 4 (event context). An event context is a specification of conditions which groups event instances so that these instances can be processed in a related manner. The event context assigns each event instance to one or more context partitions.

The context types in our work include "event interval," "fixed location," and "event distance." The detailed definitions of these contexts can be found in [8]. As an example, assume context C_1 means "within 2 km from the motel M_1," context C_2 means "within 10 km from accident A_1," and context C_3 means "traffic status of highway is traffic slow" (Traffic in a certain highway has several status values: traffic flowing, traffic slow, and traffic stationary). In this example, C_1 is an entity distance context, C_2 is an event distance context, and C_3 is a state-oriented context. The context representation in our work is based on the fuzzy ontology framework of [16] and optimized for event processing.

Definition 5 (fuzzy ontology). A fuzzy ontology O in a particular domain Δ is $O_\Delta = (C, R, P, I, A)$, where C is a set of fuzzy concepts, P is a set of fuzzy properties of concepts, I is a set of objects, R is a set of fuzzy roles that denote the relations between two objects, and A is a set of axioms expressed in a logical language.

Definition 6 (fuzzy concept). A fuzzy concept C is defined as $C = \{a_1^{v1}, a_2^{v2}, \ldots, a_n^{vn}\}$, where a_i is an object and v_i is the membership degree of object i in concept C. The degree of object a belonging to a fuzzy concept C is given by the fuzzy membership function $\mu_C : A \rightarrow [0, 1]$, where A is the set of objects.

Definition 7 (concept subsumption). For two fuzzy concepts $X = \{a_1^{w1}, a_2^{w2}, \ldots, a_n^{wn}\}$ and $Y = \{a_1^{v1}, a_2^{v2}, \ldots, a_n^{vn}\}$, where a_i is an object and w_i and v_i are the membership degrees of the objects. If for all $a_i^{wi} \in X$, $a_i^{vi} \in Y$, and $v_i \geq w_i$, then X is subsumed by Y, which is denoted by $X \subseteq Y$.

Definition 8 (fuzzy role). A fuzzy role R is a fuzzy set of binary relations between two objects in the domain. The fuzzy role is interpreted as a set of pairs of objects from the domain denoted by $R = \{\langle a_1, b_1\rangle^{w1}, \langle a_2, b_2\rangle^{w2}, \ldots, \langle a_n, b_n\rangle^{wn}\}$, where a_i and b_i are two objects and w_i is the degree of the strength of the relation and is given by the fuzzy membership function $\mu_R : A \times B \rightarrow [0, 1]$, where A and B are sets of objects. The set of objects A is regarded as the domain of the role, whereas the set of objects B is regarded as the range of the role.

Definition 9 (fuzzy role subsumption). For two fuzzy roles $S = \{\langle a_1, b_1\rangle^{w1}, \langle a_2, b_2\rangle^{w2}, \ldots, \langle a_n, b_n\rangle^{wn}\}$, and $T = \{\langle a_1, d_1\rangle^{v1}, \langle c_2, d_2\rangle^{v2}, \ldots, \langle c_n, d_n\rangle^{vn}\}$, if for all $\langle a_i, b_i\rangle^{wi} \in S$, $\langle a_i, b_i\rangle^{vi} \in T$, and $v_i \geq w_i$, then S is subsumed by T, which is denoted as $S \subseteq T$.

Definition 10 (fuzzy property). A fuzzy property P is defined as $P = R \cdot C$, where R is a fuzzy role and C is a fuzzy concept denoting the range of the fuzzy role.

Concept C is the restriction on the range of the role R in property P. This restriction requires that all objects in the range of R should be a member of the concept C. P is interpreted as a fuzzy set of pairs of fuzzy role and fuzzy object such as $(\langle a_i, b_i\rangle, b_i)^{vi}$, where $\langle a_i, b_i\rangle$ is a member of the fuzzy role R, b_i is a member of fuzzy concept C, and v_i is the degree of the object a_i possessing the property P. The degree of the object that processes the property $P = R \cdot C$ is given by the function $\mu_P : R \times C \rightarrow [0, 1]$, where R is the set of fuzzy roles and C is the set of fuzzy concepts.

Definition 11 (fuzzy property subsumption). For two fuzzy property $P_1 = \{(\langle a, c\rangle, c)^{v1i} \mid (\langle a, c\rangle, c)^{w1i} \in S, c^{y1i} \in C\}$ and $P_2 = \{(\langle a, c\rangle, c)^{v2i} \mid (\langle a, c\rangle, c)^{w2i} \in S, c^{y2i} \in D\}$, if for all $(\langle a, c\rangle, c), (\langle a, c\rangle, c)^{v1i} \in P_1$, and $(\langle a, c\rangle, c)^{v2i} \in P_2$, $v_{2i} \geq v_{1i}$, then P_1 is considered subsumed by P_2, which is denoted by $P_1 \subseteq P_2$.

Definition 12 (fuzzy property with linguistic variable). A fuzzy property can be represented by $\langle V, M\rangle$, where V is a set of linguistic variables and M is the membership function.

For example, "red car" is a fuzzy concept that subsumes another fuzzy concept, that is, "car." "Bob extremely likes a sports car" is a relation of a fuzzy role, and the degree of strength of this relation is very high (extremely). The fuzzy property "drive.speed" can be set to $\langle\{Slow, Medium, Fast\}, \{S, M, F\}\rangle$. S, M, and F are depicted in (1).

Definition 13 (fuzzy context). A fuzzy context (FC) is defined as a triple FC $= \langle N_c, N_o, N_p\rangle$, where N_c is a set of fuzzy concepts, N_o is a set of objects, and N_p is a set of fuzzy properties. Consider

$$S = \int_0^{100} \frac{((100 - x)/100)}{x},$$
$$M = \int_0^{100} \frac{(x/100)}{x} + \int_{100}^{200} \frac{((200 - x)/100)}{x}, \quad (1)$$
$$F = \int_{100}^{200} \frac{((x - 100)/100)}{x}.$$

Definition 14 (fuzzy event context). A fuzzy event context (FEC) is defined as FEC $= \{fc_1^{t1}, fc_2^{t2}, \ldots, fc_n^{tn}\}$, where fc_i^{ti} is the fuzzy context at time t_i.

The context may change during a complex event. The change in context according to time in a complex event can be modeled by using Definition 12. A part of the traffic domain fuzzy ontology is shown in Figure 2. The concept subsumption is implemented using Definitions 7 and 9. Fuzzy properties with linguistic variables are implemented based on Definitions 10, 11, and 12. The event context can be created and reasoned on the basis of this ontology.

The fuzzy ontology is represented according to Fuzzy OWL 2 (http://gaia.isti.cnr.it/~straccia/software/FuzzyOWL/), and the ontology reasoning component is created based on Fuzzy DL (http://www.straccia.info/software/fuzzyDL/fuzzyDL.html).

4. Predictive Analytic by Using Bayesian Model Averaging

In this section, we first introduce how to implement PA with single Bayesian model. Then, a Bayesian model averaging method is proposed for multiple Bayesian models. Finally, to address the model selection problem, a dynamic model selection method based on data partitioning is proposed.

4.1. Basic Predictive Analytic Method. A multilayered adaptive dynamic Bayesian network (mADBN) model for predictive analytics is designed, as shown in Figure 3. The model contains a state plane and a set of location planes. Each plane is an ADBN with two dimensions: time and space. Bayesian networks are directed acyclic graphs whose nodes represent random variables and edges represent the conditional dependences among them. In the state plane, the nodes denote the states in different time instants or spatial locations, whereas edges denote the probabilistic relations of the states. The term "dynamic" in mADBN indicates that we are modeling a dynamic system. Figure 3 shows that the state (i, t) is related to a set of states before time t. The term "adaptive" indicates that the graph structure is created on the basis of the analysis of historical data. Each location plane represents the change in the location of an object (running path). The arrow denotes the transaction probability from one place to another, and the arrow with solid line denotes the real path. The running path of an object is represented as a SEQ event.

The structure of the state plane can be created by analyzing the object location planes. Conditional probability table (CPT) is used to save and sort the conditional probability that an object proceeds to the next place. CPT is learned from massive historical data using Bayesian formula. First, we select nodes before time t that can affect the state of node (t, i) on the basis of CPT as candidates. Then, we use a search-and-score algorithm based on the Bayesian information criterion score [26] to learn the structure of the Bayesian network.

If $f_{i,t}$ represents the flow state of (i, t), $\text{pa}(i, t)$ represents the parent nodes of (i, t), and N_P denotes the number of nodes in $\text{pa}(i, t)$, then the set of flow states for $\text{pa}(i, t)$ is $F_{\text{pa}(i,t)} = \{f_{j,s} : (j, s) \in \text{pa}(i, t)\}$. According to BN theory, the joint distribution of all nodes in the flow state network can be expressed as follows:

$$p(F) = \prod_{i,t} p\left(f_{i,t} \mid F_{\text{pa}(i,t)}\right). \quad (2)$$

The conditional probability $p(f_{i,t} \mid F_{\text{pa}(i,t)})$ can be calculated as follows:

$$p\left(f_{i,t} \mid F_{\text{pa}(i,t)}\right) = \frac{p\left(f_{i,t}, F_{\text{pa}(i,t)}\right)}{F_{\text{pa}(i,t)}}. \quad (3)$$

The joint distribution $p(f_{i,t}, F_{\text{pa}(i,t)})$ can be modeled by using GMM as follows:

$$p\left(f_{i,t}, F_{\text{pa}(i,t)}\right) = \sum_{m=1}^{M} \alpha_m g_m\left(f_{i,t}, F_{\text{pa}(i,t)} \mid \mu_m, C_m\right), \quad (4)$$

where M is the number of nodes and $g_m(\cdot \mid \mu_m, C_m)$ is the mth Gaussian distribution with $(N_P + 1) \times 1$ vector of mean values μ_m and $(N_P + 1) \times (N_P + 1)$ covariance matrix C_m. Parameters $\{\alpha_m, \mu_m, C_m\}_{m=1}^{M}$ can be inferred from the historical data by using expectation-maximization (EM) algorithm [27]. Once $p(f_{i,t}, F_{\text{pa}(i,t)})$ is obtained, the conditional distribution $p(f_{i,t} \mid F_{\text{pa}(i,t)})$ can be derived, and $\widehat{f}_{i,t}$ can be estimated from $F_{\text{pa}(i,t)}$ by using minimum mean square error method.

4.2. Bayesian Model Averaging. For a set of H models, the model ensemble posterior distribution of a quantity Q (e.g., the future model predictions using new input data) given the data D can be expressed as follows:

$$p(Q \mid D) = \sum_{k=1}^{H} p(Q \mid M_k, D)\, p(M_k \mid D), \quad (5)$$

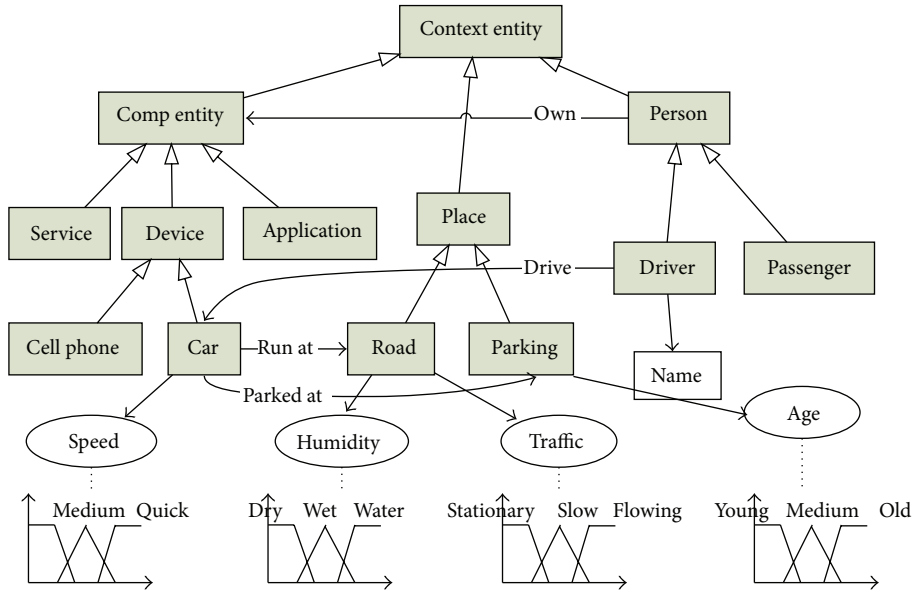


Figure 2: A simplified traffic domain fuzzy ontology.

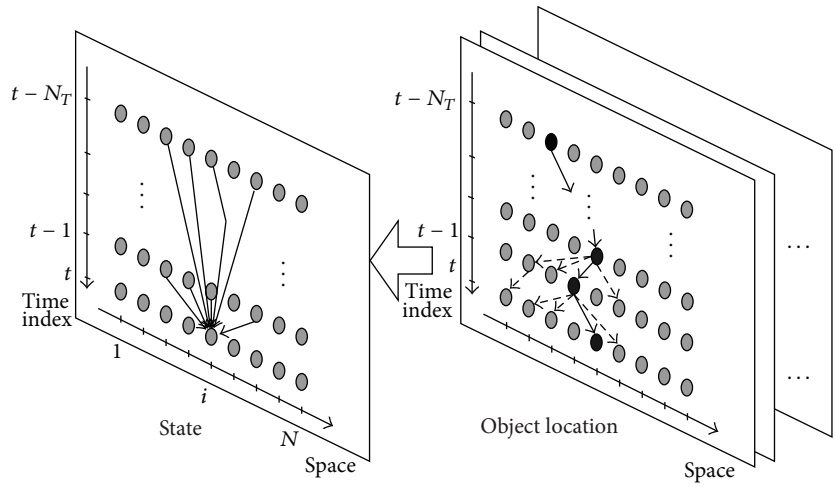


Figure 3: The mADBN model.

where $p(Q, M_k, D)$ is the posterior distribution of Q under model M_k and data D and $p(M_k \mid D)$ is the posterior model probability or model weight. This linear combination of predictions is individually performed by each model, in which the weighting coefficients are given by the posterior model probabilities. The precision of different models is compared to obtain the model weight. Assuming that model M_k has a vector of m_k parameters $\Theta_k = (\theta_{1k}, \theta_{2k}, \ldots, \theta_{mk})$ and $D = (d_1, d_2, \ldots, d_n)$ is a vector of n observation on the m_k parameters, according to Bayesian theory, we obtain

$$p(\Theta_k \mid D, M_k) = \frac{p(D \mid \Theta_k, M_k)\, p(\Theta_k \mid M_k)}{p(D \mid M_k)}, \quad (6)$$

where $p(D \mid M_k)$ is the model evidence or marginal likelihood function, which can be calculated as follows:

$$p(D \mid M_k) = \int p(D \mid \theta_k, M_k)\, p(\theta_k \mid M_k)\, d\theta_k. \quad (7)$$

The posterior probability of each model can be calculated as follows:

$$p(M_k \mid D) = \frac{p(D \mid M_k)\, p(M_k)}{p(D)}, \quad (8)$$

where $p(M_k)$ represents the prior distribution of the model k. In general, if we have no preference for any model, we assume that every model has the same prior probability. $p(D)$ is

not related to the models, suggesting that the quality of the different models is primarily determined by $p(D \mid M_k)$. To determine the weight of every model, (8) can be transformed as follows:

$$p(M_k \mid D) = \frac{p(D \mid M_k)}{\sum_{i=1}^{H} p(D \mid M_i)}. \quad (9)$$

Thus, the main issue is calculating the integration in (7). According to Gelfand and Dey [28], the model evidence $p(D \mid M_k)$ can be estimated effectively by using cross-validation distribution as follows:

$$\widehat{p}(D \mid M_k) = \prod_{i=1}^{n} p(y_i \mid M_k), (x_i, y_i) \in D, \quad (10)$$

where $p(y_i \mid M_k)$ can be calculated as follows:

$$p(y_i \mid M_k) = \int p(y_i \mid \theta_k, M_k)\, p(\theta_k \mid M_k)\, d\theta_k. \quad (11)$$

If the model is relatively complex, (11) is difficult to calculate directly. Therefore, MCMC method is used for approximate calculation. We can obtain a series of independent samples $\theta_k^{(t)} : t = 1, \ldots, T$ of θ_k from the distribution $p(\theta_k \mid M_k)$ of θ_k through sampling. Then, (11) can be approximated as follows:

$$\widehat{p}(y_i \mid M_k) = \frac{1}{T} \sum_{t=1}^{T} p\left(y_i \mid \theta^{(t)}k, M_k\right). \quad (12)$$

The primary goal in using MCMC method is to find independent series of samples. We use a Markov chain $\theta_0, \theta_1, \ldots, \theta_k$ in which every value depends only on its previous value. When certain conditions are satisfied, the series will converge to static distribution $p(\theta)$ after m iterations regardless of the original value of the series. Then, the samples $\theta^{(t)} : t = m+1, \ldots, n$ generated by the subsequent iterations can be used as independent samples of MCMC.

4.3. Dynamic Model Selection Based on Data Partitioning. Event data, even from the same event type, can be appropriate for different models when the system is in different states. We have proved this by experimentations. In our work, different states of the system can be represented by event context. We partition the event data into classes by using a context-based clustering method and then use the method presented in Section 4.2 to determine the appropriate model averaging for each class. When predicting the state for time t, we partition the data in time [t-h-1, t-1], where h is the size of the time window, into the existing classes. These data are typically portioned into one class so that we can use model averaging for that class. If the data are portioned into numerous classes, we reaverage the models of the classes according to the proportion to which the data belongs to each class.

First, we use fuzzy c-means (FCM) method to cluster the historical data. Through iteration process, FCM attempts to determine the cluster center that can minimize the following target function:

$$J_m = \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, \quad 1 \le m \le \infty, \quad (13)$$

where c_j denotes the center of the jth cluster, x_i denotes the ith training sample, and u_{ij} is the affiliation degree at which x_i affiliates to the jth data center. We can calculate u_{ij} as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left((x_i - c_j)/(x_i - c_k)\right)^{1/(m-1)}}. \quad (14)$$

In this study, the samples are complex events. Thus, we calculate the sample distance (or similarity) on the basis of the context of the events. As previously described in Section 3, the fuzzy ontology that we used to represent the event context is of hierarchical structure. The similarity between two nodes is defined based on the distance between them in the hierarchical structure.

$$\operatorname{sim}\left(C_i, C_j\right) := \frac{\alpha_i * \alpha_j}{l_{ij}}. \quad (15)$$

Here, α_i and α_j represent the weight of the context concepts C_i and C_j, respectively, and l_{ij} is the distance between the concepts in the ontology hierarchical structure. Distance is calculated by counting the number of edges in the path between two nodes. We assume the context sets of event e_1 and e_2 to be $C_{e1} = (c_{e11}, \ldots, c_{e1m})$ and $C_{e2} = (c_{e21}, \ldots, c_{e2m})$, respectively, to compare the similarity of the context set of the events. For each $c_{e1i} \in C_{e1}$, a c_{e2j} that satisfies $\max_{ce2k}(\operatorname{sim}(c_{e1i}, c_{e2k}))$ is found. The similarity between C_{e1} and C_{e2} can be calculated as follows:

$$Q(C_{e1}, C_{e2}) = \frac{n}{k+l-n} \sum_{i=1}^{m} \beta(c_{e1i}) \operatorname{sim}\left(c_{e1i}, c_{e2j}\right), \quad (16)$$

where function β means weights of similarity. Similarly, the similarity between C_{e2} and C_{e1} can be calculated as follows:

$$Q(C_{e2}, C_{e1}) = \frac{n}{k+l-n} \sum_{i=1}^{n} \beta(c_{e2i}) \operatorname{sim}\left(c_{e1i}, c_{e2j}\right). \quad (17)$$

Finally, the similarity between C_{e1} and C_{e2} can be defined as follows:

$$\operatorname{sim}(C_{e1}, C_{e2}) = \frac{Q(C_{e1}, C_{e2}) + Q(C_{e2}, C_{e1})}{2}. \quad (18)$$

A sample can belong to numerous classes because of the use of fuzzy clustering method. The following librarian criteria are used to adjust the samples that belong to multiple classes.

Librarian Criterion 1 (compactness). Copies of the same book might be placed in different shelves, allowing for multiple classifications. However, the selected classification must minimize the need for multiple copies to reduce costs.

Librarian Criterion 2 (even dimensionality). Books should be evenly distributed in various shelves.

To evaluate compactness, we first calculate the probability that an event context ce_i belongs to a class C_h as follows:

$$\widehat{p}(C_h \mid ce_i) = \frac{\operatorname{Sim}(ce_i, C_h)}{\sum_j \operatorname{Sim}\left(ce_i, C_j\right)}, \quad (19)$$

where C_j means the class that is different from C_h. Assuming that ce_i is partitioned into k classes in the original partition, we can compute the normalized entropy with

$$H_{\text{norm}}(ce_i) = \frac{-\sum_{h=1}^{k} \widehat{p}(C_h \mid ce_i) \log_2 \widehat{p}(C_h \mid ce_i)}{\log_2 k}. \quad (20)$$

According to the compactness criterion and the meaning of entropy in information theory, the closer $H_{\text{norm}}(ce_i)$ is to 0, the better is the compactness. We calculate $H_{\text{norm}}(ce_i)$ for all ce_i, and the best partition is that with minimum value.

To evaluate the even dimensionality, according to Bayesian theory, we have

$$p(ce_i \mid C_h) = \frac{p(C_h \mid ce_i)\, p(ce_i)}{p(C_h)}, \quad (21)$$

which satisfies $\sum_{i=1}^{N} p(ce_i \mid C_h) = 1$. Assuming that $p(ce_i) = 1/N$ (N is the number of classes), then

$$\sum_{i=1}^{N} p(ce_i \mid C_h) = \sum_{i=1}^{N} \frac{p(C_h \mid ce_i)\, p(ce_i)}{p(C_h)} = \frac{1}{N \cdot p(C_h)} \sum_{i=1}^{N} p(C_h \mid ce_i) = 1. \quad (22)$$

Then, we obtain

$$p(C_h) = \frac{\sum_{i=1}^{N} p(C_h \mid ce_i)}{N}. \quad (23)$$

Finally, we define the normalized entropy of C_h as follows:

$$H_{\text{norm}}(C_h) = \frac{-\sum_{h} \widehat{p}(C_h) \log_2 \widehat{p}(C_h)}{\log_2 k}. \quad (24)$$

We can calculate the average of all $H_{\text{norm}}(C_h)$ for a partition. The larger the average value, the better the distribution.

By clustering massive historical data with different granularity, we can compare the model quality of different classes of data and attempt to find the appropriate data partition and corresponding model averaging. When new events are generated, the events are classified into existing classes by using a similar method.

5. Experimental Evaluations

In this section, we report our experimental study on PABMA. A traffic simulation system is developed based on SUMO [29]. In this system, the mobility trace of cars is supported by using an OpenStreetMap [30] road map of a part of Beijing. A great number of "induction loops," which can detect cars that pass by, are placed on the roads. Virtual RFID or GPS readers are simulated by external applications that use the TraCI interface of SUMO to obtain the induction loop variables. Each induction loop covers a region. The closer a car is to the center of the region, the higher the probability that the event is detected. We select 55 junctions from the map and place 50,000 vehicles into the map. A series of rules is defined to simulate a real-life traffic system. Each vehicle has a home location and an office location. A vehicle v_i runs between home and office with probability α_i. Vehicles go to other places, such as the supermarket or the hospital, with corresponding probabilities. The routing rules are related to contexts such as weather, congestion state, and car accidents. On the basis of this simulation system, the precision and performance of PABMA are evaluated. We use five Lenovo ThinkServer RD series servers with 4 GB of memory as a cluster, and the operating system is Ubuntu Server 12. One server is used for the traffic simulating system, and the others are used for PABMA. The tasks of evaluating different models are partitioned into different servers to run in parallel.

Table 1: Deviation of different models.

	Deviation		
	Single model	SBMA	DBMA
Max	23	19	17
Min	2	3	0
Average	11.3	8.5	6.2

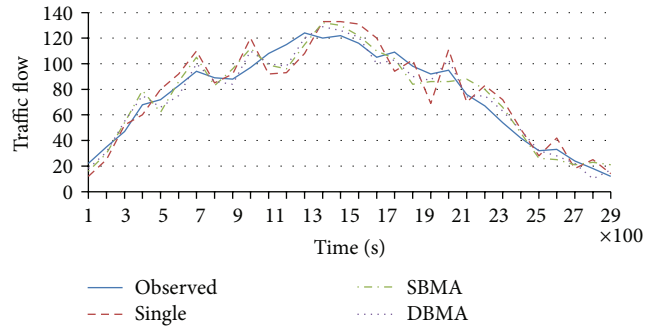


Figure 4: PA accuracy of a typical node.

First, we run the simulation several times to obtain the historical data of the vehicle paths. In the first experiment, the accuracy of the PA methods is evaluated, and the result for a typical node is shown in Figure 4 and Table 1. "Single," "SBMA," and "DBMA" denote the single Bayesian model (described in Section 4.1), the static BMA model (described in Section 4.2), and the dynamic BMA model (described in Section 4.3). Results show that through model averaging, SBMA exhibits better accuracy than the single BMA model does. DBMA has better accuracy than SBMA because an appropriate model averaging is created for different contexts.

In the succeeding experiment, the accuracy of SBMA and DBMA with different clustering granularity is evaluated, and the results are shown in Figure 5. Figure 5 shows that the maximum and average deviation decrease when the cluster number increases. However, when the cluster number increases to a certain number (10 in this experiment), the deviation no longer decreases. The reason is that, in certain extent, the increase of cluster number can make the models and clusters match better. The accuracy of the models with

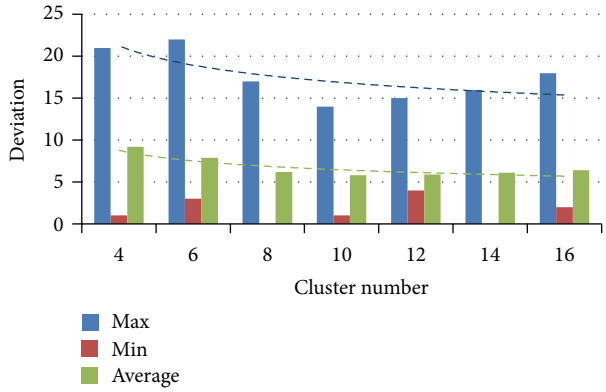


Figure 5: Accuracy for different clustering granularity with DBMA model. The training data size is 800 M, and the model number is 8.

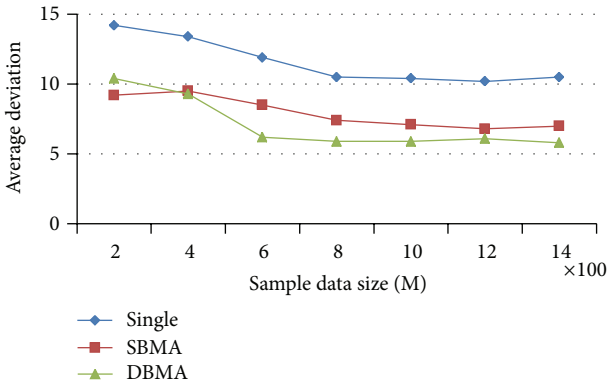


Figure 6: Accuracy for different sample data size. The model number is 8, and the cluster number in DBMA is 10.

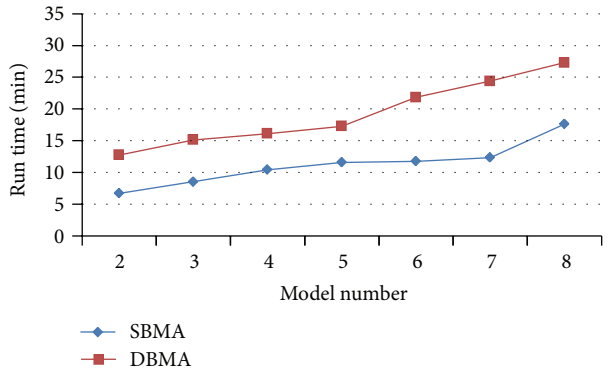


Figure 7: Performance for different model numbers. The training data size is 800 M, and the cluster number in DBMA is 10.

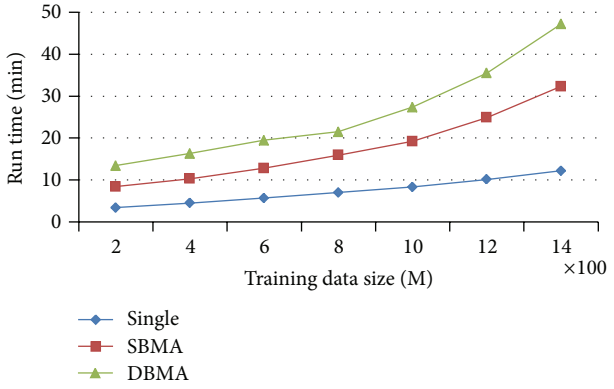


Figure 8: Performance for different training data size. The model number is 8, and the cluster number in DBMA is 10.

the sample data size is also evaluated, and the results are shown in Figure 6. The average deviation decreases when the sample data size increases. However, the decline ceases when the sample data size reaches a certain value (approximately 800 M in this experiment). The DMBA model decreases rapidly compared with the other models, suggesting that the DMBA model requires more training data to obtain better accuracy.

In the succeeding experiment, the performance of SBMA and DBMA with different model numbers is evaluated, and the results are shown in Figure 7. Figure 7 shows that the running time of both models increases linearly when the number of models increases. The performance of DBMA is lower than that of SBMA because DBMA has to evaluate models for different contexts. The performance of the three models with different training data sizes is also evaluated, and the results are shown in Figure 8. The running time for all models increases when the training data size increases. However, the running time of SBMA and DBMA increases more rapidly because the calculation becomes more complex. Although DBMA exhibits better accuracy than traditional methods, DBMA also exhibits the worst performance.

All of these experiments show that PABMA outperforms traditional methods for prediction accuracy. However, PMBA requires more training data and running time. This method trains composed Bayesian models according to different event contexts, thereby improving its accuracy. By using GMM with EM inference and MCMC in BMA, the complex calculation problem is resolved. The parallel method is also used in multicontext and multimodel training to improve performance. The performance is acceptable for normal PA applications.

6. Discussion and Conclusion

In this paper, we propose the use of DBMA to develop a high-accuracy PA method for large-scale IoT applications. Based on a multilayered ADBN model, this method uses GMM and EM inference for basic Bayesian prediction. BMA is implemented by using MCMC. DBMA is supported based on context clustering. The experimental evaluations show that this method has better accuracy compared with traditional methods. Moreover, this method exhibits acceptable performance when implemented in large-scale IoT applications.

The performance of PABMA still requires improvement. The current parallel method works only when learning the structure of models and training models for different contexts. EM inference and BMA process are not parallelized yet.

In the future, we plan to develop MapReduce algorithms to support massive historical data and complex training process.

Acknowledgments

This project was supported in part by the National Key Technology R&D Program of China (2012BAD35B07 and 2013BAD15B02) and the Hunan Provincial Natural Science Foundation (13JJ3046).

References

[1] D. C. Luckham, *The Power of Events: An Introductionto Complex Event Processing in Distributed Enterprise Systems*, Addison Wesley, Boston, Mass, USA, 2002.

[2] Y. Engel and O. Etzion, "Towards proactive event-driven computing," in *Proceedings of the 5th ACM International Conference on Distributed Event-Based Systems (DEBS '11)*, pp. 125–136, New York, NY, USA, July 2011.

[3] D. C. Martins Jr., E. A. de Oliveira, U. M. Braga-Neto et al., "Signal propagation in Bayesian networks and its relationship with intrinsically multivariate predictive variables," *Information Sciences*, vol. 225, pp. 18–34, 2013.

[4] H. C. Cho and K. S. Lee, "Nonlinear networked control systems with random nature using neural approach and dynamic bayesian networks," *International Journal of Control, Automation and Systems*, vol. 6, no. 3, pp. 444–452, 2008.

[5] A. Pascale and M. Nicoli, "Adaptive Bayesian network for traffic flow prediction," in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP '11)*, pp. 177–180, June 2011.

[6] H. Han and P. Bruce, "Bayesian averaging, prediction and nonnested model selection," *Journal of Econometrics*, vol. 167, no. 2, pp. 358–369, 2012.

[7] A. Artikis, O. Etzion, and Z. Feldman, "Event processing under uncertainty," in *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems (DEBS '12)*, pp. 32–43, Berlin, Germany, July 2012.

[8] O. Etzion and P. Niblett, *Event Processing in Action*, Manning Publications, 2010.

[9] L. R. Rabiner and B. Juang, "An introduction to hidden markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.

[10] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: probabilistic models forsegmenting and labeling sequence data," in *Proceedings of the 18th International Conference on Machine Learning (ICML '01)*, pp. 282–289, Morgan Kaufmann, 2001.

[11] C. Xu, S. Lin, and W. Lei, "Complex event detection in probabilistic stream," in *Proceedings of the 12th International Asia-Pacific Web Conference (APWeb '10)*, pp. 361–363, April 2010.

[12] H. Kawashima, H. Kitagawa, and X. Li, "Complex event processing over uncertain data streams," in *Proceedings of the 5th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 521–526, November 2010.

[13] S. Dolev, M. Kopeetsky, and A. Shamir, "RFID authentication efficient proactive information security within computational security," *Theory of Computing Systems*, vol. 48, no. 1, pp. 132–149, 2011.

[14] T. Kunz and R. Alhalimi, "Energy-efficient proactive routing in MANET: energy metrics accuracy," *Ad Hoc Networks*, vol. 8, no. 7, pp. 755–766, 2010.

[15] K. Mahbub and G. Spanoudakis, "Proactive SLA negotiation for service based systems," in *Proceedings of the 6th World Congress on Services*, pp. 519–526, July 2010.

[16] Y. Engel, O. Etzion, and Z. Feldman, "A basic model for proactive event-driven computing," in *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems (DEBS '12)*, pp. 107–118, July 2012.

[17] E. Castillo, J. M. Menéndez, and S. Sánchez-Cambronero, "Predicting traffic flow using Bayesian networks," *Transportation Research B*, vol. 42, no. 5, pp. 482–509, 2008.

[18] A. Pascale and M. Nicoli, "Adaptive Bayesian network for traffic flow prediction," in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP '11)*, pp. 177–180, June 2011.

[19] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, 2006.

[20] A. Hofleitner, R. Herring, and P. Abbeel, "Learning the dynamics of arterial traffic from probe data using a Dynamic Bayesian Network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1679–1693, 2012.

[21] R. A. Jacobs, M. I. Jordan, S. J. Nowlan et al., "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.

[22] Y. Zhou, A. M. Johansen, and J. A. D. Aston, "Bayesian model comparison via path-sampling sequential Monte Carlo," in *Proceedings of the IEEE Workshop on Statistical Signal Processing*, August 2012.

[23] A. Milias-Argeitis, R. Porreca, S. Summers, and J. Lygeros, "Bayesian model selection for the yeast GATA-factor network: a comparison of computational approaches," in *Proceedings of the 49th IEEE Conference on Decision and Control (CDC '10)*, pp. 3379–3384, IEEE, Atlanta, Ga, USA, December 2010.

[24] G. Karabatsos and S. G. Walker, "Bayesian nonparametric mixed random utility models," *Computational Statistics and Data Analysis*, vol. 56, no. 6, pp. 1714–1722, 2012.

[25] N. Tawara, T. Ogawa, S. Watanabe et al., "Fully Bayesian inference of multi-mixture Gaussian model and its evaluation using speaker clustering," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '12)*, pp. 5253–5256, March 2012.

[26] S. Samaranayake, S. Blandin, and A. M. Bayen, "Learning the dependency structure of highway networks for traffic forecast," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC '11)*, pp. 5983–5988, December 2011.

[27] J. J. Verbeek, N. Vlassis, and B. Kröse, "Efficient greedy learning of gaussian mixture models," *Neural Computation*, vol. 15, no. 2, pp. 469–485, 2003.

[28] A. Gelfand and D. Dey, "Bayesian model choice: asymptotics and exact calculations," *Journal of the Royal Statistical Society B*, vol. 56, pp. 501–514, 1994.

[29] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo—simulation of urban mobility: an overview," in *Proceedings of the 3rd International Conference on Advances in System Simulation*, pp. 63–68, Barcelona, Spain, October 2011.

[30] M. Haklay and P. Weber, "Openstreetmap: user-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 510917, 10 pages
http://dx.doi.org/10.1155/2013/510917

Research Article

Human Moving Pattern Recognition toward Channel Number Reduction Based on Multipressure Sensor Network

Zhaoqin Peng,[1] Chun Cao,[1] Jiaoying Huang,[2] and Wentao Pan[1]

[1] *School of Automation Science and Electrical Engineering, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing, China*
[2] *Science & Technology on Reliability & Environmental Engineering Laboratory, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing, China*

Correspondence should be addressed to Jiaoying Huang; huangjy@buaa.edu.cn

Received 7 August 2013; Accepted 21 September 2013

Academic Editor: Su-Qun Cao

A pair of sensing shoes for measuring foot pressure was developed. This system aims at recognizing human movement in unlimited environments. The multipressure sensor network of seven sensors on one insole was set up. Analysis for discriminating the user's movements from foot pressure distribution was carried out, considering the movements of standing, walking, going upstairs, and going downstairs. These actions were discriminated using characteristics extracted from the data of sensors. The classifier based on SVM showed highly accurate movement recognition. Specifically, to improve the classification performance, PCA based dimensionality reduction and channel reduction based data fusion were introduced. Experimental outcomes verified the testing speed of the classification function which was improved without affecting the accuracy rate. The results confirmed that this discriminant analysis can be employed for automatically recognizing human moving pattern based on foot pressure signal.

1. Introduction

Wearable robots, which can acquire physiological state and movement information, provide people with opportunities to wear sensors and intelligent devices in the daily life. This system is in the form of a mechanical structure that is combined with the exterior of a human body to enhance the power of the wearer in various environments [1]. In the field of healthcare, one major goal is to monitor activities of free-living subjects. Daily activities can provide information for doctors to accurately diagnose chronic diseases and design care plan of patients [2–5]. On the other hand, for military use, wearable robots can successfully maneuver heavy loads over unstructured terrain like forests, jungles, and deserts [6]. The first energetically autonomous lower extremity exoskeleton for soldiers, disaster relief workers, and wildfire fighters, as well as other emergency personnel to carry major loads, like food, rescue equipment, first-aid supplies, communications gear, and weaponry has been demonstrated at U.C. Berkeley [7]. Considerable kinds of sensors are employed to obtain human motion. Currently, a number of wearable robots are different in the selection of sensors, the position where sensors distribute, and the analysis of sensor data. Present researches tend to focus on daily worn wristwatch, glasses and shoes where sensors can be embedded into. With embedded sensors, noninvasive detection is available for providing action assistant. However, researchers never give up the intention to recognize users' condition to give appropriate action support. The acquisition of users' motion signal for controlling strategy of wearable robots still remains to be the key point.

Owing to the importance of data collection, various measuring techniques have already been developed to cater to the development of human-robot interaction. In many robotic systems, sensors are installed at toe or heel to recognize movements by thresholds [8, 9]. However, this method lacks of accuracy especially in identifying different moving stages [10]. Taking the advantage of multisensor technology has been a focus of interest within the last few years. Adopting the methodology of information cognition from multisensor

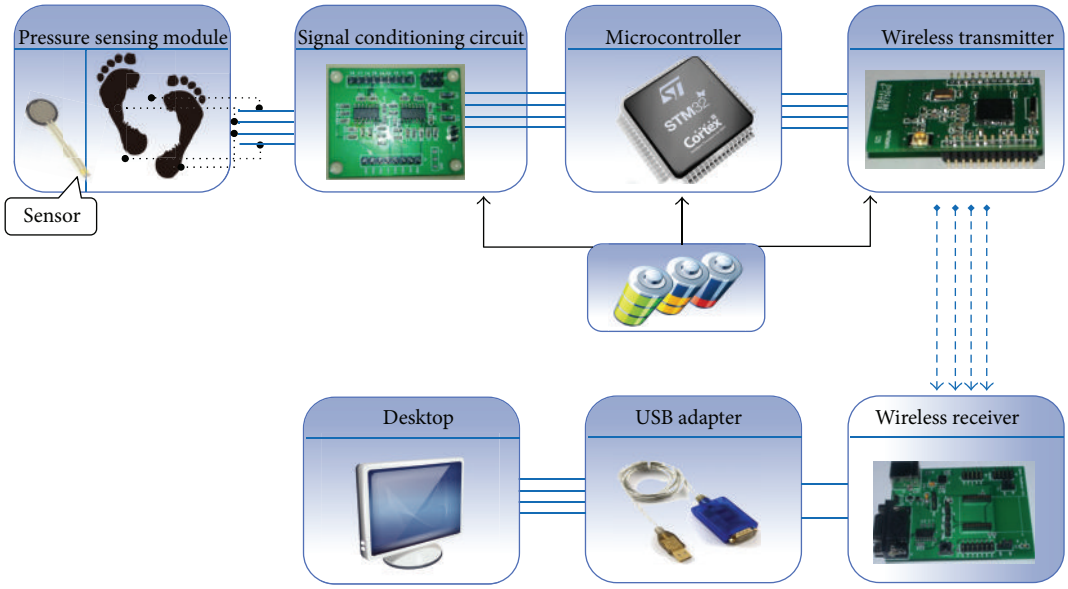


Figure 1: Overview of foot pressure sensing system.

was regarded not only efficient but also reliable [11–14]. Multisensor networks have already been deployed in a considerable number of detecting and monitoring tasks, such as aircraft structural performance detection [15], mobile health biomonitoring [16], alcohol continuous measuring within interstitial fluid [17], and forest fire alarming [18]. Studies emerged has paved a way for further researches on the distinctiveness of sensor networks in wearable robots. For example, the EMG based multisensor system is generally employed to measure muscular activity signals [19], whereas, this sensor has to be directly attached to the skin, which requires high sampling frequencies for signal collection and is difficult to quantify the signals [20].

According to the aforementioned issues, aiming at releasing physical and mental burdens on users, this paper preliminarily concentrates on developing a pair of foot pressure sensing shoes for users' movement identification. Foot pressure signal is both viable and effective for identifying behavior, for human movement and posture are well reflected in foot pressure distribution. To the best of our knowledge, plenty of researches are about foot pressure detecting system, but few studies have deeply discussed the effects of sensor distribution [21–23]. And the minority researches about the sensor positioning depend merely on foundational theoretical derivation. Therefore, in the present study, a sensor network based automatic motion signal acquisition system was analyzed to demonstrate the essential impact of sensor positioning. The relationship between recognition effect and sensor distribution was discussed to guide the design of the controlling strategy of wearable robots.

This reminder of this paper is arranged in the following order. Section 2 presented the setup of the foot pressure sensing shoes. Section 3 illustrates the development of the proposed sensor network scheme and feature representation for different motions. In Section 4, the feature selection method is depicted. Section 5 describes data fusion based movement identification algorithm. Section 6 reports the experimental results. We draw the conclusion in Section 7.

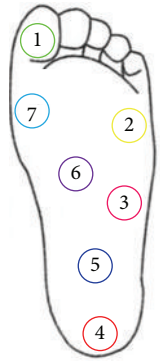


Figure 2: Measurement position.

2. System Description

The primary feature of foot pressure sensing shoes is its convenience of wearability. The wearable nature of shoes allows it to collect user's motion signal freely. The schematic of the pressure sensing system is presented in Figure 1.

Each insole of the shoe was equipped with seven pressure sensing elements respectively. The pressure sensors we employed in this system are FSR402, which are force-sensitive resistor sensor. FSR402 sensor is kind of flexible printed circuit with a thickness of 0.5 mm. It is obvious that the more sensors placed, the higher precision of plantar pressure distribution can be measured. However, we are aiming at optimizing the number of measurement points, because the number of sensors affects the amount of data processing, and power consumption. The seven sensors were installed on the insole as shown in Figure 2 and the appearance of

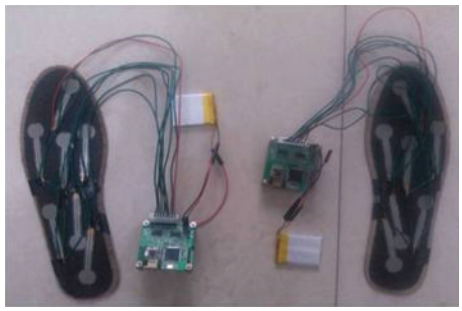


Figure 3: Appearance of insole sensors.

Table 1: Name of sensing positions.

Position number	Name
1	Great toe
2	Little ball
3	Lateral border
4	Heel
5	Posterior metatarsal
6	Anterior metatarsal
7	Great ball

the insole for measuring movements is shown in Figure 3. The seven sensors are listed in Table 1. Pressure signal was translated into voltage through an op-amp circuit whose output can be set at 0–5 V according to the pressure on the sensor. A microcontroller (STM32F107VET6) was employed for analog-digital conversion, data processing and control of data transmission. The wireless transmission module CC2530 based on Zigbee communication was for digital data sending and receiving. All these modules were driven by two 3.3 V lithium batteries. Hence, this sensing system based on the distinction of foot pressure distribution pattern can be attached to and removed from shoes whenever required.

3. Feature Extraction Based on PCA

The block diagram of foot pressure signal processing and analysis is exhibited in Figure 4, which consists of a feature abstracting module, a sensor network based data fusion part, and a classification module. These units are integrated in one system for moving pattern identification. Initially, raw pressure signals collected from the acquisition system had to be prepared for analysis. The purpose of data decomposition was to segment signal into subvectors. The normalized sample sets were applied to the feature extraction algorithm, which extract a series of feature elements from the inputs. Further, through the statistical feature transformation step, the data would be simplified into optimal form. For the purpose of optimizing the processing procedure, we applied multisensor based data fusion to reduce the number of detecting channels. With the construction of classifier, processed data were to be sent into the model for training and testing. Through this step, the recognition results were listed respectively. Consequently, the foot pressure pattern could be classified according to the utilizing demand. We preliminarily aim at holding that simple classifiers can achieve high accuracy as long as the feature is robust and representative, which is important in practical application.

Feature extraction is a major process of obtaining signal characteristics from time series data. Aiming at discriminating the reference class from other classes, feature selection can be considered as a data-compression process which removes irrelevant information and preserves relevant information from the raw data [24]. Typically, feature extraction approach applied to raw signals precedes the classification procedure.

Principal component analysis (PCA) is the most commonly technique applied to data reduction in pattern recognition and classification [25]. The basic idea of PCA is to project the data onto a lower dimensional space, where most of the information is retained. It uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components (PCs). The number of PCs should be less than or equal to that of original variables. This transformation is defined in such a way that the first PC has as high variance as possible, which accounts for as much of the variability in the data as possible. Each succeeding component has the highest variance possible under the constraint that it be orthogonal to preceding components.

Assuming that we have a set of centered input vectors of foot pressure x_t ($t = 1, \ldots, l$ and $\sum_{t=1}^{l} x_t = 0$) measured by sensor system, each of which is of m dimension $x_t = (x_t(1), x_t(2), \ldots x_t(m))^T$, PCA linearly transforms each vector x_t into a new one s_t:

$$s_t = U^T x_t, \tag{1}$$

where U is the $m \cdot m$ orthogonal matrix whose ith column u_i is the ith eigenvector of the sample covariance matrix $C = (1/l) \sum_{t=1}^{l} x_t x_t^T$. Hence, eigenvalues of C were obtained and u_i is the corresponding eigenvector:

$$\lambda_i u_i = C u_i, \quad i = 1, \ldots, m. \tag{2}$$

Based on the estimated u_i, the components of s_t were therefore calculated as the orthogonal transformations of x_t:

$$s_t(i) = u_i^T x^i, \quad i = 1, \ldots, m. \tag{3}$$

Thus, in PCA, the directions of the calculated PCs are uncorrelated with each other and computed by maximal variance. The new components we got are within a new dimensional space. By employing only a finite set of eigenvectors in the descending order of eigenvalues, the number of principal components in s_t will be reduced. Therefore, the cumulative contribution rate of the first several components would be expressed as $(1/t) \sum_{i=1}^{m} s_t(i)$. Usually the contribution rate value is over 95% to characterize the original data.

4. Pattern Recognition with SVM

During the recognition phase, the features from the foot pressure signals have been extracted by combined methods.

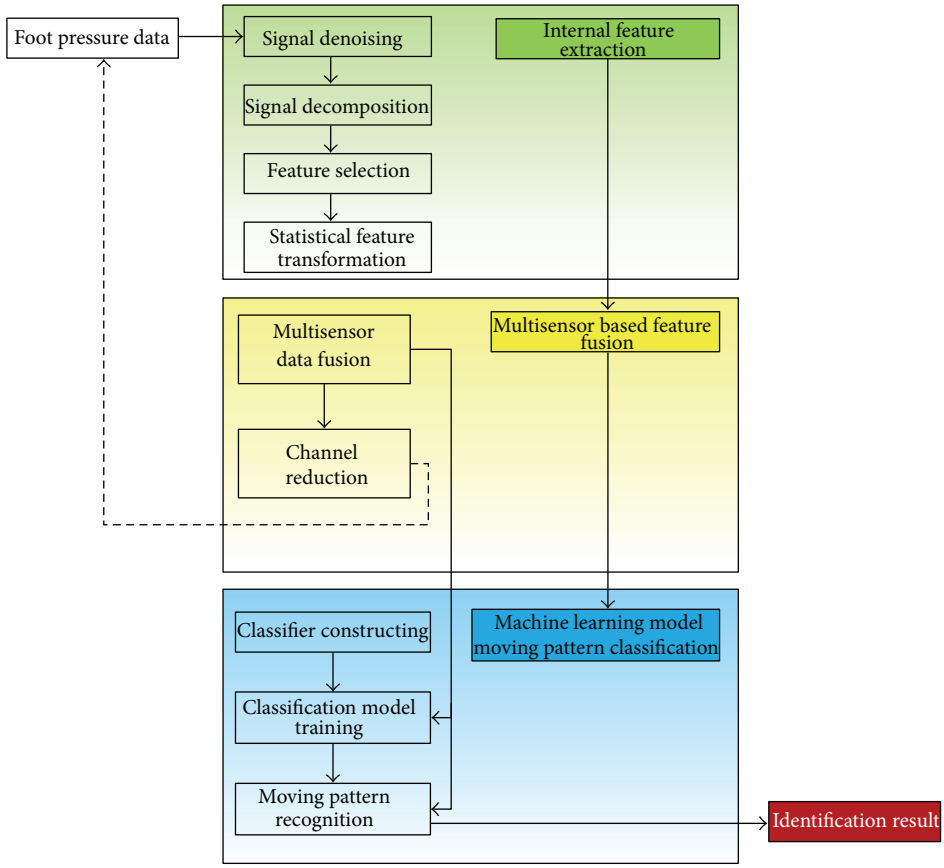


Figure 4: Process of moving pattern identification.

And they were fed into a classifier. Within this research, we took support vector machine to find the matching output by using these features. Support vector machine (SVM), which represents a major development in machine learning algorithms, is based on the foundation of statistical learning theory [26]. This algorithm is characterized by the ability to resolve problems of high-dimensional data and to represent the decision-making boundary in the form of training subsets, which is the general meaning of support vector [27, 28]. SVM has been widely applied to plenty of fields, such as pattern identification, regression analysis, and function approximating, and so forth [29, 30]. The results give the evidence that this technique cannot only satisfy from the theoretical perspective, but also can lead to high accuracy in practical applications. In this test, foot pressure detection is formulated as a classification problem, wherein the classifier is used to decide the moving pattern. Accordingly, the goal is to separate the moving pattern classes by a function which is induced from available examples.

The basic SVM is to classify a series of points into two different classes of data (Figure 5). The SVM attempts to place a linear boundary represented by a solid line between the two different classes and tries to orient the boundary in such a fashion that the distance between the boundary and the nearest data point in each class is maximal. Given the training

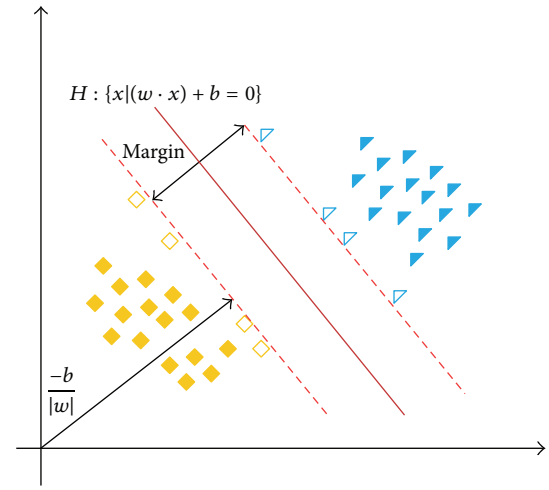


Figure 5: An example of classification of two classes by SVM.

vectors $x_i \in R^n$, $i = 1, \ldots, N$, in two classes, and an indicator $y_i \in \{-1, +1\}$, SVM constructs a linear function:

$$f(x) = w^T x + b. \tag{4}$$

(a) (b)

Figure 6: Subject going up- and downstairs with instrumented shoes.

The parameters w and b are defined according to a primal optimization problem:

$$\min J(w,\xi) = \min\left(\frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i\right), \quad \left(y_i\left(w^T\Phi(x_i)+b\right) \geq 1-\xi_i, \xi_i \geq 0,\ i=1,\ldots,N\right), \tag{5}$$

where C is for controlling the trade-off between the model complexity and empirical risk [31]. We set another parameter for the classifier G representing the kernel function for modeling. In this case, we utilize the kernel function into map input vector x_i to a higher-dimensional space through a nonlinear mapping $\Phi(x_i)$; the kernel concept can be written as:

$$K\left(x_i, x_j\right) = \Phi^T\left(x_i\right)\Phi\left(x_j\right). \tag{6}$$

Due to the possible high dimensionality of the vector variable ω, we introduced Lagrange multipliers $\alpha_i \geq 0$, $i = 1, 2, \ldots, N$, one for each of the constraints in (1), and we get the following Lagrangian:

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} y_i\alpha_i K\left(wx_i - b\right) + \sum_{i=1}^{N}\alpha_i. \tag{7}$$

Then the task is to minimize (7) with respect to w and b and to maximize it with α_i. At the very optimal point, we have

$$w = \sum_{i=1}^{N} y_i\alpha_i x_i, \qquad \sum_{i=1}^{N}\alpha_i y_i = 0. \tag{8}$$

It is shown that w is contained in the subspace spanned by x_i in (8). By substituting (8) into (7), the decision function of SVM model can be derived as

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N}\alpha_i y_i\left(x_i, x\right) + b\right). \tag{9}$$

The formula above results in an optimization problem with convex constrains, which is readily to be solved by the interior point method. For the moving pattern recognition issue, we preliminarily took the foot pressure signals of the seven points in Figure 2 as input vectors. Thus the SVM classifier was enabled to identify different kinds of movements.

5. Channel Reduction

During testing, the sensors transmit information about a common movement and the objective is to obtain an accurate identification of the moving pattern in the sensor field. Generally speaking, with more signal acquisition channels we can get more information. However, increasing the number of channel will definitely increase the complexity of computation and analysis that may lead to slow discrimination response. To overcome this limitation, we tend to find a way of using a reduced number of sensitive elements without compromising classification accuracy. On the other hand, the priority should be given to explore the relationship between the number of foot pressure sensors and the classification errors, respectively. Sensors on each insole can be regarded as a sensor network, which has independent power supply, regulator as well as signal conditioning module. For pilot phase, the information from data sets would be fused and for recognition. Data fusion of sensor network refers to the acquisition, processing, and synergistic combination of information gathered by various sensors to provide a better understanding of the phenomenon under consideration [32]. Therefore, the cooperative nature of the sensors can be exploited to improve the efficiency of resource utilization and the sensing performance [33]. Particularly, we focus on the number and distribution of pressure sensors in the sensor network. By analyzing the data fusion outcomes of sensors on different positions, the feasibility of channel reduction would be carried out.

Due to the laboratory deployment, channel number of each foot would be reduced from 7 to 4 with a decrement step of 1. All possible combinations for a reduced number of channels were to be evaluated by classification accuracy for different movement classes. Consequently, the sensor distribution that produced the lowest classification error for each number of channels would be considered as the optimal channel configurations.

6. Experimental Results and Analysis

The experiment was conducted to develop an automatic measuring system for revealing the relations between human motions and cumulative foot pressure characteristics. The experimental devices attached to the tester are shown in Figure 6. Our tests were carried out using a 24-year-old

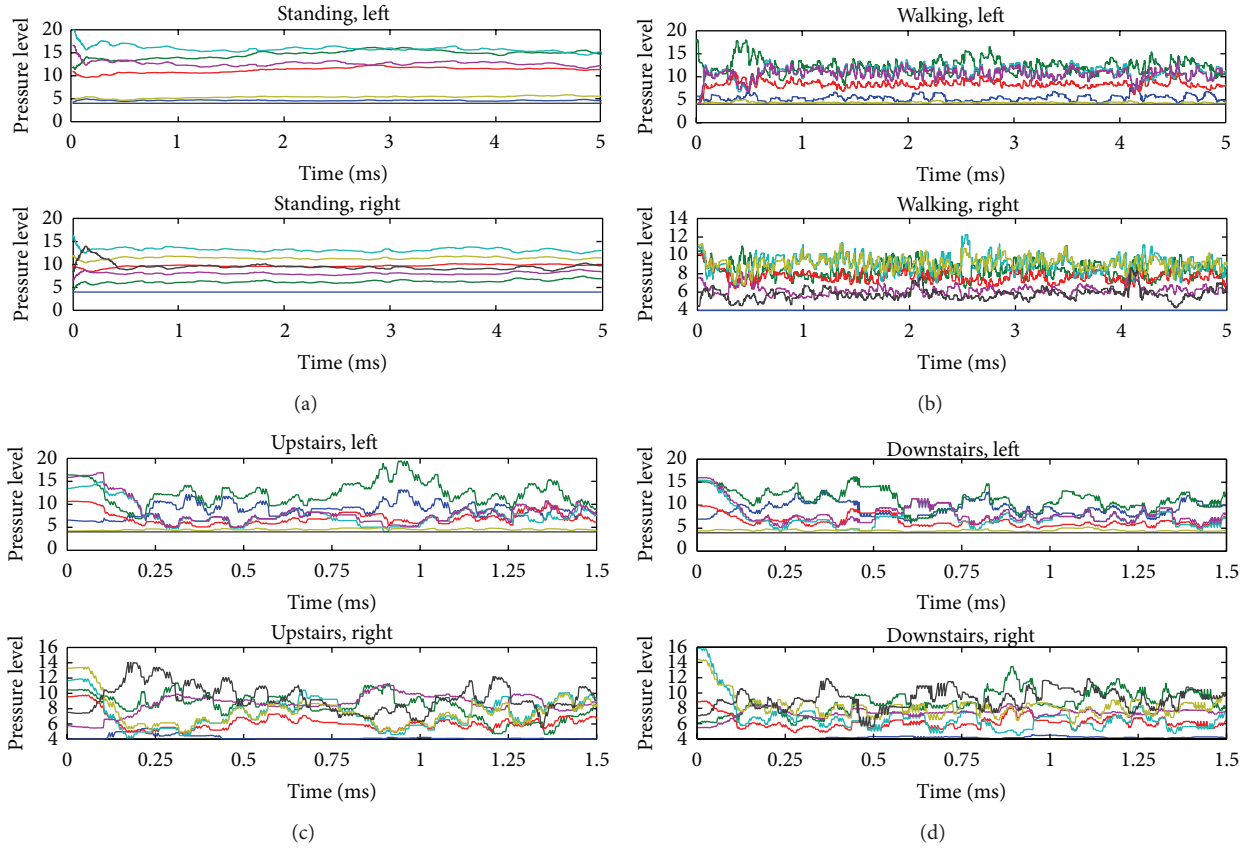


Figure 7: (a) Foot pressure distribution of standing still, (b) foot pressure distribution of walking, (c) foot pressure distribution of going upstairs, and (d) foot pressure distribution of going downstairs.

female wearer, 1.66 m tall. With the power supplied, foot pressure signals were gathered by FSR402 sensors every 40 ms and transmitted virtually through the data processing board to the computer wirelessly. The waveforms of each sensor on both feet were exhibited on the desktop simultaneously for monitoring. The signal processing procedure was implemented by Matlab 2010a, running on a PC with 2 G, 2 GHZ CPU.

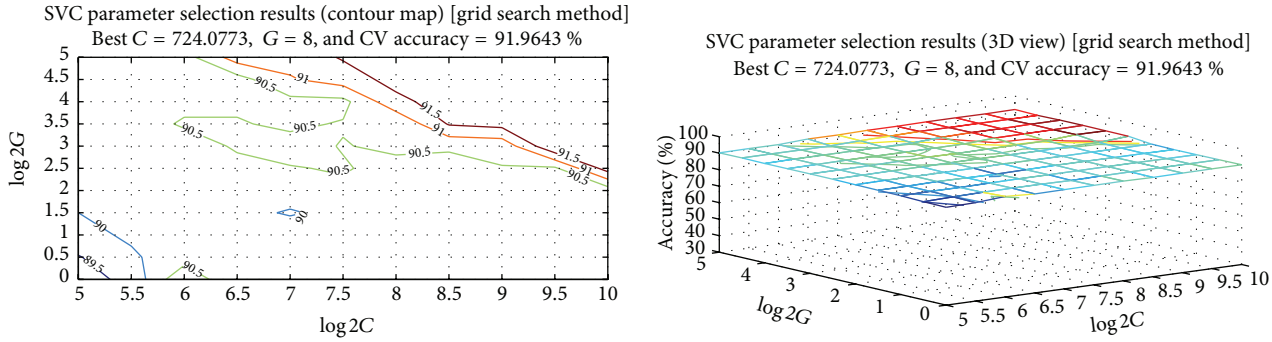
To evaluate the effectiveness of proposed methods in recognition, we choose four kinds of basic movements, which are standing still, walking, going upstairs, and going downstairs. Raw data on foot pressure distributions for each moving pattern were acquired with the developed foot pressure sensing shoes. Variation of foot pressure for each kind of movement was displayed in Figure 7. Pressure level represents the output value of digital information into which voltage is converted.

According to the figure, when not moving, the values of foot pressure basically stay constant. For walking, it is clear that the patterns change in shape and the ratio of the time, which depend on pressurization to depressurization in each step, in agreement with the way we use our feet. In case of going upstairs and downstairs, a higher peak value within each step can be observed; however the wave patterns are different for each. Therein, we are aiming at identifying these movements based on the corresponding foot pressure.

6.1. Feature and Feature Selection. In this study, five internal time-domain parameters were picked up as eigenvalue, which are average value, standard deviation, maximum value, and minimum value as well as difference deviation. Features that represented the pressure signal were memorized in matrix and sent to the classifier. We picked up 630 sets of data samples of each moving pattern: the former 420 are for classifier training and the latter for testing. The training data and training label are used to form the whole training set. For the training part, we got an optimal C of 724 and G of 8 by cross-validation. The optimization of these two parameters is for obtaining a high recognition rate based on current training samples. The radial basis function (RBF) kernel is employed. Due to the use of SVM classifier in this study, foot pressure and corresponding moving states were classified into the four categories, respectively. The cross validation result (contour map and 3D view) of parameter selection is shown in Figure 8.

The classification model was applied to predict the output category for testing samples identification. According to Figure 9, we summarize the classification performance results achieved by this SVM classifier. The average accuracy with all seven sensors is at 92.9% for all four kinds of movements and the diagnosis accuracy for each moving pattern is in Table 2.

The running time calculated by Matlab is 0.46 seconds. Compared with the preset inputs, we sent the input matrix



(a) (b)

Figure 8: (a) Contour map of parameter selection results and (b) 3D view of parameter selection results.

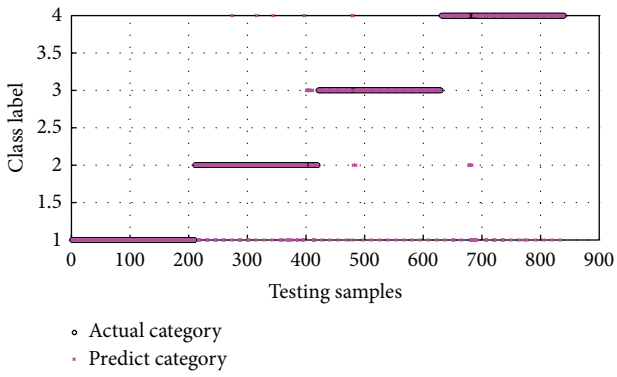
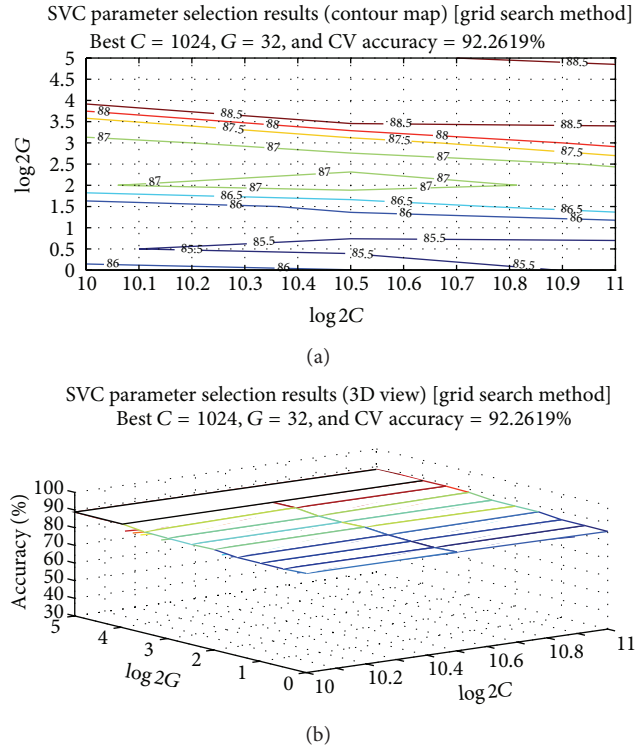


Figure 9: Actual and testing patterns of testing samples.

Table 2: Accuracy of different movements with seven sensors.

Moving pattern	Training accuracy (%)	Testing accuracy (%)
Standing still	100	100
Walking	91.2	91.0
Going upstairs	93.1	93.7
Going downstairs	90.0	86.7

to PCA processing algorithm for dimensionality reduction beforehand. Due to the aforementioned cumulative component rate in paragraph 3, we took the first three columns of PCs which occupied over 95% information of original data. A few numbers of new input eigenvectors provided sufficient information for foot pressure coding and movement recognition. The accuracy rate can be obtained when $C = 1024$ and $G = 32$, it reaches as high as 88.7% (Figure 10). The outcome of inputting the new eigenvectors in classifier is shown in Table 3. It could be noted that if a SVM classifier is used, declining recognition rate of moving patterns would be caused by PCA. Whereas, the classification time with proposed PCA algorithm did have a higher recognition speed, which was only 0.21 seconds. It decreased 0.25 seconds compared to the former classification.



(a) (b)

Figure 10: (a) Contour map of parameter selection results and (b) 3D view of parameter selection results.

Table 3: Accuracy of different movements with PCA.

Moving pattern	Training accuracy (%)	Testing accuracy (%)
Standing still	99.8	100
Walking	89.3	80.5
Going upstairs	90.8	88.6
Going downstairs	89.8	86.2

6.2. Channel Reduction Outcome. Dimensionality reduction algorithm based on PCA recorded the variation of recognition accuracy and speed but does not show the influence of sensor distribution on the insole. In this study, the impact of acquisition channels on the recognition rate and recognition

Table 4: Accuracy of different movements (four sensors).

Moving pattern	Training accuracy (%)	Testing accuracy (%)
Standing still	100	91.5
Walking	78.8	76.5
Going upstairs	88.3	85.0
Going downstairs	88.3	84.5

Table 5: Average accuracy and response time of different sensors.

Sensor distribution	Average testing accuracy (%)	Running time (ms)
FSR1, 2, 4, 7 + FSR3	89.5	68
FSR1, 2, 4, 7 + FSR5	89.1	46
FSR1, 2, 4, 7 + FSR6	90.0	64
FSR1, 2, 4, 7 + FSR3 + FSR5	89.9	63
FSR1, 2, 4, 7 + FSR3 + FSR6	91.1	60
FSR1, 2, 4, 7 + FSR5 + FSR6	89.8	61

Table 6: Accuracy of different movements (optimal set).

Moving pattern	Training accuracy (%)	Testing accuracy (%)
Standing still	100	100
Walking	88.8	83.0
Going upstairs	92.5	91.5
Going downstairs	92.3	87.0

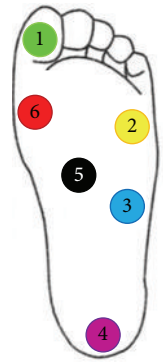


Figure 11: Rearrangement of measurement position.

time will be tested. In previous researches, sensors were generally placed at toe, little ball, great ball, and toe heel on human physiological characteristics, which are position 1, 2, 7, and 4 in Figure 2, respectively [34, 35]. We took the data of FSR1, FSR2, FSR4, and FSR7 for data fusion. And then the testing result for combing the four sensors was obtained. Using sensors of different positions could produce similar intraposition classification performance, but distinct classification accuracy. The average recognition accuracy rate is 84.4% (Table 4), which was significantly different from that of 92.9% (seven sensors). However, the classification only took 0.13 s.

With the same procedure, the number of multisensor decreasing from 6 to 4 was analyzed. The four basic sensors were FSR1, FSR2, FSR4, and FSR7 as mentioned before. Totally, the average classification accuracy across all positions is shown in Table 5.

With the increase of sensor number, the multisensor network seems to be more sensitive to different moving patterns. The SVM classifier was trained and tested using data multiple positions. For the sake of comparison, the average accuracy rate was changed with distinguished sensor data fused. When the foot pressure data from positions 1, 2, 3, 4, 6, and 7 were involved in the training set, the average classification accuracy reached a maximum value of 91.1% (Table 6). It is also noteworthy that this data set led to a shortest running time, which is only 60 ms. Therefore, using six optimally selected channels produced 91.1% average classification accuracy and 60 ms response time over four motions, compared with 92.9% and 460 ms of seven sensors. The optimal arrangement of foot pressure sensor distribution is shown in Figure 11.

The recognition error could be explained due to the finite measured data in that it was impossible to have an identical collection environment. Yet we cannot have a 100% accurate rate. Hence, this method can be encapsulated and embedded into one single program for simplifying operation, with different sorts of foot pressure signals. So we dare to say, with the ability of the SVM classifier, the detecting system applied has got more reliable results, whose service performance and effectiveness can cater to the need of monitoring in commercial use.

7. Conclusion

In this paper we suggested a laboratory setup for human moving pattern recognition based on foot pressure sensing. Pressure sensors, signal-conditioning module, microcontroller, as well as wireless transmitting devices, and so forth were integrated for signal acquisition and PCA, and data fusion of channel reduction and SVM were used to conduct the classification of several kinds of moving patterns. This laboratory setup has finished the following issues.

(1) With the design of insole equipped with pressure sensors, foot pressure was obtained exactly and freely.

(2) Specifically, to study the potential connections between moving patterns and foot pressure, we established an SVM based classifier. PCA algorithm was for feature dimension descending and multisensor data fusion for channel reduction.

(3) We compared the outcome between PCA dimensionality reduction and channel in order to address a better identification way.

(4) Data fusion technique can find an optimal number of sensors. This recognition can be better than that of PCA method for feature extraction. And the optimized SVM can improve the classification accuracy because of the best parameters.

The recognition accuracy of our optimal classification is as higher as 90.13%, which does have a fast time response at the same time. However, we have to point out that there still exists some deficiency. With the application of an advanced processor, that is, DSP, higher acquisition rate and conversion accuracy will be available, thus leading to a better classification performance. It should be noted that the testing of this study was only carried on one tester. In actual monitoring, other wearers' movement should also be trained and classified. Owing to the existing problems, further steps will be taken to focus on gait transaction, which is more applicable for the controlling strategy of wearable robots. In spite of these limitations, this study did put forward its significance. And these issues will be discussed in further studies.

Conflict of Interests

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

Acknowledgments

This research is a general project (61305131) supported by the Natural Science Foundation of China. Special appreciation should be given to the Research Centre of Fluid Power Transmission and Control, Beihang University, for the sake of their selfless help.

References

[1] C.-J. Yang, J.-F. Zhang, Y. Chen, Y.-M. Dong, and Y. Zhang, "A review of exoskeleton-type systems and their key technologies," *Proceedings of the Institution of Mechanical Engineers C*, vol. 222, no. 8, pp. 1599–1612, 2008.

[2] Z. Wang, M. Jiang, Y. Hu, and H. Li, "An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 691–699, 2012.

[3] K. Y. Chen and D. R. Bassett Jr., "The technology of accelerometry-based activity monitors: current and future," *Medicine and Science in Sports and Exercise*, vol. 37, no. 11, supplemet, pp. S490–S500, 2005.

[4] C. N. Scanaill, S. Carew, P. Barralon, N. Noury, D. Lyons, and G. M. Lyons, "A review of approaches to mobility telemonitoring of the elderly in their living environment," *Annals of Biomedical Engineering*, vol. 34, no. 4, pp. 547–563, 2006.

[5] F. Pitta, M. Y. Takaki, N. H. D. Oliveira et al., "Relationship between pulmonary function and physical activity in daily life in patients with COPD," *Respiratory Medicine*, vol. 102, no. 8, pp. 1203–1207, 2008.

[6] P. Neuhaus and H. Kazerooni, "Industrial-strength human-assisted walking robots," *IEEE Robotics and Automation Magazine*, vol. 8, no. 4, pp. 18–25, 2001.

[7] A. Zoss, H. Kazerooni, and A. Chu, "On the mechanical design of the Berkeley Lower Extremity Exoskeleton (BLEEX)," in *Proceedings of the International Conference on Intelligent Robots and Systems*, pp. 3465–3472, 2005.

[8] P. H. Veltink, C. Liedtke, E. Droog, and H. van Der Kooij, "Ambulatory measurement of ground reaction forces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 3, pp. 423–427, 2005.

[9] C. Liedtke, S. A. W. Fokkenrood, J. T. Menger, H. van der Kooij, and P. H. Veltink, "Evaluation of instrumented shoes for ambulatory assessment of ground reaction forces," *Gait and Posture*, vol. 26, no. 1, pp. 39–47, 2007.

[10] E. S. Sazonov, T. Bumpus, S. Zeigler, and S. Marocco, "Classification of plantar pressure and heel acceleration patterns using neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '05)*, vol. 5, pp. 3007–3010, 2005.

[11] T. Becker, M. Kluge, J. Schalk et al., "Autonomous sensor nodes for aircraft structural health monitoring," *IEEE Sensors Journal*, vol. 9, no. 11, pp. 1589–1595, 2009.

[12] C. J. Stam, M. Breakspear, A.-M. van Cappellen van Walsum, and B. W. van Dijk, "Nonlinear synchronization in EEG and whole-head MEG recordings of healthy subjects," *Human Brain Mapping*, vol. 19, no. 2, pp. 63–78, 2003.

[13] S. L. Bressler and J. A. S. Kelso, "Cortical coordination dynamics and cognition," *Trends in Cognitive Sciences*, vol. 5, no. 1, pp. 26–36, 2001.

[14] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," *IEEE Sensors Journal*, vol. 2, no. 2, pp. 107–119, 2002.

[15] K. Sampigethaya, R. Poovendran, L. Bushnell, M. Li, R. Robinson, and S. Lintelman, "Secure wireless collection and distribution of commercial airplane health data," *IEEE Aerospace and Electronic Systems Magazine*, vol. 24, no. 7, pp. 14–20, 2009.

[16] F. Hu, Y. Xiao, and Q. Hao, "Congestion-aware, loss-resilient bio-monitoring sensor networking for mobile health applications," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 450–465, 2009.

[17] M. Venugopal, K. E. Feuvrel, D. Mongin et al., "Clinical evaluation of a novel interstitial fluid sensor system for remote continuous alcohol monitoring," *IEEE Sensors Journal*, vol. 8, no. 1, pp. 71–80, 2008.

[18] B. C. Arrue, A. Ollero, and J. R. Martinez de Dios, "An intelligent system for false alarm reduction in infrared forest-fire detection," *IEEE Intelligent Systems and Their Applications*, vol. 15, no. 3, pp. 64–73, 2000.

[19] J. Rosen and J. C. Perry, "Upper limb powered exoskeleton," *International Journal of Humanoid Robotics*, vol. 4, no. 3, pp. 529–548, 2007.

[20] H. Lee, W. Kim, J. Han, and C. Han, "The technical trend of the exoskeleton robot system for human power assistance," *International Journal of Precision Engineering and Manufacturing*, vol. 13, no. 8, pp. 1491–1497, 2012.

[21] R. M. Queen, A. N. Abbey, J. I. Wiegerinck, J. C. Yoder, and J. A. Nunley, "Effect of shoe type on plantar pressure: a gender comparison," *Gait and Posture*, vol. 31, no. 1, pp. 18–22, 2010.

[22] M. J. Hessert, M. Vyas, J. Leach, K. Hu, L. A. Lipsitz, and V. Novak, "Foot pressure distribution during walking in young and old adults," *BMC Geriatrics*, vol. 5, article 8, 2005.

[23] M. M. Rodgers, "Dynamic foot biomechanics," *Journal of Orthopaedic and Sports Physical Therapy*, vol. 21, no. 6, pp. 306–316, 1995.

[24] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.

[25] R. Polikar, "Pattern recognition in bioengineering," in *Wiley Encyclopedia of Biomedical Engineering*, M. Akay, Ed., vol. 4, pp. 2695–2716, New York, NY, USA, Wiley Interscience, 2006.

[26] O. Ivanciuc, "Applications of support vector machines in chemistry," *Reviews in Computational Chemistry*, vol. 23, pp. 291–400, 2007.

[27] C. H. Huang, "A reduced support vector machine approach for interval regression analysis," *Information Sciences*, vol. 217, pp. 56–64, 2012.

[28] X. Peng and D. Xu, "Twin Mahalanobis distance-based support vector machines for pattern recognition," *Information Sciences*, vol. 200, pp. 22–37, 2012.

[29] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[30] Y.-X. Lai, C.-F. Lai, Y.-M. Huang, and H.-C. Chao, "Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home," *Information Sciences*, vol. 230, pp. 39–55, 2013.

[31] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[32] P. K. Varshney, "Multisensor data fusion," *Electronics and Communication Engineering Journal*, vol. 9, no. 6, pp. 245–253, 1997.

[33] Y.-W. P. Hong, K.-U. Lei, and C.-Y. Chi, "Channel-aware random access control for distributed estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2967–2980, 2008.

[34] C. M. Senanayake and S. M. N. A. Senanayake, "Computational intelligent gait-phase detection system to identify pathological gait," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1173–1179, 2010.

[35] E. Parikesit, T. L. R. Mengko, and H. Zakaria, "Wearable gait measurement system based on accelerometer and pressure sensor," in *Proceedings of the International Conference on Instrumentation, Communication, Information Technology and Biomedical Engineering (ICICI-BME '11)*, pp. 395–398, Bandung, Indonesia, November 2011.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 791560, 15 pages
http://dx.doi.org/10.1155/2013/791560

Research Article

Peer Selection Strategy Using Mobile Agent and Trust in Peer-to-Peer Streaming Media System

He Xu,[1,2,3] Han-Chen Huang,[4] Ruchuan Wang,[1,2,3] and Jun Dong[5]

[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, Jiangsu 210003, China*
[3] *Key Lab of Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Ministry of Education Jiangsu Province, Nanjing 210003, China*
[4] *Department of Leisure Management, Yu Da University, Miaoli 36143, Taiwan*
[5] *Institute of Intelligent Machines, Hefei Institute of Physical Science Chinese Academy of Sciences, Chinese Academy of Sciences, Hefei 230031, China*

Correspondence should be addressed to He Xu; xuhe@njupt.edu.cn

Received 12 August 2013; Accepted 26 September 2013

Academic Editor: Deguang Le

The aim of this study is to improve the quality of service (QoS) for peer-to-peer (P2P) streaming media system with the proposed peer selection strategy. P2P technology is promising for large scale streaming media files distribution over the Internet. As peers contribute their resources over multipeer transmission paths, it is particularly challenging to achieve consistently high streaming QoS such as low start-up delay, low jitter, low packet loss ratio, and high bandwidth assurance. Recently, there are a lot of researchers studying peer selection strategy in peer-to-peer streaming system. In this paper, a novel P2P streaming media peer selection strategy based on mobile agent and trust (MATPS) is proposed. The proposed approach is compared with bandwidth first peer selection strategy (BFPSS), GridMedia's peer selection (GridPS), trust-based peer selection strategy (TrBPS), and SmartPeerCast peer selection strategy (SPCPS), and the simulation results show that the proposed strategy improves the streaming QoS.

1. Introduction

In recent years, peer-to-peer (P2P) technology has become very popular and has drawn the attention of both academia and industry. According to [1], P2P file transfer traffic is occupying 86.7% of the total file transfer traffic. The basic principle of P2P technology is that the peer in P2P network acts as both a client and a server, which gets the resources from P2P network and also shares its own resources to other peers. P2P greatly reduces the server's load of the Client/Server (C/S) model and raises user satisfaction with the services on the network. At the beginning, applications of P2P networks were mostly focused on non-real-time data transmission such as file sharing. As the demand for multimedia services rapidly increases, especially as traditional centralized multimedia systems cannot scale well to accommodate such an explosive growth of requests, the highly scalable P2P streaming schemes naturally become a feasible solution to address the issues. Currently, Internet video broadcast and video on demand systems are mostly P2P-based architecture, such as PPLive, PPStreaming, and SopCast. Because of its advantages, P2P technology is widely used in the field of media streaming applications [2–5], having obtained a huge success. However, P2P networks are dynamic, given that any node in the network can join or leave at any time. This fact requires that special attention is devoted to the peer selection strategy, in order to minimize service failure. P2P systems build overlay networks on top of physical networks to facilitate peers' organization for content distribution. But the traditional Internet protocol (IP) is based on the best effort packet switching technology and it does not guarantee any quality of service (QoS), such as transfer delay, jitter, loss, and bandwidth [6]. P2P streaming systems require QoS performance guarantee in terms of bounded transfer delay and jitter, low packet loss, and bandwidth guarantee. This paper presents mobile agent based architecture and

uses the trust mechanism to choose super-peer for peer selection in P2P streaming system. From the jitter rate, start-up delay, packet loss ratio, network load, and network scale, compared with Overcast's bandwidth first peer selection strategy (BFPSS) [7], GridMedia's peer selection (GridPS) [8], trust-based peer selection strategy (TrBPS, only using the trust scheme to select peer) [9], and SmartPeerCast's peer selection strategy (SPCPS) [10], the simulation results show that the proposed strategy improves the streaming QoS and also show that the proposed strategy is feasible to adapt to changes in network scale and has expansion and stability.

The remainder of the paper is structured as follows. In Section 2, we discuss the related work. In Section 3, firstly, we introduce the concepts of the mobile agent, and then we present the main modules of our proposed peer selection strategy based on mobile agent and super-peer selection strategy based on trust for P2P streaming media system. Section 4 is devoted to the description of the simulation environment with agent used for the assessment of the new peer selection strategy. In Section 5, simulation results are presented. Finally, the overhead of using MATPS is discussed, and concluding remarks and ideas for future work are drawn in Section 6.

2. Related Work

MATPS is related to some previous works in the context of using agent techniques in P2P network and trust-based peer selection scheme. Many research efforts were done to integrate multiagent with P2P architectures [6, 11–14]. Some recent works propose the use of P2P networks and agent technology together for content delivery network (CDN) to support large scale streaming transmission. The main idea is that as users in the same interests or regions often have the same content to get, thus those contents can be shared among users who have common interests. To support a large number of concurrent users, P2P network is used for easing the server loading, and agent is used for resource searching [11, 12]. Trust management is another critical issue in P2P system. How to build the trust relationship within P2P system is studied by many researchers [9, 15–18]. In [15], the authors divide the previous research work on building trust in P2P systems into two broad categories, that is, reputation based and trade based. Our proposed trust-based super-peer selection scheme is the combination of reputation based and trade based.

There is a huge amount of work that has been done in efficient peer selection in P2P system. A multiagents based P2P-SIP real time stream sharing system was proposed by Yang et al. [14], called MPSS. The peer selection strategy of this system does not decrease the backbone's overall throughput and is unstable for playback. Cluster based architecture was proposed in [19], where peers are clustered based on semantic routing in P2P networks. They present the scenarios under which k-means clustering is more useful than randomly selecting the cluster centroids, but the quality of the clusters formed is not guaranteed, and the effect of dimensionality reduction on query efficiency and result quality is not discussed. Huang et al. in [20] propose a network-aware P2P file sharing architecture, which has a resource provider selection algorithm that can select a new resource provider for mobile peers experiencing broken connections in wireless mobile networks, but this resource selection strategy cannot be used in the P2P streaming system. Sun et al. in [21] propose an optimization method using particle swarm optimization algorithm for neighbor selection in P2P networks. Yao et al. in [22] indicate that PPLive achieves high ISP level traffic locality spontaneously with its decentralized, latency based, and neighbor referral peer selection strategy, which provide some new insights for better understanding and optimizing the network- and user-level performance in practical P2P live streaming systems. Ciullo et al. in [23] assess what level of "network awareness" has been embedded in the most successful P2P-TV systems, namely, PPLive, SopCast, and TVAnts. They describe what parameters mainly drive the peer selection and data exchange and believe that next-generation P2P live streaming applications definitively need to improve the level of network awareness, to better localize the traffic in the network and thus increase their network friendliness as well.

Overcast [7] is designed to offer an on-demand delivery of noninteractive, bandwidth demanding, and video streaming service to a self-similar community of users through overlay network based multicast. The design goal of Overcast is to build single source multicast trees that maximize bandwidth availability from the root to each overlay node without knowing the details of the underlying network topology. Through a centralized lookup mechanism, a client wishing to receive a video streaming service finds a root node of a multicast tree that distributes a desired content. The client checks the availability of bandwidth from the root through each one of every descendant and determines which overlay nodes can offer the same amount of bandwidth as the nominal download bandwidth. The most distant node, in terms of tree hierarchy, from the root that satisfies the bandwidth requirement will be selected as the node to which the client will attach itself in the multicast tree. We named Overcast's peer selection strategy as BFPSS.

"GridMedia" is a well-know system offering P2P-based IPTV services [8]. "GridMedia" organizes the peers in unstructured overlay networks and implements a push-pull based approach to fetch the media contents from the neighbor peers [24]. It implements a block scheduling mechanism for efficient contents sharing among the peers. The neighbor peers are selected on the random basis. We named GridMedia's peer selection strategy as GridPS.

The peers of SmartPeerCast [10] are grouped into different clusters according to their uploading bandwidth. Three different ALM trees are generated based on the clusters to broadcast different quality streams to eliminate the bandwidth bottleneck between the heterogeneous peers. And the "transrating engine" is used between the peers in different ALM trees to fully utilize the leaf nodes' resources which are in the high quality ALM trees. SmatPeerCast uses the receiving peer to select the sending peer according to its uploading QoS records [10]. We named SmartPeerCast's peer selection strategy as SPCPS.

Another similar work is called P4P [25], which is a simple, light-weight framework to allow more effective cooperative

traffic control between applications and network providers. Our proposed MATPS is a good application inspired by P4P. The focus in this paper is to explore peering strategy along with using mobile agent and trust. Our approach attempts to provide a scalable, robust, and reliable method ensuring QoS peer selection amongst peers in P2P streaming system. We propose a novel mobile agent based and trust value rank based peer selection in order to choose peers so as to improve stream QoS. From the jitter, start-up delay, packet loss ratio, network load, and network scale, compared with BFPSS, GridPS, TrBPS, and SPCPS, the results show that the proposed strategy is feasible.

Inspired by Raghuram and Munindar who proposed agent based service selection [26] and by Fortino and Russo who using P2P and agent technologies together for the development of content distribution networks [11], we introduce agent technology to the selection of peers for P2P streaming system. Aberer and Despotovic in [18] introduced that a simple, yet robust method that shows a solution to the problem of trust assessment based on reputation management in P2P environment. In the literature [18], they indicated that mutual trust allows agents to cooperate in a game-theoretic situation that corresponds to the repeated prisoners' dilemma and leads in the long term to an increased aggregated utility for the participating agents. And our proposed peer selection scheme is the combination of using agent and trust.

This paper makes the following main contributions for peer selection in P2P streaming media system. One, it shows how to use agent technology to support service for peer selection. Two, it proposed a novel trust-based super-peer selection strategy. Three, it proposes a novel simulation platform based on agent for evaluating peer selection. Four, it shows how the peer selections considered compare along different metrics of QoS requirements. The main benefits of our approach are as follows.

(i) Reducing the jitters of the receiving peers. By using mobile agent, we will be able to communicate to various peers to get required resources in local network. This will reduce the number of remote transmissions and interference, in turn decreasing the time needed for data transmissions in the system.

(ii) Guaranteeing the fairness of the peers' contribution by using trust value rank mechanism. Our trust scheme is the combination of reputation based and trade based. Peers that do not contribute/transmit stream enough will get low trust value, so they are periodically dropped in order to search/discover new potentially better peers. This will ensure that even with the lack of centralized information, if network environment changes, better new network topologies can be discovered.

3. P2P Streaming Peer Selection Model

P2P streaming system is required to encompass the key functions of object lookup, peer-based aggregated streaming, and dynamic adaptations to network and peer conditions. The quality of a peer depends on its availability, offered uplink bandwidth, and machine performance. We believe that a better peer selection for P2P streaming can offer a high streaming QoS. In this section, we will describe our proposed model for peer selection in P2P streaming system with using agent technology and trust mechanism.

3.1. Mobile Agent Based Peer Selection. Agent based computing is an effective paradigm for developing applications in complex domains as it supports the design and implementation of applications in terms of autonomous software entities, or "agents," which can achieve their goals flexibly by interacting with each another through high-level protocols and languages [11]. There is no accepted definition on what an agent is, but there is a set of characteristics which are commonly recognized by the community as endowing agents. There is a strong agreement that the following six characteristics form common features of agents [13].

(i) Autonomy: agents can operate without any direct human or any outside program intervention and incorporate some control over their actions and internal state.

(ii) Reactivity: agents perceive the environment and its changes and react to it as required by their goals.

(iii) Social ability: agents interact with other agents (possibly, with humans) by means of a communication language (either with the goal of cooperation and/or competition), giving rise to multiagent systems or community systems.

(iv) Proactivity: agents, in addition to the ability to react to the changing conditions in their environment, can also initiate actions in order to reach a specific goal.

(v) Adaptation/learning: agents learn by adapting their behaviors after a set of experiences for the purpose of improving the efficiency and efficacy of a given task and/or of performing new tasks which were hereto beyond their capabilities.

(vi) Mobility: agents have the ability to move from one to another node in a network.

Agents can be endowed with a mobility which enables them to roam across the network, thus potentially saving network resources, improving performance, and enabling dynamic configurability [27]. Agents are essentially autonomous programs that can take decisions based on their own history and the environment. In this paper, the mobile agent technology is exploited to make our peer selection mechanism adaptive to a dynamic P2P environment. Mobile agent has some benefits as follows [28, 29].

(1) A mobile agent can reduce the network load and latency by dispatching the mobile agents including the required services or data to remote nodes. Then, the services or data are locally executed at the remote nodes.

(2) A mobile agent can solve frequent and intermittent disconnection. Once a mobile agent is dispatched to

a destination node, it does not require direct connection with a user any more. Therefore, the mobile agent on behalf of a user is performed asynchronously and autonomously, even though a user (i.e., mobile device) is disconnected from the network.

(3) A mobile agent enables dynamic service customization and software deployment because it encapsulates some services or protocols into mobility entity.

(4) A mobile agent can adapt to heterogeneous environment and dynamic changes because it is computer- and transport-independent and also reacts autonomously according to its current execution environment.

In this paper, a framework by using mobile agent is proposed for peer selection. A peer selection architecture based on mobile agent is built and it is shown in Figure 1. Our proposal strategy not only keeps the simplicity and easy maintaining features of the P2P streaming system but also utilizes trust value rank to improve startup time delay and playback jitters performance of the P2P streaming system. As Figure 1 shows, the framework we propose in this paper is named as Mobile Agent and Trust based Peer Selection (MATPS), which is composed of three types of nodes: client, Mobile Agent Manage Server (MAMS), and peer. The client creates mobile agent to exchange information with MAMS. MAMS manages a number of peers' information (such as peer's trust value, file information). Besides the client node, the peer node also creates mobile agent to exchange information with other agents. In addition, our approach is based on a network of super-peers. A super-peer is a node in a peer-to-peer network that operates both as a server to a set of clients and as an equal in a network of super-peers. Super-peer based networks strike a balance between the inherent efficiency of centralized search and the autonomy, load balancing, and robustness to attacks provided by distributed search. Furthermore, they take advantage of the heterogeneity of capabilities (e.g., bandwidth, processing power) across peers. The basic steps of the proposed mechanism based on mobile agent are as follows.

(1) Client generates a control mobile agent (CA).

(2) CA generated in Step 1 can acquire query resources key from the client, also distribute a search agent (SA), and provide search key and search termination conditions (in case of resources are not found and the maximum number of queries sent to other peers).

(3) Then, SA generated in Step 2 migrates to remote management server MAMS. MAMS stores a large number of nodes information and keys. The selection of the MAMS can be made by the ISP or it can be automatically done using a dedicated super-peer selection. Super-peer selection strategy based on trust will be introduced in the following section.

(4) MAMS extracts the query key from SA and searches it in its own database. When finishing the inquiry, MAMS will create an information agent (IA) for the exchange of information with the SA. An IA returns the information included with the resources information, the trust value of peer information, and sorted list of peers (ranked peers), which contain attributes about whether they belong to local network.

(5) Then, SA will make a decision after receiving the information; if IA has the local peers, SA will return to this local peers; if IA returns no local peers, SA will move to ranked peers generated in Step 4.

(6) If Step 4 does not return the search peer information or reach the termination conditions, SA will return to its distributed client to announce that the resource was not found in this network.

In the P2P network, a fundamental distinction is made among unstructured and structured P2P systems for resource location. In unstructured P2P systems in principle peers are unaware of the resources that neighboring peers in the overlay networks maintain. Typically they resolve search requests by flooding techniques. In contrast, in structured P2P systems peers maintain information about what resources neighboring peers offer. Thus, queries can be directed and in consequence substantially fewer messages are needed. This comes at the cost of increased maintenance efforts during changes in the overlay network as a result of peers joining or leaving. The most prominent class of approaches to structured P2P systems is distributed hash tables (DHT) [30], for example, Chord. Figure 2 is the overlay structure of MAMS. All MAMS use a structured self-organizing P2P protocol to organize a structured P2P overlay network. Structured P2P uses DHT storage information (KeyID, PeersList), where KeyID is the identifier to find a shared resource; PeersList is the list of peers who share resources in P2P network. Request and response are the two messages between MAMS. When the index information of resources cannot be found on the local MAMS, MAMS will send a request message, while the request message format is Request (KeyID); while another MAMS receives this message, checklist is maintained by itself whether there is index information of resources needed; if it has, it will give a response information, the response message format is Reply (KeyID, PeersList), if not, put Request (KeyID) message forwarded to other MAMS; until it finds the resources needed or reaches the SA termination conditions, then it will return to its distributed client to announce that the resource was not found in this network.

There are some advantages to make use of mobile agents and also some problems to be resolved in our proposed peer selection strategy.

(1) A mobile agent can decrease the overhead of P2P system by performing peer selection, fault tolerance, and replication algorithms in a decentralized way. The peer selection mobile agents are distributed to groups. Then, they autonomously conduct a peer selection, fault tolerance, and replication algorithms in each super-peer group. Accordingly, the overhead system does not undergo the overhead any more.

(2) A mobile agent can adapt to a dynamical P2P environment. In a P2P environment, peers can join and

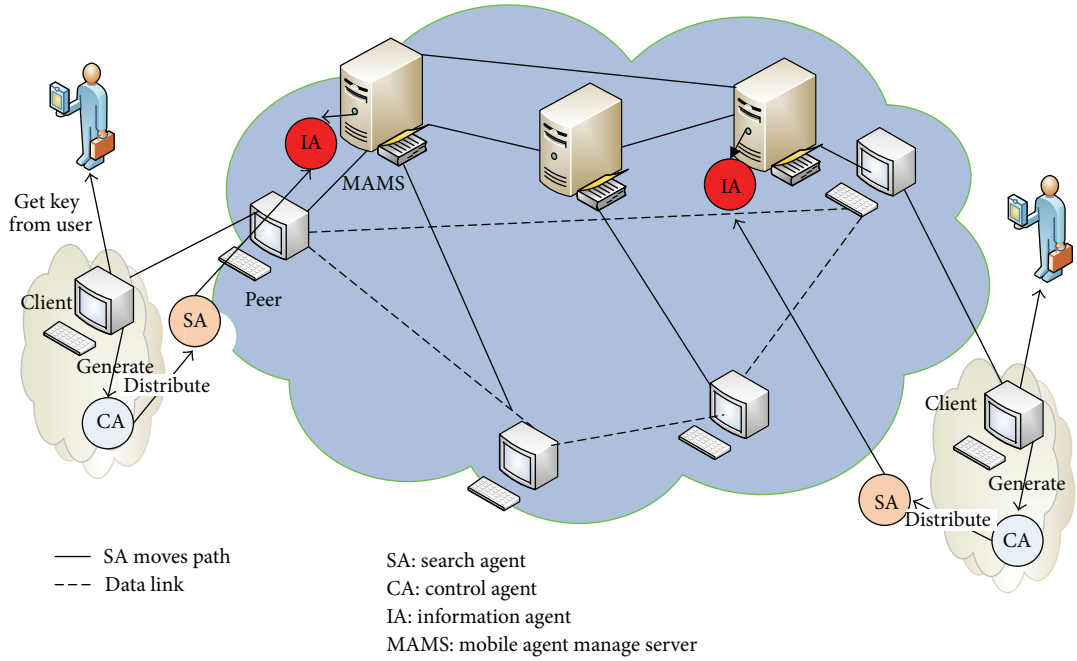


Figure 1: Mobile agent based peer selection architecture.

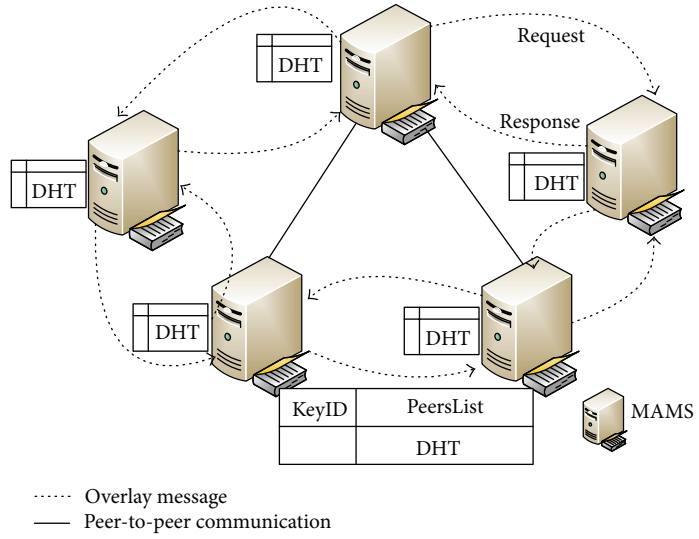


Figure 2: The architecture of MAMS's overlay.

leave at any time. In addition, they have heterogeneous properties such as capabilities (e.g., CPU, storage, or network bandwidth), location, availability, and credibility. These environmental properties change over time. A mobile agent can be performed asynchronously and autonomously coping with the changes. It can also tolerate the peers' autonomy failures by using migration and replication functionalities that a mobile agent itself provides.

(3) Various peers can be classified into different groups according to the properties of peers, and various peer selection mechanisms can be performed at a time in different groups. For example, these peers' selection mechanisms are implemented as mobile agents (e.g., peer selection mobile agents). After peers are classified into super-peer groups, the other peer selection mobile agent for a specific super-peer group is assigned according to its property. Existing P2P system, however, cannot apply various peer selection mechanisms at a time because only one peer selection mechanism is performed.

3.2. Trust Based Super-Peer Selection. Trust management is any mechanism that allows establishing mutual trust.

Reputation is a measure that is derived from direct or indirect knowledge on earlier interactions of agents and is used to assess the level of trust an agent puts into another agent. Thus, reputation-based trust management is one specific form of trust management. There are three issues that are important to address in any P2P reputation system. (1) The system should not assign any profit to newcomers. That is, reputation should be obtained by consistent good behavior through several transactions, and it should not be advantageous for malicious peers with poor reputations to continuously change their opaque identifiers to obtain newcomers status. (2) The system should have minimal overhead in terms of computation, infrastructure, storage, and message complexity. (3) The system should be robust to malicious collectives of peers who know one another and attempt to collectively subvert the system. An important example of successful reputation management is the online auction system eBay [31]. In eBay's reputation system, buyers and sellers can rate each other after each transaction, and the overall reputation of a participant is the sum of these ratings over the last 6 months. Inspired by eBay, we define that time is a factor influencing our trust value.

In the P2P streaming system, super-peer not only needs to transfer its own streaming media data but also to help other streaming media client to access resources and support information search and data forwarding, which require that super-peer has a high performance requirements, so the peer's performance (e.g., CPU computing power, memory size, disk space, available upload bandwidth, etc.) is a factor influencing our trust value in our proposed trust based strategy.

We propose a mechanism for super-peer selection strategy based on the trust management. In order to address above three problems in P2P system, we define that the trust value of each peer is calculated as follows:

$$Ti = C * \alpha T_{\text{time}} + \beta * \sum_{i,j \in N} Tj * Tji - \gamma D. \quad (1)$$

In the formula, Ti is the trust value of streaming media peer i, C is the peer's capacity points, including peer's CPU computing power, memory size, disk space, available upload bandwidth, and whether it has the public network IP. And C is divided into three grades, that is, excellent, middle, and poor, each level is divided into 1000 points, 500 points, and 100 points, separately. T_{time} is the peer's online time duration points; while T_{time} is longer, their trust value will be higher. In our experiment, we set that T_{time} will be added by 1 points when the peer is online at 1 hour. D is the penalty points; that is, when the resources peer denied service to other peers, D's points will be reduced. In our experiment, we set that D will be reduced by 10 points when the peer refuses to support service once. In a real system, we can cut down the D's points by using antivirus software to detect if the downloaded file is a virus. This mechanism ensures that the peer with less contributing to others will be punished fairly. $\sum_{i,j \in N} Ti * Tji$ is the recommended point, Tj indicates that the node j's trust value, and Tji indicates trust value that the node j trust the node i; when the node j support service for the node i once, the Tji will be added; in our experiment we set 1 point once. α and β are the factors for equity, and γ is the penalty factor. α and β are needed to be set to adequate values according to the strategy of trust and the characteristics of P2P streaming system. The Ti is updated in MAMS 1 hour once or when the stream is transferred completely between node i and j. The recommended points are obtained as follows: peer A establishes trust of peer C through peer B's recommended ($T_{AC} = T_B * T_{AB}$); while there are n recommenders, $T_{AC} = 1/n \sum_{i=1}^{n} T_B * T_{AB_i}$. The initial point of the recommended points is set as 1 and the initial point of the penalty points is set as 0.

Table 1: Parameter values description.

Parameter	Emphasis
α	While the online time is longer, the trust value of peer is bigger.
β	The bigger the value, the bigger the weight of the recommended trust among peers.
γ	When the peer denies serving other peers or fails to exchange resources, the peer's trust value will be reduced.

In the formula, for setting the size of different parameters, the trust measure metrics are different, as specified in Table 1. We chose the maximum trust value of peers as the best super-peer in the network, that is, as MAMS. Mobile agent based peer selection in Step 4 returns a list of sorted peers according to the trust value of peers, and the peer, which is standing in the front, is the priority to be chosen for resource links.

The benefits of using trust management to select super-peer are (1) the trust scheme is the combination of reputation based and trade based. Peers that do not contribute/transmit stream enough will get low trust value, so they are periodically dropped in order to search/discover new potentially better peers. This will ensure that even with the lack of centralized information, if network environment changes, better new network topologies can be discovered. (2) From the given formula (1), we can see that the trust value is difficult to get but easy to lose, which is guaranteeing the fairness of the peers' contribution by using trust value rank mechanism. The formula has no complicated iterative calculation and has good convergence of computing and scalability. (3) The trust mechanism is an incentive mechanism to make peer online longer and share their resources to other peers. When the online time is longer, the trust value of peer is bigger.

4. Experimental Environment

To prove and verify the MATPS performance, the OverSim [32] simulator is used to run and check the proposed MATPS protocol implementation. The experiment's target is to show that MATPS is a stable solution with good performance for improving QoS for P2P streaming system.

Peer-to-peer (P2P) systems are characterized by many annoying real life complexities, such as a high turnover of peers, download and connection failures, large numbers of stochastically behaving peers. A simulation is an attempt

to model a system in order to study it scientifically [33]. Analytical approaches are being applied to P2P systems, but they often require simplifying assumptions that break down under real world conditions. In addition, the intention of many P2P systems is to scale to large numbers of peers; and testing that scalability in the actual system may be impractical. As a result P2P simulation studies for P2P network have been growing in numbers, many simulators have been developed for simulating P2P, such as PlantSim [34], PeerNS [35], PeerSim [36, 37], P2PAM [38], P2PStutio [39], SimGrid [40], ChunkSim [41], OPSS [42, 43], and OverSim [32]. We choose OverSim because it is an open-source overlay and P2P network simulation framework for the OMNeT++ [44] simulation environment. The simulator contains several models for structured (e.g., Chord, Kademlia, and Pastry) and unstructured (e.g., GIA) P2P systems and overlay protocols.

Experimental environment is set up in the server DELL business desktop PC OptiPlex 780 and the simulation runs on it. The server's configurations are as follows: CPU is Intel Core 2 Duo E7500 2.93 GHz, RAM is 2 GB, the operating system is Windows XP, using the OMNeT++, INET [45], and OverSim simulation platform. OMNeT++ is an object-oriented modular discrete event network simulation framework. It has a generic architecture, so it can be used in various problem domains. The INET underlay model is derived from the INET framework of OMNeT++, which includes simulation models of all network layers from the MAC layer. It is useful for simulations of complete backbone structures. The INET underlay model includes simulation models for all network layers. OverSim is a new simulation framework based on OMNeT++. It has been developed in the ScaleNet project at the University of Karlsruhe (Germany). It can simulate up to 100,000 nodes and its design is thoroughly documented. The main reasons for using this tool include the ease of use, a highly flexible and modular architecture, and an open-source code base. Both OMNeT++ and OverSim are built as hierarchical architectures. Simulation is done using the chord protocol of OverSim.

In order to verify the proposed peer selection strategy based on mobile agent and trust (MATPS), in our simulation experiment we had to extend P2P protocol based on mobile agent and the improvement module structure of OverSim based on agent is shown in Figure 3. In Figure 3, agent runtime environment is to support mobile agent running on OverSim; agent creation module is primarily responsible for creating the content distribution agent, content search agent, and content download agent; agent migration module will create the agent for moving to other peers; P2P protocol simulator layer is according to OverSim's P2P network protocol to organize peers and routing, and it is used as the underlying network; P2P application layer mainly provides peers login, resources release, and download functions.

The details of how peers distribute content and download content will be discussed in the following descriptions. Figure 4 is the process diagram of content distribution and Figure 5 is the process diagram of content download.

The steps of content distribution are as follows.

(1) Client creates a content distribution agent (DA).

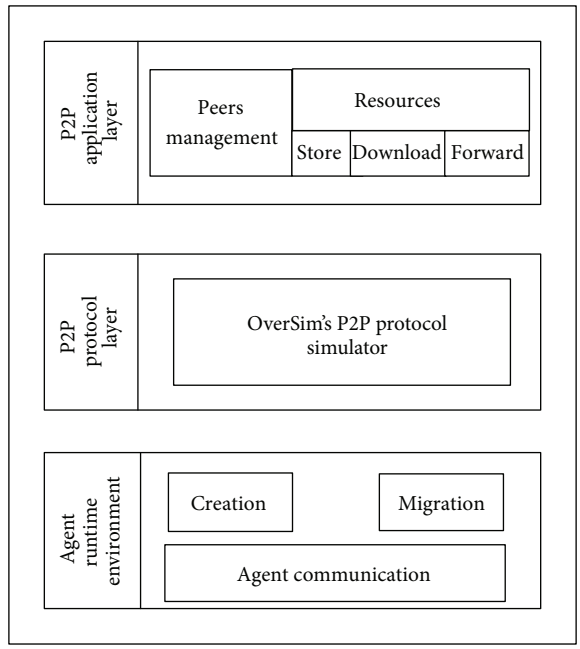


Figure 3: Mobile agent based on OverSim's P2P protocol layer.

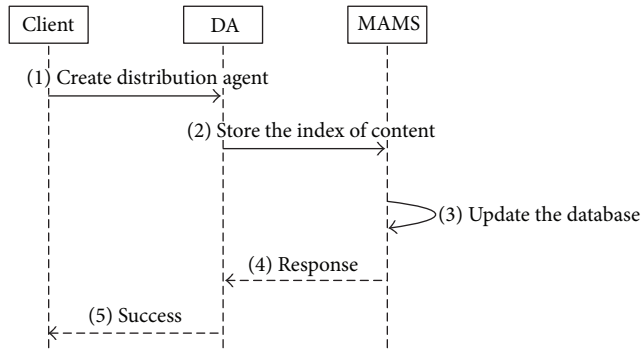


Figure 4: Diagram of content distribution.

(2) DA extracts the index information of shared resources, such as file name, file size, and file type.

(3) DA moves to MAMS. MAMS will create an information agent (IA) for the exchange of information with the DA. And IA will extract the index information from DA; if the index of content does not exist in database, the database will update it into database.

(4) IA will give a response message to inform DA whether the content is distributed or not.

(5) If the content is distributed, then DA will give a success message to its origin client.

The steps of content download are as follows.

(1) Client creates a content download agent (DownA).

(2) DownA moves to MAMS.

(3) MAMS will create an information agent (IA) for the exchange of information with the DownA. IA will

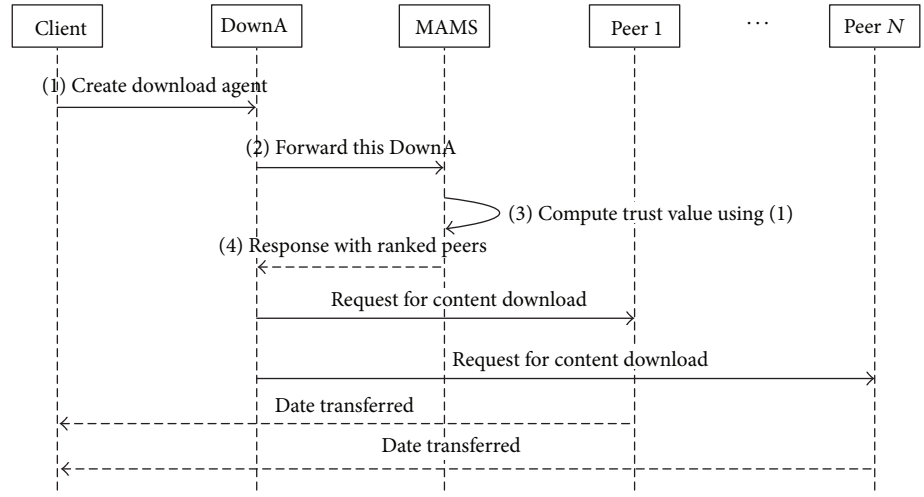


Figure 5: Diagram of content download.

extract the download file KeyID information from DownA, and then MAMS will search for it in its own database. When finishing the search, MAMS will give a list of peers who have the resource and compute each peer's trust value using (1). After that IA will respond by message to DownA with ranked peers list according each peer trust value, which also contains attributes about whether it belongs to local network;

(4) DownA receives the ranked peers and requests those peers for downloading content.

In the experiments, we have set the request issuing rate between 0 and 100 per minute, the nodes churn using the Pareto model of OverSim (OverSim provides several models for generating churn, which includes a lifetime based churn model supporting different distribution functions, such as Weibull, Pareto, or Exponential. Alternatively, a scenario or trace file containing join and leave events can be used to model churn behavior. It is possible to use more than one churn generator at the same time to simulate groups of nodes with different churn behaviors. For each churn generator different node configurations and overlay parameters can be specified, which allows to easily generate complex scenarios with heterogeneous node behavior.), download bandwidth of each node is 64 kbps–3 Mbps, the maximum upload bandwidth is 1 Mbps, the overall system has 500 media files, media playback rate is 256 kbps, the average length of each file is 1800 seconds, and each node has one of the 10~20 media files randomly. After the peers join the system, they can make player operation anytime. The number of nodes that are allowed to perform on-demand operations is controlled by a software script. Detailed simulation parameters are shown in Table 2. Simulation results and performance analysis are shown in the following section.

Table 2: Simulation parameters.

Category	Value
Peers' download bandwidth	64 kbps~3 Mbps
Peers' upload bandwidth	0~1 Mbps
Number of media files	500
Play rate	256 kbps
The average length of media files	30 minute
Frequency of queries issued	0~100 query/minute
Peers	10000
Peers loss	Pareto
Simulation duration	500 minutes

5. Performance Evaluation

The following section first finds the optimal values for α, β, and γ, then evaluates the effectiveness of the proposed method with BFPSS, GridPS, TrBPS, and SPCPS, and gives comparisons of the impact of their jitter, start-up delay, packet loss ratio, network load, and network scale.

5.1. Finding Optimal Values for α, β, and γ. We have described earlier how to determine the best super-peer using (1). We conduct simulations for different values of α, β, and γ (ranges from 0.1 to 0.9) to compute the overall streaming media client's traffic throughput and the overall packet drop ratio for each set of values.

A brief description of the results for different sets of parameters is preset in Table 3. The network topology of simulation experiment for finding optimal α, β, and γ is based on Figure 6 with random nodes attached to super-peers. Link capacities are also chosen randomly on uplink and download links (uplink is from 0 to 1 Mbps, and download is from 64 kbps to 3 Mbps). There are five machines in the LAN, which are running a number of popular P2P applications, such as PPLive, Bittorrent, eMule, and Xunlei. Because OMNET++ is a discrete event simulation environment, in order to simulate large-scale P2P network, we use this small LAN to run OMNET++ simultaneously. In current network, we set up a flow collector, which uses NETMATE [46] software to capture packets and calculate the client's traffic

Table 3: Parameter values.

	$\alpha = 0.1$ $\beta = 0.1$ $\gamma = 0.8$	$\alpha = 0.2$ $\beta = 0.4$ $\gamma = 0.6$	$\alpha = 0.3$ $\beta = 0.5$ $\gamma = 0.7$	$\alpha = 0.4$ $\beta = 0.4$ $\gamma = 0.7$	$\alpha = 0.5$ $\beta = 0.7$ $\gamma = 0.6$	$\alpha = 0.6$ $\beta = 0.6$ $\gamma = 0.8$	$\alpha = 0.7$ $\beta = 0.8$ $\gamma = 0.9$	$\alpha = 0.8$ $\beta = 0.9$ $\gamma = 0.4$	$\alpha = 0.9$ $\beta = 0.9$ $\gamma = 0.7$
Traffic throughput (kbps)	923.57	901.83	1028.36	**1289.25**	1123.01	1127.63	999.34	972.93	1212.49
Packet drop ratio (%)	6.22	6.91	4.06	**3.25**	5.68	7.02	8.52	4.90	5.11

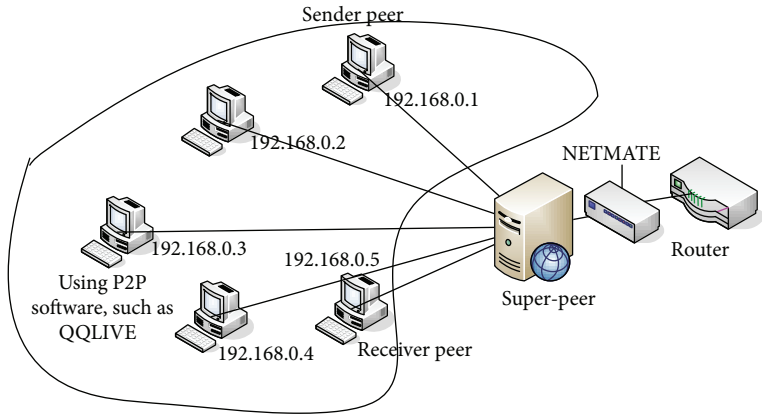


Figure 6: The network topology of experiment for finding optimal α, β, and γ.

throughput and the overall packet drop ratio according to its source address or destination address.

The results presented in the table are averages over 729 runs of simulations and only some of the results are presented due to the layout restrictions of the paper. We noticed that the streaming client (Receiver) receives maximum traffic throughput and low packet drop ratio for the set of values ($\alpha = \beta = 0.4$, and $\gamma = 0.7$) for calculating the trust value Ti. Thus, we give a weight of 0.4 to α and β, and 0.7 to γ.

5.2. The Jitter Performance Results. In the experiments, different P2P clients are used as separate MATPS, BFPSS, GRIDPS, TrBPS, and SPCPS simulator machines according to Figure 6. The jitter and start-up delay are the two performance parameters of the P2P streaming application. Jitter is introduced into the traffic as a result of unexpected delay experienced by the flow on network routers [47]. Assuming that the packet arrives at the receiver node at time t, if the expected arrival time of packet is e, $t - e$ is the jitter. Figure 7 shows the jitter caused by the node failure rate changes in the case of the request rate of 60 (query/min). From Figure 7, we can see that in the beginning of the system start-up stage, the five strategy's jitter rates are unstable, and after 10 minutes, the jitter rates are stable, but the jitter of MATPS and SPCPS strategies remains at a low level, the jitter of BFPSS, GridPS, and TrBPS strategies changes severely. MATPS strategy jitter rate is 0.45% lower than SPCPS strategy. As BFPSS strategy is the bandwidth first used by the selection strategy and does not take into account the stability of the node, GridPS uses random peer selection, TrBPS uses only trust based peer selection, while the MATPS strategy is comprehensive in consideration of the node bandwidth, performance, and so forth and select online time as a standard for peer choice, so the longer time the node in the network takes, the probability of being selected to support service for other peers is increasing. So the average jitter rate of the MATPS is lower than that of the other peer selection strategies.

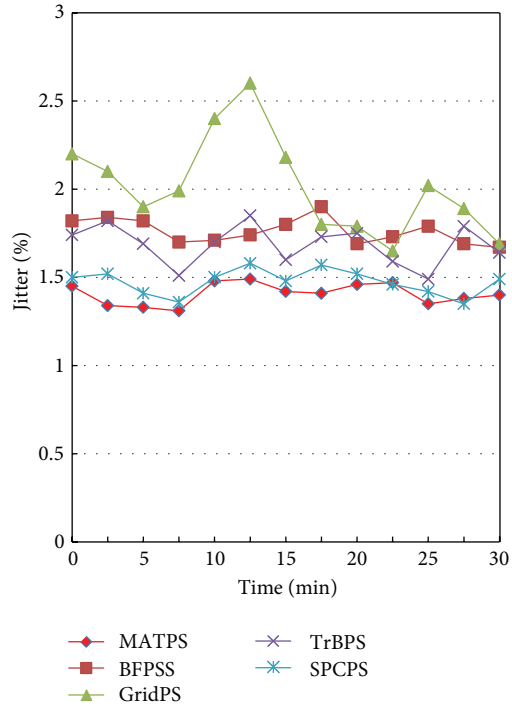


Figure 7: The jitter rates change over time.

5.3. The Start-Up Delay Performance Results. The start-up delays incurred by the clients and the number of channels

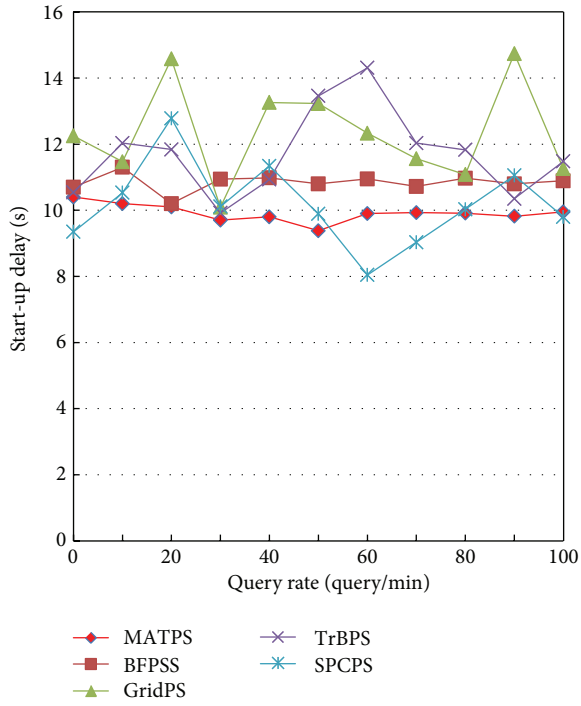


Figure 8: The start-up delay of different request rates.

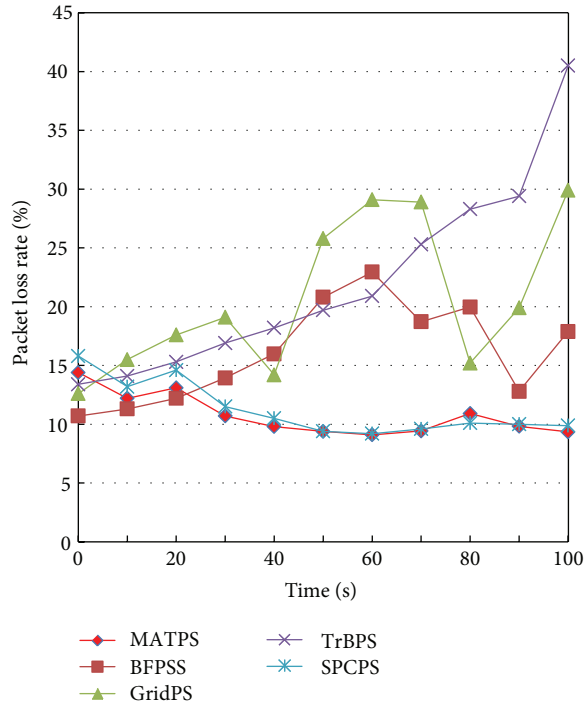


Figure 9: The packet loss ratio performance.

(bandwidth) needed to serve all the requests [48]. In P2P streaming media system, resource query delay and streaming media transmission delay are the main factor of affecting the client start-up delay. Figure 8 is the client start-up delay as the node request rate variation diagram, while the simulation platform to run for some time and the number of nodes in P2P network is stability. From Figure 8, we can see that at the beginning of the playing among 10 seconds, because resources are distributed on each node irregularity, and the number of nodes join P2P network is not enough, the query requests sent but most of resources are not find, which led to a larger search resources delay and influences the client to play. With the increasing number of requests and the stabilized network size, a large number of requests by the network are responded. At the same time, the proposed MATPS selects high bandwidth and local node as a resource provider node, which greatly reduces the network transmission time of media resources; thus, the MATPS has the better start-up delay time performance.

5.4. The Packet Loss Ratio Performance Results. The packet loss ratio is another key performance metric besides the start-up delay in the P2P streaming application [10]. The packet loss will cause the receiving peer to have playback jitters. The playback jitter becomes worse with more packets lost in the receiving peer when necessary stream frames are dropped. In our experiment, (2) as defined below is used to measure the packet loss ratio at the receiving peer. Equation (2) calculates the receiving peer's packet loss ratio by dividing n with N:

$$\text{packet loss ratio} = \frac{n}{N}. \tag{2}$$

In (2), n is defined as the number of the total dropped packets, and N is defined as the total expected receiving packets in the duration of a stream connection session. The following gives the value of N:

$$N = \frac{T_w}{T_p}. \tag{3}$$

In (3), T_w presents the total time that the receiving peer is watching and staying with the stream, and T_p is the playback duration for a packet. In our experiments, T_w is a fixed number in the receiving peer and it always equals thirty minutes in every experiment. T_p is also a constant value in our experiments because it is assumed that the stream is encoded in CBR and the stream is divided to packets with the fixed playback duration in the source node. The packets have variable data size in different quality streams, but they always have the same playback duration. According to the definition of T_w and T_p, it is concluded that N is also a constant value. Thus, the packet loss ratio defined in (3) depends on the value of n.

Figure 9 shows the packet loss experiment results for the five peer selection strategies. MATPS has smaller packet loss ratio than that of the other four strategies. Because the results are based on PC simulation, in a real deployed system the MATPS still has the chance to get lower pack loss ratio performance than the results shown in Figure 9 by increasing the receiving peer's buffer size.

5.5. Network Load. From Figure 10, we can see that with the rate request increasing the MATPS and SPCPS strategy backbone utilization decreased, while the other three

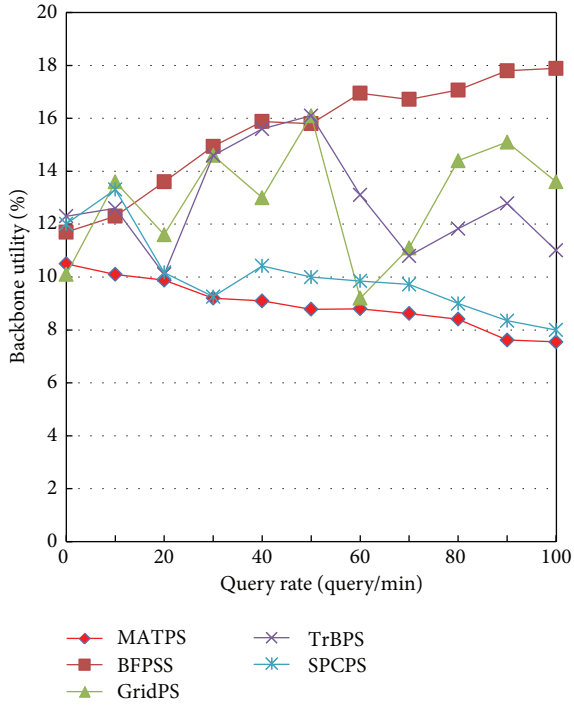


Figure 10: The backbone utilization rates of different requests.

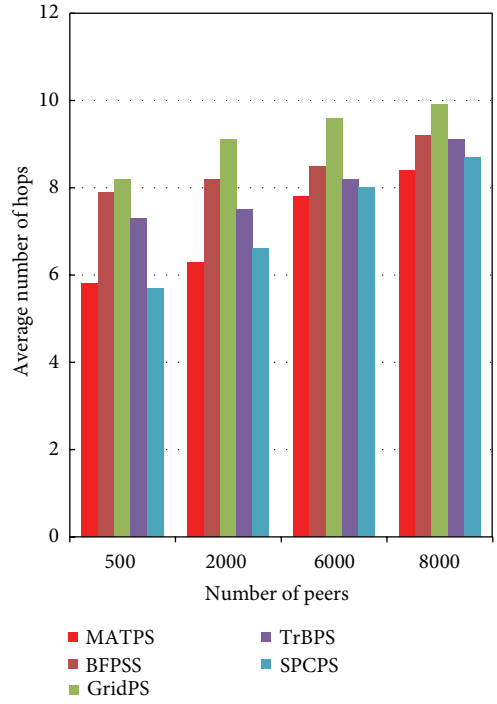


Figure 11: The average number of hops compared with network scale.

strategies' backbone usage increased gradually. In the request rate of 100 (query/min) the backbone utilization of MATPS strategy has fallen to 7.5%, while the backbone utilization rate of BFPSS, GridPS, and TrBPS strategy has increased. Due to the number of requests increasing, more and more traffic of the media data needs to be transferred through the upper network in BFPSS strategy, GridPS, and TrBPS, which makes backbone network usage increasing. While in MATPS strategy, because it uses the way of the local node based on trust preferred, when a growing number of local node requests the same content, which also means more and more nodes become nodes of local and follow-up requests for resource providers, so the more the requester, the more the local resources, and thus reduce the local nodes to exchange resources through the upper network, continuously reduce the backbone network usage. The MATPS makes lower backbone utility than SPCPS.

5.6. The Impact of Network Scale. The maximum number of hops can specify how many times a query is allowed to be transferred in P2P streaming system, which determines how much the network will be flooded by a single query. Generally, the larger the network is, and the greater the cost is. By measuring and comparing the number of hops a query sent to different nodes, we can analyze the performance of the selection algorithm. The network scale is carried out in four conditions: 500, 2000, 6000, and 8000. The average number of hops results is shown in Figure 11. From Figure 11, with the network size increasing, in MATPS node selection algorithm, the hops change little, while in the other four strategies the average number of hops rises. This is because MATPS node selection strategy is based on trust mechanism, and it takes into account local resources, network bandwidth, node performance, and other factors, which can be a good adaptive to change in network size. With the increasing of the network scale, the adaptive choice peers of optimized node information make inquiries in localization, so the average hop remains more stable than other strategies.

Figure 12 shows the average jitter rates of the five strategies with the increase of the network scale. In general, the average jitter rates of the five strategies decrease with the increase of the peers. Figure 13 shows the average start-up delays of the five strategies with the increase of the network scale. In general, the average start-up delays of the five strategies decrease with the increase of the peers. Figure 14 shows the average packet loss rates of the five strategies with the increase of the network scale. In general, the average packet loss rates of the five strategies increase with the increase of the peers. As shown in Figure 14, the average packet loss rates of BFPSS, GridPS and TrBPS increase very quickly. On the contrary, the average packet loss rates of MATPS and SPCPS go up slowly. The average packet loss rate of MATPS is 1 s lower than that of SPCPS strategy. Figure 15 shows the average backbone utilities of the five strategies with the increase of the network scale. In general, the average backbone utilities of the five strategies increase with the increase of the peers. As shown in Figure 15, the average backbone utilities of BFPSS, GridPS and TrBPS increase very quickly, while the average backbone utilities of MATPS and SPCPS go up slowly. The average backbone utility of MATPS is 1% lower than that of SPCPS strategy.

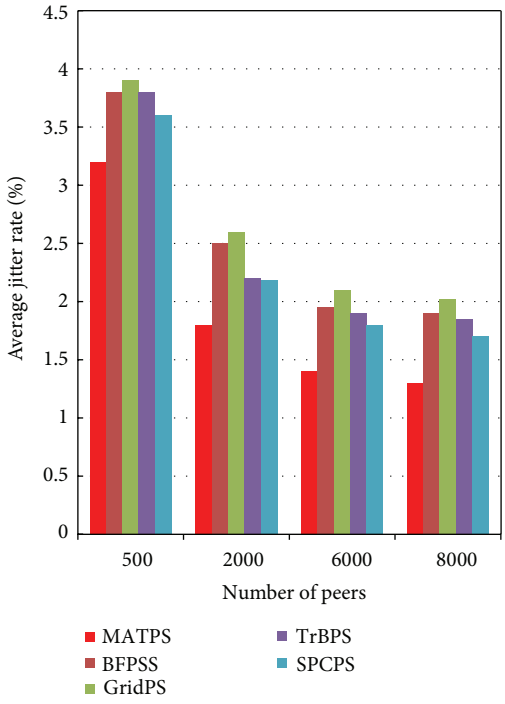


Figure 12: The average jitter rate compared with network scale.

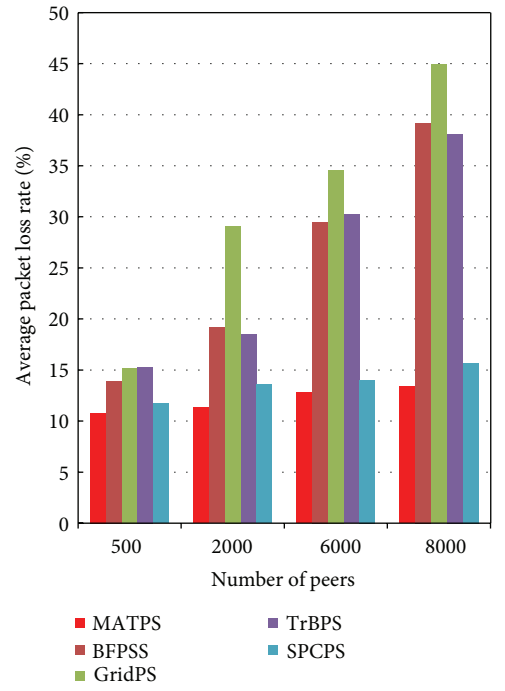


Figure 14: The average packet loss rate compared with network scale.

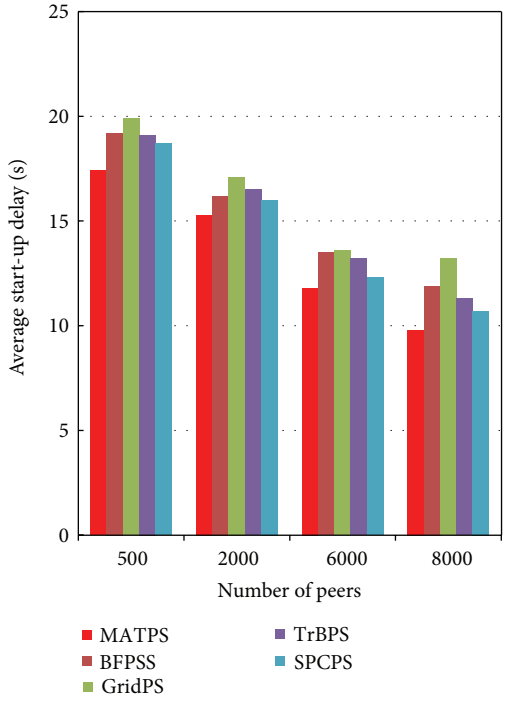


Figure 13: The average start-up delay compared with network scale.

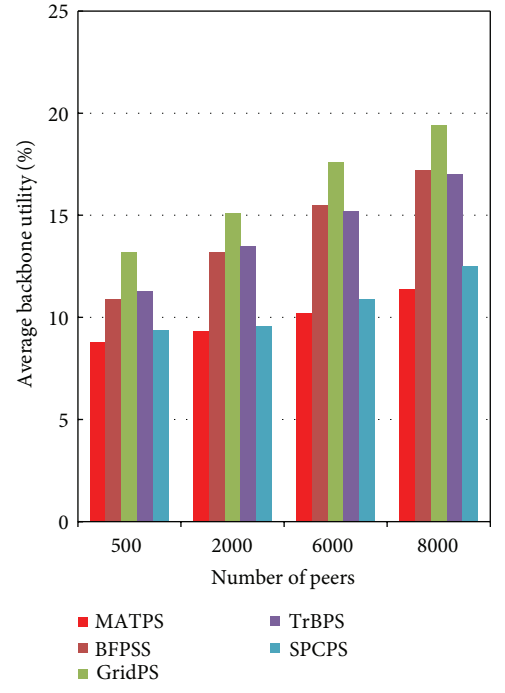


Figure 15: The average backbone utility compared with network scale.

From Figures 11–15, we can find that MATPS strategy has better average indicators than other four strategies, and we can make a conclusion that the proposed MATPS strategy can improve the streaming QoS effectively. In addition, we also select some other simulation parameters to evaluate the performance, and then we found that the performance indicators of other parameters' values of alfa, beta, and gamma are worse than those of optimal parameters, and because we have already done more experiments for finding the optimal values for α, β, and γ in Section 5.1, we have not given the results of the other performance indicators with different simulations parameters of alfa, beta, and gamma.

Table 4: Message's overhead of using MATPS.

	Mobile agent		Trust value	
	Use	No	Use	No
Total packets sent	134235	129894	145069	142872
Message's overhead	3.34%		1.54%	

6. Discussion, Conclusions, and Future Work

6.1. Discussion

Message's Overhead of Using MATPS. When peers login or quit, and the trust value of each peer is updating, refresh messages are distributed through creating agents. As for the message's overheads with using mobile agent and trust or not, from Table 4 we can see that in our simulation experiment, the message's overhead of using mobile agent is negligible, that is, only a percentage of 3.34 for using mobile agent and 1.54 for updating trust value.

6.2. Conclusions. In this paper, we have presented a peer selection strategy for improving the QoS requirements of P2P streaming system. The basic idea is to use agent to select peer and to choose peer as super-peer which has high trust value. Also a new architecture to simulate P2P network is presented, which uses agent technology to OverSim platform. The proposed MATPS scheme aims at improving the QoS of P2P streaming system. Simulation results show that our scheme can effectively select good peer to support sources. Our experiments results also show that the Internet backbone utilization of the P2P streaming system has been decreased and the QoS requirements of P2P streaming system can be improved comparing with other peer selection strategies.

6.3. Future Work. Future research can proceed in the following directions. First, we are developing the P2P streaming system implementing MATPS. How to design the system architecture of such a system and how to enhance the current available streaming protocol to support MATPS are both challenging work. Second, in this paper we point out that the value of α, β, and γ is optimized set through experiment, but if these parameters are really optimal in all use cases we need to do more experiments. A technique for selecting these parameters automatically by the system is also worth further study. Third, in real life, some contents are typically very popular and some are not; the cache technology is needed to be integrated with MATPS in the future. Finally, the client behind the firewall may not be able to allow the outside clients to get the stream, and this paper did not consider the security issues of mobile agent. We will study mechanisms to overcome these questions in the future.

Conflict of Interests

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

Acknowledgments

The subject is sponsored by the National Natural Science Foundation of P. R. China (nos. 61170065, 61103195, 61203217, 61202355), the Natural Science Foundation of Jiangsu Province (BK2011755, BK2012436, and BK20130882), Scientific & Technological Support Project of Jiangsu Province (nos. BE2011189, BE2012183, and BE2012755), Scientific Research & Industry Promotion Project for Higher Education Institutions (JHB2012-7), Jiangsu Provincial Research Scheme of Natural Science for Higher Education Institutions (12KJB520009), Doctoral Fund of Ministry of Education of China (20103223120007 and 20123223120006), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and Talent Introduction Foundation of Nanjing University of Posts and Telecommunications (NY213034). The authors would like to thank the editors and the anonymous reviewers, who provided insightful and constructive comments for improving this paper.

References

[1] "Cisco Systems, Inc., Cisco Visual Networking Index: Forecast and Methodology, 2012–2016," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf.

[2] A. Bikfalvi, J. García-Reinoso, I. Vidal, and F. Valera, "A peer-to-peer IPTV service architecture for the IP Multimedia Subsystem," *International Journal of Communication Systems*, vol. 23, no. 6-7, pp. 780–801, 2010.

[3] N. Efthymiopoulos, S. L. Tompros, A. Christakidis, K. Koutsopoulos, and S. Denazis, "Enabling live video streaming services realization in telecommunication networks using P2P technology," *International Journal of Communication Systems*, vol. 24, no. 10, pp. 1354–1374, 2011.

[4] C. Lin, "Enhancing P2P live streaming performance by balancing description distribution and available forwarding bandwidth in P2P streaming network," *International Journal of Communication Systems*, vol. 24, no. 5, pp. 568–585, 2011.

[5] M. Amad, A. Meddahi, D. Aïssani, and G. Vanwormhoudt, "GPM: a generic and scalable P2P model that optimizes tree depth for multicast communications," *International Journal of Communication Systems*, vol. 25, no. 4, pp. 491–514, 2012.

[6] A. Radu and G. Baudoin, "Intelligent agents for quality of service evaluation in multimedia services," in *Proceedings of the IADIS International Conference (WWW/Internet '03)*, pp. 1219–1224, Algarve, Portugal, 2003.

[7] J. Jannotti, D. K. Gifford, K. L. Johnson, M. F. Kaashoek, and J. W. O'Toole Jr., *Overcast: Reliable Multicasting with an Overlay Network*, USENIX, Berkeley, Calif, USA, 2000.

[8] L. Zhao, J. Luo, J. Luo et al., "Gridmedia: a practical peer-to-peer based live video streaming system," in *Proceeding of IEEE 7th Workshop on Multimedia Signal Processing (MMSP '05)*, November 2005.

[9] S. Marti and H. Garcia-Molina, "Taxonomy of trust: categorizing P2P reputation systems," *Computer Networks*, vol. 50, no. 4, pp. 472–484, 2006.

[10] W. Y. Wang and Y. W. Chen, "SmartPeerCast: a Smart QoS driven P2P live streaming framework," *Multimedia Tools and Applications*, vol. 54, no. 2, pp. 445–471, 2011.

[11] G. Fortino and W. Russo, "Using P2P, GRID and Agent technologies for the development of content distribution networks," *Future Generation Computer Systems*, vol. 24, no. 3, pp. 180–190, 2008.

[12] V. Gorodetskiy, O. Karsaev, V. Samoilov, and S. Serebryakov, "Agent-based service-oriented intelligent P2P networks for distributed classification," in *Proceedings of the International Conference on Hybrid Information Technology (ICHIT '06)*, pp. 224–233, November 2006.

[13] G. Moro, A. M. Ouksel, and C. Sartori, "Agents and Peer-to-Peer computing: a promising combination of paradigms," in *Proceedings of the 1st International Workshop: Agents and Peer-to-Peer Computing (AP2PC '02)*, Lecture Notes in Artificial Intelligence, pp. 1–14, Springer, Berlin, Germany, July 2002.

[14] D. G. Yang, H. Wang, C. R. Wang, and Y. Gao, "MPSS: A multi-agents based P2P-SIP real time stream sharing system," in *Proceedings of the 9th Pacific Rim International Workshop on Multi-Agents (PRIMA '06)*, vol. 4088 of *Lecture Notes in Computer Science*, pp. 398–408, Springer, Berlin, Germany, 2006.

[15] B. Zhu, S. Jajodia, and M. S. Kankanhalli, "Building trust in peer-to-peer systems: a review," *International Journal of Security and Networks*, vol. 1, no. 1-2, pp. 103–112, 2006.

[16] X. Jin, Q. Xia, and S.-H. G. Chan, "Building a monitoring overlay for peer-to-peer streaming," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '06)*, December 2006.

[17] L. Xiong and L. Liu, "PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 7, pp. 843–857, 2004.

[18] K. Aberer and Z. Despotovic, "Managing trust in a peer-2-peer information system," in *Proceedings of the 10th International Conference on Information and Knowledge Management (CIKM '01)*, pp. 310–317, Atlanta, Ga, USA, November 2001.

[19] M. Eisenhardt, W. Müller, A. Henrich, D. Blank, and S. El Allali, "Clustering-based source selection for efficient image retrieval in peer-to-peer networks," in *Proceedings of the 8th IEEE International Symposium on Multimedia (ISM '06)*, pp. 823–828, San Diego, Calif, USA, December 2006.

[20] C. M. Huang, T. H. Hsu, and M. F. Hsu, "Network-aware P2P file sharing over the wireless mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 1, pp. 204–210, 2007.

[21] S. Sun, A. Abraham, G. Zhang, and H. Liu, "A particle swarm optimization algorithm for neighbor selection in peer-to-peer networks," in *Proceedings of the 6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM '07)*, pp. 166–170, Elk, Poland, June 2007.

[22] L. Yao, G. Lei, L. Fei, and C. Songqing, "A case study of traffic locality in Internet P2P live streaming systems," in *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems (ICDCS '09)*, pp. 423–432, Montreal, Canada, June 2009.

[23] D. Ciullo, M. A. Garcia, A. Horvath et al., "Network awareness of P2P live streaming applications: a measurement study," *IEEE Transactions on Multimedia*, vol. 12, no. 1, pp. 54–63, 2010.

[24] M. Zhang, J. G. Luo, L. Zhao, and S. Yang, "A Peer-to-Peer network for live media streaming- using a push-pull approach," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*, pp. 287–290, 2005.

[25] H. Y. Xie, A. Krishnamurthy, and L. Popkin, "P4P: provider portal for (P2P) applications," in *Proceedings of the ACM SIGCOMM Conference*, 2008.

[26] R. M. Sreenath and M. P. Singh, "Agent-based service selection," *Web Semantics*, vol. 1, no. 3, pp. 261–279, 2004.

[27] M. Luck, P. McBurney, and C. Preist, "A manifesto for agent technology: towards next generation computing," *Autonomous Agents and Multi-Agent Systems*, vol. 9, no. 3, pp. 203–252, 2004.

[28] D. Wong, N. Paciorek, and D. Moore, "Java-based mobile agents," *Communications of the ACM*, vol. 42, no. 3, pp. 92–102, 1999.

[29] S. Choi, M. Baik, and C. Hwang, "Location management & message delivery protocol in multi-region mobile agent computing environment," in *Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS '04)*, pp. 476–483, March 2004.

[30] "Kademlia: a Peer-to-Peer information system based on the XOR metric," in *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, P. Druschel, F. Kaashoek, and A. Rowstron, Eds., vol. 2429 of *Lecture Notes in Computer Science*, pp. 53–65, 2002.

[31] eBay, http://www.ebay.com/.

[32] I. Baumgart, B. Heep, and S. Krause, "OverSim: a flexible overlay network simulation framework," in *Proceedings of 10th IEEE Global Internet Symposium (GI '07)*, pp. 79–84, Anchorage, Alaska, USA, May 2007.

[33] A. M. Law, W. D. Kelton, and W. D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, NY, USA, 1991.

[34] J. Pujol-Ahulló and P. García-López, "PlanetSim: an extensible simulation tool for peer-to-peer networks and services," in *Proceedings of the IEEE 9th International Conference on Peer-to-Peer Computing (P2P '09)*, pp. 85–86, Piscataway, NJ, USA, September 2009.

[35] S. K. Dhurandher, S. Misra, M. S. Obaidat, I. Singh, R. Agarwal, and B. Bhambhani, "Simulating peer-to-peer networks," in *Proceedings of the 7th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA '09)*, pp. 336–341, Piscataway, NJ, USA, May 2009.

[36] M. Jelasity, A. Montresor, G. P. Jesi et al., "The Peersim simulator," 2010, http://peersim.sourceforge.net/.

[37] E. J. L. Lu, Y. T. Jang, and G. W. Hwang, "PeerSim Cooker: a guide for PeerSim," in *Proceedings of the International Computer Symposium (ICS '08)*, pp. 267–272, Tamkang University, New Taipei, Taiwan, 2008.

[38] M. Agosti, F. Zanichelli, M. Amoretti et al., "P2PAM: a framework for peer-to-peer architectural modeling based on PeerSim," in *Proceedings of the 1st International Conference on Simulation Tools and Techniques for Commuications, Networks and Systems & Workshops*, pp. 1–10, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, Belgium, 2008.

[39] N. Kotilainen, M. Vapa, A. Auvinen, M. Weber, and J. Vuori, "P2PStudio-monitoring, controlling and visualization tool for peer-to-peer networks research," in *Proceedings of the ACM international workshop on Performance monitoring, measurement, and evaluation of heterogeneous wireless and wired networks (PM2HW2N '06)*, pp. 9–12, ACM, Cologne, Germany, October 2006.

[40] H. Casanova, A. Legrand, and M. Quinson, "SimGrid: a generic framework for large-scale distributed experiments," in *Proceedings of the 10th International Conference on Computer*

Modelling and Simulation (EUROSIM/UKSim '08), pp. 126–131, IEEE, Piscataway, NJ, USA, April 2008.

[41] J. Kangasharju, U. Schmidt, D. Bradler et al., "ChunkSim: simulating peer-to-peer content distribution," in *Proceedings of the 2007 Spring Simulaiton Multiconference*, pp. 25–32, Society for Computer Simulation International, San Diego, Calif, USA, 2007.

[42] L. Bracciale, F. L. Piccolo, D. Luzzi et al., "OPSS: an overlay Peer-to-peer streaming simulator for large-scale networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 35, no. 3, pp. 25–27, 2007.

[43] L. Bracciale, F. L. Piccolo, S. Salsano et al., "Simulation of Peer-to-peer streaming over large-scale networks using OPSS," in *Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools*, pp. 1–10, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, Belgium, 2007.

[44] "OMNet++," http://www.omnetpp.org/.

[45] "INET," http://inet.omnetpp.org/.

[46] "NETMATE 0.9.4," http://www.ip-measurement.org/tools/netmate/.

[47] J. Khan and R. Zaghal, "Jitter and delay reduction for time sensitive elastic traffic for TCP-interactive based world wide video streaming over ABone," in *Proceedings of the 12th IEEE International Conference on Computer Communications and Networks (ICCCN '03)*, pp. 311–318, Dallas, Tex, USA, 2003.

[48] A. Bar-Noy, J. Goshi, and R. E. Ladner, "Off-line and on-line guaranteed start-up delay for Media-on-Demand with stream merging," *Journal of Discrete Algorithms*, vol. 4, no. 1, pp. 72–105, 2006.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 875342, 13 pages
http://dx.doi.org/10.1155/2013/875342

Research Article

An Improved Centralized Cognitive Radio Network Spectrum Allocation Algorithm Based on the Allocation Sequence

Jianli Zhao and Jinsha Yuan

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, 071003 Hebei, China

Correspondence should be addressed to Jianli Zhao; jianlizhao891@126.com

Received 20 May 2013; Revised 28 July 2013; Accepted 30 July 2013

Academic Editor: Deguang Le

The demand of the wireless communications service is increasing in recent years, and the demand of the wireless communication rate is increasing too. However, the static spectrum allocation mechanism results in lots of free spectrum resources in space, and the spectrum is wasted. The cognitive radio technology with intelligent searching and efficient utilization of the idle spectrum resources just provides the opportunity to use the free spectrum resources. Thus, the spectrum allocation technology of the cognitive radio has also been more and more broadly concerned. In order to avoid the interference, a mathematical model of spectrum allocation for cognitive radio spectrum allocation algorithm is established through a graph theory. The spectrum allocation algorithms with different allocation objectives based on graph theory are studied according to the different network structures. After the basic ideas and allocation processes of these algorithms are described, a centralized network spectrum allocation algorithm based on the allocation sequence is proposed. Simulation results show that the improved algorithm can significantly reduce the time overhead and can have better fairness compared with other existing Collaborative-Max-Sum-Reward (CSUM), Collaborative-Max-Min-Reward (CMIN), and Collaborative-Max-Proportional-Fair (CFAIR) algorithms under the same design criterion.

1. Introduction

Wireless technology is playing a more and bigger fundamental role in the Internet than it has nowadays. The radio frequency spectrum has been chronically regulated with static spectrum allocation policies since the early 20th century. With the fast growing services and devices based on the spectrum, the remaining available spectrum for future wireless services is being exhausted, known as the spectrum scarcity problem [1, 2]. The current fixed spectrum allocation scheme leads to significant spectrum white spaces, where many allocated spectrum blocks are used only in certain geographical areas. Recognizing that the traditional spectrum management process can stifle innovation, and it is difficult to provide a certain quality of service (QoS) for systems operated in unlicensed spectrum [3–6], Federal Communications Commission (FCC) has proposed new spectrum management models and the use of a measure of interference temperature. Current spectrum management methods include command and control, exclusive usage based on license, commons, interference temperature, and fast command and control [7–9].

Existing cognitive radio network architecture model mainly has distributed list coloring spectrum allocation and centralized coloring spectrum allocation [10–12]. Dr. Joseph Mitola and Jondral Professor are the founder of cognitive radio; they proposed the centralized spectrum pooling system [13–15], and Bell Laboratories and Stevens Institute proposed DIMSUMnet (Dynamic Intelligent Management of Spectrum for Ubiquitous Mobile-access network) system [16]. IEEE 802.22 Working Group studied WRAN (Wireless Regional Area Network). Microsoft Asia Research Institute and University of California-Santa Barbara cooperates distributed spectrum sharing Nautilus project that is designed to emphasize distributed coordination enabled spectrum sharing,without relying on centralized control [17]. The United States Army's DARPA XG program aims to implement the policy based intelligent radios known as cognitive radios [18]. Professor Brodersen proposed CORVUS (cognitive radio approach for usage of virtual unlicensed spectrum) system [19]. The European OverDRiVE (Spectrum Efficient Uni- and Multicast Services Over Dynamic Radio Networks in Vehicular Environments) project aims at UMTS enhancements

and coordination of existing radio networks into a hybrid network to ensure spectrum efficient provision of mobile multimedia services [20].

Among the graph theory methods of spectrum allocation, sequence-based spectrum allocation algorithm is proposed to solve the problem of spending too much time overhead to maximize the minimum income criteria by the spectrum allocation algorithm based on graph theory centralized network algorithm. Firstly, establish the mathematical model of graph theory methods to solve the problem of spectrum allocation. Secondly, analyze the basic idea of the different network structures under the existing graph theory algorithms and summarize their advantages and disadvantages. Finally, CSGC algorithm maximizing the minimum income criteria is not well adapted to the needs of the centralized network, and it must improve the time overhead, so this paper proposed an allocation sequence based on the spectrum allocation algorithm.

2. Graph Theory Modeling of Spectrum Allocation

The cognitive radio network uses the spectrum sensing function to find an available idle spectrum, and then the idle spectrum resources are divided into many channels allocated to the cognitive users of the network, and the cognitive users can communicate. Therefore, the key issue of affecting spectrum allocation algorithm design is how to know the spectrum sensing function detected for the communication services of the cognitive users. Currently, there are two main models to determine each cognitive users' rights: {0, 1} model and interference temperature model. It can also be understood from the foregoing cognitive users on the use of spectrum to the two models. Overlay is corresponding to {0, 1} model, and underlay is corresponding to the interference temperature model.

2.1. {0, 1} *Model.* {0, 1} model with two states 0 and 1 indicates the cognitive users' availability. Cognitive users possessing spectrum sensing function can detect whether the authorized user is using each band in a broad band. If the band is being used, then the cognitive user cannot hire this band, and this band is marked "0" by the cognitive users. If the users are not authorized using this band, this band is the optional band selected by cognitive users, and it can be labeled "1" by the cognitive users. When each detected band is detected, each cognitive user gets the availability form of the spectrum resource, and authorized users' geographical location uses the different states as well as the limit of their perceived ability. The form of the cognitive users is also different.

The available form of the spectrum resource is just the start of the real communication business to obtain the results of the cognitive users' first screening channel, and thereafter also to consider the interference between other cognitive users, and the second screening process is done. After two rounds of screening, cognitive users really learn using the communication channel resources. In addition, in the process of the communication service, if the authorized users have the communication needs, the cognitive users need to immediately stop the use of the band, and the band is marked as "0" by the cognitive users, which are no longer available. After the authorized users have finished using the band, the cognitive users' some detection cycle is marked as "1," which is the available frequency bands.

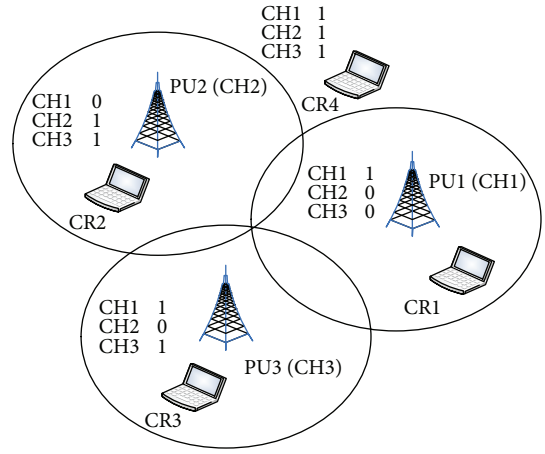


Figure 1: {0, 1} model diagram.

The model is applied to the actual operation, and the authorized users' power received by the measured cognitive users' spectrum detecting function is less than the power threshold, and the spectrum used by the authorized users is the idle spectrum resources of the cognitive users. In the spectrum allocation algorithm design, the distance of the cognitive users and authorized users is greater than the distance threshold, and the band used by the authorized users can be used as the available band. A specific example can refer to three cells of the authorized network and multiple cognitive users' scene shown in Figure 1. The authorized users (PU1, PU2, and PU3) are three base stations, respectively, using the channels CH1, CH2, and CH3, and four cognitive users' (CR1, CR2, CR3, and CR4) optional spectrum list is also illustrated in Figure 1.

2.2. Interference Temperature Model. The interference temperature concept is proposed by the United States Federal Communications Commission (FCC), whose aims are to measure the size of the interference and to effectively manage interference [21]. The interference here refers to the interference of PU, and the interference is the interference of the reception power. Interference temperature model is to ensure that the interference of PU to cognitive user (SU) cannot exceed the interference temperature limit. SUs based on the measured current interference situation do real-time adjustment of the transmitter parameters such as operating frequency and transmit power in order to meet this requirement. There is a noise temperature, and its definition is equivalent to the interference temperature, and they are the powers measuring the interference. The corresponding units are Kevin, which is defined as in (1):

$$T_I(f_c, B) = \frac{P_I(f_c, B)}{kB}, \quad (1)$$

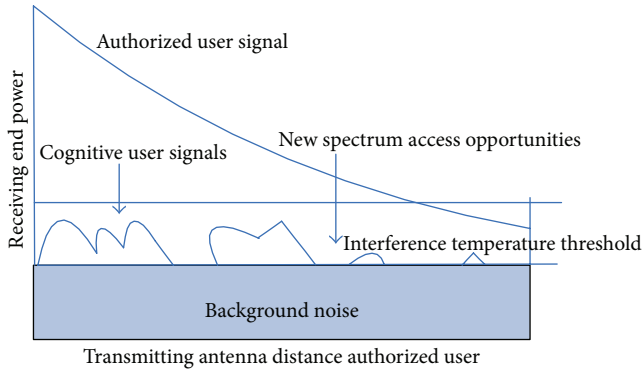


Figure 2: Interference temperature model.

where $P_I(f_c, B)$ is watt (W) having a center frequency f_c of the bandwidth B with the average interference level, $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $T_I(f_c, B)$ is the interference temperature.

Another term related to the interference temperature is interference temperature threshold T_L, and it is used to determine the size of the tolerate cognitive users' interference to determine whether cognitive users can use the authorized users' band as shown in Figure 2.

The literature [11] defined the ideal model and the generic model of the interference temperature and has a detailed analysis and comparison of two models, and the signal and the interference as well as the relationship of bandwidth and center frequency are explicitly discussed.

(1) Ideal Model. When considering the background noise and interference signals from cognitive users, the cognitive restrictions and requirements of the users' receiver parameters are defined in the authorized users' bandwidth. Assume that the average power of cognitive users is P, and the center frequency is f_c, and the signal band width is B. There are n bands with the width B_i in the bandwidth B; the center frequency f_i of the authorized users has the following formula (M_i is the fading coefficient from the cognitive users to the authorized users, which is in (0, 1) range):

$$T_I(f_i, B_i) + \frac{M_i P}{kB_I} \leq T_i(f_i). \quad (2)$$

An approximation of the interference temperature is as follows ($P(f)$ is the power at the frequency f; τ is the safe redundancy with the unit kHz):

$$T_I(f_c, B) \approx \frac{P(f_c - B/2 - \tau) + P(f_c - B/2 + \tau)}{2kB}. \quad (3)$$

(2) Common Model. Consider the background noise, authorized users, and other cognitive users; the restriction of the cognitive users' transmitter parameters is defined in the cognitive user bandwidth:

$$T_I(f_c, B) + \frac{MP}{kB} \leq T_L(f_c). \quad (4)$$

The main difference between the two models is that the latter cannot distinguish between the signal of authorized users and interference, and the interference temperature model is for the entire frequency band. And the monitoring point of the former is in the authorized user, so the authorized signals are easily distinguished from the interference and noise, and interference temperature model is for the specific authorized users.

However, the interference temperature is still a vague concept to the cognitive equipment manufacturer. So the literature [12] proposed the prediction and assessment model with the small interference to the authorized users, and the model has guiding significance for using broadcast television band for the unauthorized users. But this model is impacted by the cognitive users' signal modulation parameters, antenna parameters, and power control and detects authorized user channel capacity and many other factors.

Interference temperature model for cognitive radio spectrum allocation algorithm usually considers cognitive users using the same spectrum of the authorized users, and the interference from cognitive users to authorized users is limited by the interference temperature. The literature [20] pointed out that this model can be efficiently used for dynamic spectrum access. Due to the problem of spectrum allocation for cognitive radio networks, the literature [18] designed cognitive wireless mesh network channel selection algorithm based on the interference temperature model.

Finally, the cognitive radio network may run on the licensed band and the nonauthorized band depending on the type of free spectrum resources, and it can also be run in a mixed manner. When the cognitive users' idle spectrum is unlicensed spectrum such as ISM (Industrial Scientific and Medical) band, it is referred to as "horizontal sharing" spectrum; when the cognitive users' idle spectrum is unlicensed spectrum using the licensed spectrum, it is referred to as "vertical sharing" spectrum.

Assume that in the cognitive radio network area there are a P-authorized user (PU) and N cognitive users (SU), and the system available spectrum is divided into M same bandwidth and completely orthogonal sub channels. In order to share the channel available information of the respective nodes and the central node as well as the delivery channel allocation message, the system also requires the specialized transmission control channel.

In the coloring theory, the available channel is determined according to the distance of the cognitive users and the authorized users, and whether there is interference according to the distance between the cognitive users and other cognitive users. As shown in Figure 3, assume that each user uses the omnidirectional antenna in the system, and the transmission range of the authorized users is the circular area for the radius R_P, and the cognitive users' transmission range is the circular area for the radius R_S. If the distance between the cognitive users and an authorized user is less than $R_S + R_P$, then the cognitive user will not be able to use this spectrum of authorized users; otherwise you can use it. If the distance between the cognitive users and the authorized user is less than $2R_S$, then there exists the interference between the two cognitive users; otherwise there is no interference. In order to

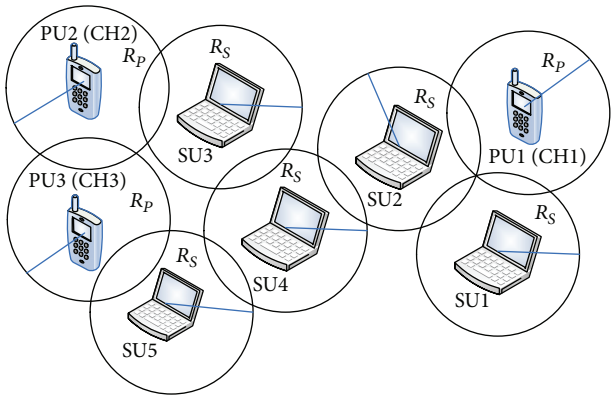


Figure 3: Cognitive radio spectrum allocation scene.

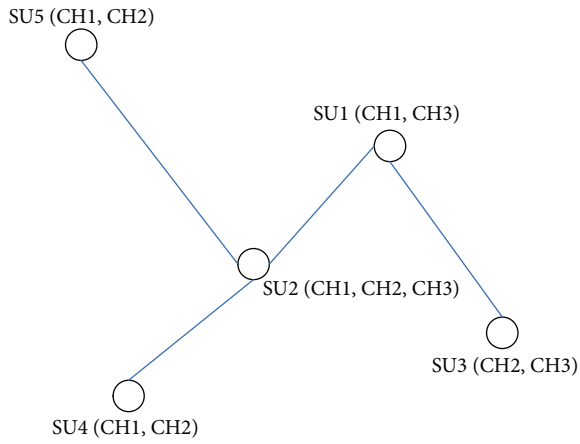


Figure 4: Graph topology of cognitive radio network.

use graph theory to analyze, all cognitive users are abstracted to the vertices of the graph, and the interference between the cognitive users is understood as an edge between vertex and vertex. Thus, Figure 3 can get the network topology diagram of the cognitive radio network graph theory assignment model in Figure 4, where CH1, CH2, and CH3 represent the available subchannels from authorized users. Because the edges among each vertex are the undirected segments, so Figure 4 is an undirected graph.

It should be noted that the network topology of Figure 4 represents just the mutual interferences of one moment cognitive user between the cognitive radio networks. Because the authorized users' spectrum usage status is constant conversion between idle and busy in reality, the available channel has the time-varying characteristics. In addition, authorized users and cognitive users must consider the impact of mobility on the interference between them. The spectrum allocation algorithm of the graph theory model is generally assumed that cognitive users' position is unchanged and the cognitive users' available spectrum resources are unchanged in the process of implementation of the allocation algorithm; namely, the graph topology of the cognitive radio network is unchanged.

The graph theory topology of Figure 4 can be further described by mathematical method, and the undirected graph of Figure 4 is recorded as $G = (V, E, L)$, where and V, E, and L, are respectively, call vertex vector, interfere matrixed, and channel availability matrix. In addition, the channel allocation result can be represented by conflict-free channel allocation matrix A.

(1) Vertex Vector. V is the vertex of the graph, which represents N cognitive users.

(2) Interference Matrix. $E = \{e_{ij} \mid e_{ij} \in \{0, 1\}, i, j = 1, 2, \ldots N\}$ is a matrix with N rows and N columns, which represents all edges of the graph, and $e_{ij} = 1$ represents the mutual interferences of cognitive user i and cognitive user j, and there is an edge between the corresponding vertices; $e_{ij} = 0$ indicates no interference between cognitive user i and cognitive user j, and there is no edge between the corresponding vertices.

(3) Channel Availability Matrix. $L = \{l_{ik} \mid l_{ik} \in \{0, 1\}, i = 1, 2, \ldots N; k = 1, 2, \ldots M\}$ is a matrix with N rows and M columns, which represents the availability of N cognitive users to M subchannels, and $l_{ik} = 1$ indicates that cognitive user i can use the kth subchannel, and $l_{ik} = 0$ indicates that cognitive user i cannot use the kth subchannel due to the interference to the authorized users.

(4) Conflict-Free Channel Assignment Matrix. $A = \{a_{ik} \mid a_{ik} \in \{0, 1\}, i = 1, 2, \ldots N; k = 1, 2, \ldots M\}$ is a matrix with N rows and M columns, which represents the final distribution of N cognitive users to M subchannels, and $a_{ik} = 1$ indicates that cognitive user i shares the kth subchannel, and $a_{ik} = 0$ indicates that cognitive user i cannot use the kth subchannel.

Obviously, if $l_{ik} = 0$, then $a_{ik} = 0$. This is because only when the channel is available, this channel may be assigned. Further, since any two connection nodes cannot allocate the same channel, it is possible to obtain the spectral distribution of the interference constraints [13]:

$$a_{ik} a_{jk} e_{ij} = 0, \mid \forall i, j = 1, 2 \ldots, N; \; k = 1, 2, \ldots, M. \tag{5}$$

3. Existing Algorithms

3.1. Distributed List Coloring Spectrum Allocation. Three cognitive radio spectrum allocation algorithms were proposed in the case of distributed network [13]: Distributed Greedy Algorithm (DGA), Distributed Fairness Algorithm (DFA), and Randomized Distributed Algorithm (RDA). And the spectrum allocation using mathematical methods is described by formula (6) as follows:

$$\begin{aligned} a_{ik} &\le l_{ik}, \\ a_{ik} a_{jk} e_{ij} &= 0, \\ a_{ik} &= 0, 1, \end{aligned} \tag{6}$$

where $a_{ik} = 0$ is that cognitive user i cannot use the kth subchannel and vice versa, e_{ij} represents interference between

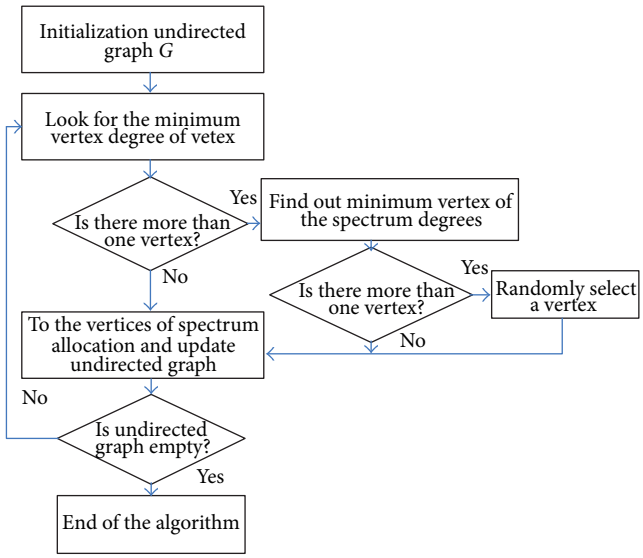


Figure 5: Distributed greedy algorithm.

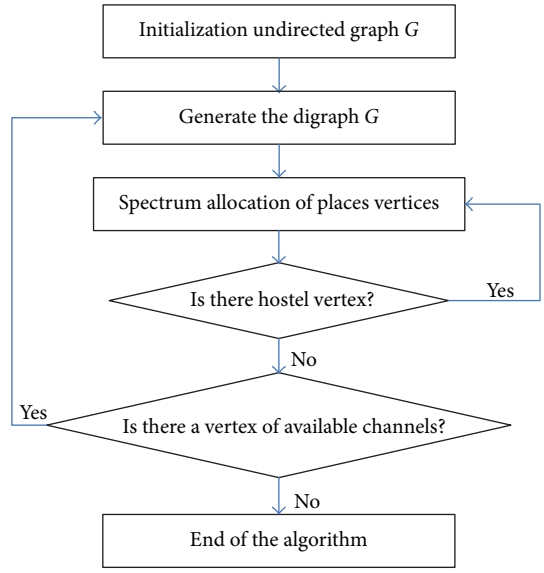


Figure 6: Distributed fairness algorithm.

the cognitive user i and the cognitive user $j, i, j = 1, 2 \ldots, N$; $k = 1, 2, \ldots, M$, this is a list coloring problems with three constraints, and it can be seen as M available subchannels to assign colors.

(1) Distributed Greedy Algorithm. DGA's goal is to achieve the maximizing spectrum utilization of formula (7) in the case of satisfying formula (6) in the three constraints; that is, the results of the spectrum allocation are to let the maximum of the number of subchannels shared by all the cognitive users:

$$\max \sum_{i=1}^{N} \sum_{k=1}^{M} a_{ik}. \tag{7}$$

In order to facilitate describing the algorithm, the number of vertices d_i of each vertex is defined as the sum of the vertex v_i with the interference of other vertices, that is, formula (8):

$$d_i = \sum_{j=1, j \neq i}^{N} e_{ij}. \tag{8}$$

The number of vertices d_i may also be referred to the number of connections, and it is the number of neighbor vertices of the vertex v_i, that is, the number of vertices with the mutual interference of the vertex v_i in graph G.

The idea of distributed greedy algorithm is to deal with all the colors one by one, and the goal of the greedy algorithm is to maximize the use of the color for each color k. The specific approach is when color k is assigned, color k is given priority to the smallest vertex degree d_i. Vertex degrees of all vertices are the same in the graph, and then it will be the color k assigned to the least number of color vertices. If all vertices have obtained the number of colors that are the same, color k is assigned to any one of the vertices using a random approach. When vertex v_i gets color k, the conflict-free channel is allocated to matrix A for $a_{ik} = 1$. Then the graph G is updated topologically including channel availability matrix L updating and vertex vector V updating. That is, there is $l_{jk} = 0$ for the vertex v_j of $e_{ij} = 1$, so $a_{jk} = 0$. The vertex will be deleted from the graph G when the vertex has no available colors for the channel availability matrix L. When the graph G is empty, all vertices are no longer available for allocation, and the distributed greedy algorithm finishes the allocation. Distributed greedy algorithm flow is shown in Figure 5.

Distributed greedy algorithm considers the constraint conditions between the cognitive users based on the vertex degrees of all vertices to maximize the spectrum utilization. But the cognitive user minimally disturbed by the other users with small vertex degree always has the highest priority of the subchannel allocation to always get the spectrum. The users with large vertex degree always are allocated less spectrum or even are not allocated the spectrum, so it is less fairness.

(2) Distributed Fairness Algorithm. Distributed fairness algorithm goal is to satisfy three constraints of formula (6) to achieve the fairness of spectrum allocation. The spectrum allocation of the distributed greedy algorithm would lead to spectrum allocation unfairness, so distributed fairness algorithm makes more equitable distribution of the spectrum through the establishment of a directed graph approach.

In the distributed fairness algorithm, the spectrum degree s_i of the vertex is defined as the available total number of the vertex v_i, and the mathematical expression is for formula (9). Obviously, the different vertex v_i has the different spectral degree s_i:

$$s_i = \sum_{k=1}^{M} l_{ik}. \tag{9}$$

Distributed fairness algorithm is divided into the following three steps (Figure 6).

First, the undirected graph G is modified to the directed graph G based on the concept of vertex degree and spectrum

degree. This requires the undirected graph G edge at one end add to an arrow from a vertex point to another vertex. Increasing the rules of the arrow is if the spectrum degree s_i and the spectrum degree s_j are not the same between vertex v_i and vertex v_j, and $s_i > s_j$, the edge between vertex v_i and vertex v_j is from vertex v_i to vertex v_j; if the spectrum degree is $s_i = s_j$ between vertex v_i and vertex v_j and the relationship between vertex degree d_i and vertex degree d_j of vertex v_i and vertex v_j is $d_i > d_j$, the edge between vertex v_i and vertex v_j is from vertex v_i to vertex v_j; if the vertex degree between vertex v_i and vertex v_j is $d_i = d_j$, the edge between vertex v_i and vertex v_j is random decision.

The interference matrix E can reflect the point to the edge of the graph G, and E matrix is modified as follows: if $e_{ij} = 1$ and vertex v_i points to vertex v_j, let e_{ij} remain changed and e_{ij} is modified to -1. In short, in the new interference matrix E, $e_{ij} = 1$ is the edge from vertex v_i to vertex v_j existing between vertex v_i and vertex v_j; $e_{ij} = -1$ is the edge from vertex v_j to vertex v_i existing between vertex v_i and vertex v_j. In the new interference matrix E, if any vertex v_i has $e_{ij} \geq 0$ ($j = 1, 2, \ldots, N$) and e_{ij} is not identically 0, vertex v_i only has the output edge called the source vertex; if any vertex v_i has $e_{ij} \leq 0$ ($j = 1, 2, \ldots, N$) and e_{ij} is not identically 0, vertex v_i only has the input edge called the places vertex. The directed graph G can have more than one source vertex and multiple places vertices.

Second, the places vertices have the highest priority of the spectrum allocation at the cycle of the algorithm, and only the places vertex does the spectrum allocation to gradually allocate spectrum to the source vertex. Multiple places vertices at the same time select a subchannel, and the subchannels are the least number of channels for all available channels of the places vertices of all neighbors. The places vertex selects the available channel, and issues set-color flag to notify all its neighbors vertices which cannot use the channel. When a nonplaces vertex receives set-color flag issued by all neighbors of the vertex, all output sides of the nonplaces vertices are deleted to become a new places vertex, and in the next cycle participate in spectrum allocation. In addition, the vertex that have not available spectrum should quit spectrum allocation.

Finally, if all vertices are no longer available subchannels, the algorithm ends; otherwise, enter the first step to start a new spectrum allocation from the places vertex to the source vertex. First, the source vertex issues reset flag to all its downstream neighbor vertices, only when a vertex received reset flag from the upstream neighbor vertex, the vertex delivered the reset flag to all of its downstream neighbor vertices. When all nonsource vertices receive the reset flag, the system is restarted. Then, enter the first step of the algorithms. Start from the places vertex to check whether there is available spectrum; if available spectrum exists, the spectrum is allocated in accordance with the second step of the method; if there is no available spectrum, exit the operation. Finally, if there is no vertex available spectrum, the spectrum allocation algorithm ends.

Distributed fairness algorithm not only considers the vertex degree, but also considers the spectrum degrees relative to the best fairness spectrum allocation in the other two algorithms. It realized the fairness spectrum allocation, but consumed more communication overhead in the same time. This is mainly due to the need to sort vertex degree and spectrum degrees.

(3) Random Distributed Algorithm. Random distributed algorithm goal is to meet three constraint conditions of formula (1), and to reduce communication overhead. When the number of vertices and color is very big, distributed greedy algorithm and distributed fairness algorithm need large number of iterations, and therefore random distributed algorithm was proposed to reduce communication latency and overhead.

If a vertex in the loop of the algorithm and the competition channel of the neighbor vertices fail in random distributed algorithm, they will increase their contention window. The next contention channel will have bigger winning probability so that we can reduce the probability of the vertex competitive resource conflict.

The basic idea of the random distributed algorithm is that each vertex i of the undirected graph G sets its own contention window $\text{Window}_i = \text{Window}$ in the cycle of the algorithm, and the available colors generate the corresponding number of uniformly distribution in the interval $[0, \text{Window}_i]$. If the random number of a vertex in the color k ($k = 1, 2, \ldots, M$) is larger than its vertices in the same colors available on the random number for all neighbors, let the vertex be assigned to the color k; that is, $a_{ik} = 1$; this color of all neighbor vertices of the vertex is not available, that is, $a_{jk} = 0$, where j is the neighbor vertex of vertex i. If all vertices have available colors in undirected graph G, they need to continue the cycle of the next round of the algorithm until all vertices do not have available colors. Random distributed algorithm flow is shown in Figure 7. Note that the color vertex i obtained in this loop should be adjusted to half of the original of its own contention window Window_i to reduce their own spectrum competitiveness; if the competition fails and color vertex j is not obtained, their own contention window Window_j should be adjusted to twice of the original to increase their own competitiveness.

Random distributed algorithm achieves cognitive radio spectrum allocation by the way of comparing the cognitive users and the random numbers generated by the neighbor cognitive users. In addition, the fairness of the cognitive user spectrum allocation is realized by adjusting the size of the contention window. This algorithm significantly reduces the complexity of the spectrum allocation, while reducing the overhead of the communication.

3.2. Centralized Coloring Spectrum Allocation. The coloring spectrum allocation algorithm described in the previous section reflects that the interference matrix E is a two-dimensional matrix between the users. That is, in the different subchannels, the mutual interferences between the characteristics of the users are the same. If two cognitive users interference cannot be ignored in a subchannel, then another subchannel must be mutual interference. The number of any subchannel interference neighbors is the same in the undirected graph G and the directed graph G.

In fact, because the interference of the cognitive users and authorized users is the same, the cognitive users' available

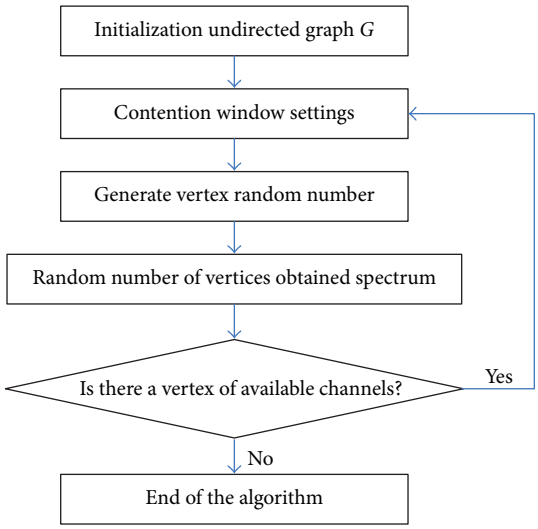


Figure 7: Random distributed algorithm.

subchannels also have differences. So taking into account the difference of the different channels, the edges between two vertices in the undirected graph G cannot simply think that all subchannels all exist the edges, but they exist with the subchannels. That is to say on a different subchannel, in which the number of interfere neighbors is different, which is the problems of interference spectrum difference.

Therefore, the literature [14] proposed Color-Sensitive Graph Coloring (CSGC) spectrum allocation algorithm against the interference spectrum differences and spectral quality differences. Figure 3 needs to be amended as the undirected graph with repeated edges shown in Figure 8.

After undirected graph is modified as the undirected graph with repeated edges, in order to reflect the interference differences, interference matrix E is also required to have the spectrum characteristics to become three-dimensional interference matrix E', and then any two cognitive users in each subchannel interference on the relationship can easily be seen:

$$E' = \left\{ e'_{ijk} \mid e'_{ijk} \in \{0,1\}, e'_{ijk} = e_{ij} l_{ik} l_{jk}, \; i,j = 1,2\ldots,N; \; k = 1,2,\ldots,M \right\}. \tag{10}$$

Then, the channel gain matrix B is defined as a set of all cognitive users on each subchannel gain. The gains generally refer to the maximum bandwidth or throughput, and so on.

$$B = \left\{ b_{ik} \mid b_{ik} > 0, \; i,j = 1,2\ldots,N; \; k = 1,2,\ldots,M \right\}. \tag{11}$$

The main goal of color sensitive graph theory coloring spectrum allocation algorithm is to achieve a variety of spectrum allocations to maximize the revenue function based on the interference conditions according to the different requirements of cognitive radio network plan, and the revenue function mainly includes the following three forms.

(1) Max Sum Reward (MSR). When the cognitive radio networks do not consider the fairness of the cognitive user

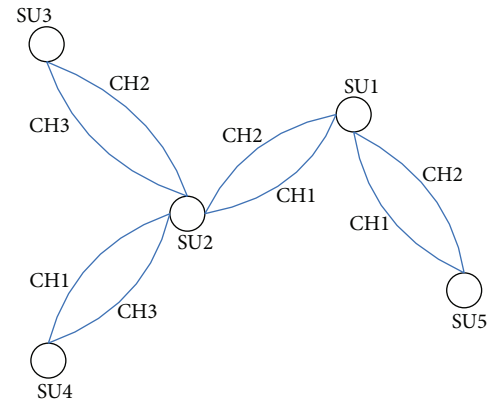


Figure 8: Spectrum allocation undirected graph with interference difference.

spectrum allocation, max sum reward makes the spectrum utilization maximum. The revenue function of the cognitive radio network can be expressed by the following formula:

$$U_{\text{sum}} = \sum_{i=1}^{N} \alpha_i = \sum_{i=1}^{N} \sum_{K=1}^{M} a_{ik} b_{ik}. \tag{12}$$

(2) Max Min Reward (MMR). Min is the cognitive users that allocated the least spectrum at the end of the spectrum allocation algorithm. The revenue function of Min can be expressed by the following formula:

$$U_{\min} = \min_{1 \le i \le N} \sum_{k=1}^{M} a_{ik} b_{ik}. \tag{13}$$

The revenue function lets the spectrum allocation worst users always be able to get more spectrum gains with priority on the premise of ensuring that the network spectrum resources are not wasted, and it is a simple form of the spectrum allocation fairness.

(3) Max Proportional Fairness (MPF). The revenue function design goal of the max proportional fairness is to find fairness distribution results. Any other form of distribution with respect to the allocation of revenue in the form of proportional change is negative in the allocation result. The revenue function of the max proportional fairness can be expressed by the following formula:

$$U_{\text{fair}} = \sum_{i=1}^{N} \log_{10} \left(\sum_{k=1}^{M} a_{ik} b_{ik} \right), \tag{14}$$

or using the form of formula (15):

$$U_{\text{fair}} = \left(\prod_{i=1}^{N} \sum_{k=1}^{M} a_{ik} b_{ik} \right)^{1/N}. \tag{15}$$

The basic idea of CSGC spectrum allocation algorithm is to design the appropriate vertex labeling rule in order to achieve the above revenue function needs. The size of

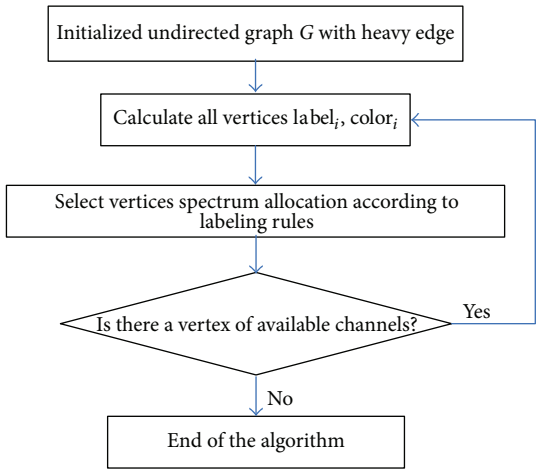


Figure 9: CSGC spectrum allocation algorithm.

the tag reflects the contribution of the vertex to the revenue function. Each label corresponds to a lot of color values. Each selection of CSGC algorithm has the largest contribution to the vertex of the revenue function and chooses the biggest color of the color values. When a vertex is colored, all other adjacent vertices are necessary to remove the color from their available channels, and the channel availability matrix is updated with heavy edge, and the undirected graph G has also been updated. If the graph G is not empty, then continue to the next algorithm cycles until all vertices are no longer available colors, and the algorithm ends. CSGC spectrum allocation algorithm flow is shown in Figure 9.

CSGC algorithm is divided into collaborative and noncollaborative according to the vertex whether to consider the impact on the neighbors. Collaborative CSGC spectrum allocation algorithm is necessary to consider the system spectrum utilization level, but it also considers the impact on the neighbors of each color choices, so the labeling rules of CSGC algorithm can be divided into the following six forms.

(1) Collaborative-Max-Sum-Reward (CSUM). The number of conflicts is an important concept in the collaborative CSGC spectrum allocation algorithm, which is defined as a subchannel k, and the number of the neighbors of cognitive user i can be calculated by the following formula:

$$d_{ik} = \sum_{j=1, j\neq i}^{N} e_{ijk} l_{ik} l_{jk}. \tag{16}$$

So, the number of conflicts matrix can be defined as $D = \{d_{ik} \mid i = 1, 2 \ldots, N;\ k = 1, 2, \ldots, M\}$. CSUM is calculated for the size of each vertex label and the size of each corresponding color value rule:

$$\text{label}_i = \max_k \frac{b_{ik}}{d_{ik} + 1}, \tag{17}$$

$$\text{color}_i = \arg\max_k \frac{b_{ik}}{d_{ik} + 1}. \tag{18}$$

(2) Noncollaborative-Max-Sum-Reward (NSUM). NSUM does not consider the impact on the neighbor color selection, while the size of the tags for each vertex and the corresponding size of each color are calculated by the following method:

$$\text{label}_i = \max_k b_{ik}, \tag{19}$$

$$\text{color}_i = \arg\max_k b_{ik}. \tag{20}$$

(3) Collaborative-Max-Min-Reward (CMIN). CMIN is different from CSUM and NSUM rules, and its label features are related to the previous spectrum allocation revenue. The vertex of the smallest gains on spectrum allocation stage has the spectrum priority at this stage. Each spectrum allocation phase labeling rule is as follows:

$$\begin{aligned} \text{label}_i &= -\sum_{k=1}^{M} a_{ik} b_{ik}, \\ \text{color}_i &= \arg\max_k \frac{b_{ik}}{d_{ik} + 1}. \end{aligned} \tag{21}$$

(4) Noncollaborative-Max-Min-Reward (NMIN). NMIN does not consider the interference problem of the vertices, which still has priority access to the spectrum at this stage. The vertex of the smallest gains has the spectrum priority. Each spectrum allocation phase labeling rule is as follows:

$$\begin{aligned} \text{label}_i &= -\sum_{k=1}^{M} a_{ik} b_{ik}, \\ \text{color}_i &= \arg\max_k b_{ik}. \end{aligned} \tag{22}$$

(5) Collaborative-Max-Proportional-Fairness (CFAIR). CFAIR is to achieve the fairness of the spectrum allocation described by formula (14). The size of the label is not only influenced by the number of conflicts, but it is also related to the spectrum allocation in the last phase. The labeling rules of each stage are as follows:

$$\begin{aligned} \text{label}_i &= \frac{\max_k b_{ik} / (d_{ik} + 1)}{\sum_{k=1}^{M} a_{ik} b_{ik}}, \\ \text{color}_i &= \frac{\arg\max_k b_{ik}}{d_{ik} + 1}. \end{aligned} \tag{23}$$

(6) Noncollaborative-Max-Proportional-Fairness (NFAIR). NFAIR is the noncooperation of CFAIR, and it is no longer considering the impact on the size of the label. The labeling rules of each spectrum allocation phase are as follows:

$$\begin{aligned} \text{label}_i &= \frac{\max_k b_{ik}}{\sum_{k=1}^{M} a_{ik} b_{ik}}, \\ \text{color}_i &= \arg\max_k b_{ik}. \end{aligned} \tag{24}$$

When each subchannel gain is identical, that is, b_{ik} is a constant, NFAIR and NMIN are consistent.

4. Improved Spectrum Allocation Algorithm Based on Allocation Sequence

4.1. Overhead of CSGC Spectrum Allocation Algorithm. The list coloring spectrum allocation algorithm and CSGC spectrum allocation algorithm are the spectrum allocation for the static cognitive radio network topology, that is in the process of the spectrum allocation algorithm execution; either undirected or directed cognitive radio network topology graphs always remain the same. After the execution of the spectrum allocation algorithm, if there is a new cognitive radio nodes join the network, the system must re-construct graphical topology including all of the network nodes, the global optimization algorithm increase the algorithm's time overhead. CSGC algorithm overhead is formula (25), which increases with the increase of the number of users and the number of available channels:

$$T_{\cos t} = \sum_{i=1}^{N} \sum_{K=1}^{M} a_{ik}. \tag{25}$$

In order to reduce the allocation overhead, the literature [15] proposed a distributed local bargaining spectrum allocation algorithm, and it started from the reduction in the number of users involved in spectrum allocation point of view and took full advantage of the previous allocation result, which only does spectrum allocation for the local network topology graph of the newly added network nodes at bargaining, thus reducing the time overhead of the spectrum allocation algorithm. The literature [16] proposed parallel spectrum allocation algorithm from the point of view of the amount of spectrum to further reduce the time overhead of CSGC spectrum allocation algorithm. The overhead of the parallel spectrum allocation algorithm is formula (26):

$$T_{\cos t} = \max_{1 \le k \le M} \sum_{i=1}^{N} a_{ik}. \tag{26}$$

However, the parallel spectrum allocation algorithm currently only improves CSGC spectrum allocation algorithm based on MSR and MPF guidelines. In order to meet the requirements of the MMR guidelines, new spectrum allocation algorithm must be designed to further reduce the time overhead of the CSGC algorithm.

4.2. Improved CSGC Algorithm Based on MMR Criterion. MMR criterion is represented by formula (13) in CSGC algorithm to maximize the cognitive users' bandwidth with the least number of the allocation spectrums. To ensure that the network spectrum resources are not wasted, the spectrum allocation worst user is always able to get more spectrum gains priority with the spectrum allocation fairness. However, the time overhead is the key factors that affect the actual application in CSGC algorithm.

MMR criterion requires that the goal of the spectrum allocation is to maximize $U_{\min}$ based on formula (13) under CSGC algorithm. It can be seen that $U_{\min}$ is not only related to the number of the spectrum shared by the cognitive users, but is also related to the channel gains b_{ik} brought by the spectrums.

Equations (17)~(20) explain that CSGC algorithm has the following main features in both forms of cooperation or noncooperation.

(1) Start point of the spectrum allocation is random. In CSGC algorithm of MMR guidelines, the first cycle of the algorithm randomly allocates spectrum to a user. Because any user does not get the spectrum at the first time, the calculation result of $-\sum_{k=1}^{M} a_{ik} b_{ik}$ is 0; that is, a user is randomly selected to allocate.

(2) The number of channel allocation opportunities is equal; in the form of cooperation and noncooperation of CSGC algorithm, the more the amount of spectrum allocations is, the smaller the opportunity of cognitive users again shared the spectrum is; the less the amount of spectrum allocations is, the greater the opportunity of cognitive users again shared the spectrum is.

(3) The opportunity of the channel gains is equal. Channel allocation sequence always begins from the channel giving cognitive user maximum channel gain. However, from the results of the final allocation, the shared channel may be more than one. Both can make it to get the maximum channel gain channel, but also make it to obtain the minimum revenue channel. Visible is to make the channel gains close to the average allocation among the cognitive users so that bottlenecks user has more bandwidth.

(4) The number of conflicts is the main reason affecting the income of bottlenecks user bandwidth.

Based on the above four points, it is necessary to improve the fairness of the distribution for bottleneck users, and to reduce the execution time of CSGC algorithm in a centralized network. In this paper, spectrum allocation algorithm based on the allocation sequence is proposed in the MMR criterion. Design the tag rules shown in formula (27) and formula (28) considering the separation of the time overhead and cognitive users' historical allocation information:

$$\text{num}_{ik} = \left(1 + \frac{r_{ik}}{N}\right) \frac{1}{\sum_{k=1}^{M} l_{ik} b_{ik}} \frac{1}{d_{ik} + 1}, \tag{27}$$

$$\text{num}_{ik} = \left(1 + \frac{r_{ik}}{N}\right) \frac{1}{\sum_{k=1}^{M} l_{ik} b_{ik}}, \tag{28}$$

where i is the cognitive user number, k is the channel number, $k = 1, 2, \ldots, M$, $\sum_{k=1}^{M} l_{ik} b_{ik}$ is cognitive user i in all available channels to get the revenue of full bandwidth. d_{ik} is the number of conflict of cognitive user i on channel k calculated in formula (16), where N is the total number of cognitive users, r_{ik} is randomly generated by the algorithm. Formula (27) is a cooperative labeling rule, and formula (28) is a noncooperative labeling rule.

Specifically, in each channel, all the users' r_{ik} is based on the channel gains, the number of channels to get the idea of equalization opportunities, in accordance with $1 \sim N$

random ordering disposable generates N-dimensional row vectors $(r_{1K}, r_{2K}, \ldots r_{NK})$, and the directed graph generated independently. r_{ik} $(i = 1, 2, \ldots N)$ are elements of the N-dimensional row vectors $(r_{1k}, r_{2k}, \ldots r_{Nk})$, a total of $N!$ Different $1 \sim N$ random arrangement can be used for spectrum allocation algorithm.

Process of spectrum allocation algorithm based on the allocation sequence is as follows:

(1) Undirected graph G with repeated edges generates M undirected graphs $(1, 2, \ldots, k, \ldots, M)$ according to M subchannels. Because the frequencies of these undirected graphs are not the same, so the spectrum distributions of M undirected graphs may be performed simultaneously, and the results of each allocation do not affect each other. Then, each undirected graph k, respectively, randomly generates the N-dimensional allocation sequence $(r_{i1}, r_{i2}, \ldots r_{iN})$ and calculates all the user's label value num_{ik} in channel k.

(2) Compare N cognitive users' num_{ik} values and select user i^* with the largest num_{ik} value assigned to the channel k to get all cognitive users tag values num_{ik} calculated in any undirected graph k.

(3) Find out all users interfering with cognitive user i^*, and the availability of the kth channel given by these cognitive users is marked as "0" in the available channel matrix L. Then, these users with cognitive user i^* are deleted from the undirected graphs so that undirected graph k gets update.

(4) To determine whether there is an available channel for cognitive users. If the result is "1," the algorithm returns to (2); if the result is not "1," the whole spectrum allocation algorithm ends.

A more intuitive process of spectrum allocation algorithm based on allocation sequence is shown in Figure 10.

Table 1: Parameter list.

The number of available channel M	10
The cognitive number of users N	10~40
Scope of the map	1000×1000
Cognitive coverage	100
Channel availability matrix L	Uniform random 0, 1 matrix
Three dimensional interference matrix E	Calculated
Conflict matrix D	Obtained from L, E calculated
Channel gain matrix B	1~6 randomly generated

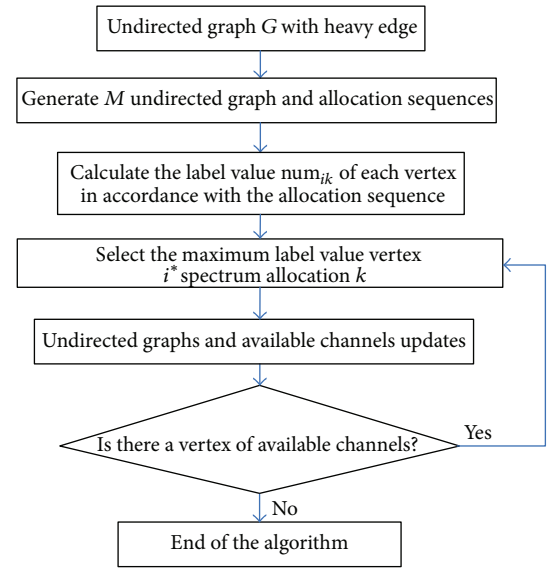


Figure 10: Spectrum allocation algorithm based on allocation sequence.

5. Simulation and Analysis

5.1. Parameter Settings. This section mainly studies the performance of the spectrum allocation algorithm based on the allocation sequence through MATLAB simulation, and the simulation parameters are shown in Table 1.

5.2. Simulation Results. In the form of cooperation spectrum allocation through MATLAB simulation, the spectrum allocation algorithm based on the allocation sequence proposed in this paper does the comparison with the CSGC spectrum allocation algorithm meeting MSR, MMR, and MPF (called as CSUM, CMIN, and CFAIR algorithms, resp.). In this paper, we consider the measure of spectrum allocation algorithm performance indicators including cognitive users' minimum income, cognitive users' average income, fairness of algorithm allocation, and time overhead of the algorithm.

(1) Cognitive Users' Minimum Income. First of all, the MMR guidelines using formula (14) analyze the performance in the cognitive user minimum income to verify whether the algorithm is in line with the MMR guidelines. The relationship of the cognitive user minimum income and of the number of cognitive users cognitive users is shown in Figure 11 in a fixed amount of spectrum and $M = 10$. Minimum income in Figure 11 is calculated completely in accordance with formula (14). In addition, due to CSUM label rule, it will not be assigned to the case of the spectrum, so the simulation of the algorithm considers that the user has obtained the minimum bandwidth gains.

As can be seen from Figure 11, CMIN algorithm can get more bandwidth income characteristics of best protected cognitive users with the minimum income. When the cognitive users are enough, the algorithm proposed in this paper will be very close to CMIN algorithm. CSUM algorithm and CMIN algorithm do not consider the minimum bandwidth gains as much cognitive users bandwidth gains. And with the increase of cognitive users, CSUM algorithm and CFAIR algorithm are increasingly unable to meet the bandwidth minimum income that cognitive users' channel needs.

(2) Cognitive Users' Average Income. Secondly, cognitive users' average income can also evaluate the performance of

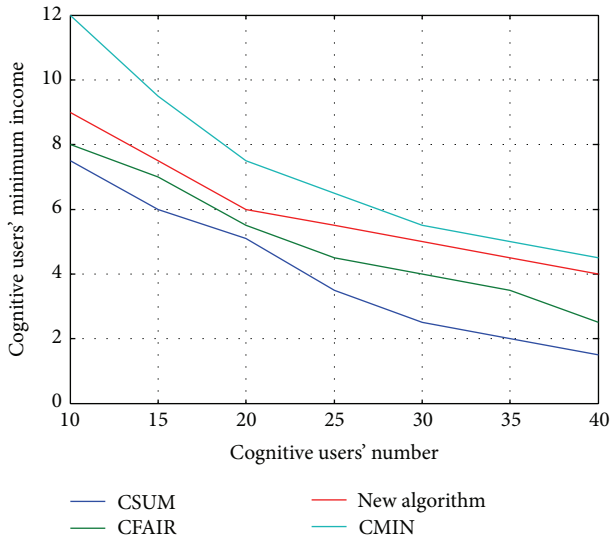


Figure 11: Comparison of cognitive users' minimum income.

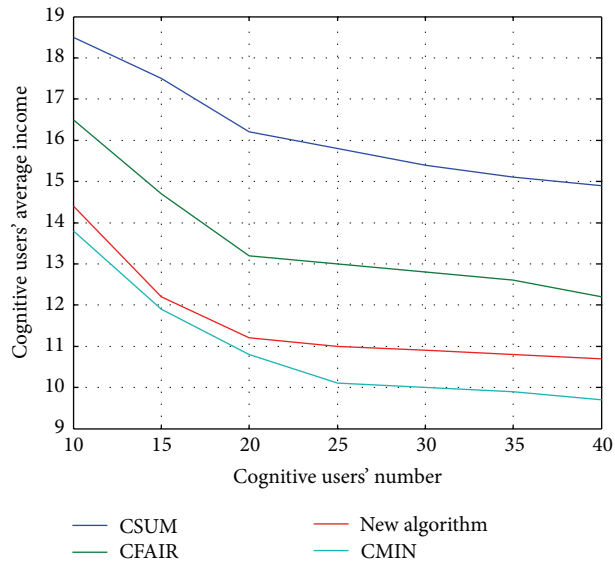


Figure 12: Comparison of cognitive users' average income.

the proposed algorithm. The average of the total income of formula (13) based on the MSR criteria can get the average income in the form of formula (28) as the standard of the average income of the evaluation of cognitive users. The relationship between cognitive users' average income and the number of cognitive users in a fixed amount of spectrum and $M = 10$ is shown in Figure 12,

$$U_{\text{mean}} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{M}a_{ik}b_{ik}. \tag{29}$$

It can be seen from Figure 12 that CMIN algorithm meeting MMR guidelines is at a disadvantage compared to the proposed algorithm in this paper and other criteria algorithms, and the channel bandwidth gains of the two spectrum allocation algorithms are lower than the other spectrum allocation algorithms. However, the bandwidth income based on spectrum allocation sequence algorithm is higher than the existing CMIN algorithm in the same MMR guidelines.

(3) Fairness of Algorithm Allocation. The fair evaluation criteria of the algorithm use formula (14) of MPF guidelines. The relationship between the fairness of algorithm allocation and the number of cognitive users in a fixed amount of spectrum and $M = 10$ is shown in Figure 13.

The comparison of the fairness of algorithm allocation in Figure 13 illustrates that the fairness of spectrum allocation is better than CMIN algorithm under the same MMR guidelines, but the performance is still below the CFAIR algorithm meeting the fair MPF guidelines. In addition, with the increase of the cognitive user network, the proposed algorithm and CMIN algorithm simultaneously increase in fairness, while the fairness of CSUM algorithm will increase slowly.

(4) Time Overhead of the Algorithm. Finally, in order to evaluate the time overhead proposed by this paper, MATLAB

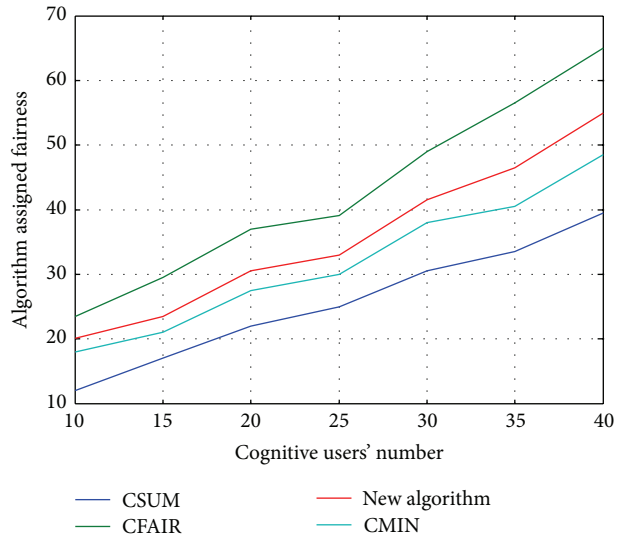


Figure 13: Comparison of the fairness of algorithm allocation.

program run time is directly used to measure with other algorithms, and the time overhead of existing algorithm and CSGC algorithm is compared to each other. The curve of the time overhead of the spectrum allocation algorithm with the changes of the number of cognitive users is shown in Figure 14 in a fixed amount of spectrum and $M = 10$; the curve of the time overhead of the spectrum allocation algorithm with the changes of the number of spectrum allocations is shown in Figure 15 in a fixed amount of spectrum and $M = 40$.

As seen from Figure 14, the time overhead of the algorithm proposed in this paper is significantly lower than existing CSUM algorithm, CMIN algorithm, and CFAIR algorithm in the same amount of spectrum. With the increase in the number of cognitive users, the time overhead of the

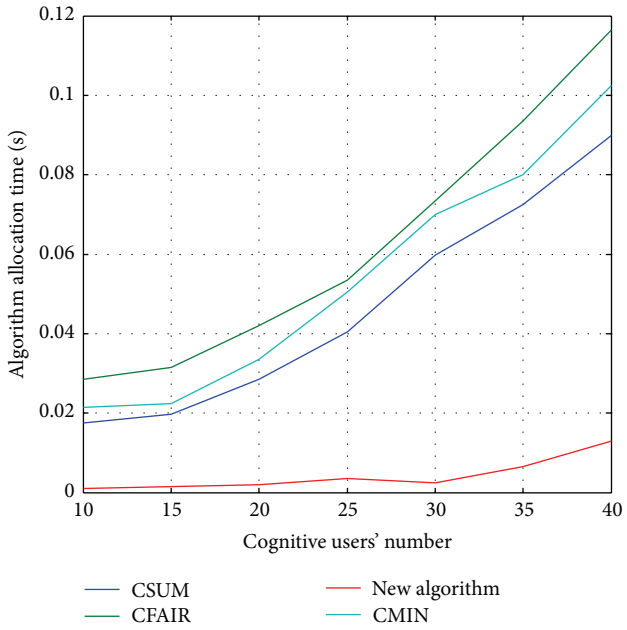


FIGURE 14: Time overhead and cognitive users' number.

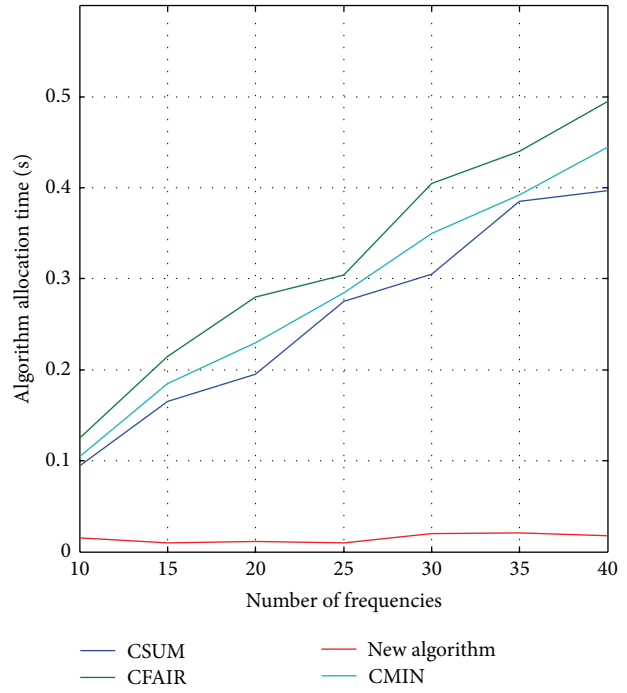


FIGURE 15: Time overhead and number of frequencies.

algorithm proposed in this paper will increase, but it is not obvious, and the time overhead is still very small. It is very important to the centralized cognitive radio networks, because the rapid convergence of the algorithm is able to increase the flexibility and applicability of cognitive radio network.

Figure 15 illustrates the relationship between the time overhead of the algorithm and the amount of spectrum in the same number of cognitive users. As can be seen, with the increase in the number of the spectrum, the execution time of the algorithm in this paper does not increase, but fluctuates in a range, and the time complexity of the algorithm depends only on the allocation for the longest time undirected graph.

In short, the simulation results show that the spectrum allocation algorithm proposed in this paper based on the allocation sequence is not only able to meet the design requirements of cognitive radio networks based on MMR guidelines but also able to reduce the spectrum allocation algorithm execution time in centralized cognitive radio network. And the time is only related to the number of users, is not related to the amount of spectrum. In addition, the spectrum allocation is closer to CFAIR algorithm that is done the best in fairness compared with other algorithms.

6. Conclusion

This paper establishes the mathematical model for the cognitive radio spectrum allocation based on the graph theory and the existing spectrum allocation algorithms based on the distributed cognitive radio networks and the centralized cognitive radio networks, which include the basic idea of the algorithms and the specific process of the algorithms. The improved algorithm based on the central cognitive radio network spectrum allocation algorithm is proposed. The simulation results show that the improved algorithm can significantly reduce the allocation time of the algorithm compared to the existing graph theory algorithms.

References

[1] J. Mitola and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] Q. D. La, Y. H. Chew, and B. H. Soong, "An interference minimization game theoretic subcarrier allocation algorithm for OFDMA-based distributed systems," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pp. 1–6, Honolulu, Hawaii, USA, December 2009.

[4] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri, "Towards real-time dynamic spectrum auctions," *Computer Networks*, vol. 52, no. 4, pp. 879–897, 2008.

[5] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 517–528, 2007.

[6] T. Feng and Y. Zhen, "A new algorithm for weighted proportional fairness based spectrum allocation of cognitive radios," in *Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD '07)*, pp. 531–536, Qingdao, China, August 2007.

[7] D. Niyato and E. Hossain, "Competitive pricing for spectrum sharing in cognitive radio networks: dynamic game, inefficiency of nash equilibrium, and collusion," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 192–202, 2008.

[8] Y. Peng, J. Peng, X. Zheng, Z. Liu, and H. Long, "A cross-layer architecture for OFDMA-based cognitive radio network," in *Proceedings of the WRI World Congress on Software Engineering (WCSE '09)*, pp. 129–133, Xiamen, China, May 2009.

[9] K. E. Morsy, F. F. Digham, M. H. Nafie, and A. Y. Elezabi, "Cross-layer adaptive resource allocation algorithm for wireless communications networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '08)*, pp. 1–5, New Orleans, La, USA, December 2008.

[10] R. Wang, V. K. N. Lau, L. Lv, and B. Chen, "Joint cross-layer scheduling and spectrum sensing for OFDMA cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2410–2416, 2009.

[11] Y. Chen, Z. Feng, and X. Chen, "Cross-layer resource allocation with heterogeneous QoS requirements in cognitive radio networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '11)*, pp. 96–101, Cancun, Mexico, March 2011.

[12] L. Ding, T. Melodia, S. N. Batalama, J. D. Matyjas, and M. J. Medley, "Cross-layer routing and dynamic spectrum allocation in cognitive radio Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1969–1979, 2010.

[13] T. A. Weiss and F. K. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency," *IEEE Communications Magazine*, vol. 42, no. 3, pp. S8–S14, 2004.

[14] B. S. Manoj, R. R. Rao, and M. Zorzi, "CogNet: a cognitive complete knowledge network system," *IEEE Wireless Communications*, vol. 15, no. 6, pp. 81–88, 2008.

[15] S. Bahramian and B. H. Khalaj, "Joint dynamic frequency selection and power control for cognitive radio networks," in *Proceedings of the International Conference on Telecommunications (ICT '08)*, pp. 1–6, St. Petersburg, Russia, June 2008.

[16] J. Jia, C. Wang, Z. Zhang, D. Shi, and Y. Li, "A spectrum allocation algorithm to maximize the system access amount in cognitive radio system," in *Proceedings of the 3rd IEEE International Conference on Communication Software and Networks (ICCSN '11)*, pp. 251–255, Xi'an, China, May 2011.

[17] Nautilus Project Website, http://www.cs.ucsb.edu/~htzheng/cognitive/nautilus.html.

[18] DARPA XG WG, The XG Architectural Framework V1. 0, 2003.

[19] D. Niyato and E. Hossain, "A game-theoretic approach to competitive spectrum sharing in cognitive radio networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 16–20, Kowloon, Hong Kong, March 2007.

[20] D. Grandblaise, D. Bourse, K. Moessner, and P. Leaves, "Dynamic spectrum allocation (DSA) and reconfigurability," in *Proceedings of the Software-Defined Radio (SDR) Forum*, November 2002.

[21] L. Chen, S. Iellamo, M. Coupechoux, and P. Godlewski, "An auction framework for spectrum allocation with interference constraint in cognitive radio networks," in *Proceedings of the IEEE INFOCOM 2010*, pp. 1–9, San Diego, Calif, USA, March 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 307489, 8 pages
http://dx.doi.org/10.1155/2013/307489

Research Article

A Microseismic/Acoustic Emission Source Location Method Using Arrival Times of PS Waves for Unknown Velocity System

Longjun Dong and Xibing Li

School of Resources and Safety Engineering, Central South University, Lushan South Road, Changsha 410083, China

Correspondence should be addressed to Longjun Dong; rydong001@csu.edu.cn

Received 28 July 2013; Accepted 30 August 2013

Academic Editor: Yong Jin

To eliminate the location error of MS/AE (microseismic/acoustic emission) monitoring systems caused by the measurement deviations of the wave velocity, a MS/AE source location method using P-wave and S-wave arrivals for unknown velocity system (PSAFUVS) was developed. Arrival times of P-wave and S-wave were used to calculate and fit the MS/AE source location. The proposed method was validated by numerical experimentations. Results show that the proposed method without the need for a premeasured wave velocity has a reasonable and reliable precision. Effects of arrival errors on location accuracy were investigated, and it shows location errors enlarged rapidity with the increase of arrival errors. It is demonstrated the proposed method can not only locate the MS/AE source for unknown velocity system but also determine the real time PS waves velocities for each event in rockmass.

1. Introduction

Geophysical methods, statistical analysis, microseismic monitoring, and in-seam seismic techniques, have shown an increased significance in rock physical mechanics and mining engineering in recent decades [1–21]. In particular, the developments of seismic monitoring provide a scientific basis for controlling the rockburst and seismic hazards.

The seismic location, one of the most important parameters for a seismic monitoring, is very important for rockburst control and warning in a mine through the seismological method of the prediction of the areal rockbursts, which could improve the safety performance of deep geotechnical engineering. Researchers have developed many acoustic emission or seismic source location techniques [1, 2, 7, 22–32]. The problem of locating a signal source using time difference of arrival (TDOA) measurements also has numerous applications in aerospace, surveillance, structural health, navigation, industrial process, speaker location, machine condition, and the monitoring of nuclear explosions [1, 33–45]. And some of which are mature technologies and are widely used in the location of acoustic emission or seismic source.

However, for the existing technologies based on AE or seismic source, a given wave velocity or practical pre-measured wave velocity of the propagation medium is required. It is well known that the wave velocity is influenced by the materials, size, and surface conditions of transmission media and other factors. When the input wave velocity is different from the real wave velocity of the measured object, an error would occur in the system [46]. The average wave velocity is different from that of various regions, and the actual location of the occurrence of rockburst is not necessarily in the predetermined wave velocity area. The measured wave velocity is affected significantly by the distance between sensors; the measured P-wave velocity of the general container is between 2800 and 3100 $m \cdot s^{-1}$ when the distance is large while that is about 5000 to 6000 $m \cdot s^{-1}$ when the distance is small [41]. The large location errors can be induced by inaccurate average wave velocity. Both of these conditions result in some errors between the pre-measured wave velocity as an input in the positioning system and the actual wave velocity of the area where the rockburst occurs; hence, it would result in a large positional error [41]. To quantitatively study the location errors induced by deviation of sonic speed, the line and plane location tests were carried out in [47], and some interesting conclusions were summarized: the results show that for line positioning, the maximum error of absolute distance is about 0.8 cm. With the speed difference of 200 m/s, the average value of absolute difference from the position error is about

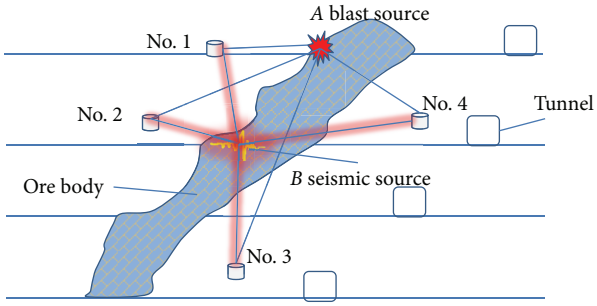


Figure 1: Location schematic diagram for different trace and different time.

0.4 cm; for the plane positioning, in the case of the sensor array of 30 cm, the absolute positioning distance is up to 8.7 cm. It shows the sonic speed seriously impacts on the plane positioning accuracy; the plane positioning error is larger than the line positioning error, which means that when the line position can satisfy the need in practical engineering, it is better to use the line position instead of the plane location, and the plane positioning error with the diagonal speed is the minimum one. Dong et al. analyzed above the drawbacks systematically in [48]. Velocities of P-wave and S-wave are obtained by blast experiments in existed methodologies. Two clear disadvantages were discussed. Firstly, active times are different because the premeasured velocity is obtained before the real-time seismic event. The fact is that the velocity in rockmass is changing all the time, and the velocity in different time would be not the same. Secondly, propagation traces are different (i.e., the traces of the wave propagation for blast experiments are always different from the wave propagation for real-time events), which would induce a big error because of the different average velocities in different propagation traces. The above drawbacks are further discussed in Figure 1, and the traces between blast source A and sensors were expressed as A_1, A_2, A_3, and A_4, and the average velocity of the four traces is supposed as V_A; the traces between seismic source B and sensors were expressed as B_1, B_2, B_3, and B_4, and the average velocity of the four traces is supposed as V_B. Considering the active time, V_A is not equal to V_B at different time; while considering the active space, propagation traces are different, and V_A is also not equal to V_B.

In present work, we proposed a an innovative MS/AE source location method without the need for premeasured P-wave and S-wave velocities, which can calculate MS/AE source locations in unknown velocity networks.

2. Methodology

The microseismic n sensors are placed in a location area, and they are not in the same plane. Three-dimensional coordinates of the microseismic sensors are known, which are (x_1, y_1, z_1), (x_2, y_2, z_2), …, and (x_i, y_i, z_i). When the microseismic event occurs, the sensor i receives the MS/AE source signals and arrival times of PS waves are recorded as $t_{\alpha i}$ and $t_{\beta i}$, respectively. The distance between MS/AE source and received station (sensor) i is R_i. The average velocities of PS waves (P-wave and S-wave) are expressed as α and β, respectively.

The distance from MS/AE source to sensor i can be expressed as

$$R_i = \alpha t_{\alpha i} = \beta t_{\beta i}. \tag{1}$$

The time difference of arrival times between PS waves is given by

$$t_{\beta i} - t_{\alpha i} = \frac{R_i}{\beta} - \frac{R_i}{\alpha} = R_i\left(\frac{1}{\beta} - \frac{1}{\alpha}\right). \tag{2a}$$

Supposing an equivalent parameter consider

$$\widetilde{v} = \left(\frac{1}{\beta} - \frac{1}{\alpha}\right)^{-1} = \frac{\alpha\beta}{\alpha - \beta}. \tag{2b}$$

Equation (2a) and can be rewritten as

$$t_{\beta i} - t_{\alpha i} = \frac{R_i}{\beta} - \frac{R_i}{\alpha} = \frac{R_i}{\widetilde{v}}. \tag{2c}$$

Then, the distances between MS/AE source and received stations i, as well as the distance between MS/AE source and received stations j, are obtained as follows:

$$R_i = \widetilde{v}\left(t_{\beta i} - t_{\alpha i}\right), \tag{3a}$$

$$R_j = \widetilde{v}\left(t_{\beta j} - t_{\alpha j}\right). \tag{3b}$$

According to the distance formula of two point in the space, for sensors i and j, (4) and (5) are given as follows:

$$\left(x_0 - x_i\right)^2 + \left(y_0 - y_i\right)^2 + \left(z_0 - z_i\right)^2 = R_i^2, \tag{4}$$

$$\left(x_0 - x_j\right)^2 + \left(y_0 - y_j\right)^2 + \left(z_0 - z_j\right)^2 = R_j^2. \tag{5}$$

The left side and the right side of (4) added the left side and the right side of (5), it can be expressed as

$$\begin{aligned}&\left(x_0 - x_i\right)^2 + \left(y_0 - y_i\right)^2 + \left(z_0 - z_i\right)^2\\ &\quad + \left(x_0 - x_j\right)^2 + \left(y_0 - y_j\right)^2 + \left(z_0 - z_j\right)^2 = R_i^2 + R_j^2.\end{aligned} \tag{6}$$

Submitting (3a) and (3b) into (6), it can be rewritten as

$$\begin{aligned}&\left[\left(x_0 - x_i\right)^2 + \left(y_0 - y_i\right)^2 + \left(z_0 - z_i\right)^2\right.\\ &\quad \left.+\left(x_0 - x_j\right)^2 + \left(y_0 - y_j\right)^2 + \left(z_0 - z_j\right)^2\right] \times \left(\widetilde{v}^2\right)^{-1}\\ &\quad = \left(t_{\beta i} - t_{\alpha i}\right)^2 + \left(t_{\beta j} - t_{\alpha j}\right)^2.\end{aligned} \tag{7}$$

Equation (7) can be expressed as

$$\begin{aligned}\Delta t_{ij} &= \left(t_{\beta i} - t_{\alpha i}\right)^2 + \left(t_{\beta j} - t_{\alpha j}\right)^2\\ &= f\left(x_i, y_i, z_i, x_j, y_j, z_j, x_0, y_0, z_0, \widetilde{v}\right).\end{aligned} \tag{8}$$

In (8), sensor coordinates (x_i, y_i, z_i), (x_j, y_j, z_j), and dependent variable $t_{\alpha i}, t_{\beta i}, t_{\alpha j}, t_{\beta j}$ are known parameters, and x_0, y_0, z_0 and $\widetilde{v}$ are unknown parameters. There are four unknown parameters; therefore, it needs more than four sensors to get the solutions by solving the nonlinear equations. This method was called PSAFUVS (MS/AE source location method using P-wave and S-wave arrivals for unknown velocity system).

Equation (8) is a nonlinear fitting problem with a single dependent variable. According to all the observed data $(x_i, y_i, z_i; x_j, y_j, z_j)$, (8) can determine a regression value

$$\Delta\widehat{t}_{ij} = f\left(x_i, y_i, z_i, x_j, y_j, z_j, x_0, y_0, z_0, \widetilde{v}\right). \tag{9}$$

The difference between $\Delta\widehat{t}_{ij}$ and Δt_{ij} can describe the degree of deviation between the regression value and observed values.

Due to $(x_i, y_i, z_i; x_j, y_j, z_j)$, if the degree of deviation between Δt_{ij} and $\Delta\widehat{t}_{ij}$ is smaller, it shows that the fitted line and experimental points could fit better. The sum of squared deviations of all observations and fitted values is

$$Q\left(x_0, y_0, z_0, \widetilde{v}\right) = \sum_{i,j=1}^{n}\left[\Delta\widehat{t}_{ij} - \Delta t_{ij}\right]^2. \tag{10}$$

Equation (10) describes the deviations between all observed and experimental values. Therefore, the MS/AE source parameters $(x_0, y_0, z_0, \widetilde{v})$ can have values such that the $Q(x_0, y_0, z_0, \widetilde{v})$ will reach a minimum as follows:

$$Q\left(x_0, y_0, z_0, \widetilde{v}\right) = \sum_{i,j=1}^{n}\left[\Delta\widehat{t}_{ij} - \Delta t_{ij}\right]^2 = \min. \tag{11a}$$

$P_{ij}(x_0, y_0, z_0, \widetilde{v})$ is a parameter to express the fitting error for each set of observations and fitted values as follows:

$$P_{ij}\left(x_0, y_0, z_0, \widetilde{v}\right) = \left(\Delta\widehat{t}_{ij} - \Delta t_{ij}\right)^2 = \min_{ij}. \tag{11b}$$

Equations (11a) and (11b) are the nonnegative function of $x_0, y_0, z_0, \widetilde{v}$, so it always has the minimum. The coordinates of the MS/AE source location and the equivalent velocity $(x_0, y_0, z_0, \widetilde{v})$ can be obtained by solving (11a) and (11b). According to $\alpha = \sqrt{3}\beta$, (2b) can be rewritten as

$$\widetilde{v} = \frac{\sqrt{3}\beta^2}{\sqrt{3}\beta - \beta} = \frac{\sqrt{3}\beta}{\sqrt{3} - 1}, \tag{12}$$

α and β can be obtained by resolving (12).

For a simple MS/AE source location problem, only x_0, y_0, z_0 should be solved. Equation (11a) is a nonlinear fitting problem with a single dependent variable. The commonly used nonlinear fitting methods include Levenberg-Marquardt (LM), Simplex Method (SM), Quasi-Newton Method (QN), Max Inherit Optimization (MIO), Self-Organizing Migrating Algorithms (SOMA), and Global Optimization (GO) [27, 31, 48]. Dong et al. discussed the location error of the different method [48], and analyzed results show that it is easy to trap

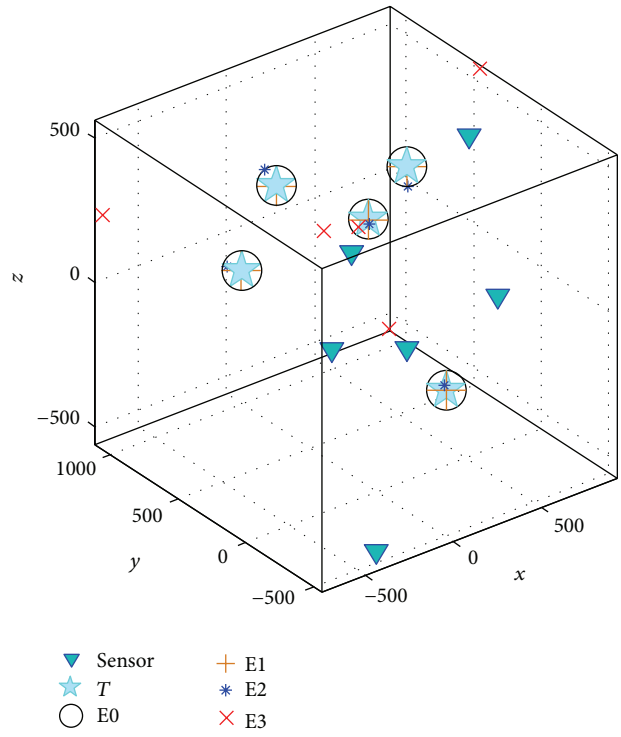


Figure 2: Spatial location of sensors and MS/AE sources (unit: meter).

a local optimum value using MIO, QN, and LM simply. It is suggested that the selection of MIO, QN, and LM simply should be careful. The joint use of GO as well as MIO, QN, and LM can significantly improve positioning accuracy. The SOMA, SM coupled with GO, MIO coupled with GO, and LM coupled with GO, which are more stable with the high positioning accuracy, should be recommended preferred methods to locate MS/AE source coordinates. Therefore, in present work, The SOMA coupled with GO was used to fit the MS/AE source locations using (11a) and (11b).

3. Validation with Numerical Experimentations and Discussion

In numerical experimentations, the location network includes 6 sensors S_1, S_2, S_3, S_4, S_5, and S_6. Their coordinates are (600, 100, 500), (300, −500, 200), (400, 300, −250), (−300, −200, 400), (200, 600, −300), and (−350, −450, −550), respectively. The average P-wave velocity and S-wave velocity in the medium are 5500 and 3175 m/s, respectively. The original times of MS/AE sources are supposed as 0 s. According to the relationship between distance formula, the coordinates of MS/AE sources A, B, C, D, and E as well as their arrivals are obtained and listed in Table 1. The sensor and the spatial distribution of MS/AE source locations are shown in Figure 2 (all coordinates are in the length unit: meter). The MS/AE source locations were calculated by the PSAFUVS using arrivals in Table 1 and coordinates of sensors. The arrivals is accurate result with a accuracy

Table 1: Arrival times of MS/AE sources.

Sensors	A (100, 200, 300)		B (200, −250, −180)		C (−350, 550, 120)		D (400, 300, 380)		E (−280, 380, 450)	
	P	S	P	S	P	S	P	S	P	S
S_1	0.099586	0.172511	0.156923	0.271835	0.20323	0.352052	0.055863	0.09677	0.16815	0.291283
S_2	0.133609	0.231448	0.084677	0.146685	0.225	0.389763	0.150195	0.260181	0.196944	0.341162
S_3	0.115351	0.19982	0.107165	0.18564	0.158703	0.274919	0.114545	0.198425	0.178033	0.308404
S_4	0.104447	0.180931	0.139527	0.2417	0.14584	0.252637	0.156448	0.271013	0.105908	0.183463
S_5	0.132366	0.229295	0.156078	0.270371	0.126151	0.218529	0.139941	0.242417	0.166768	0.288889
S_6	0.211058	0.365612	0.125889	0.218075	0.218855	0.379119	0.25648	0.444295	0.236629	0.409909

Table 2: Location results and error of MS/AE source coordinates under different reading errors of arrivals.

Sources	E0			E1			E2			E3		
	x	y	z	x	y	z	x	y	z	x	y	z
Locations												
A	99.84	200.11	300.01	99.62	197.74	298.60	104.59	196.01	282.83	23.38	169.48	298.38
B	200.13	−250.03	−179.98	199.68	−250.14	−178.76	191.63	−244.53	−161.62	6.42	−76.88	26.21
C	−349.90	549.74	120.03	−351.57	552.78	119.51	−392.80	604.08	127.11	−752.04	1051.38	254.27
D	399.69	299.74	379.51	398.40	301.27	379.58	374.79	265.48	327.16	930.39	449.65	549.21
E	−280.25	380.36	450.18	−279.72	379.75	447.11	−310.42	433.43	495.68	−186.53	156.04	340.15
Errors												
A	−0.16	0.11	0.01	−0.38	−2.26	−1.4	4.59	−3.99	−17.17	−76.62	−30.52	−1.62
B	0.13	−0.03	0.02	−0.32	−0.14	1.24	−8.37	5.47	18.38	−193.58	173.12	206.21
C	0.10	−0.26	0.03	−1.57	2.78	−0.49	−42.8	54.08	7.11	−402.04	501.38	134.27
D	−0.31	−0.26	−0.49	−1.6	1.27	−0.42	−25.21	−34.52	−52.84	530.39	149.65	169.21
E	−0.25	0.36	0.18	0.28	−0.25	−2.89	−30.42	53.43	45.68	93.47	−223.96	−109.85

0.000001 s. The location results calculated by accurate arrivals were expressed as E0 (the error is approximately equal to 0) and listed in Table 2. Compared results between authentic values and fitting results of dependent variables were listed in Table 3. It shows that the calculated coordinates (C) of A, B, C, D, and E using PSAFUVS are consistent with authentic results (T), and the fitting values are approximately equal to authentic values. To further verify the proposed method, the α and β were recalculated using (12), and the calculated results were compared with the authentic α and β. $\bar{v}$ is 7510.767. α and β are 5497.88 and 3174.30 m/s which are approximately equal to authentic values 5500 and 3175 m/s, respectively. Results prove that the proposed PSAFUVS is reasonable and reliable with a high precision for the unknown velocity system.

As we all know, it is always difficult to read the accurate arrivals because of limitations of technologies, personal error, or machine error. In present work, the errors are considered as three levels within E1 [0.000001, 0.00001], E2 [0.00001, 0.0001], and E3 [0.0001, 0.001], which are randomly generated by random function of Microsoft Excel. The calculated location results of the three levels were also listed in Table 2. Compared results between authentic values (T) and fitting results of MS/AE sources A, B, C, D, and E for error levels E0, E1, E2 and E3 are shown in Figure 3. From Table 2 and Figure 3, we can conclude that the location errors enlarged rapidity with the increase of arrival errors. The error E1 results in errors within 3 m, and the minimum and maximum are −0.28 and −2.98 m, respectively. The error E2 results in errors within 55 m, and the minimum and maximum are −3.99 and 54.08 m, respectively. The error E3 could induce errors more than 500 m, and the minimum and maximum are −1.62 and 530.39 m, respectively. Figure 3 also shows that the differences between fitted values and authentic values increased with the enlarging of arrival errors. The induced location errors were further investigated by absolute distance errors between real MS/AE sources and fitted MS/AE sources in Figure 4. It also clearly shows the absolute distance errors enlarging exponentially with the increase of arrival errors, and the location errors for A, B, C, D, and E are increased from 0.19, 0.13, 0.28, 0.64, and 0.47 m to 82.49, 331.61, 656.54, 576.49, and 266.39 m, respectively.

4. Conclusions

A new MS/AE source location method PSAFUVS was developed to address the difficult problem of the location errors for microseismic monitoring system induced by wave velocity measurement deviation. The proposed method was validated by numerical experimentations. Results show that the proposed method without the need for a premeasured wave velocity has a reasonable and reliable precision. It is demonstrated that the proposed method can not only locate the MS/AE source for unknown velocity network but

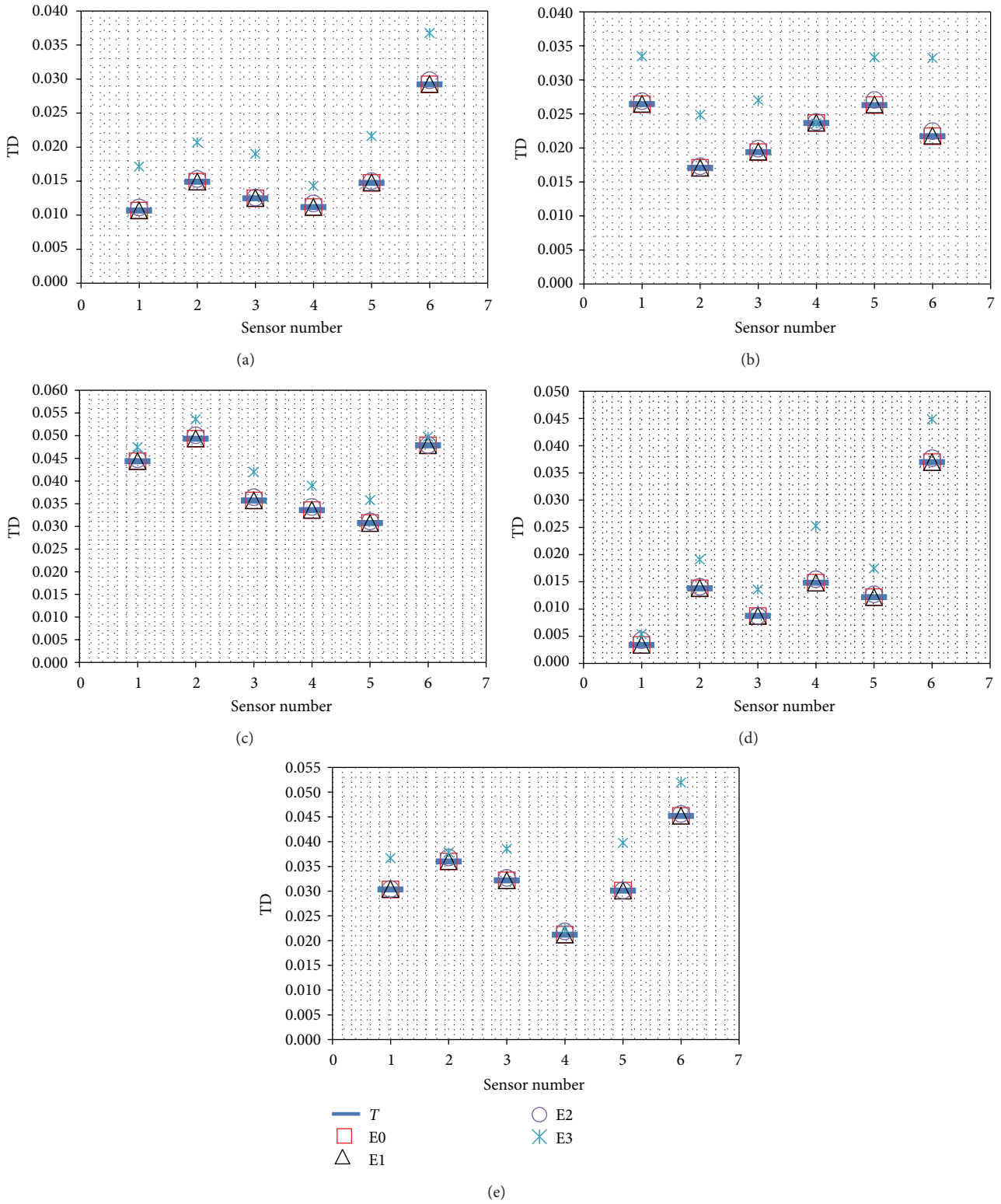


FIGURE 3: Comparisons between authentic values and fitting results: (a) source A, (b) source B, (c) source C, (d) source D, and (e) source E.

Table 3: Comparisons between authentic values and fitting results for the E0 condition.

Sensor $i(j = 1)$	A		B		C		D		E	
	T	C	T	C	T	C	T	C	T	C
1	0.010636	0.010636	0.02641	0.02641	0.044296	0.044296	0.003347	0.003347	0.030323	0.030323
2	0.014891	0.014891	0.01705	0.01705	0.049295	0.049295	0.01377	0.01377	0.035961	0.035961
3	0.012453	0.012453	0.019363	0.019363	0.035654	0.035654	0.008709	0.008709	0.032158	0.032158
4	0.011168	0.011168	0.023644	0.023644	0.033554	0.033553	0.014799	0.014798	0.021177	0.021176
5	0.014713	0.014713	0.026268	0.026268	0.030682	0.030682	0.012175	0.012175	0.030075	0.030075
6	0.029205	0.029205	0.021703	0.021703	0.047833	0.047833	0.036948	0.036948	0.045188	0.045188

Note: 1, 2, 3, 4, 5, and 6 indicating the values calculated using (8), and i is equal to 1, 2, 3, 4, 5 and 6, while j is equal 1.

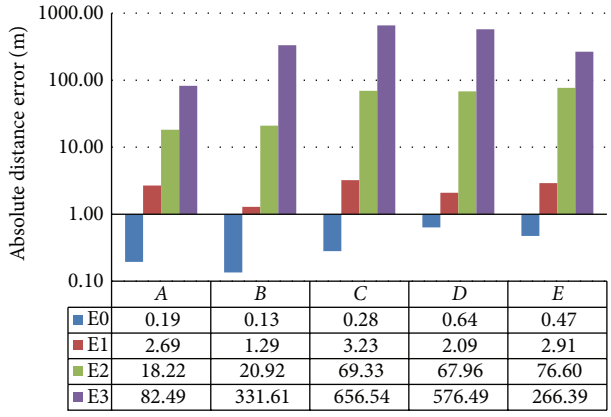


Figure 4: Comparisons of absolute distance error for different arrival errors.

also determine the PS wave real-time velocities for each event. Effects of arrival errors on location accuracy were investigated using PSAFUVS, and it shows location errors enlarged rapidity with the increase of arrival errors. The absolute distance errors between real MS/AE sources and fitted MS/AE sources were further discussed, and results show the absolute distance errors enlarging exponentially with the increase of arrival errors.

Conflict of Interests

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

Acknowledgments

The authors gratefully acknowledge the financial support of National Natural Science Foundation of China (50934006,41272304), National Basic Research (973) Program of China (2010CB732004), China Scholarship Council (CSC), Doctoral Candidate Innovation Research Support Program by Science & Technology Review (kjdb201001-7), Scholarship Award for Excellent Doctoral Student from Ministry of Education of China (105501010), and Support Program for Cultivating Excellent Ph.D. Thesis of Central South University.

References

[1] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1843–1854, 2002.

[2] M. Ge, H. R. Hardy Jr., H. Wang, and J. Wang, "Development of a simple mechanical impact system as a non-explosive seismic source," *Geotechnical and Geological Engineering*, vol. 29, no. 1, pp. 137–142, 2011.

[3] P. Senatorski, "Apparent stress scaling for tectonic and induced seismicity: model and observations," *Physics of the Earth and Planetary Interiors*, vol. 167, no. 1-2, pp. 98–109, 2008.

[4] J. Šílený and A. Milev, "Source mechanism of mining induced seismic events—resolution of double couple and non double couple models," *Tectonophysics*, vol. 456, no. 1-2, pp. 3–15, 2008.

[5] N. Xu, C. Tang, S. Wu, G. Li, and J. Yang, "Optimal design of microseismic monitoring networking and error analysis of seismic source location for rock slope," *Advanced Materials Research*, vol. 163, pp. 2991–2999, 2011.

[6] M. Hudyma and Y. H. Potvin, "An engineering approach to seismic risk management in hardrock mines," *Rock Mechanics and Rock Engineering*, vol. 43, no. 6, pp. 891–906, 2010.

[7] L. Dong, X. Li, and K. Peng, "Prediction of rockburst classification using Random Forest," *Transactions of Nonferrous Metals Society of China*, vol. 23, pp. 472–477, 2013.

[8] L. Dong, X. Li, and Z. Zhou, "Nonlinear model-based support vector machine for predicting rock mechanical behaviors," *Advanced Science Letters*, vol. 5, pp. 806–810, 2012.

[9] X. B. Li, L. J. Dong, G. Y. Zhao et al., "Stability analysis and comprehensive treatment methods of landslides under complex mining environment—a case study of Dahu landslide from Linbao Henan in China," *Safety Science*, vol. 50, no. 4, pp. 695–704, 2012.

[10] L. J. Dong, X. B. Li, M. Xu, and Q. Y. Li, "Comparisons of random forest and support vector machine for predicting blasting vibration characteristic parameters," *Procedia Engineering*, vol. 26, pp. 1772–1781, 2011.

[11] L.-J. Dong and X.-B. Li, "Interval non-probabilistic reliability method for surrounding jointed rock mass stability of underground caverns," *Chinese Journal of Geotechnical Engineering*, vol. 33, no. 7, pp. 1007–1013, 2011.

[12] L.-J. Dong and X.-B. Li, "Interval parameters and credibility of representative values of tensile and compression strength tests on rock," *Chinese Journal of Geotechnical Engineering*, vol. 32, no. 12, pp. 1969–1974, 2010.

[13] L. Dong, X. Li, G. Zhao, and F. Gong, "Fisher discriminant analysis model and its application to predicting destructive effect of masonry structure under blasting vibration of open-pit mine," *Chinese Journal of Rock Mechanics and Engineering*, vol. 28, no. 4, pp. 750–756, 2009.

[14] L.-J. Dong, X.-B. Li, and Y.-F. Bai, "Fisher discriminant analysis model for classifying top coal cavability of the steep seam," *Journal of the China Coal Society*, vol. 34, no. 1, pp. 58–63, 2009.

[15] L.-J. Dong, X.-B. Li, and F.-Q. Gong, "Unascertained average clustering method for classification of grade of shrink and expansion for expansive soils and its application," *Journal of Central South University*, vol. 39, no. 5, pp. 1075–1080, 2008.

[16] L.-J. Dong, G.-J. Peng, Y.-H. Fu, Y.-F. Bai, and Y.-F. Liu, "Unascertained measurement classifying model of goaf collapse prediction," *Journal of Coal Science and Engineering*, vol. 14, no. 2, pp. 221–224, 2008.

[17] L. Dong, D. Hu, Y. Bai, and Y. Liu, "Unascertained average grade model for surrounding rock classification on hydraulic tunnels," in *Proceedings of the 2008 International Symposium on Safety Science and Technology: Progress in Safety Science and Technology, Volume VII, Parts A and B*, 2008.

[18] L. Dong, Y. Fu, Y. Liu, and Y. Bai, "A Fisher discriminant analysis model for classification of rocks surrounding in tunnel," in *Proceeding of the Information Technology and Environmental System Sciences (ITESS '08)*, vol. 2, 2008.

[19] L. Dong and F. Wang, "Comprehensive evaluation seismic stability of slopes based on unascermined measurcmem," *The Chinese Journal of Geological Hazard and Control*, vol. 18, pp. 74–78, 2007.

[20] L. Dong, X. Li, C. Ma, and W. Zhu, "Comparisons of Logistic regression and Fisher discriminant classifier to seismic event identification," *Disaster Advances*, vol. 6, p. 8, 2013.

[21] L. Dong and X. Li, "Comprehensive models for evaluating rockmass stabilitybased on statisticalcomparisons of multiple classifiers," *Mathematical Problems in Engineering*, vol. 2013, Article ID 395096, 9 pages, 2013.

[22] B. L. N. Kennett, K. Marson-Pidgeon, and M. S. Sambridge, "Seismic source characterization using a Neighbourhood Algorithm," *Geophysical Research Letters*, vol. 27, no. 20, pp. 3401–3404, 2000.

[23] Y. Tian and X. F. Chen, "Review of seismic location study," *Progress in Geophysics*, vol. 17, no. 1, pp. 147–155, 2002.

[24] P. Nivesrangsan, J. A. Steel, and R. L. Reuben, "Source location of acoustic emission in diesel engines," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 1103–1114, 2007.

[25] F. Waldhauser and W. L. Ellsworth, "A double-difference Earthquake location algorithm: method and application to the Northern Hayward Fault, California," *Bulletin of the Seismological Society of America*, vol. 90, no. 6, pp. 1353–1368, 2000.

[26] M. Knapmeyer, "Location of seismic events using inaccurate data from very sparse networks," *Geophysical Journal International*, vol. 175, no. 3, pp. 975–991, 2008.

[27] L. Dong, X. Li, L. Tang, and F. Gong, "Mathematical functions and parameters for microseismic source location without premeasuring speed," *Chinese Journal of Rock Mechanics and Engineering*, vol. 30, no. 10, pp. 2057–2067, 2011.

[28] L. J. Dong and X. B. Li, "Three-dimensional analytical solution of acoustic emission or microseismic source location under cube monitoring network," *Transactions of Nonferrous Metals Society of China*, vol. 22, pp. 3087–3094, 2012.

[29] L. Geiger, "Herdbestimmung bei Erdbeben aus den Ankunftszeiten. Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 331–349. 1912 transliterated in English by FWL Peebles & AH Corey: probability method for the determination of earthquake epicenters from the arrival time only," *Bulletin of St. Louis University*, vol. 8, pp. 60–71, 1910.

[30] L. Dong, X. Li, and F. Gong, "An analytical location method for acoustic emission and microseismic sources," China Patent, 2012.

[31] X. Li and L. Dong, "A location method for acoustic emission and microseismic sources," China Patent, 2012.

[32] L. Dong and X. Li, "Closed-form solutions of acoustic emission source location for minimal element monitoring arrays of rectangular pyramid and circular column," *Disaster Advances*, vol. 6, no. 13, p. 11, 2013.

[33] K. R. Anderson, "Robust earthquake location using *M*-estimates," *Physics of the Earth and Planetary Interiors*, vol. 30, no. 2-3, pp. 119–130, 1982.

[34] K. R. Anderson, "Epicentral location using arrival time order," *Bulletin of the Seismological Society of America*, vol. 71, pp. 541–545, 1981.

[35] M. Antolik, G. Ekström, and A. M. Dziewonski, "Global event location with full and sparse data sets using three-dimensional models of mantle *P*-wave velocity," in *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Sourse Location*, pp. 291–317, Springer, Berlin, Germany, 2001.

[36] R. Belchamber and M. Collins, "Acoustic emission source location using simplex optimization," *Journal of Acoustic Emission*, vol. 9, no. 4, pp. 271–276, 1990.

[37] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, 1997.

[38] M. P. Flanagan, S. C. Myers, and K. D. Koper, "Regional travel-time uncertainty and seismic location improvement using a three-dimensional a priori velocity model," *Bulletin of the Seismological Society of America*, vol. 97, no. 3, pp. 804–825, 2007.

[39] M. Ge, "Source location error analysis and optimization methods," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 4, pp. 1–10, 2012.

[40] B. L. N. Kennett, "Non-linear methods for event location in a global context," *Physics of the Earth and Planetary Interiors*, vol. 158, no. 1, pp. 46–54, 2006.

[41] X. B. Li and L. J. Dong, "Comparison of two methods in acoustic emission source location using four sensors without measuring sonic speed," *Sensor Letters*, vol. 9, no. 5, pp. 2025–2029, 2011.

[42] G. Mellen, M. Pachter, and J. Raquet, "Closed-form solution for determining emitter location using time difference of arrival measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 3, pp. 1056–1058, 2003.

[43] T. Nicholson, "Application of 3D empirical travel times to routine event location," *Physics of the Earth and Planetary Interiors*, vol. 158, no. 1, pp. 67–74, 2006.

[44] T. Nicholson, M. Sambridge, and O. Gudmundsson, "Hypocenter location by pattern recognition," *Journal of Geophysical Research B*, vol. 107, no. 6, pp. ESE 5-1–ESE 5-19, 2002.

[45] L. Dong and X. Li, "Hypocenter relocation for Wenchuan Ms 8. 0 and Lushan Ms 7. 0 earthquakes using TT and TD methods," *Disaster Advances*, vol. 6, no. 13, p. 10, 2013.

[46] G. L. Pavlis and J. R. Booker, "The mixed discrete-continuous inverse problem: application to the simultaneous determination of earthquake hypocenters and velocity structure," *Journal of Geophysical Research*, vol. 85, pp. 4801–4810, 1980.

[47] Q.-Y. Li, L.-J. Dong, X.-B. Li, Z.-Q. Yin, and X.-L. Liu, "Effects of sonic speed on location accuracy of acoustic emission source in rocks," *Transactions of Nonferrous Metals Society of China*, vol. 21, no. 12, pp. 2719–2726, 2011.

[48] L. Dong, X. Li, and L. Tang, "Main influencing factors of micro-seismic source location accuracy," *Science and Technology Review*, vol. 31, pp. 26–32, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 829804, 8 pages
http://dx.doi.org/10.1155/2013/829804

Research Article

Monte Carlo Based Personalized PageRank on Dynamic Networks

Zhang Junchao,[1,2] Chen Junjie,[1] Jiancheng Song,[2] and Rong-Xiang Zhao[3]

[1] *College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China*
[2] *Shanxi Key Laboratory of Coal Mining Equipment and Safety Control, Taiyuan 030024, China*
[3] *Shanxi Taiyuan Tideflow Electronic Technology Co., Ltd., Taiyuan 030024, China*

Correspondence should be addressed to Jiancheng Song; sjc6018@163.com

Received 24 June 2013; Accepted 25 August 2013

Academic Editor: Deguang Le

In large-scale networks, the structure of the underlying network changes frequently, and thus the power iteration method for Personalized PageRank computation cannot deal with this kind of dynamic network efficiently. In this paper, we design a Monte Carlo-based incremental method for Personalized PageRank computation. In a dynamic network, first, we do a random walk starting from each node and save the performed walks into a fingerprint database; second, we update the fingerprint database in a fixed time interval with our proposed update algorithm; finally, when a query is issued by a user, we estimate the Personalized PageRank vector by our proposed approximation algorithm. Experiments on real-world networks show that our method can handle multichanges of the underlying network at a time and is more efficient than related work, so it can be used in real incremental Personalized PageRank-based applications.

1. Introduction

Large-scale networks are ubiquitous in today's world, such as World Wide Web, online social networks, and huge search and query-click logs regularly collected and processed by search engines. PageRank [1], which is the stability distribution of a random walk in a Markov chain, has emerged as an effective measure for ranking on large-scale networks. In addition, Personalized PageRank [2, 3], which is the stability distribution of a random walk with restart in a Markov chain, has been proved to be an effective method of personalized ranking in many applications, such as link prediction [4–6], friend recommendation [7–9], and personalized searching [10, 11]. Given a network, there are two main methods for computing the PageRank or Personalized PageRank vector: one is power iteration applying the linear algebra proposed by Page et al. [1] and the other is the Monte Carlo approximation methods proposed by Litvak [12] and Fogaras and Rácz [13].

In practice, the structure of the underlying network is dynamic, such that Webpages join or leave frequently in World Wide Web and following relations appear or disappear frequently in an online social network. Recomputing PageRank or Personalized PageRank for every change, that appears or disappears in an edge, of the structure is impossible in large-scale networks, so incremental PageRank or Personalized PageRank algorithms are necessary for this situation.

1.1. Background. Given a graph $G = (V, E)$, where V is the set of nodes, E is the set of edges, and the number of nodes and edges is n and m, respectively; the adjacent matrix M of the graph G is

$$m_{i,j} = \begin{cases} 1 & (i, j) \in V, \\ 0 & (i, j) \notin V, \end{cases} \tag{1}$$

where $i, j \in V$, and (i, j) are a directed edge from i to j.

The transition possibility matrix P of the graph G is

$$p_{i,j} = \frac{m_{i,j}}{\sum_{j=1}^{n} m_{i,j}}, \tag{2}$$

where the weights on the outgoing edges of each node sum up to 1.

PageRank is the stability distribution of a random walk in a Markov chain. The main idea of random walk on a graph is that starting from a source node, at each step, jump to a random node with a probability ε (usually called the teleport probability), and follow a random outgoing edge from the current with the probability $1 - \varepsilon$. If the number k of random steps tends to be infinite, that is, $k \to +\infty$, then the probability that the random walk stops at each node is a constant, and, no matter how we choose the source node, the probability of stopping at each node does not change.

Personalized PageRank is the same as PageRank, except that the random walk jumps to the source node with a probability ε at each step. Given a source (or called personalized) node u, the Personalized PageRank score of node v is

$$\pi_u(v) = \varepsilon\delta_u(v) + (1-\varepsilon)\sum_{i=1}^{n} p_{i,v} \cdot \pi_u(i), \quad (3)$$

where $\delta_u(v) = 1$ if and only if $u = v$, and $\delta_u(v) = 0$, otherwise.

1.2. Problem Statement. In this paper, we study the incremental computation of Personalized PageRank based on the Monte Carlo method; that is, given a graph, we do a random walk of fixed length starting from each node update the performed walks incrementally in fixed time interval, and approximate Personalized PageRank with the performed walks whenever needed.

Here, the core of the problem is designing an incremental update algorithm and a Personalized PageRank approximation algorithm.

1.3. Our Contribution. First, we design an approximation method for Personalized PageRank based on Monte Carlo method. We do a random walk of fixed length starting from each node, and save the performed walks in a fingerprint database.

Second, we propose an incremental algorithm, which updates the performed walks in fixed time interval, while the structure of underlying graph changes.

Third, we propose an approximation algorithm for computing Personalized PageRank vector based on the performed walks.

Finally, we do some experiments on real world graphs, and the experiments show that our proposed method is more efficient than related researches in estimating Personalized PageRank on dynamic networks.

2. Related Work

Matrix Computation. Matrix computation, which applies power iteration of linear algebra to compute the PageRank vector [1], is the most popular method for PageRank computation. In order to accelerate the convergence of power iteration, Borgs et al. [14] propose a sublinear time algorithm for PageRank computations. Because the graphs in real applications are usually sparse, del Corso et al. [15] and Fujiwara et al. [16] all apply sparse matrices to speed up the PageRank computation. Though these algorithms can accelerate the Personalized PageRank computation, they cannot be used in dynamic networks.

Monte Carlo Method. The Monte Carlo approximation methods [12, 13] are efficient for estimating Personalized PageRank on large-scale graphs. The basic idea is that starting from the source node, do a number R of random walks (called fingerprints); at each step, stop the random walk with a probability ε, and follow a random outgoing edge from the current node with the probability $1 - \varepsilon$. The lengths of the R random walks from each node are subject to the geometrical distribution Geom(ε). Then, use the frequencies of visit to different nodes in the performed walks to approximate the Personalized PageRank vector. The lengths of the performed walks conform to geometrical distribution, and thus there should be some long walks, so Sarma et al. [17] propose to do all random walks with length $1/\varepsilon$. In addition, Bahmani et al. [18] study the Personalized PageRank problem on MapReduce [19], and propose an optimal algorithm for constructing single random walk on MapReduce. In dynamic networks, Bahmani et al. [20] propose to store the performed walks in database and update them when an edge joins or leaves the underlying network. They consider that there is only one joining or leaving edge at a time, and update the fingerprints whenever the underlying network changes. However, in this paper, we study the same problem as [20], but we assume that there should be multichanges (multijoining and leaving edges) in a network.

Personalized PageRank Application. Personalized PageRank and other random walk with restart-based measures are the most well-known graph computation problems, and they have proved to be very effective in a variety of applications, such as link prediction [5, 21] and friend recommendation [22, 23], Webpage retrieval [1–3, 24–27], and trust-based recommendation systems [28–31] in large-scale networks. A similar work is [32], which uses dynamic GPS data to recommend an individual with interesting locations or travel sequences matching her travel preferences. But, the difference is that although they use the dynamic dataset, they do not mention how they handle this dynamic dataset incrementally. However, in this paper, we focus on the incremental update of recommendations, and the results can be used in any Personalized PageRank-based applications.

3. Monte Carlo-Based Incremental Personalized PageRank Computation

In this section, we describe our proposed system for incremental Personalized PageRank computation and illustrate the system in detail.

3.1. System Overall Design. Our system answers users' Personalized PageRank queries or recommend Personalized PageRank results to the specific users, and the architecture of the system is in Figure 1. First, we initialize the fingerprints and save them in a fingerprint database, and then, in a fixed time interval, we update the fingerprint database. These two

Input: G, the input graph, and λ, the length of fingerprints.
Output: $W = \{w_i \mid 1 \le i \le n \wedge i \in V \wedge |w_i| = \lambda\}$, where for each node $i \in V$, return a random walk w_i of length λ.

```
(1) For each u ∈ V do
(2)   Let last = u, w_u = {}.append(last);
(3)   For i = 1 to n do
(4)     current = last.randomSuccessor();
(5)     w_u = w_u.append(current)
(6)     last = current;
(7)   End for;
(8) End for;
(9) Return W;
```

Algorithm 1: Initialization (G, λ).

phases are done automatically by the system. When a query is issued by a personalized user, the searching engine retrieves related fingerprints from the fingerprint database, estimates the Personalized PageRank vector, and then returns nodes with highest score to the user. Some of the main phases of the system work as follows.

(i) *Initializing Fingerprint Database*. Given a graph $G = (V, E)$, we first do a random walk (fingerprint) with fixed length λ starting from each node in the graph, and save all the performed walks in the fingerprint database.

(ii) *Updating Fingerprint Database*. The structure of the graph changes frequently, and it is impossible to update the fingerprint database with every change, so we update the fingerprint database with a fixed time interval t.

(iii) *Computing Personalized PageRank*. When a query is issued from a user, the searching engine generates SQL queries according to the specified user, retrieves related results from the fingerprint database, computes the Personalized PageRank vector, and returns a top-k list to the user.

3.2. Initializing Fingerprint Database. Given a graph $G = (V, E)$, for each node $u \in V$, do a random walk w_u of length λ starting from the node u, and the details are given in Algorithm 1.

After doing the performed random walks, we save the performed walks in the fingerprint database.

3.3. Updating Fingerprint Database. In dynamic networks, Bahmani et al. [20] propose a model to update the fingerprint database whenever there is a joining edge or leaving edge. However, the structure of the underlying graph changes frequently, so it is impossible to update the fingerprint database for very change. In this paper, we propose an update algorithm that handles multichanges while updating the fingerprint database.

In a time interval t, there will be k changes in the structure of the underlying network. These k changes include joining edges and leaving edges in any order. We can apply the algorithm proposed in [20] k times, but this will consume lots of computation resources. In this paper, we propose an update algorithm, which could handle multichanges at a time, and the details of the algorithm are in Algorithm 2.

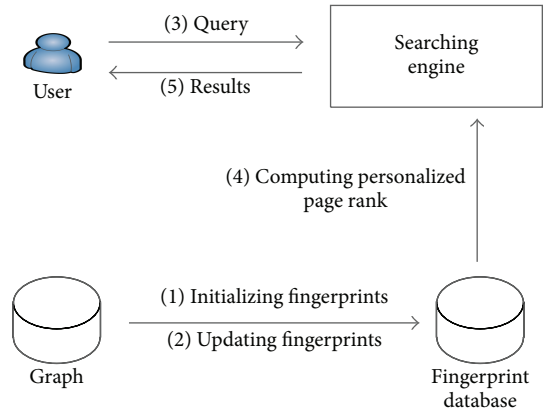


Figure 1: System architecture.

Given a series of changes $k = [\underbrace{+, -, -, +, -, \ldots, +}_{k}]$, where + means a joining edge, and − means a leaving edge we can reorder them into $k' = [\underbrace{-, \ldots, -}_{k_1}, \underbrace{+, \ldots, +}_{k_2}]$, where $k = k_1 + k_2$.

For the k_1 leaving edges, we search in the fingerprint database for the fingerprints which contain one of these leaving edges, and the results are presented with L. For each $l \in L$, if it contains one leaving edge e from its beginning, then we delete e and its following edges but keep the start node of e.

For the k_2 joining edges, we collect all start nodes in S, search in the fingerprint database for the fingerprints which contain any node $s \in S$, and the results are presented with J. For each $j \in J$, if it contains one node s from its beginning, then, with a probability $1/(d + 1)$, we delete the following nodes of s.

Finally, for both L and J, the lengths of fingerprints are smaller than λ, so we go on walking randomly on $G' = G - k_1 \cup k_2$ until their lengths reach λ.

Input: G, the input graph, k, the changes of graph structure, and W, the set of fingerprints.

(1) Reorder $k = \left[\underbrace{+, -, -, +, -, \ldots, +}_{k}\right]$ in order to obtain $k' = \left[\underbrace{-, \ldots, -}_{k_1}, \underbrace{+, \ldots, +}_{k_2}\right]$, where $k = k_1 + k_2$;
(2) Let L = null;
(3) **For** each $e \in k_1$ **do**
(4) Let $le \leftarrow$ select all fingerprints which contain e in W;
(5) $L = L \cup le$
(6) **End for**;
(7) **For** each $l \in L$ **do**
(8) **For** each $e \in k_1$ **do**
(9) **If** e in l **then**
(10) Let $l \leftarrow$ cut off l at e, and keep the start node of e;
(11) **End if**;
(12) **End for**;
(13) **End for**;
(14) Let J = null;
(15) Let $S \leftarrow$ all start nodes in k_2;
(16) **For** each $s \in S$ **do**
(17) Let $ls \leftarrow$ select all fingerprints which contain s in W;
(18) $J = J \cup ls$;
(19) **End for**;
(20) **For** each $j \in J$ **do**
(21) **For** each $s \in S$ **do**
(22) **If** s in j **then**
(23) Let $j \leftarrow$ cut off j after s;
(24) **End if**;
(25) **End for**;
(26) **End for**;
(27) **For** each $w \in L \cup J$ **do**
(28) Let $w \leftarrow$ go on walking randomly from the end node of w until the length reaches λ;
(29) **End for**;
(30) Update W with $L \cup J$;

Algorithm 2: Update (G, k, W).

3.4. Computing Personalized PageRank. In this subsection, we propose an approximation method for estimating Personalized PageRank vector based on the performed walks, and the details are in Algorithm 3.

For each fingerprint $w \in W$, we find the subfingerprints which start from the personalized node u, and have the length of $1/\varepsilon$. In order to make these subfingerprints independent of each other, we must make sure that these subfingerprints do not overlap. After finding out those subfingerprints, we can estimate the Personalized PageRank vector with visit frequencies of the nodes.

4. Experiments

In this section, we do some experiments to validate the efficiency of our proposed method. We use the method proposed by the Bahmani et al. [20] as baseline algorithm, and compare it with our method in execution time.

Table 1: Details of datasets.

Dataset	n	m	n/m
Dianping	204074	926720	4.5
Gowalla	196591	1900654	9.7

4.1. Dataset. To validate the performance of both algorithms, we use two real-world datasets of networks, Dianping [33] and Gowalla [34]. The details of the datasets are in Table 1.

4.2. Experimental Setup. Our experimental environment is a personal computer with a 4-core processor and 4 GB memory. All algorithms are implemented by Java on the Ubuntu operation system. Our toolkit for dealing with graph is Jung [35], and the database for storing fingerprints is MySQL.

Input: W, the set of fingerprints, u, a personalized node for Personalized PageRank, and ε, the teleport probability.
Output: π_u, the Personalized PageRank vector for u.
(1) Let $result = \{\}$, $\pi_u(v) = 0$ for $1 \le v \le n$;
(2) **For** each $w \in W$ **do**
(3) Scan w, find all sub-fingerprints of length $1/\varepsilon$ starting from u, and add these sub-fingerprints to $result$;
(4) **End for**;
(5) **For** each $w \in result$ **do**
(6) **For** each node $v \in w$ **do**
(7) $\pi_u(v) = \pi_u(v) + 1$;
(8) **End for**;
(9) **End for**;
(10) **For** $v = 1$ to n **do**
(11) $\pi_u(v) = \dfrac{\pi_u(v)}{\sum_{v=1}^{n} \pi_u(v)}$;
(12) **End for**;
(13) Return π_u;

Algorithm 3: Computing PPR (W, u, ε).

4.3. Experimental Results. For each graph G, we choose 90% of the dataset as already known graph A, and the other 10% as B. When an edge joins A, we randomly choose an edge e from B, delete it from B, and add it to A. When an edge leaves A, we randomly choose an edge e from A, delete it from A, and add it to B. The probabilities of joining and leaving are equal. We build an extra thread that changes the structure of the underlying graph once per ten seconds.

First, we compare the update time between our method with the method proposed by Bahmani et al. We update the fingerprint database in a fixed time interval from 1 to 10 second, and the results for two datasets are in Figures 2 and 3. From these two figures we can see that our update algorithm executes only once, but the Bahmani's algorithm need to update the fingerprint database multitimes that is proportional to the time interval. Moreover, we find that the time used in updating the fingerprint database contains a nearly constant time retrieving the related fingerprints.

Second, we compare the execution time of Personalized PageRank approximation between the proposed method and Bahmani's method. Both of the algorithms retrieve fingerprints from the fingerprint database first and then estimate the Personalized PageRank vector with the fingerprints. The Bahmani's algorithm retrieves related fingerprints starting from the personalized node, and there are millions of fingerprints, so the searching process takes a lot of time. Our algorithm returns all fingerprints, and it needs only one pass of scanning the whole fingerprints to compute the Personalized PageRank vector. We randomly choose 10 personalized nodes and compare the average approximation time of Personalized PageRank, and the details are in Figure 4.

Third, we compare the overall performance between the two methods. The update time and Personalized PageRank approximation time of our method are all shorter than the Bahmani's method, and thus it is obvious that the overall time used by our method should be shorter. Here,

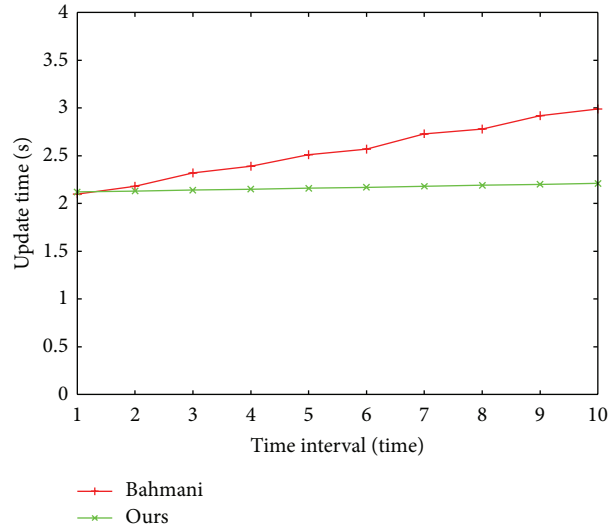


Figure 2: Comparison of update time (Dianping).

using the multithreading technology of Java, we compare the throughput of the two methods, and the results are in Figure 5. In this experiment, one thread is used to generate changes of the structure of underlying graph, one thread is used to update fingerprints, and other threads are generated dynamically to compute the Personalized PageRank vector. We use the number of finished Personalized PageRank vectors per minute as the throughput. From the figure we can see that our proposed method is more efficient than the Bahmani's method in throughput.

Finally, we compare the accuracy of the two algorithms. To measure the quality of the results, for a node u and a value k, we define Z_u^k to be the set of nodes with the k largest Personalized PageRank scores for u. In most Personalized PageRank-based applications, only nodes in Z_u^k

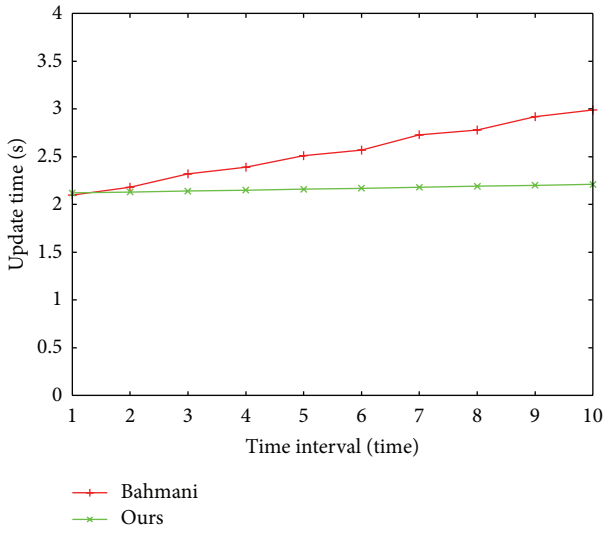


Figure 3: Comparison of update time (Gowalla).

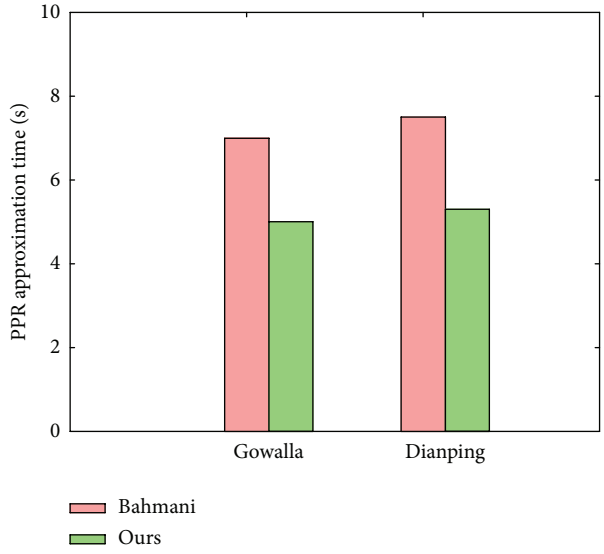


Figure 4: Average approximation time comparison of Personalized PageRank (PPR) approximation.

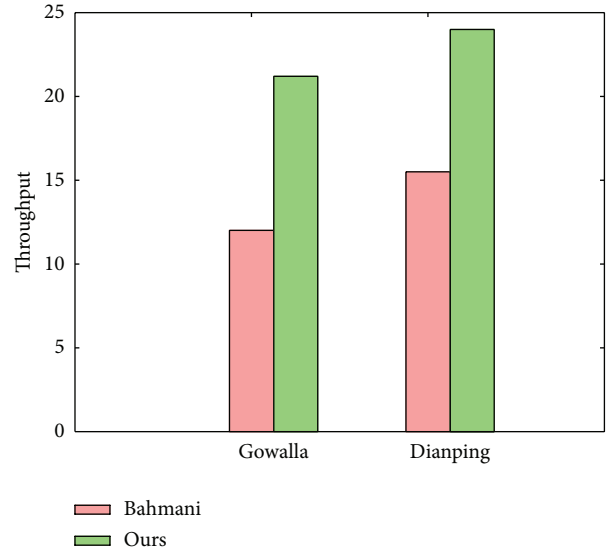


Figure 5: Comparison of throughput.

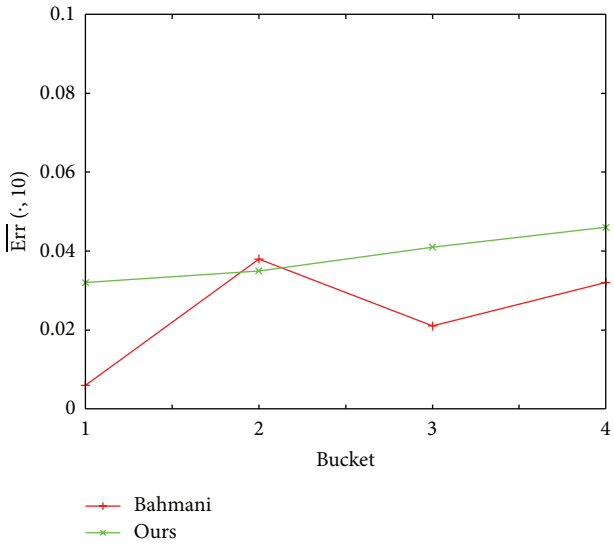


Figure 6: Bucket versus $\overline{\text{Err}}(\cdot, 10)$, Dianping.

are of interest. For example, in recommending top-k friends in online social networks or returning top-k documents for a query, Personalized PageRank usually acts as one of the most important features. In this experiment, we assume Personalized PageRank as standard and compare how much approximation methods approximate it by the following equation:

$$\text{Err}(u,k) = \frac{\sum_{u \in Z_u^k} \left| \pi_u(v) - \widehat{\pi}_u(v) \right|}{\sum_{u \in Z_u^k} \pi_u(v)}, \quad (4)$$

where $\widehat{\pi}_u(v)$ is the approximate Personalized PageRank by either Bahmani's or our algorithm. For the ground truth of $\pi_u(v)$, we use 100 iterations of power iteration. For testing the efficiency of the algorithms for different nodes, we divide nodes in a graph into 4 buckets, where bucket i ($1 \leq i \leq 4$) contains nodes with degree in $[10^{i-1}, 10^i)$. There is no node with degree larger than 10000 in our datasets, so bucket 4 equals to $[1000, +\infty)$. We randomly choose 10 nodes from each bucket and average the error defined by (4). We return top 10 and 100 results for each node and observe how approximation error changes along with the degree of nodes. Figures 6 and 7 illustrate the top-10 recommender results on both datasets, and Figures 8 and 9 illustrate the top-100 recommender results on both datasets. From these four figures we can see that, for both top 10 and top 100, our algorithm is more accurate in approximating Personalized PageRank than Bahmani's algorithm.

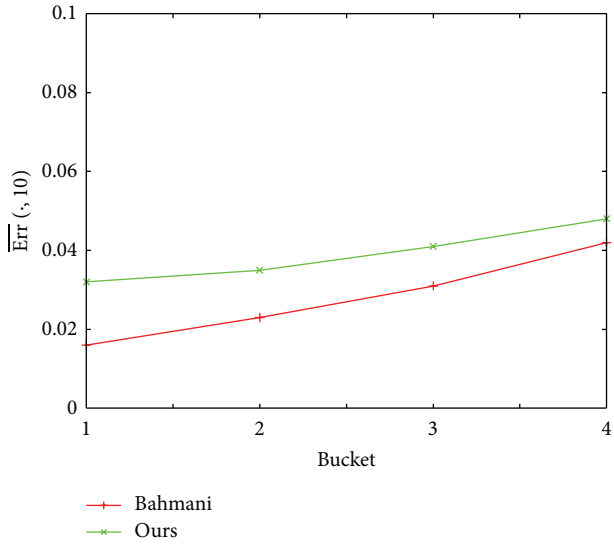


Figure 7: Bucket versus $\overline{\text{Err}}(\cdot, 10)$, Gowalla.

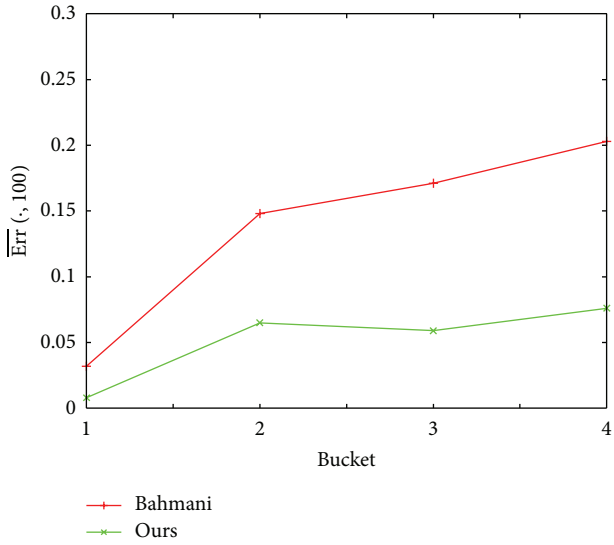


Figure 8: Bucket versus $\overline{\text{Err}}(\cdot, 100)$, Dianping.

5. Conclusion

Personalized PageRank is very important for complex networks, and it has been proved to be effective in many real applications, such as link prediction, entity recommendation, and social searching. However, in these applications, the underlying networks are usually dynamic, and traditional power iteration method cannot deal with them efficiently.

In this paper, we design a Monte Carlo-based incremental method for Personalized PageRank approximation in dynamic networks. We propose an efficient update algorithm to handle changes of the structure of the underlying network, and an efficient Personalized PageRank approximation algorithm with the performed walks. Experiments on real-world networks show that our method can handle multichanges of the underlying network structure at one time and is more efficient than related work, so can be used in real incremental Personalized PageRank-based applications.

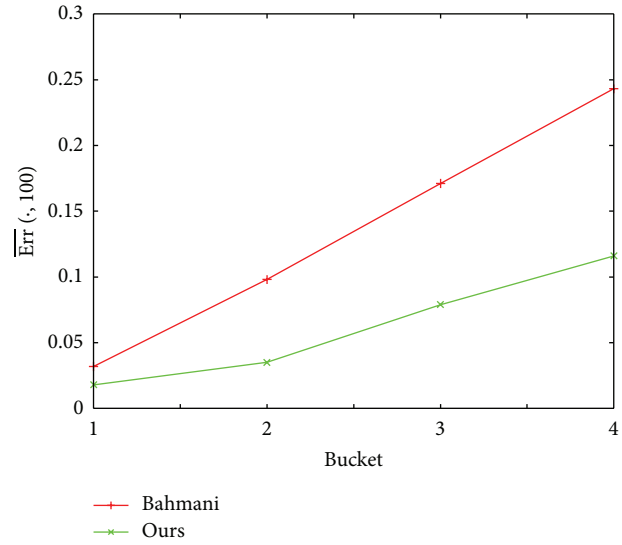


Figure 9: Bucket versus $\overline{\text{Err}}(\cdot, 100)$, Gowalla.

Acknowledgments

This work was supported in part by the National Significant Science and Technology Special Project of China and the National Natural Science Foundation of China.

References

[1] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: bringing order to the web," Tech. Rep., Stanford InfoLab, 1999.

[2] T. H. Haveliwala, "Topic-sensitive PageRank," in *Proceedings of the 11th International Conference on World Wide Web (WWW '02)*, pp. 517–526, May 2002.

[3] T. Haveliwala, S. Kamvar, and G. Jeh, "An analytical comparison of approaches to personalizing PageRank," Tech. Rep., Stanford, 2003.

[4] W. Liu and L. Lü, "Link prediction based on local random walk," *Europhysics Letters*, vol. 89, no. 5, Article ID 58007, 2010.

[5] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[6] H. H. Song, T. W. Cho, V. Dave, Y. Zhang, and L. Qiu, "Scalable proximity estimation and link prediction in online social networks," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference*, pp. 322–335, 2009.

[7] F. Jiang and Z. Wang, "Pagerank-based collaborative filtering recommendation," in *Information Computing and Applications*, pp. 597–604, Springer, 2010.

[8] J. H. Su, B. W. Wang, and V. S. Tseng, "Effective ranking and recommendation on web page retrieval by integrating association mining and PageRank," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*

and Intelligent Agent Technology (WI-IAT Workshops '08), pp. 455–458, December 2008.

[9] M. Ramezani, "Improving graph-based approaches for personalized tag recommendation," *Journal of Emerging Technologies in Web Intelligence*, vol. 3, no. 2, pp. 168–176, 2011.

[10] S. Ye, S. Yan, C. Yang, K. Yu, H. Ding, and Y. Yang, "Research on micro-blog search and sorting algorithm based on improved PageRank," in *Proceedings of the International Conference on Information Engineering and Applications (IEA '13)*, pp. 811–817, 2013.

[11] B. Bhushan and N. Kumar, "Searching the most authoritative & obscure sources from the web," *International Journal of Computer Science and Network Security*, vol. 10, no. 8, pp. 149–153, 2010.

[12] N. Litvak, *Monte Carlo Methods of PageRank Computation*, Department of Applied Mathematics, University of Twente, 2004.

[13] D. Fogaras and B. Rácz, "Towards scaling fully personalized pagerank," in *Algorithms and Models for the Web-Graph*, pp. 105–117, Springer, 2004.

[14] C. Borgs, M. Brautbar, J. Chayes, and S. Teng, "A sublinear time algorithm for pagerank computations," in *Algorithms and Models for the Web Graph*, pp. 41–53, Springer, 2012.

[15] G. M. del Corso, A. Gullí, and F. Romani, "Fast PageRank computation via a sparse linear system," *Internet Mathematics*, vol. 2, no. 3, pp. 251–273, 2005.

[16] Y. Fujiwara, M. Nakatsuji, M. Onizuka, and M. Kitsuregawa, "Fast and exact top-k search for random walk with restart," *Proceedings of the VLDB Endowment*, vol. 5, no. 5, pp. 442–453, 2012.

[17] A. D. Sarma, S. Gollapudi, and R. Panigrahy, "Estimating Pagerank on graph streams," in *Proceedings of the 27th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '08)*, pp. 69–78, June 2008.

[18] B. Bahmani, K. Chakrabarti, and D. Xin, "Fast personalized PageRank on MapReduce," in *Proceedings of the ACM SIGMOD and 30th PODS 2011 Conference*, pp. 973–984, June 2011.

[19] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[20] B. Bahmani, A. Chowdhury, and A. Goel, "Fast incremental and personalized pagerank," *Proceedings of the VLDB Endowment*, vol. 4, no. 3, pp. 173–184, 2010.

[21] W. Liu and L. Lü, "Link prediction based on local random walk," *Europhysics Letters*, vol. 89, no. 5, Article ID 58007, 2010.

[22] J. Weng, E. P. Lim, J. Jiang, and Q. He, "TwitterRank: finding topic-sensitive influential twitterers," in *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM '10)*, pp. 261–270, February 2010.

[23] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, pp. 591–600, April 2010.

[24] S. Brin, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks*, vol. 30, no. 1–7, pp. 107–117, 1998.

[25] W. Xing and A. Ghorbani, "Weighted PageRank algorithm," in *Proceedings of the 2nd Annual Conference on Communication Networks and Services Research*, pp. 305–314, May 2004.

[26] S. Dominich and A. Skrop, "PageRank and interaction information retrieval," *Journal of the American Society for Information Science and Technology*, vol. 56, no. 1, pp. 63–69, 2005.

[27] H. Taneja and R. Gupta, "WEB information retrieval using query independent Page Rank algorithm," in *Proceedings of the International Conference on Advances in Computer Engineering (ACE '10)*, pp. 178–182, June 2010.

[28] Y. A. Kim and H. S. Song, "Strategies for predicting local trust based on trust propagation in social networks," *Knowledge-Based Systems*, vol. 24, no. 8, pp. 1360–1371, 2011.

[29] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*, pp. 203–210, July 2009.

[30] M. Jamali and M. Ester, "TrustWalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 397–405, July 2009.

[31] S. Meyffret, L. Médini, and F. Laforest, "Trust-based local and social recommendation," in *Proceedings of the 4th ACM RecSys Workshop on Recommender Systems and the Social Web*, pp. 53–60, 2012.

[32] Y. Zheng and X. Xie, "Learning travel recommendations from user-generated GPS traces," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 1, article 2, 2011.

[33] Z. Jin, D. Shi, Q. Wu, H. Yan, and H. Fan, "LBSNRank: personalized pagerank on location-based social networks," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 980–987, 2012.

[34] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*, pp. 1082–1090, August 2011.

[35] Jung, http://jung.sourceforge.net/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 281707, 8 pages
http://dx.doi.org/10.1155/2013/281707

Research Article

Distributional Fractal Creating Algorithm in Parallel Environment

Shuai Liu,[1,2] Weina Fu,[1,3] Huimin Deng,[1] Caihe Lan,[1] and Jiantao Zhou[1]

[1] *College of Computer Science, Inner Mongolia University, Hohhot 010012, China*
[2] *School of Physical Science and Technology, Inner Mongolia University, Hohhot 010012, China*
[3] *Department of Computer Science and Technology, Hohhot University of Nationalities, Hohhot 010012, China*

Correspondence should be addressed to Shuai Liu; cs_liushuai@imu.edu.cn

Received 3 June 2013; Revised 9 August 2013; Accepted 19 August 2013

Academic Editor: Yong Jin

Nowadays, the fractal is used widely everywhere. Then, its creating time becomes an important study area for complex iteration functions because the escape-time algorithm (ETA), which is the most used algorithm in fractal creating, performs not so well in this condition. In this paper, in order to solve this problem, we improve ETA into the parallel environment and reach well performance. At first, we provide a separation method of ETA to reform it into a SIMC-MC2 grid. Secondly, we prove its correctness and compute the complexity of this novel parallel algorithm. Meantime, we separate an improved ETA which we have presented into the same parallel environment and compute its complexity. Additionally, theoretical and experimental results show the characteristics of this novel algorithm. Finally, the computational result shows that a novel environment is needed to decrease large manual allocation strategies, which block the improved benefit.

1. Introduction

Since Mandelbrot presented the Mandelbrot set (M set), which is the dictionary of all Julia sets [1], the fractal has been soon used everywhere for 30 years. Nowadays, thinking and algorithms, which are generated by fractal, have been used in many science domains. Today, two algorithms are generalized and used when fractals are created [2]. One is the escape-time algorithm (ETA), and the other is the iterated function system (IFS). Although IFS is a fast algorithm, it can be only used in fractals with a certain iteration strategy. However, admittedly, a lot of fractal iteration functions are uncertain in practical applications. So the ETA becomes the most universal and effective algorithm in creating fractal where the iteration functions are complex.

In order to decrease the creating time of ETA, we reform ETA into the parallel environment to improve its speed and effectiveness in this paper. At first, we should show the ETA by Algorithm 1 [2].

In this way, we know that ETA is such an algorithm that uses the escape threshold and the max iteration number to create fractals. Then, it colors all points in the displayed area with different colors. These colors are different with different iteration times under a given color strategy. Finally, if $m_z > M$, the point z is convergent, and if $m_z \leq M$, the point z is divergent.

Earlier on, to solve some defects of ETA, we have presented an improved algorithm (IA) to remedy the following two defects [3].

> Defect a. The convergent points have to perform max iterations.
>
> Defect b. When the iteration midresults are in the computed regions, ETA wastes existing computations.

Although IA remedies these two defects of ETA, its performance is also not so well when the iterated function is more complex and the displayed area is larger. For example, we use two iteration functions to create fractals with IA: one is in [4, 5], and the other is the one we created in [6]. We have these two results, and we show them in Tables 1 and 2. We can see that the fractal creating times enlarge greatly when we enlarge the displayed area and compute more points. From the bold data in Tables 1 and 2, we can see that these data are so

Table 1: Runtimes of the two functions with ETA.

Iteration functions	$f(z) = (1/2)z3^{\sin^2(\pi z/2)}$	$f(z) = (1/2)(z3^{\sin^2(\pi z/2)} + \sin^2(\pi z/2))$
Displayed area/points' number	$(-1, 1) \times (-1, 1)/640 \times 640$	
Runtime(s)	16.887	139.242
Displayed area/points' number	$(-2, 2) \times (-2, 2)/1280 \times 1280$	
Runtime(s)	67.353	**573.460**
Displayed area/points' number	$(-4, 4) \times (-4, 4)/2560 \times 2560$	
Runtime(s)	**278.441**	**2318.005**

Table 2: Runtimes of the two functions with IA.

Iteration functions	$f(z) = (1/2)z3^{\sin^2(\pi z/2)}$	$f(z) = (1/2)(z3^{\sin^2(\pi z/2)} + \sin^2(\pi z/2))$
Displayed area/points' number	$(-1, 1) \times (-1, 1)/640 \times 640$	
Runtime(s)	3.165	6.287
Displayed area/points' number	$(-2, 2) \times (-2, 2)/1280 \times 1280$	
Runtime(s)	12.914	24.102
Displayed area/points' number	$(-4, 4) \times (-4, 4)/2560 \times 2560$	
Runtime(s)	52.411	**101.578**

large that we cannot wait for a so long time for these fractals in practice. Then, we know that there are more images that need more points to compute. So we have to reach a novel method to solve them.

We analyze these large runtimes in Tables 1 and 2, and we find that the phenomenon of these results is due to the large extra space, which stores the midresult of IA by an $n \times n$ matrix ($n \times n$ is the computational points number). So the extra matrix becomes too large to compute when we enlarge the computational points (the displayed area).

In order to consider a novel method to decrease computational complexity, we admit that the parallel environment is a suitable way to solve this problem [7, 8]. Furthermore, it is more suitable in fractals [9]. This is because many fractals have highlighted characteristics and fit into parallel environments. For example, symmetrical characteristics are owned by many generalized Mandelbrot sets [10, 11], Julia sets [12], and other fractals [13].

So, in this paper, we separate and improve ETA and IA into parallel environments with a separation method at first. Moreover, we prove their correctness and compute their complexity. Then, we use some well-known generation method in the experiment. Finally, we discuss the experiment result by our conclusion.

The remainder of this paper is organized as follows. We present a separation method of both ETA and IA in Section 2. Moreover, we have their parallel experimental results and analyses in Section 3. Finally, Section 4 summarizes the main results of this paper and presents the following study objectives.

2. Separation Method of ETA and IA

In order to reform ETA and IA into parallel environments, our separation method can be given by 4 steps.

Step 1 (dividing computations into tasks). In this step, we divide computations into a class of tasks. The strategy makes all processes busy as far as possible without much supervisory cost. In this way, parallel structure is constructed.

Step 2 (distributing tasks among processes). The objective of this step is to load balance between processes in task distribution. The balanced load contains computation, input/output, data access and communication, and so forth

Step 3 (applying coordination, communication, and synchronization of data). Coordination decreases the cost of communication and synchronization. Moreover, in order to increase locality of data access, this strategy must finish those tasks earlier, which depends many tasks.

Step 4 (mapping processes into processors). From Steps 1–3, we get a complete parallel algorithm which can control mappings from processes into processors. Otherwise, this work is executed by OS. Of course, this mapping is for a certain system or environment.

So, in order to run ETA and IA in parallel environments, we should distribute them in parallel forms. In this paper, we use the SIMC environment to distribute them and the separation method of ETA and IA. We show the separation method in Algorithms 2 and 3. To simplify without loss of solution, we divide the displayed area into $p(n)$ parts by horizontal lines.

After these two separation methods are presented, we process our experiment by constructing a parallel environment with the same 9 PCs (personal computers). In our experiment, one PC is a primary node, and other 8 PCs are task nodes. Then, all subresults are connected to the primary node as the final result.

Step 1. Assuming N as the escape threshold number, M as the max iteration number, $m_z = 0$ as the iteration number of point z and $N_z = z$ as records of $f^n(z)$.
Step 2. For all points z in complex plane,
While $m_z \le M$ and $|N_z| < N$
Let $m_z = m_z + 1$. $N_z = f(N_z)$.
Step 3. Color all points z with color strategy of m_z.
Algorithm finished.

Algorithm 1: The escape time algorithm.

Primary node
For all task nodes
Send group message (x_{start}, x_{end}, y_{start_i}, y_{end_i}, f) to ith task nodes;
--$y_{start\_i} = y_{start} + (i-1) * \text{int}[(y_{end} - y_{start})/p(n)]$
--$y_{end\_i} = y_{start\_i+1} - 1$
--x_{start}, x_{end}, y_{start_i}, y_{end_i} are all pixel numbers (positive integer), f is iterated function/mapping
While all task nodes send its result
Connect all sub-images and display it;
Finished;
Task nodes
If primary node send message (x_{start}, x_{end}, y_{start_i}, y_{end_i}, f)
Use ETA to create sub-image by compute all points in this area with iterated function f;
Send sub-image to primary node;
Finished;

Algorithm 2: Separation method of ETA.

In our experiment, in order to decrease communicational cost, we find that the communications are mostly between the task nodes; that is, there is no need for the primary node.

3. Parallel Experiment and Analysis

3.1. Parallel Parameters. At first, we have some parallel evaluating indicators of a parallel algorithm. In the following evaluating indicators, n is the scale of the problem:

(a) running time $t(n) = t_r + t_c$, where t_r is routing time of data by network or memorizer and t_c is the computing time of arithmetic and logic in processors;

(b) number of processors $p(n)$, which obeys the exponential distribution by defining $p(n) = n^{1-e}$ $(0 < e < 1)$;

(c) parallel cost $c(n) = t(n) \cdot p(n)$, which is called the best cost when the time complexity of executing cost is same between the existing parallel algorithm and the worst serial algorithm.

(d) speedup ratio $\text{SR}(n) = \text{WS}(n)/\text{WP}(n)$, where $\text{WS}(n)$ is the worst time cost with the best serial algorithm and $\text{WP}(n)$ is the worst time cost of the parallel algorithm with the same problem; easily saying that $1 \le \text{SR}(n) \le p(n)$ and that the parallel algorithm is better when $\text{SR}(n)$ is larger;

(e) parallel efficiency $\text{PE}(n) = \text{SR}(n)/p(n)$, which is used to measure utilization efficiency of processors in a parallel algorithm;

(f) parallel flexibility measures, which is the relation between n and $\text{PE}(n)$ by a steady $p(n)$, calling an algorithm flexible when $\text{PE}(n)$ increases linearly by n.

3.2. Experiments with $T(z)$ and $B(z)$. Then, we execute fractals of (1) and (2) as our experiments. The creating fractals are given in Figures 1 and 2.

In our experiments, we left the primary node free for computations, and we only granted it with the authority, which is the connection with another 8 worker nodes. The connection is done under the SIMC strategy with ETA and the MIMC strategy with IA. This is because ETA needs fewer communications. In ETA, the primary node only sends the computational area to the task nodes. Meantime, the task nodes do not have any other communications. In this way, ETA can be executed with single instruction. On the contrary, IA needs many communications by the worker nodes. So we use multiple instructions in it.

We have the running time of every fractal image, which are created by (1) and (2) in Tables 3 and 4 with the different

Primary node

For all task nodes
Send group message $(x_{\text{start}}, x_{\text{end}}, y_{\text{start\_}i}, y_{\text{end\_}i}, f)$ to ith task nodes;

$$--y_{\text{start\_}i} = y_{\text{start}} + (i-1) * \text{int}[(y_{\text{end}} - y_{\text{start}})/p(n)]$$
$$--y_{\text{end\_}i} = y_{\text{start\_}i+1} - 1$$

While all task nodes send its result
Connect all sub-images and display it;
Finished;

Task nodes

If primary node send message $(x_{\text{start}}, x_{\text{end}}, y_{\text{start\_}i}, y_{\text{end\_}i}, f)$
Use IA to create sub-image by compute all points in this area with iterated function f;
{
$\Delta x + i\Delta y = f(x + iy)$;
While $\Delta y \in (y_{\text{start\_i}}, y_{\text{end\_}i})$ and $m_{x+iy} < M$;
$f(x + iy) = f(\Delta x + i\Delta y)$, $m_{x+iy} = m_{x+iy} + m_{\Delta x+i\Delta y}$;
If $\Delta y \notin (y_{\text{start\_i}}, y_{\text{end\_}i})$ and $m_{x+iy} < M$;
Send message $(\Delta x, \Delta y, i, \text{Need})$ to jth node;
$$--j = \text{int}[\Delta y/(y_{\text{end\_}i} - y_{\text{start\_}i} + 1)]$$
If receive message (x, y, j, Need)
Send message $(\Delta x, \Delta y, m_{\Delta x+i\Delta y})$ to jth node;
$$--\Delta x + i\Delta y = f(x + iy)$$
If receive message $(\Delta x, \Delta y, m_{\Delta x+i\Delta y})$
$f(x + iy) = f(\Delta x + i\Delta y)$, $m_{x+iy} = m_{x+iy} + m_{\Delta x+i\Delta y}$
If $m_{x+iy} > N$
$m_{x+iy} = N$;
}
Send sub-image to primary node;
Finished;

Algorithm 3: Separation method of IA.

algorithms ETA and IA. Every running time contains both serial time and parallel time:

$$f_1(z) = \frac{1}{2} z 3^{\sin^2(\pi z/2)}, \tag{1}$$

$$f_2(z) = \frac{1}{2}\left(z 3^{\sin^2(\pi z/2)} + \sin^2\frac{\pi z}{2}\right). \tag{2}$$

We compare the running times of ETA and IA with iteration functions (1) and (2). Then, we compute the evaluating indicators of these two algorithms. We use $p(n) = 8$ as the constant. So we only discuss $c(n)$, SR(n), PE(n), and PF (parallel flexibility) in our paper.

We know that $c(n) = t(n) \cdot p(n)$ and $p(n) = 9$. So we have Figure 3 to show their costs individually. Moreover, we show SR(n) in Figure 4 using the times WS(n) and WP(n) in experiments. Then, we need not compute PE(n) because it is only a constant coefficient different from the SR(n). Finally, we use Figure 5 to present parallel flexibility with equation SR(n)/n.

Then, with these results, we have that the speedup ratio of ETA is better than that of the IA (in Figure 4), and that of the running time of IA is smaller than that of the ETA (in Figure 3). Admittedly, the upper bound of the speedup is $p(n) = 9$. It is to say that parallel ETA still has space in speedup ratio. Contrarily, the speedup ratio of IA is lower because there are too many communications between worker nodes in parallel IA. However, it reduces t_r, but it increases t_c. Then, we have to say that speedup ratio is smaller when worker nodes are more.

Moreover, from Figures 4 and 5, we can see that the two absolute flexibilities of the two algorithms are not so well. Especially, flexibility of IA is worse. But this does not this mean that their complexities are nearly linear. So the two relative flexibilities of the two algorithms are well.

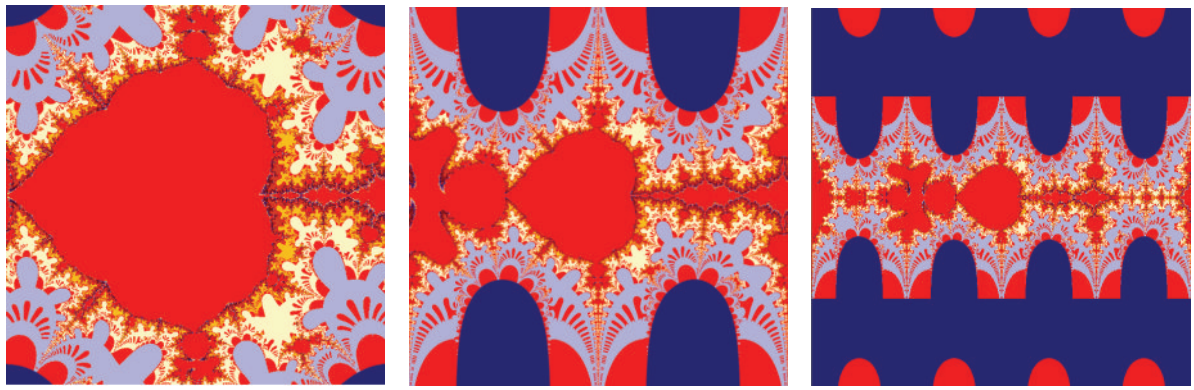
Finally, the created fractals are the same as in Figures 1 and 2. So we do not present them with additional figures.

3.3. Experiments with Generalized M Sets. After the experiments with $T(z)$ and $B(z)$, which are two iteration functions about the generalized $3x + 1$ functions, we reach the final experiment of the generalized Mandelbrot sets with exponent k (k-M set). Similarly, in order to validate effectiveness of the novel algorithm, we process a k-M set with $k = 2.47$ in serial ETA (SA), serial IA (SIA), parallel ETA (PA), and parallel IA (PIA). The fractal image of k-M set is given in Figure 6.

Then, we use Figure 7 to present the eight subresults of the k-M set. It also validates the correctness of PIA.

(a) $(-1, 1) \times (-1, 1)$ with 640×640 (b) $(-2, 2) \times (-2, 2)$ with 1280×1280 (c) $(-4, 4) \times (-4, 4)$ with 2560×2560

Figure 1: Fractals of (1).



(a) $(-1, 1) \times (-1, 1)$ with 640×640 (b) $(-2, 2) \times (-2, 2)$ with 1280×1280 (c) $(-4, 4) \times (-4, 4)$ with 2560×2560

Figure 2: Fractals of (2).

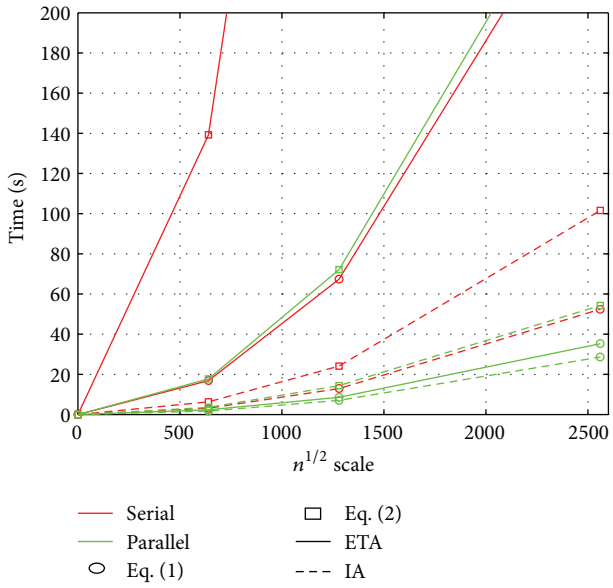


Figure 3: Running time to compare between ETA and IA with these two iteration functions into these two environments.

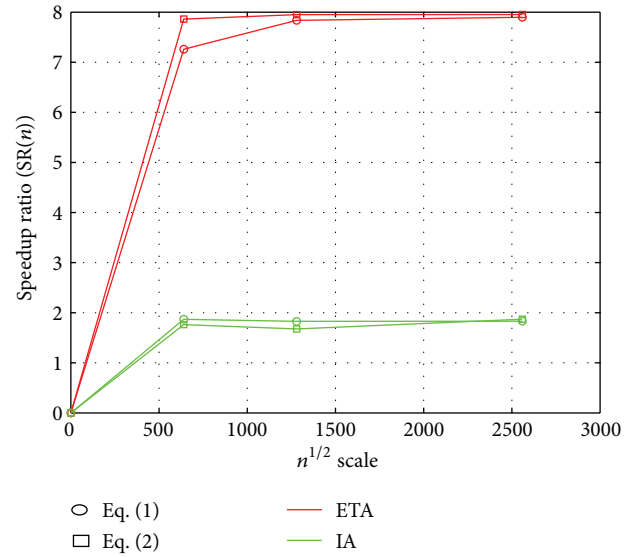


Figure 4: Speedup ratio to compare between ETA and IA with these two iteration functions.

Table 3: Running time of ETA.

Iteration functions	$f(z) = (1/2)z3^{\sin^2(\pi z/2)}$	$f(z) = (1/2)(z3^{\sin^2(\pi z/2)} + \sin^2(\pi z/2))$
Displayed area/points' number	$(-1,1) \times (-1,1)/640 \times 640$	
Runtime(s)		
Serial	16.887	139.242
Parallel	2.326	17.710
Displayed area/points' number	$(-2,2) \times (-2,2)/1280 \times 1280$	
Runtime(s)		
Serial	67.353	573.460
Parallel	8.594	72.135
Displayed area/points' number	$(-4,4) \times (-4,4)/2560 \times 2560$	
Runtime(s)		
Serial	278.441	2318.005
Parallel	35.260	291.462

Table 4: Running time of IA.

Iteration functions	$f(z) = (1/2)z3^{\sin^2(\pi z/2)}$	$f(z) = (1/2)(z3^{\sin^2(\pi z/2)} + \sin^2(\pi z/2))$
Displayed area/points' number	$(-1,1) \times (-1,1)/640 \times 640$	
Runtime(s)		
Serial	3.165	6.287
Parallel	1.694	3.556
Displayed area/points' number	$(-2,2) \times (-2,2)/1280 \times 1280$	
Runtime(s)		
Serial	12.914	24.102
Parallel	7.063	14.381
Displayed area/points' number	$(-4,4) \times (-4,4)/2560 \times 2560$	
Runtime(s)		
Serial	52.411	101.578
Parallel	28.623	54.279

In fact, the k-M set is a fast creating fractal. In this way, we do not present the creating time for these four methods (SA, SIA, PA, and PIA) because they are mixed and hard to discuss. In this paper, we only compute SR = WS(SA)/WP(PA) and flexibility = SR/n of them. We present them in Figures 8 and 9. In Figure 8, we present SR of the k-M set, and, in Figure 9, we present the flexibility of the k-M set.

In Figures 8 and 9, we also use $n = 640 \times 640$, 1280×1280, and 2560×2560.

In Figures 8 and 9, we also use $c(n) = t(n) \cdot p(n)$ and $p(n) = 9$ with any n. So we have $SR_1(n) = WS(SA)/WP(PA) = c(SA)/c(PA)$ and $SR_2(n) = WS(SIA)/WP(PIA) = c(SIA)/c(PIA)$ in Figure 4 by using the cost time instead of WS and WP in experimental results. Then, we use Figure 9 to present parallel flexibility with equation $SR(n)/n$.

Then, with these results, we also have that speedup ratio of ETA is better than that of IA (in Figure 8). Then, we find that the SR of the k-M set is smaller than those of $T(z)$ and $B(z)$. This is because the creating time of the k-M set is small. Meantime, the distributional time comes to a larger ratio. So SRs are decreasing. But we also find that the SR of IA is larger. This is because of the local attractiveness of the k-M set. Moreover, it is similar that the speedup ratio of IA is lower than that of ETA. This is also because there are too many communications between worker nodes in PIA.

Moreover, from Figure 8, we compute flexibilities of the two algorthims, and we find that both of them are not so well. Then, we can also find that two relative flexibilities of the two algorithms are well.

4. Conclusion

We constructed a parallel environment to run the distribution fractal creating algorithms ETA and IA by static task distribution strategy. Then, we compared these two parallel algorithms by some parallel evaluating indicators. Although there were many communication redundancies in IA, experiment results show that both ETA and IA can increase their speed and efficiency with a better environment.

The next step is to avoid communication redundancies, so we will transplant ETA and IA into a cloud environment. We will process this transformation since we have computed a universal computational complexity of fractal creating methods by iteration conditions. It is a new way

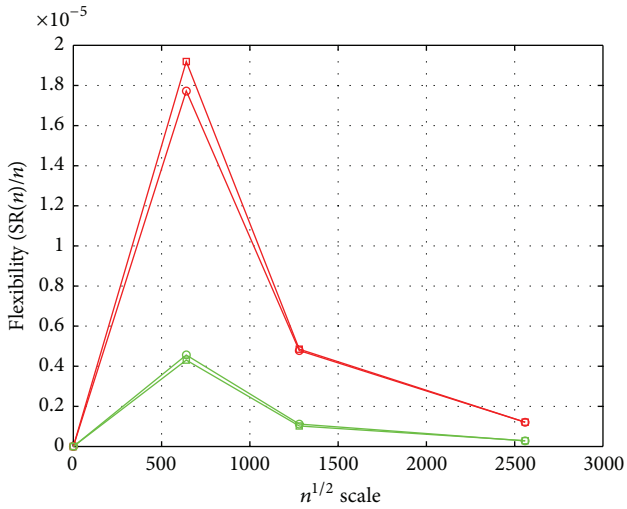


Figure 5: Flexibility to compare between ETA and IA with these two iteration functions into these two environments.

Figure 6: Fractal image of the k-M set with $k = 2.47$.

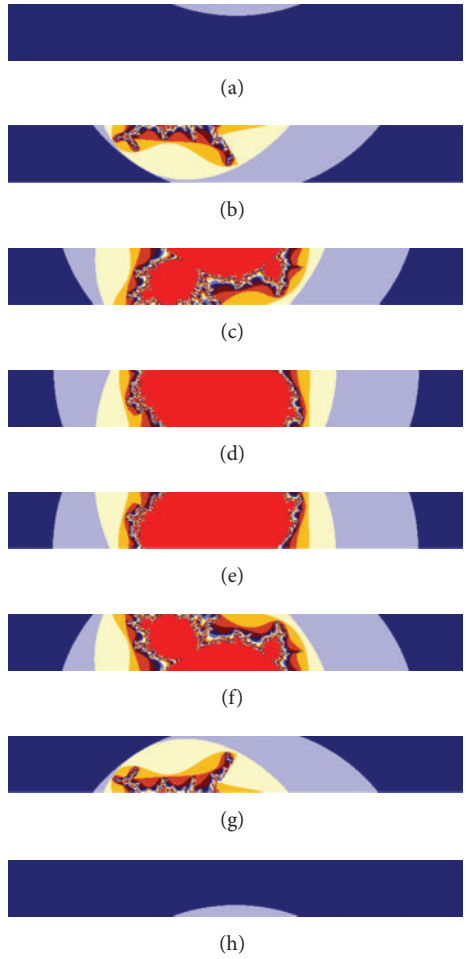


Figure 7: Subfractal image of the k-M set with $k = 2.47$.

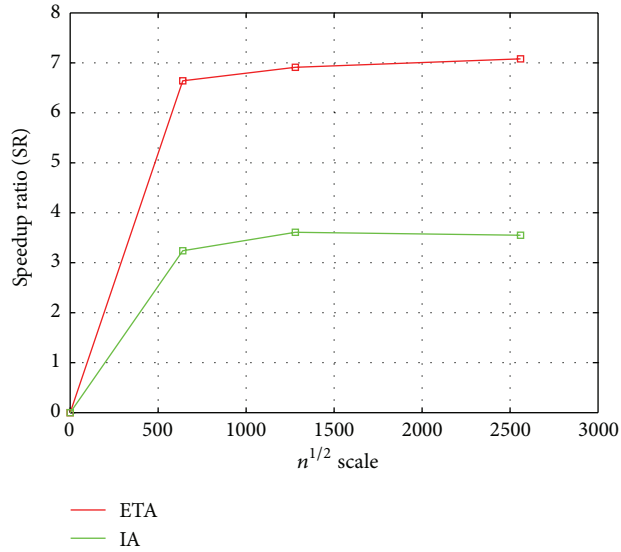


Figure 8: Speedup ratio to compare between ETA and IA with the k-M set, where $k = 2.47$.

to use inner mechanism in the cloud environment to avoid outer communication. Furthermore, we will present a special novel algorithm for the generalized Mandelbrot sets with rational number exponent when we have its structural characteristics.

Acknowledgments

This work is supported by grants from the Program of Higher-Level Talents of Inner Mongolia University (nos. 125126 and 115117), National Natural Science Foundation of China (nos. 61261019 and 61262082), the key Project of Chinese Ministry of Education (no. 212025), the inner Mongolia Science Foundation for Distinguished Young Scholars (2012JQ03) and Scientific projects of higher school of Inner Mongolia (no. NJZY13004). The authors would like to thank

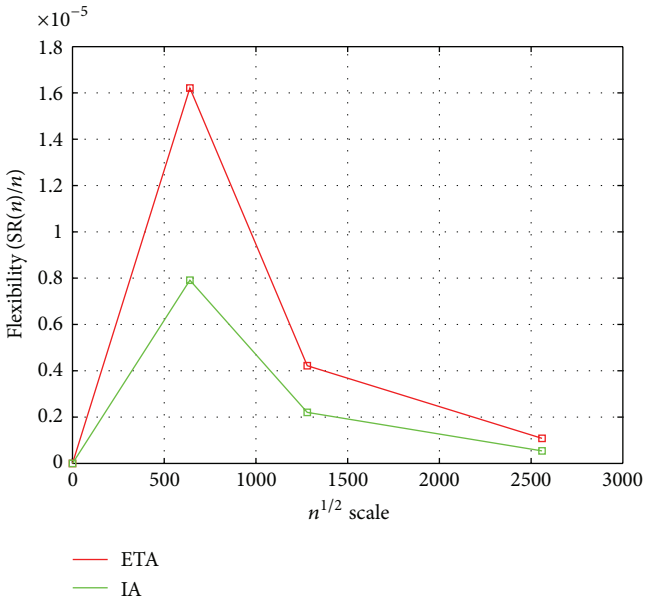


Figure 9: Flexibility to compare between ETA and IA with the k-M set, where $k = 2.47$.

the anonymous reviewers for their helpful comments in reviewing this paper.

References

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, San Fransisco, Calif, USA, 1982.

[2] J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2003.

[3] S. Liu, X. Che, and Z. Wang, "Improvement of escape time algorithm by no- escape-point," *Journal of Computers*, vol. 6, no. 8, pp. 1648–1653, 2011.

[4] J. P. Dumont and C. A. Reiter, "Visualizing generalized 3x+1 function dynamics," *Computers and Graphics (Pergamon)*, vol. 25, no. 5, pp. 883–898, 2001.

[5] S. Liu, X.-J. Che, and Z.-X. Wang, "Existence domain analysis and numerical algorithm of fixed point for generalized 3x+1 function Tx," *Acta Electronica Sinica*, vol. 39, no. 10, pp. 2282–2287, 2011.

[6] S. Liu and Z. Wang, "Fixed point and fractal images for a generalized approximate 3x+1 function," *Journal of Computer-Aided Design and Computer Graphics*, vol. 21, no. 12, pp. 1740–1744, 2009.

[7] Q. Wu, M. Zhang, R. Zheng et al., "A Qos-satisfied prediction model for cloud-service composition based on a hidden markov model," *Mathematical Problems in Engineering*, vol. 2013, Article ID 387083, 7 pages, 2013.

[8] C. B. Guure and N. A. Ibrahim, "Bayesian analysis of the survival function and failure rate of weibull distribution with censored data," *Mathematical Problems in Engineering*, vol. 2012, Article ID 329489, 18 pages, 2012.

[9] M. Liu, S. Liu, W. Fu et al., "Distributional escapetime algorithm based on generalized fractal sets in cloud environment," *Chinese Journal of Electronics*. In press.

[10] S. Liu, X. Cheng, C. Lan et al., "Study in fractals of generalized M-set with rational number exponent," *Applied Mathematics and Computation*, vol. 220, pp. 668–675, 2013.

[11] I. Andreadis and T. E. Karakasidis, "On numerical approximations of the area of the generalized mandelbrot sets," *Applied Mathematics and Computation*, vol. 219, no. 23, pp. 10974–10982, 2013.

[12] Y. Sun, X. Zhao, and K. Hou, "Calculation of julia sets by equipotential point algorithm," *International Journal of Bifurcation and Chaos*, vol. 23, no. 1, 2013.

[13] R. D. D. Díaz, L. H. Encinas, and J. M. Masqué, "A fractal sets attached to homogeneous quadratic maps in two variables," *Physica D*, vol. 245, no. 1, pp. 8–18, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 216478, 11 pages
http://dx.doi.org/10.1155/2013/216478

Research Article

High Reliable Relay Selection Approach for QoS Provisioning in Wireless Distributed Sensor Networks

Yong Jin

School of Computer Science & Engineering, Changshu Institute of Technology, Changshu 215500, China

Correspondence should be addressed to Yong Jin; jinyong@cslg.cn

Received 11 July 2013; Revised 13 August 2013; Accepted 15 August 2013

Academic Editor: Gelan Yang

In wireless distributed sensor networks, one open problem is how to guarantee the reliable relay selection based on the quality of services diversity. To address this problem, we focus on the reliable adaptive relay selection approach and adaptive QoS supported algorithm, based on which we present a Markov chain model, in consideration of different packet states and error control algorithm assignment. The mathematical analyses and NS-2 simulation results show that the proposed relay selection approach could perform better in terms of saturation throughput, reliability, and energy efficiency, compared with the traditional approaches. More importantly, the quality of real-time multimedia streaming is improved significantly, in terms of decodable frame ratio and delay.

1. Introduction

It is well known that delivering and transmitting various data over wireless distributed and dynamic networks is a very challenging task. Wireless communication link is a dynamic environment. Data transmission over wireless channel suffers from the limited bandwidth, high packet loss rate, dynamic changes of channel state, sensor nodes mobility, channel competition, and so forth [1]. These constraints and challenges, in combination with the delay and loss of sensitive nature of multimedia applications [2], make QoS provisioning over wireless sensor networks a challenging proposition.

The data delivery of different applications inherently has different QoS requirements [3, 4] such as throughput, real time performance, playable frame rate, reliability and energy efficiency, and so forth. There is an open problem in the relay selection approach design: how to take advantage of the requirements information of QoS diversity in the wireless distributed sensor networks to find out the optimal relay for forwarding data packets from the potential relays.

In this paper, we investigate how to select the optimal relay nodes to provide the reliable and effective QoS provision and satisfy the diversity requirements of application services.

The rest of the paper is organized as follows. The related work and our work are elaborated in Section 2. Section 3 describes the system models and problem description. In Section 4, we design a Markov chain model to describe process of data communication in wireless distributed sensor network and give the theoretical analysis model to show the achieved performance. The supported scheme of QoS diversity is used to realize reliable, efficient, and robust communication in wireless distributed sensor network, which is shown in Section 5. Section 6 proposes the QoS supported adaptive relay selection (ECA-Q-RS) approach, followed by the details of implementation. Simulation and mathematics results are given in Section 7. Finally, we conclude the paper in Section 8.

2. Related Work

Recent research shows that the traditional cooperative strategies, such as amplify-and-forward (AF), decode-and-forward (DF), or estimate-and-forward (EF), work well in networks of static channel state information (CSI) but have scalability limitations that degrade performance in larger or rapidly moving networks. So, research on AF and DF has drawn researchers' attention which focuses on outdated channel estimates [5], outdated CSI [6], outage probability and bit error probability [7], and mobile relays [8], in order to reduce

the complexity in management about mobile sensor nodes and simplify some essential procedures such as reliability guarantee and wireless resource allocation. In additon, error control scheme is usually used to improve the reliability for wireless distributed sensor networks, which can be realized by automatic repeat request (ARQ) [9], forward error correction (FEC) [10], or hybrid ARQ (HARQ) [11, 12].

How to combine error control protocol and cooperative technology in wireless distributed sensor networks has been fairly studied in the literature [13–15]. Lo et al. [13] present a decentralized relay selection protocol, which supports HARQ transmission where relay nodes forward parity information to the destination node in the event of a decoding error. A high energy efficiency adaptive cooperative error control mechanism was proposed in [14], which predicts frame loss rate using the model of GM (1,1) and adjusts the FEC parameter according to the energy efficiency of the sensor nodes. In addition, Wang et al. [15] designed a deterministic linear network coding scheme for reliable communication against multiple failures in wireless sensor networks. Moreover, the energy of sensors depends on the battery mainly, which is limited and difficult to replace or recharge. Therefore, a data forwarding scheme termed dynamic energy-based relaying proposed in [16] tries to forward data packets toward these sensor nodes which are closer to the data sink with optimal distance.

Besides, QoS requirements [17] play a major part in relay selection mechanism for real-time multimedia applications since they can be applied to find relay nodes under application's diversity requirements.

As we know, the static scheme has several drawbacks. That means that the data packets will be treated with the same way, which may introduce a great inefficiency due to the characteristics of ARQ, FEC, and HARQ. Hence, an adaptive error control mechanism was proposed in [18], which can improve the performance of wireless sensor network by scheduling the optimal error control protocols according to the feature of distance between sending node and next-hop receiving node with the energy efficiency.

Existing relay selection protocols do not have explicit consideration of cross-layer design and interactions, as well as combination of adaptive error control schemes and QoS diversity guarantees for different application services. This paper proposes an efficient and robust QoS supported adaptive relay selection protocol (ECA-Q-RS) for service diversity in wireless distributed sensor networks. The purpose is to explore a relay selection protocol with satisfying QoS diversity. The proposed ECA-Q-RS introduces a series of improved error control mechanisms. First, we present a Markov chain model based on three different states of the wireless data packets and two different error control algorithms. Second, a reliability aware QoS analysis model to detect the performance is suggested. The aim is to estimate whether we should update the error control scheme by the threshold of distance or signal-to-noise ratio (SNR) or not. Finally, considering the characteristics of QoS diversity requirement, we introduce an adaptive QoS supported strategy into relay selection protocol, based on cross-layer design, to alleviate the congestion and optimize the limited resource allocation.

3. System Model

In this section we present the wireless distributed sensor network model which is composed of Mica2 sensor nodes, which are comprised of ATmega128L processor and CC1000 radio module. In order to analyze the variation characteristics of SNR between hops, the log-distance path loss model is used, which is able to approximate the effects of the signal propagating through the wireless channel. In this model, d denotes the distance between the sending node and receiving node. The received power of transmitter node with d is given by

$$P_r(d) = P_t - P(d_0) - 10\beta lg\left(\frac{d}{d_0}\right). \tag{1}$$

Here, let β denote the path loss exponent, and assume that it is set to 3. The close-in reference distance d_0 is determined by the physical attribution of the sensor node. And the SNR at the receiver node can be calculated as

$$\mathrm{SNR} = P_r(d) - P_n, \tag{2}$$

where P_n is the noise power and Mica2 node is implemented with no coherent FSK modulation scheme. So the bit error rate of the wireless channel is given by

$$P_b = \frac{1}{2}e^{\mathrm{SNR}(B_N/2R_{\mathrm{radio}})}, \tag{3}$$

where B_N is the noise bandwidth and the data rate of CC1000 is set to R_{radio}.

In this network model, HARQ or ARQ scheme is applied in sending nodes, receiving nodes, and relay nodes. Let P_{ARQ} denote the packet error rate with ARQ, which could be derived as follows

$$P_{\mathrm{ARQ}} = 1 - (1 - P_b)^{l_{\mathrm{DATA}}}, \tag{4}$$

where the sum of data link layer frame header length and frame check sequence size is α and the packet length of ACK is l_{ACK}, so the data packet size l_{DATA} can be given as α plus l_{ACK}. Here, the value of l_{DATA} and l_{ACK} could be obtained according to the MAC protocol in sensor node.

Furthermore, let P_{HARQ} denote the packet error rate with HARQ, which could be obtained by

$$P_{\mathrm{HARQ}} = 1 - \left(\sum_{l=K}^{M}\binom{M}{l}(1 - P_b)^l {P_b}^{M-l}\right)^{l_{\mathrm{DATA}}}. \tag{5}$$

4. Reliability Aware QoS Analysis Model

4.1. Markov Chain Model. When the packet has been sent into the wireless channel, it may be in three states: (1) received state, (2) drop state, or (3) aware state. Let P_{EC} denote the packet error rate, and let $N_{\max}$ be the maximum retransmission time. N is called retransmission time. Any packet would be in the aware state if it has been sent by the sender node. In particular, there are two error control mechanisms, HARQ and ARQ, in our work. When HARQ is used,

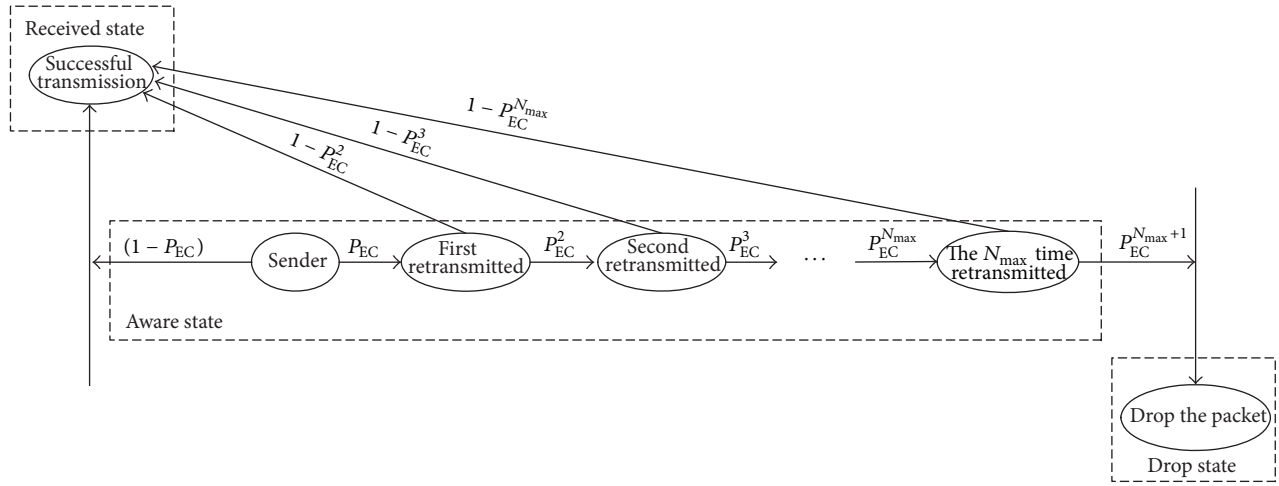


Figure 1: Markov chain model for the error control scheme.

N is set to N_{HARQ}. When ARQ is used, N is set to N_{ARQ}. Because the stochastic processes of outage of wireless channel and packet error are independent each other, the future outage of wireless link has nothing to do with the past damaged packets. Hence, it is able to model the process with the discrete-time Markov chain depicted in Figure 1.

In this Markov chain, the nonnull one-step transition probabilities from aware state into received state are listed as follows:

$$P\{(\text{Aware state}, N) \longrightarrow \text{Received state}\} = 1 - P_{\mathrm{EC}}, \quad N = 0, \tag{6}$$

$$P\{(\text{Aware state}, N) \longrightarrow \text{Received state}\} = 1 - P_{\mathrm{EC}}^{N}, \quad 0 < N \le N_{\max}. \tag{7}$$

The probability of packet retransmitted by the relay nodes is given by following formula:

$$P\{(\text{Aware state}, N) \longrightarrow (\text{Aware state}, N+1)\} = P_{\mathrm{EC}}^{N}, \quad 0 < N \le N_{\max}. \tag{8}$$

For the transition probabilities from other states into drop state, we have

$$P\{(\text{Aware state}, N_{\max}) \longrightarrow \text{Drop state}\} = P_{\mathrm{EC}}^{N_{\max}+1}, \quad 0 < N \le N_{\max}. \tag{9}$$

When the physical characteristics, distance, and the value of $N_{\max}$ are all known, we can calculate the transition probabilities between three states, which are unique. Therefore, the analytical model of saturation throughput, packet dropping rate, average delay, and energy efficiency would be given and discussed in Section 4.2.

4.2. QoS Analysis Model. Let S_{EC} denote the normalized saturation throughput, which is the ratio of the payload length to all the lengths of data packets after these packets have been retransmitted at the N time successfully. So, we have

$$\begin{aligned} S_{\mathrm{EC}} &= \frac{l_{\mathrm{payload}}}{l_{\mathrm{DATA}}}\left((1 - P_{\mathrm{EC}}) + P_{\mathrm{EC}}(1 - P_{\mathrm{EC}}) + P_{\mathrm{EC}}^{2}(1 - P_{\mathrm{EC}}) + \cdots + P_{\mathrm{EC}}^{N_{\max}}(1 - P_{\mathrm{EC}})\right) \\ &= \frac{l_{\mathrm{payload}}}{l_{\mathrm{DATA}}}\left(1 - P_{\mathrm{EC}}^{N_{\max}+1}\right). \end{aligned} \tag{10}$$

Here, l_{payload} is the length of payload. As shown in Figure 1, the data packet would be dropped only when N is larger than $N_{\max}$. Consequently, the packet dropping rate after the error control process could be achieved by the following equation:

$$P_{\mathrm{EC}} = P_{\mathrm{EC}}^{N_{\max}+1}. \tag{11}$$

The average retransmission time $N_{\text{avg-EC}}$ can hence be given as

$$N_{\text{avg-EC}} = P_{\mathrm{EC}} + P_{\mathrm{EC}}^{2} + \cdots + P_{\mathrm{EC}}^{N_{\max}}. \tag{12}$$

According to (11), we can obtain the solution of average delay, which is shown as (12):

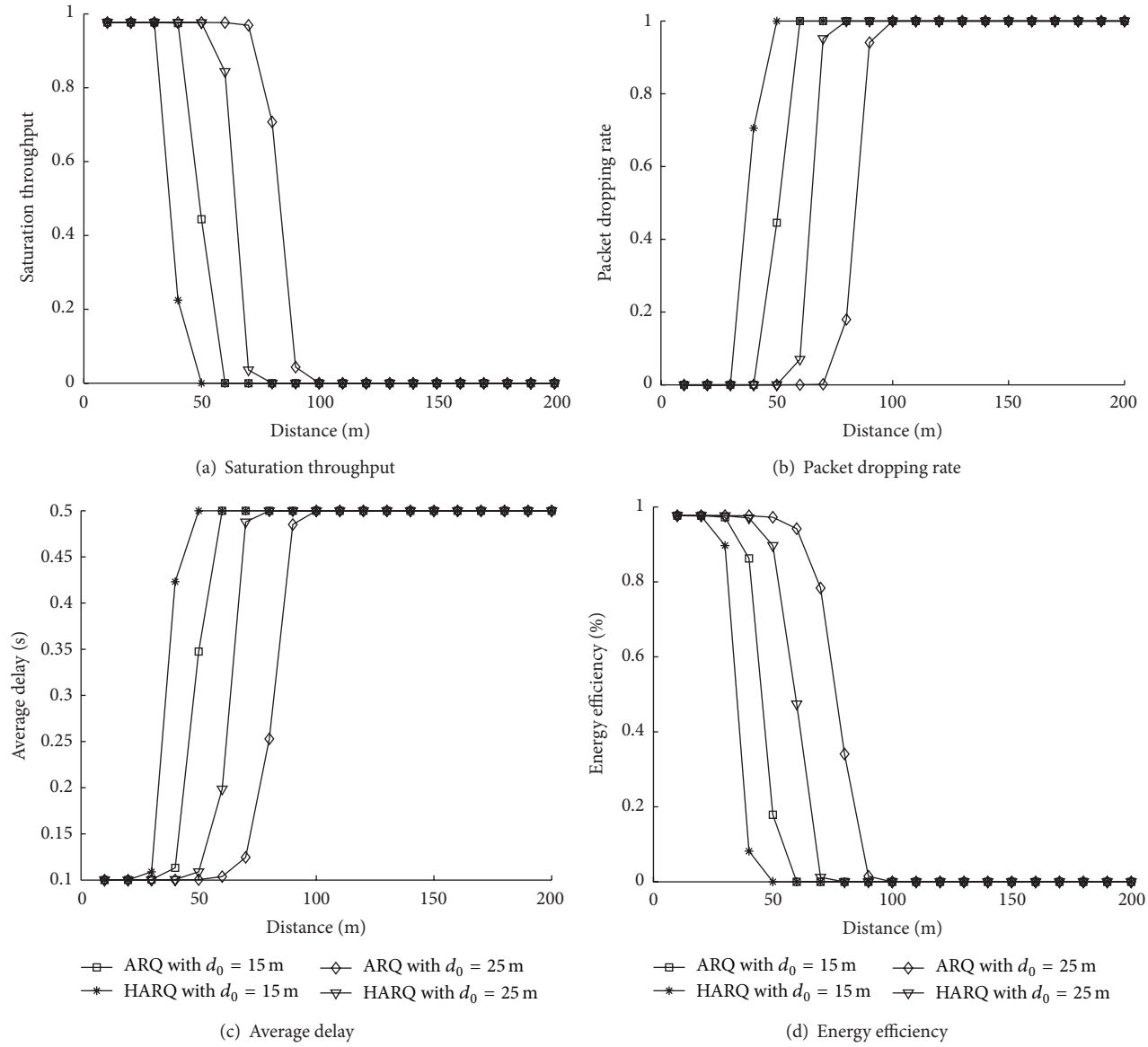
$$T_{\mathrm{ARQ}} = T\frac{1 - P_{\mathrm{EC}}^{N_{\max}+1}}{1 - P_{\mathrm{EC}}}. \tag{13}$$

Here, let T denote the round trip time of one packet. The energy efficiency η is defined as (14), which is a suitable metric that could capture the energy and reliability constraints:

$$\eta = \frac{E_{\mathrm{effi}}}{E_{\mathrm{total}}}(1 - P). \tag{14}$$

η means the ratio of the energy of payload E_{effi} to the total energy consumption E_{total}. Therefore, the energy efficiency of error control scheme η_{EC} is given as

$$\eta_{\mathrm{EC}} = \frac{E_{\mathrm{effi}}^{\mathrm{ARQ}}}{E_{\mathrm{total}}^{\mathrm{ARQ}}}(1 - P_{\mathrm{EC}}) = \frac{l_{\mathrm{payload}}}{l_{\mathrm{DATA}}}(1 - P_{\mathrm{EC}}). \tag{15}$$



(a) Saturation throughput

(b) Packet dropping rate

(c) Average delay

(d) Energy efficiency

Figure 2: Performance analyses of ARQ and HARQ with different distances and d_0.

4.3. Results and Discussion. In our previous research achievements [18], the performance of ARQ or HARQ scheme has a close relationship with distance and d_0 at first. Then, the variation characteristic of ARQ or HARQ scheme performance is presented by mathematical analyses in different SNR. All the above evaluations are based on the equations from (10) to (15).

On the one hand, Figure 2 shows four performance metrics as a function of distance. We only consider two different d_0, which are 15 m and 25 m. As illustrated in Figure 2, because the channel state is perfect and the frame error rate is extremely small, the performance of ARQ scheme is maintained relatively well with the increasing of the distance. However, the saturation throughput and energy efficiency of ARQ decrease to zero rapidly. Meanwhile, the packet dropping rate and average delay of this scheme increase quickly when the distance between sender node and next-hop receiver node is greater than 40 m, and d_0 is 15 m. Hence, the distance between sender node and next-hop receiver node has a constant value dependent on variation of d_0, which is 40 m for ARQ scheme. It is clearly noticed that there is a distance threshold value of ARQ scheme, which is 40 m, when d_0 is 15 m. Likewise, the distance threshold value is 70 m, 30 m, and 60 m, respectively, for ARQ (d_0 is 25 m), HARQ (d_0 is 15 m), and HARQ (d_0 is 25 m). Above all, there is always constant distance threshold value of different error control mechanisms with different d_0. Consequently, we can obtain the number of relay nodes, and select the best relay node according to this result.

On the other hand, the change trend of four performance metrics as a function of SNR is shown in Figure 3. It was obvious that the performance of ARQ is maintained gradually

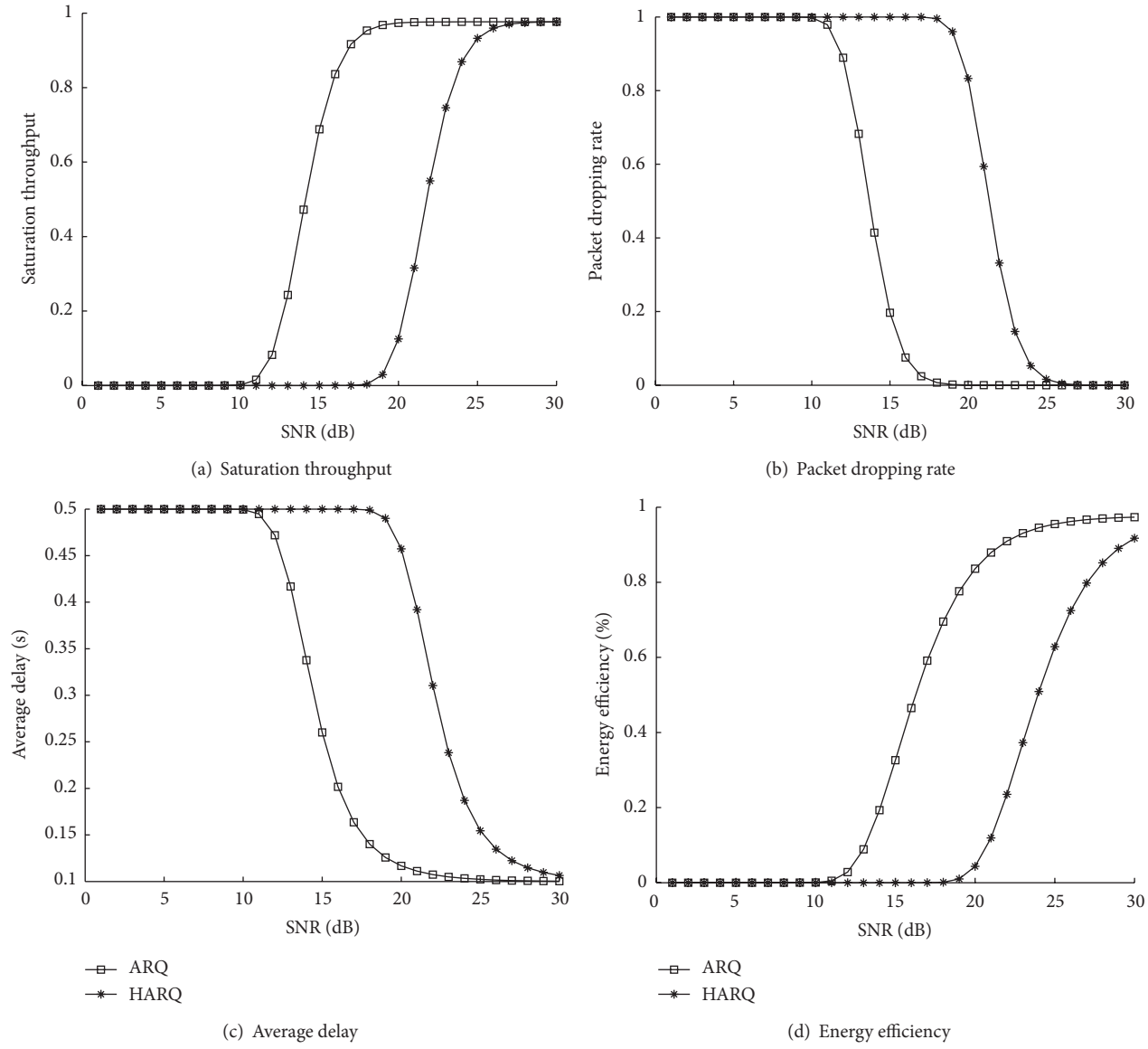


Figure 3: Performance analyses of ARQ and HARQ with SNR.

perfect with increasing of SNR. Specially, the saturation throughput and energy efficiency of ARQ increase to 1 rapidly; meanwhile, the packet dropping rate and average delay of this scheme decrease fast when the SNR between sender node and next-hop receiver node is greater than 16 dB. Hence, SNR between sender node and next-hop receiver node has a constant value, which is 16 dB for this scheme. There is apparently a SNR threshold value of ARQ scheme. Likewise, the SNR threshold value is 20 dB for HARQ. Therefore, different error control mechanisms have always one constant SNR threshold value. We can choose the best relay node as the next-hop receiving node according to this conclusion.

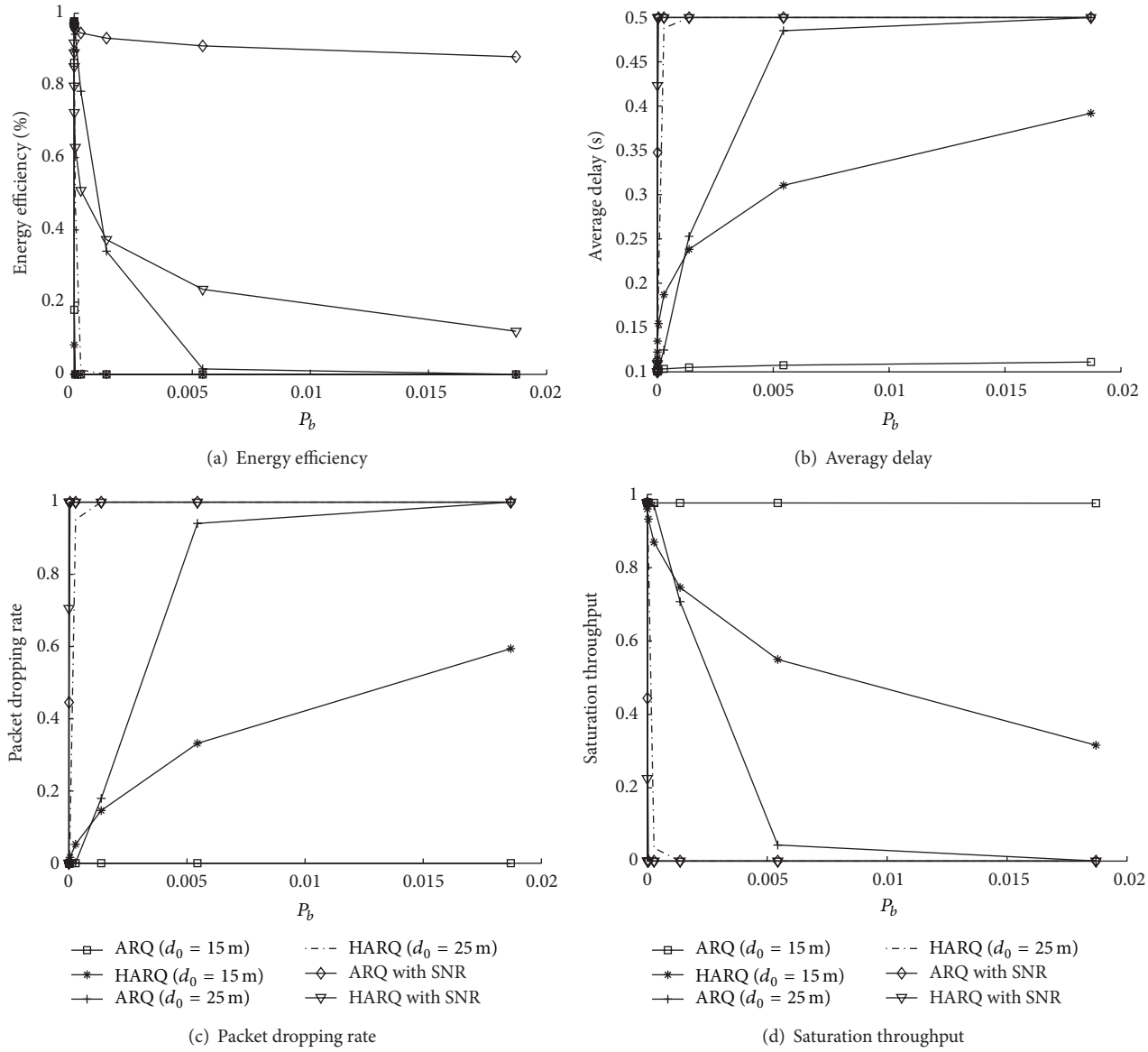
5. Diversity of QoS Supported Scheme

In this section, we define six error control schemes: (1) ARQ with distance ($d_0 = 15$), (2) ARQ with distance ($d_0 = 25$), (3) HARQ with distance ($d_0 = 15$), (4) HARQ with distance ($d_0 = 25$), (5) ARQ with SNR, and (6) HARQ with SNR. On the basis of the above QoS analysis model based on error control, the variation trend of saturation throughput, packet dropping rate, average delay, and energy efficiency is illustrated in Figure 4.

From Figure 4, we discover the relationship of QoS supported capacity with the above error control scheme, which is listed as follows.

(1) Energy efficiency: ARQ with SNR > HARQ with SNR > ARQ with distance ($d_0 = 25$) > HARQ with distance ($d_0 = 25$) > ARQ with distance ($d_0 = 15$) > HARQ with distance ($d_0 = 15$).

(2) Average delay: HARQ with distance ($d_0 = 15$) < ARQ with distance ($d_0 = 15$) < ARQ with distance ($d_0 = 25$) < HARQ with distance ($d_0 = 25$) < HARQ with SNR < ARQ with SNR.



(a) Energy efficiency

(b) Averagy delay

(c) Packet dropping rate

(d) Saturation throughput

Figure 4: Analytical research of four performance metrics between six different error control schemes.

(3) Packet dropping rate: ARQ with SNR > HARQ with SNR > ARQ with distance (d_0 = 25) > HARQ with distance (d_0 = 25) > ARQ with distance (d_0 = 15) > HARQ with distance (d_0 = 15).

(4) Saturation throughput: ARQ with SNR > HARQ with SNR > ARQ with distance (d_0 = 25) > HARQ with distance (d_0 = 25) > ARQ with distance (d_0 = 15) > HARQ with distance (d_0 = 15).

We found out that the trend of saturation throughput, packet dropping rate, and energy efficiency is quite opposite to the one of average delay. Therefore, the application service could select the optimal QoS supported scheme to satisfy the diversity requirement from Table 1. Here, QoS scheme I is composed of the guarantee ability of saturation throughput, packet dropping rate, and energy efficiency. In addition, QoS scheme II is comprised of real-time guarantee ability. Particularly, the more the digit of the error control scheme in Table 1, the better its guarantee capacity.

Table 1: QoS supported scheme.

Error control scheme	Scheme I	Scheme II
ARQ with SNR	6	1
HARQ with SNR	5	2
ARQ with distance (d_0 = 25)	4	3
HARQ with distance (d_0 = 25)	3	4
ARQ with distance (d_0 = 15)	2	5
HARQ with distance (d_0 = 15)	1	6

6. High Reliable Relay Selection Approach

6.1. Relay Selection Algorithm. If we can obtain the end to end distance d, the optimal relay node should be selected according to distance threshold, and also the number of relay nodes is given, which are as shown in the following equations:

$$d = \begin{cases} D_{\mathrm{TE}}, & d < D_{\mathrm{TE}},\ H_{\max} = 1, \\ H_{\max} D_{\mathrm{TE}}, & d \bmod D_{\mathrm{TE}} = 0,\ H_{\max} = \dfrac{d}{D_{\mathrm{TE}}}, \\ H_{\max} D_{\mathrm{TE}}, & d \bmod D_{\mathrm{TE}} \neq 0,\ H_{\max} = \left[\dfrac{d}{D_{\mathrm{TE}}}\right] + 1, \end{cases} \tag{16}$$

where D_{TE} is the distance threshold value; let mod denote the calculation of solving the modulus operator. $[d/D_{\mathrm{TE}}]$ indicates the rounding calculation. That means that the integer part is intercepted and the decimal fraction is dropped. The cooperative hops $H_{\max}$ can be obtained if d is known; the number of relay nodes is $H_{\max}$ minus 1 simultaneously.

In a word, the work flow of relay selection based on distance is as follows.

(1) To obtain the value of d_0 with sensor node and d according to the actual measurement.

(2) Calculate D_{TE} according to the error control aware analytical model.

(3) Get $H_{\max}$ and the number of relay nodes on the basis of (16).

(4) If the distance is less than or equal to D_{TE}, it would be selected as the relay node.

If we can obtain the SNR of wireless channel, the optimal relay node should be chosen according to SNR threshold. The process of relay selection based on SNR is as follows.

(1) The SNR of the wireless channel is acquired with real-time detection.

(2) Acquire the SNR threshold $\mathrm{SNR}_{\mathrm{TE}}$ based on the QoS analytical model.

(3) If the SNR of one sensor node is less than or equal to $\mathrm{SNR}_{\mathrm{TE}}$, it would be selected as the relay node.

Particularly, relay selection based on SNR should be used first when the wireless channel is stable and perfect; otherwise, relay selection based on distance has to be employed promptly.

6.2. Implementation of Adaptive Relay Selection. In this subsection, the reliability aware and diversity of QoS supported adaptive relay selection mechanism (ECA-Q-RS) is proposed, which is implemented in wireless distributed sensor network. Because the relay node is selected based on characteristic of error control in ECA-Q-RS, the performance of the proposed mechanism can be evaluated by the following formula:

$$Q_{\text{ECA-Q-RS}} = \sum_{t=1}^{H_{\max}} Q_{\mathrm{C}}(t) + q, \tag{17}$$

where $Q_{\text{ECA-Q-RS}}$ denotes the performance metrics of ECA-Q-RS, which include saturation throughput, packet dropping rate, average delay, and energy efficiency. Let $Q_{\mathrm{C}}(t)$ denote the above four performance metrics when d is larger than D_{TE} or SNR is greater than $\mathrm{SNR}_{\mathrm{TE}}$. q is used to record these metrics when d is less than D_{TE} or SNR is less than $\mathrm{SNR}_{\mathrm{TE}}$.

Afterwards, we present the basic idea of the ECA-Q-RS and its implementation at sender node, receiver node, and relay nodes in detail, which is illustrated as follows.

Sender Node

(1) Carry out the QoS supported scheme. The guarantee priority of saturation throughput, packet error rate, and energy efficiency, or average delay is appointed according to the requirement of application service.

(2) At the data link layer, the optimal error control and QoS supported scheme are chosen on the basis of Table 1. Moreover, the value of D_{TE} and $H_{\max}$ is obtained based on the QoS analytical model.

(3) Relay selection mechanism based on distance is implemented when d and d_0 are known, or the wireless channel is unstable and poor. Relay selection mechanism based on SNR is implemented when channel state information is known, or the wireless link is perfect.

(4) Starting to send data packets. If ACK packet is received, new data packets are sent, and retransmission time-out is used for each packet at the same time.

(5) When the timer matures, or NACK packet is received, the optimal error control scheme is started according to the QoS scheme.

Relay Node

(6) Select the optimal relay node from candidate nodes based on distance or SNR threshold.

(7) Steps (3), (4), and (5) are implemented repeatedly in proper order until the data packet is received successfully or discarded actively.

Receiver Node

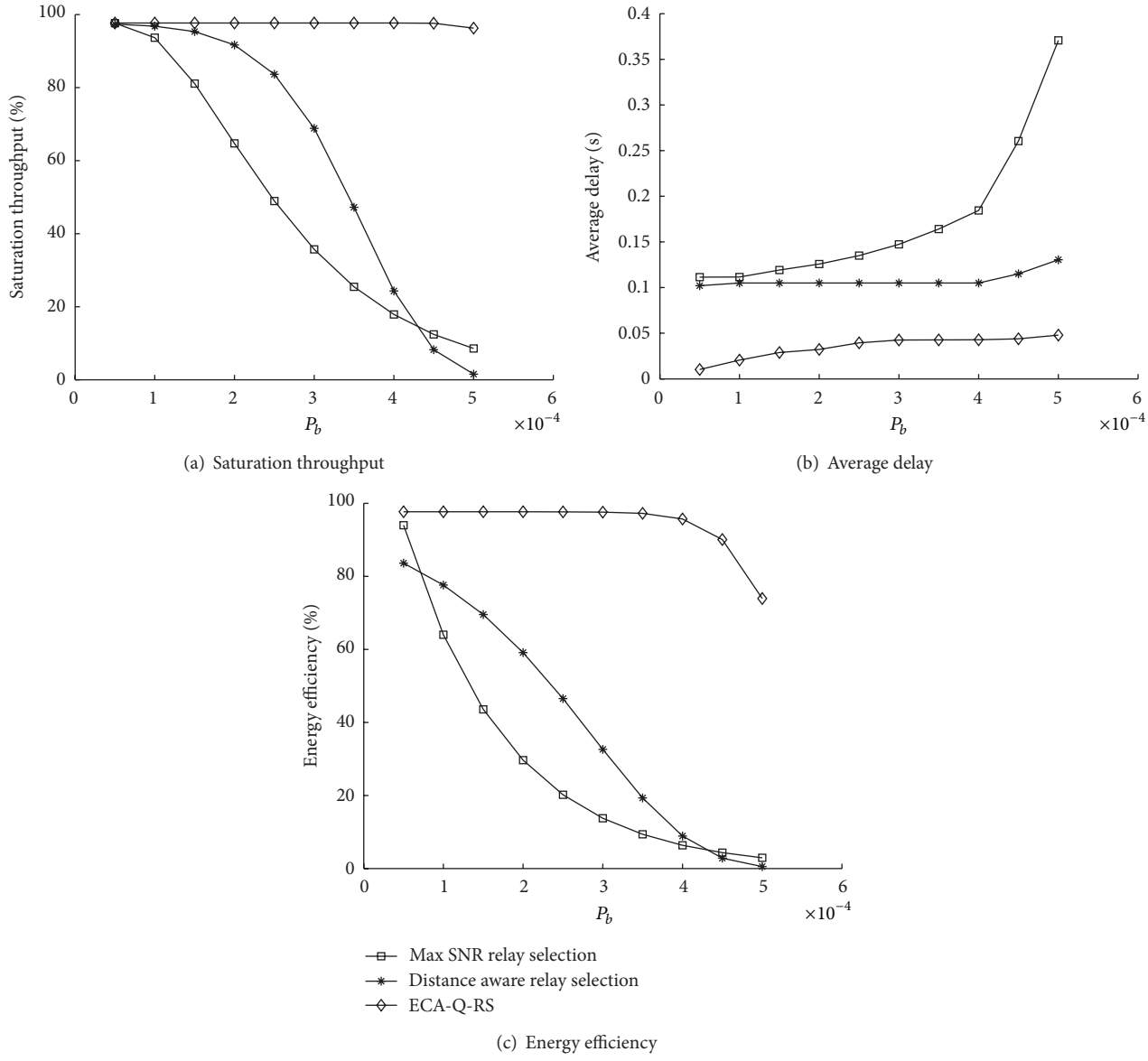
(8) If ARQ scheme is used, checksum is calculated and tested. If HARQ scheme is used, checksum testing and FEC encoding are implemented.

(9) If the result obtained is right, the data packet is accepted, and ACK packet is sent simultaneously; otherwise, it is rejected, and NACK packet is sent at the same time.

(10) Deliver the correct data packet to the upper layer.

7. Performance Evaluation

In this work, we use NS-2 and VC++6.0 to simulate, analyze and evaluate the performance of ordinary data transmission



(a) Saturation throughput

(b) Average delay

(c) Energy efficiency

Figure 5: Performance with varying channel state (P_b).

and multimedia streaming using ECA-Q-RS, compared with max SNR relay selection and distance aware relay selection through two group experiments. The experimental data is the average value after 100-time simulation and mathematical analysis.

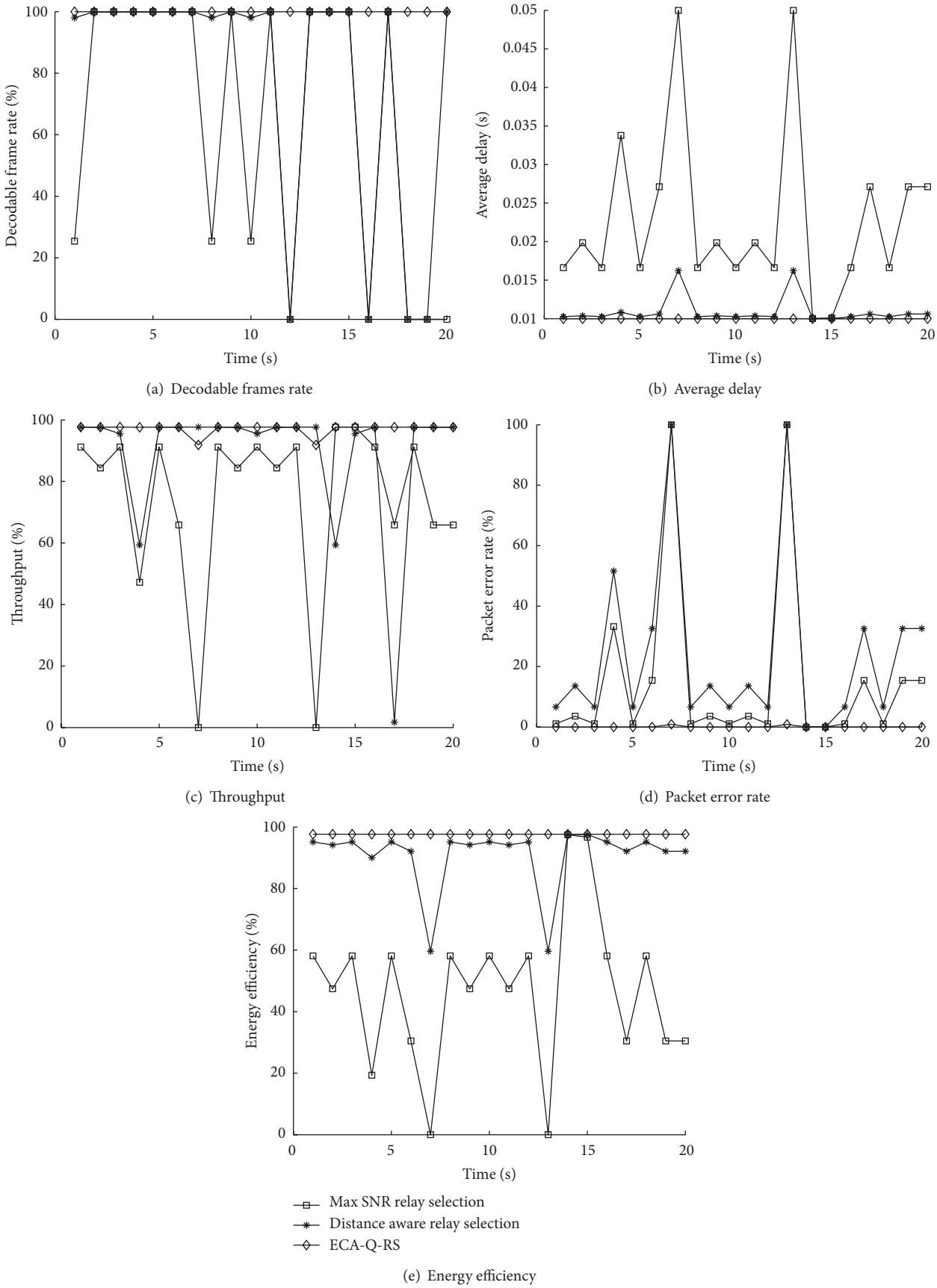
7.1. Simulation Parameters. In experiment 1, 50 sensor nodes move in an 800 m × 800 m rectangular region. α is 11 bytes. l_{ACK} is set to 7 bytes. Noise bandwidth B_N is 30 kHz. The data rate of CC1000 R_{radio} is equal to 38.4 kBaud. Round trip time of one frame transmission at data link layer T is 0.01 second. The mobility model is the random waypoint model. Each sensor node moves towards the destination node at a speed uniformly chosen between the minimal speed 0 m/s and maximal speed 10 m/s. The MAC protocol is IEEE 802.15.4 protocol. We change d_0 from 15 m to 30 m to investigate the impact of node physical attribution on performance.

For multimedia traffic of experiment 2, we use a medium quality MPEG4 video clip from the movie forman_qcif.yuv, which consists of 400 video frames. The structure of the group of pictures is IBBPBBPBBPBB ($N = 12$, $M = 3$). The video frame rate is 25 frames per second. The maximum transmission unit is 1024 bytes.

7.2. Simulation Results. Two case studies are designed and conducted, with the variation of bit error rate P_b and the simulation time, respectively. The main metrics for performance evaluation of the ECA-Q-RS for supporting the application services of QoS diversity are listed as follows.

(1) Decodable frames rate: it considers the dependency between different MPEG4 video frames.

(2) Average delay: we only calculate the end to end delay of the decodable frames.



(a) Decodable frames rate

(b) Average delay

(c) Throughput

(d) Packet error rate

(e) Energy efficiency

Figure 6: Performance with multimedia communication.

(3) Throughput: the total size of data packets received successfully by the receiver node.

(4) Energy efficiency: the number of obtained decodable video frames per unit of energy consumption.

7.2.1. Experimental Results with Channel State. Figure 5 shows three performance metrics as a function of P_b in experiment 1. The result indicates that channel state has a significant impact on quality of data transmission in wireless sensor networks. We can observe tremendous improvement of performance using ECA-Q-RS, as compared with other two mechanisms. Even at a more poor state of the wireless channel, the quality using ECA-Q-RS is maintained in a good condition, while the saturation throughput and energy efficiency using max SNR relay selection and distance aware relay selection reduce and become close to zero gradually. The reasons are that not only the supported mechanism of QoS diversity minimizes the average delay but also the adaptive error control strategy improves the energy efficiency; as a result, the network throughput utilization increased.

7.2.2. Experimental Results with QoS Supported Capacity. Figure 6 shows the five performance metrics as a function of different relay selection schemes in experiment 2.

Figure 6(a) shows the result of the decodable frame rate. As the transmission rate of multimedia data increases, the collision probability of data packets transmission increases significantly, leading to an unstable and dynamic decreasing tendency of decodable frame rate. The result demonstrates tremendous improvement of decodable frames rate with ECA-Q-RS as compared with other mechanisms. On this basis, we determine that ECA-Q-RS can prime accommodate the poor and dynamic wireless network environment. This obvious improvement depends on their stable relay selection scheme based on QoS diversity supported mechanism and adaptive error control strategy.

The real-time performance is shown in Figure 6(b). The average delay using ECA-Q-RS achieves a tremendous improvement. The main reason is that the error control aware adaptive relay selection strategy enhances the bandwidth and frequency spectrum utilization, as well as propagation path. As shown in Figure 6(c), the ECA-Q-RS always performs better than the other two mechanisms. The throughput gain we have is in the improvement of hop-by-hop throughput and reduction in bandwidth consumption.

Figure 6(d) provides packet error rate for each video frame in different mechanisms. When using ECA-Q-RS, the packet error rate fluctuates a little, and the average is the lowest. These adaptive schemes in ECA-Q-RS are not only able to achieve higher decodable frame rate and network throughput but also improve the stability of the video transmission. Figure 6(e) shows the energy utilization efficiency of different mechanisms. The enhancement of data exchange gain and adaptive error control can greatly reduce energy consumption for data transmission. Comparing with the other two schemes, ECA-Q-RS achieves significant improvement in energy utilization efficiency.

8. Conclusions and Future Work

There is a great potential for relay selection and error control to enhance the performance of QoS diversity services and resource utilization. The purpose of this work is to overcome all kinds of limitations; we propose a reliability aware and QoS supported relay selection (ECA-Q-RS).

The main contributions in our work are as follows. First, considering the characteristics of different error control schemes, we introduce the Markov chain model based on ARQ and HARQ to study the characteristics of QoS. Second, a QoS supported strategy with dynamic priority is designed to ensure that packets have a better chance to be forwarded. Finally, we present adaptive relay selection algorithm, working together with the above QoS supported mechanism, in order to reduce the ratio of potential damaged or lost opportunities.

The mathematics and simulation results demonstrate that, compared with the existing typical relay selection mechanisms, ECA-Q-RS greatly improves the data transmission quality and achieves significant gains in terms of throughput and energy efficiency. As a result, we determine that the proposed mechanism is feasible for data diversity communication in wireless distributed sensor network.

Our future work focuses on the implementation and validation of the proposed adaptive relay selection algorithms on the prototype system. Besides, one possible future direction is to study the channel prediction approach and the tradeoff between performance and complexity.

Conflict of Interests

The authers declare that they have no financial and personal relationships with other people or organizations that can inappropriately influence their work; there is no professional or other personal interest of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in, or the review of the paper.

References

[1] F. Xia, "QoS challenges and opportunities in wireless sensor/actuator networks," *Sensors*, vol. 8, no. 2, pp. 1099–1110, 2008.

[2] G. W. Bai, K. Oladosu, and C. Williamson, "Performance benchmarking of wireless web servers," *Ad Hoc Networks*, vol. 5, no. 3, pp. 392–412, 2007.

[3] G. W. Bai, J. J. Tao, and H. Shen, "A link-lifetime based dynamic source routing protocol (LTDSR) for multimedia over MANETs," *Journal of the Chinese Institute of Engineers*, vol. 33, no. 5, pp. 761–768, 2010.

[4] H. Shen, G. W. Bai, L. Zhao, and Z. M. Tang, "An adaptive opportunistic network coding mechanism in wireless multimedia sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 565604, 13 pages, 2012.

[5] D. S. Michalopoulos, H. A. Suraweera, G. K. Karagiannidis et al., "Amplify-and-forward relay selection with outdated channel estimates," *IEEE Transactions on Communications*, vol. 60, no. 5, pp. 1278–1290, 2012.

[6] Y. B. Li, "On the design of relay selection strategies in regenerative cooperative networks with outdated CSI," *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 3086–3097, 2011.

[7] V. N. Q. Bao and H. Y. Kong, "Performance analysis of multihop decode-and-forward relaying with selection combining," *Journal of Communications and Networks*, vol. 12, no. 6, pp. 616–623, 2010.

[8] K. Miranda, E. Natalizio, and T. Razafindralambo, "Adaptive deployment scheme for mobile relays in substitution networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 128904, 9 pages, 2012.

[9] G. Fairhurst and L. Wood, "Advice to link designers on link Automatic Repeat reQuest (ARQ)," RFC 3366, IETF, August 2002.

[10] M. Watson, M. Luby, and L. Vicisano, "Forward Error Correction (FEC) Building Block," RFC 5052, IETF, August 2007.

[11] D. Chase, "Code Combining: a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 385–393, 1985.

[12] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, "Energy efficiency of non-collaborative and collaborative Hybrid-ARQ protocols," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 326–335, 2009.

[13] C. K. Lo, R. W. Heath, and S. Vishwanath, "The impact of channel feedback on opportunistic relay selection for hybrid-ARQ in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1255–1268, 2009.

[14] Y. Jin and G. W. Bai, "A cooperative FEC based on the model of GM (1, 1) and IPv6 for wireless multimedia sensor networks," *Journal of Convergence Information Technology*, vol. 7, no. 18, pp. 230–239, 2012.

[15] J. Wang, X. M. Wang, and S. K. Zhang, "An efficient reliable communication scheme in wireless sensor networks using linear network coding," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 605494, 11 pages, 2012.

[16] J. Deng, Scott C.-H. Huang, and Y. S. Han, "An online relay selection scheme in power controllable wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 213598, 13 pages, 2012.

[17] S.-R. Cho, C. Wan, and K. B. Huang, "QoS provisioning relay selection in random relay networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 6, pp. 2680–2689, 2011.

[18] Y. Jin, D.-G. Le, G.-W. Bai, and J.-Y. Chang, "Adaptive error control scheme for wireless sensor networks based on hops and communication distance," *Control Theory and Applications*, vol. 28, no. 4, pp. 596–600, 2011.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 846791, 6 pages
http://dx.doi.org/10.1155/2013/846791

Research Article

A Novel PROMSIS Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks

Shengmei Liu, Zhongjiu Zheng, and Su Pan

Broadband Wireless Communication and Sensor Network Technology Key Laboratory of Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Correspondence should be addressed to Shengmei Liu; smliu@njupt.edu.cn

Received 17 July 2013; Revised 16 August 2013; Accepted 18 August 2013

Academic Editor: Gelan Yang

A novel preference ranking organization method by similarity to ideal solution (PROMSIS) vertical handoff algorithm is proposed for heterogeneous wireless networks, and its essential idea includes the preference structure of the PROMETHEE and the concept of Euclid distance of the TOPSIS. Four 3GPP defined traffic classes are considered in performance evaluation. An attribute matrix is constructed considering some major attributes. Handoff decision meeting multiattribute QoS requirement is made according to the traffic features. The weight relation of decision elements is determined with the least square (LS) approach. The final decision is made using the proposed PROMSIS algorithm based on the attribute matrix and weight vector. The simulation results have manifested that the proposed PROMSIS algorithm can provide satisfactory vertical handoff performance, and the LS-PROMSIS algorithm can be fit to the characteristics of the traffic.

1. Introduction

The architecture of the beyond 3rd generation (B3G) or 4th generation (4G) wireless networks aims at integrating various heterogeneous wireless access networks over an IP based backbone. To provide seamless mobility, one of the design issues is the vertical handoff support [1, 2], a multiple attributes decision subject. Since the handoff may happen in different RATs and management domains, handoff decision will depend on the combination of multiple attributes rather than a single parameter.

In general, the vertical handoff process can be divided into three main steps, namely, system discovery, handoff decision, and handoff execution. During the phase of system discovery, the networks may advertise the supported data rates and quality-of-service (QoS) parameters for different services. Because the users are mobile, the available collocated networks depend on the location of the user. The traffic load in each network may also change with time. Thus, this phase may be periodically invoked.

Various vertical handoff decision mechanisms have been proposed. In [3], a combining SINR based vertical handoff (CSVH) algorithm is proposed. It studies the combined effects of SINR in different access networks, that is, in the source network and the equivalents in the target networks, compared with the RSS based vertical handoff algorithm. Further on, a multidimensional adaptive SINR based vertical handoff (MASVH) algorithm is proposed in [4]. In addition to the combined effects of SINR, it also takes account of the user required bandwidth, traffic cost, and resource utilization in the participating access networks. A parameter k is used in the MASVH algorithm to adjust the weight of multiple attributes. Nevertheless, no discussion elaborates on the determination of the optimal k value under different conditions, the relation between multiple attributes, the relative importance of each attribute, and the impact of system load. There are also some researchers focusing on solving the ping-pong effect. In [5], the user movement information is considered, and the residence time in a base station is estimated to avoid the unnecessary handoff.

In this paper, a novel algorithm, namely, the PROMSIS algorithm, is proposed to be applied in the vertical handoff decision technology based on the preference ranking organization method for enrichment evaluation (PROMETHEE) [6, 7] and the TOPSIS [8]. In TOPSIS, the level of the decision maker's participation is rather low in the process

of decision making, and the decision maker's preference information is not integrated into the method. So, we introduce the preference function associated with each criterion in PROMETHEE, integrate the preference structure of PROMETHEE into TOPSIS, and obtain PROMSIS as a result. The scenario analyzed is referred to in [4]. It considers multiple attributes, concluding in the combined effects of SINR in WLAN and WCDMA, the required bandwidth, service cost, and available bandwidth of the participating access networks, to make handoff decisions meeting multiattribute QoS requirement. An attribute matrix of alternative networks is established. An appropriate weight factor is assigned to each criterion to account for its importance. In the weight determining process, four 3GPP defined traffic classes [8] are considered and the least square (LS) weighted approach method [9] is adopted. Finally, how the connections are contained or rerouted is decided by the PROMSIS (or LS-PROMSIS) algorithm according to the attribute matrix and the weight vector.

2. PROMSIS and LS-PROMSIS Vertical Handoff Algorithm

The handoff metrics and QoS parameters are categorized into different groups (e.g., bandwidth, latency, power, price, security, reliability, availability, etc.). Some representative metrics approaches are considered in this paper.

Assuming that there are a BSs and b Aps, all candidate BSs and APs for the user can be indexed by 1 to $a + b$ in the set

$$\mathbf{P} = \left[\mathrm{BS}_1, \mathrm{BS}_2, \ldots, \mathrm{BS}_a; \mathrm{AP}_1, \mathrm{AP}_2, \ldots, \mathrm{AP}_b\right]. \tag{1}$$

For each handoff event, the best BS or AP from the candidate set $\mathbf{P}$ for each user will be determined by the handoff algorithm considering the following criteria: SINR, the required bandwidth, traffic cost, and network available bandwidth.

2.1. Attribute Matrix. Let us presume R_{AP} and R_{BS} as the maximum achievable downlink data rates of WLAN and WCDMA. According to Shannon capacity, we have

$$\begin{aligned} R_{\mathrm{AP}} &= W_{\mathrm{AP}} \log_2\left(1 + \frac{\gamma_{\mathrm{AP}}}{\Gamma_{\mathrm{AP}}}\right), \\ R_{\mathrm{BS}} &= W_{\mathrm{BS}} \log_2\left(1 + \frac{\gamma_{\mathrm{BS}}}{\Gamma_{\mathrm{BS}}}\right), \end{aligned} \tag{2}$$

where γ_{AP} and γ_{BS} are the receiving SINR values from the coexisting networks. When the networks offer the same downlink data rate to the user; that is, $R_{\mathrm{AP}} = R_{\mathrm{BS}}$, we can solve the equation and get the relationship between γ_{AP} and γ_{BS} as

$$\gamma_{\mathrm{BS}} = \Gamma_{\mathrm{BS}}\left(\left(1 + \frac{\gamma_{\mathrm{AP}}}{\Gamma_{\mathrm{AP}}}\right)^{W_{\mathrm{AP}}/W_{\mathrm{BS}}} - 1\right), \tag{3}$$

where the carrier bandwidth is 22 MHz for WLAN W_{AP} and 5 MHz for WCDMA W_{BS}. Γ_{AP} is equal to 3 dB for WLAN, and Γ_{BS} is equal to 16 dB for WCDMA.

It is assumed that a BS is transmitted to merely one user via the HSDPA channel at a time, with the maximum power to achieve the optimal physical rate. The SINR $\gamma_{\mathrm{BS}_j,i}$ received by user i from WCDMA BS_j can be represented as

$$\gamma_{\mathrm{BS}_j,i} = \frac{G_{\mathrm{BS}_j,i} P_{\mathrm{BS}_j,i}}{P_O + \sum_{k=1,k\neq j}^{m} \left(G_{\mathrm{BS}k,i} P_{\mathrm{BS}k}\right) + G_{\mathrm{BS}_j,i}\alpha\left(P_{\mathrm{BS}_j} - P_{\mathrm{BS}_j,i}\right)}, \tag{4}$$

where P_{BS_j} is the total transmitting power of BS_j, $P_{\mathrm{BS}_j,i}$ is the transmitting power of BS_j to user i, $G_{\mathrm{BS}_j,i}$ is the channel gain between user i and BS_j, α is the orthogonality factor equal to 0.4, and P_O is the thermal noise power equal to −99 dBm.

For WLAN, the SINR $\gamma_{\mathrm{AP}_j,i}$ received by user i from WLAN AP_j can be represented as

$$\gamma_{\mathrm{AP}_j,i} = \frac{G_{\mathrm{AP}_j,i} P_{\mathrm{AP}_j}}{P_B + \sum_{k=1,k\neq j}^{n} \left(G_{\mathrm{AP}k,i} P_{\mathrm{AP}k}\right)}, \tag{5}$$

where P_{AP_j} is the transmitting power of AP_j, $G_{\mathrm{AP}_j,i}$ is the channel gain between user i and AP_j, and P_B is the background noise power equal to −86 dBm.

A macrocell propagation model for urban and suburban areas [3] is adopted, and for an antenna height of 15 meters the path loss is

$$\begin{aligned} \text{Path loss (dB)} = {} & 58.8 + 21\log_{10}(f) \\ & + 37.6\log_{10}(R) + \log(F), \end{aligned} \tag{6}$$

where f is the carrier frequency (2 GHz for WCDMA and 2.4 GHz for WLAN), R is the distance in meters between the user and the BS or AP, and $\log(F)$ is the log-normal distribution shadowing with standard deviation $\sigma = 10$ dB.

Using (3), the SINR received from APs ($\mathbf{S}_{\mathrm{AP},i}$) is converted to the equivalent SINR ($\mathbf{S}'_{\mathrm{AP},i}$) to achieve the same data rate via BS.

The set of the SINR value $\mathbf{S}_i$ of all BSs and APs in the candidate set $\mathbf{P}$ for the user i can be represented by

$$\mathbf{S}_i = \mathbf{S}_{\mathrm{BS},i} \cup \mathbf{S}'_{\mathrm{AP},i}. \tag{7}$$

For a required bandwidth R_i for a user i, the minimum receiving SINR from BS ($\gamma_{\min,i}$) can be calculated in Shannon equation.

Let us suppose $\mathbf{C}$ to be the system cost vector. In order to directly associate the cost value with the SINR value, the cost per bit is converted to cost per SINR ($\mathbf{C}_{\mathrm{SINR}}$).

Let us suppose $\mathbf{U}$ as the network available bandwidth vector represented by the available capacity of each candidate BS and AP.

Then, the attribute matrix is as follows:

$$\mathbf{R}_a = \begin{bmatrix} \mathbf{S} - \gamma_{\min} \\ \dfrac{1}{\mathbf{C}_{\mathrm{SINR}}} \\ \mathbf{U} \end{bmatrix}^T = \begin{array}{c} \\ A_1 \\ A_2 \\ \vdots \\ A_m \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & \cdots & X_n \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & & \\ & \cdots & \cdots & \\ & & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \end{array} = \left(x_{ij}\right)_{m\times n}, \tag{8}$$

where $A_1, A_2, \ldots, A_m$ are feasible alternatives, $m = a+b$, and $X_1, X_2, \ldots, X_n$ are evaluation attributes, $n = 3$. Here, x_{ij} is the performance rating for alternative A_i under attribute X_j.

2.2. Handoff Decision. The proposed PROMSIS is also a multicriteria analysis approach, and its essence includes the preference structure of the PROMETHEE and the concept of Euclid distance of the TOPSIS. In the first place, it performs the comparison between every pair of solutions (a_i, a_r) using a preference function $p(d)$, and $d = f_j(a_i) - f_j(a_r)$ is the difference between the evaluations of two alternatives. Reference [7] contains many types of preference functions. This function $p(d)$ reflects the preference level of a_i over a_r in the interval [0, 1], in such a way that if $p(d) = 0$, then a_i is indifferent to a_r; if $p(d) = 1$, then a_i is strictly preferred to a_r.

PROMSIS consists of the following steps.

(1) Construct the decision matrix $\mathbf{R}_a$ and the weight vector $\mathbf{w}$.

(2) Define the preference function for each attribute.

(3) Define the preference index for each couple of alternatives:

$$n = V(a_i, a_r) = \sum_{j=1}^{n} w_j \cdot p_j\left(f_j(a_i) - f_j(a_r)\right). \tag{9}$$

The preference index is given in the intensity of preference of the decision maker for a_i over a_r. We have $0 \le V(a_i, a_r) \le 1$. The matrix $\mathbf{V} = (v_{ir})_{m\times n}$ is obtained to calculate the weighted Euclidean distances.

(4) The preference concept from PROMETHEE is presented as above, and now we use the concept of Euclid distance in TOPSIS to continue. Define the positive ideal point and the negative ideal point, and calculate the distance between each scheme and the positive/negative ideal point. Calculate the distance S_i^+ between each scheme and the positive ideal point and the distance S_i^- between each scheme and the negative ideal point:

$$S_i^+ = \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^+\right)^2}, \quad i \in m,$$
$$S_i^- = \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^-\right)^2}, \quad i \in m. \tag{10}$$

(5) Calculate the relative approach degree C_i^+ of each scheme to the ideal points:

$$C_i^+ = \frac{S_i^-}{\left(S_i^+ + S_i^-\right)}, \quad 0 < C_i^+ < 1,\ i \in m. \tag{11}$$

(6) Ranke the schemes based on C_i^+. The larger is the C_i^+, the better is the scheme.

Above all, the proposed PROMSIS combines the qualitative and quantitative analysis. Preference comparison corresponds to the qualitative analysis. The Euclidean distance can describe the degree of preference through the quantity. So, the utility of the network can be achieved in both qualitative and quantitative aspects. And the final decision will be appropriate based on subjective and objective factors.

2.3. Weight Vector. There are a variety of weight methods, such as analytic hierarchy process (AHP) and the information Entropy weight method, some are subjective and others are objective. We can choose the appropriate weight method according to actual conditions. In the weight determining process, we apply the LS [9] to estimate the weights of decision elements introduced by Chu in 1979.

Firstly, the comparison matrix $\mathbf{G}_c = (g_{ij})_{n\times n}$ is defined according to the relative importance. The judgments are ranked on a 9-point scale [7]. Numbers 1 to 9 are used to represent equal, weakly moderate, moderate, moderate plus, strong, strong plus, very strong, very very strong, and extremely important to the objective, respectively. When an element is less important than another, the comparison result equals the reciprocal of one of the numbers. So for the matrix $\mathbf{G}_c$ to be the diagonal elements are observed 1, demonstrating the elements' self-comparisons. The other entries in the matrix are symmetric with respect to the diagonal, as a result of the inverted comparisons.

Four traffic classes defined by 3GPP are taken into consideration, namely, the conversational, streaming, interactive, and background classes. Based on the traffic requirements, the comparison matrices for the four traffic classes according to the 9-point scale can be established.

The element g_{ij} of matrix $\mathbf{G}_c$ shall be considered and desired to determine the weights w_i, such that, given g_{ij}, $g_{ij} \approx w_i/w_j$.

The weights can be obtained by solving the constrained optimization problem

$$\min \quad \mathbf{S} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(g_{ij} w_j - w_i\right)^2 \tag{12}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} w_i = 1, \quad w_i > 0. \tag{13}$$

In order to minimize $\mathbf{S}$, form the sum

$$\mathbf{S}' = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(g_{ij} w_j - w_i\right)^2 + 2l \sum_{i=1}^{n} w_i, \tag{14}$$

where l is the Lagrange multiplier. Differentiating $\mathbf{S}'$ with respect to w_z $(z = 1, 2, \ldots, n)$, the following set of equations is obtained:

$$\sum_{i=1}^{n} \left(g_{iz} w_z - w_i\right) g_{iz} - \sum_{j=1}^{n} \left(g_{zj} w_z - w_z\right) + l = 0. \tag{15}$$

Equations (15) and (13) form a set of $n+1$ inhomogeneous linear equations with $n + 1$ unknowns.

By the way, using the numerical method to solve mathematical problems, due to almost inevitable rounding errors, the results obtained are generally inaccurate. Some other measures can be applied to estimate the error, but we will not go into detail here due to space limitations.

3. Simulation Results

In this research, we concentrate on the downlink traffic, since it normally requires higher bandwidth than uplink, especially for multimedia services such as video streaming through the HSDPA channel while connected to WCDMA.

The performance of different vertical handoff algorithms has been evaluated with the scenario illustrated in Figure 1, in which there are 7 BS and 12 AP placed at each WCDMA cell boundary. The WCDMA cell radius is 1200 meter. 200 randomly generated UEs are used inside the simulation area, whose position changes in the time interval depending on their moving speed and direction. The direction is uniformly distributed in the range of $[0, 2\pi]$, and the speed change rate is 5 per 100 seconds. The maximum user's moving speed is 80 km/hour. In the traffic generator module, for a mean session duration of 60 seconds and a certain given mean session arrival rate, user traffic is randomly generated with a Poisson arrival distribution. As revealed in Figure 1, the direction of the arrow represents the user's moving direction. The length of the arrow corresponds to the moving distance of the user in 20 seconds (supposing that the moving direction is fixed in the 20 seconds), so the larger one indicates the faster moving speed.

The V-shape with indifference criterion type preference function in PROMETHEE was adopted here. In case of this type, the thresholds of indifference q $(q = 0.1)$ and strict preference p $(p = 0.5)$ have to be separately selected.

The system performance for different session arrival rates is shown in Figure 2. The simulated algorithms include

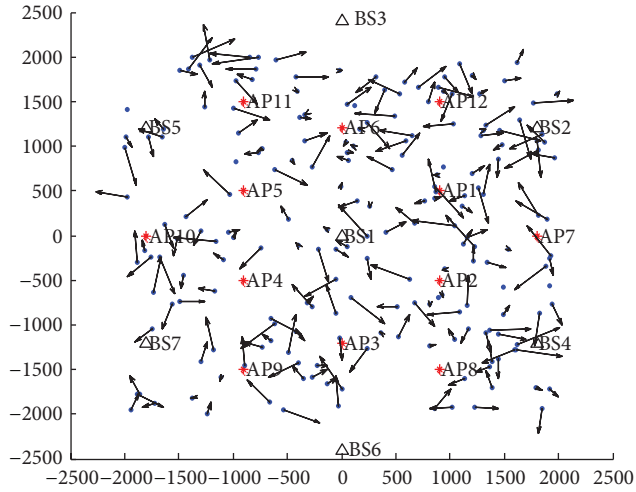
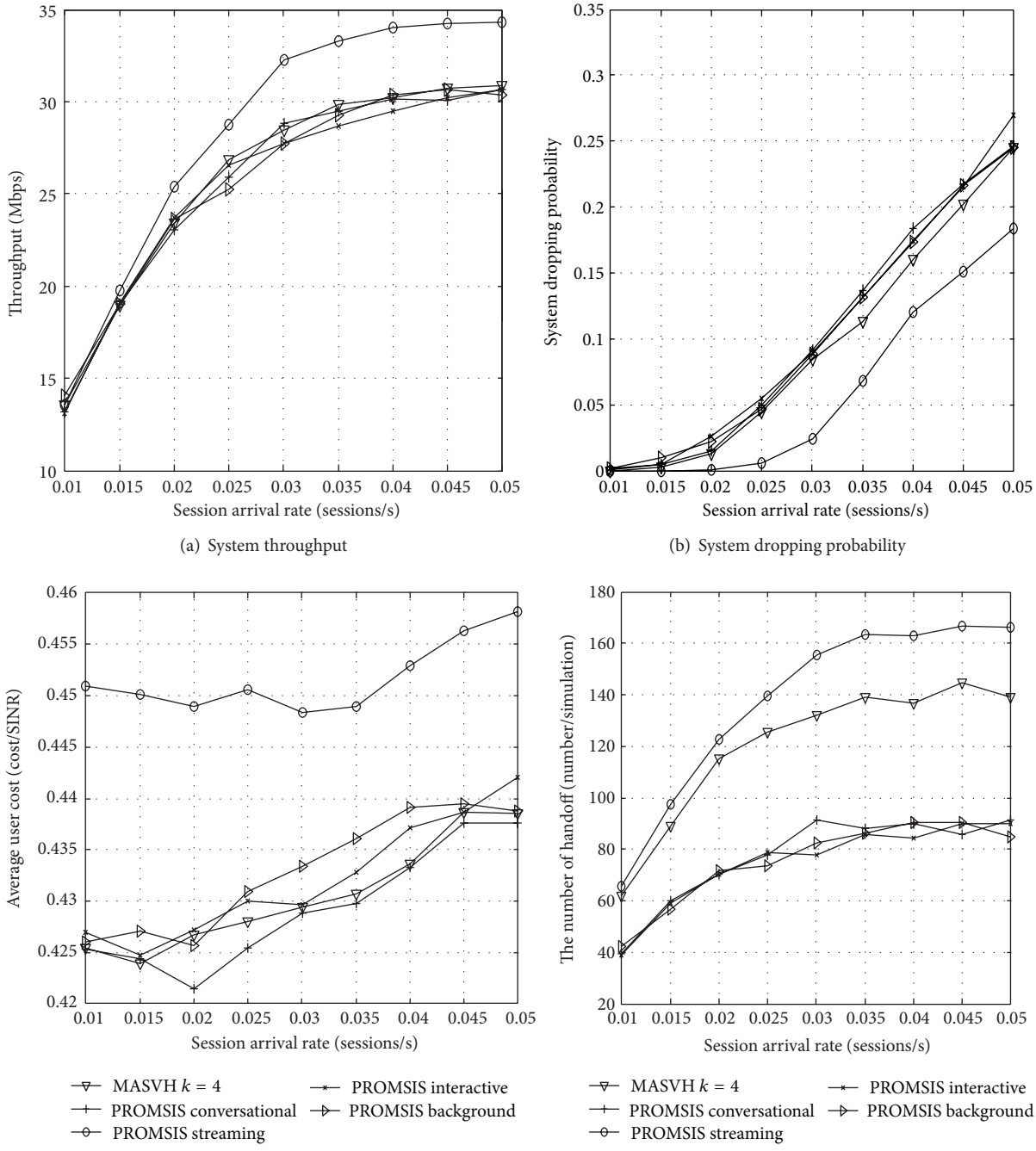


Figure 1: Simulation scenario.

the proposed LS-PROMSIS algorithm and the MASVH ($k = 4$) algorithm [6]. The downlink system throughput of each algorithm is measured and shown in Figure 2(a). It demonstrates that the LS-PROMSIS algorithm for streaming traffic class achieves the highest throughput performance because "the available bandwidth" attribute has the greatest weight in the handoff criterion, so the network of the largest available bandwidth is selected considering the load balance. Likewise, the system dropping probability of this algorithm is the lowest. The dropping probability performance of each algorithm is exhibited in Figure 2(b). The average user traffic cost performance is presented in Figure 2(c). It is noted that the cost of the LS-PROMSIS algorithm for streaming traffic class is quite high, but the cost of the LS-PROMSIS algorithm for conversational traffic is the lowest. Since the attribute of user traffic cost covers the highest proportion in the handoff criterion, the network of the lowest cost is inclined to be selected. Figure 2(d) indicates the number of vertical handoff. It is revealed that the number of vertical handoffs of the LS-PROMSIS algorithm for streaming traffic class is the highest and that of the MASVH algorithm finishes the second. By the way, if the unnecessary handoff needs to be further decreased, the proposed MADM algorithm can be combined with the mobile prediction technique to mitigate the impact of the ping-pong effect.

4. Conclusions

In this paper, a novel PROMSIS vertical handoff algorithm is proposed and compared with the existing CSVH and MASVH algorithms. The vertical handoff of heterogeneous networks is a multiple attributes decision subject. With regard to the relations between all the attributes, the observed objects are the four 3GPP defined traffic classes. According to the features of diverse traffic classes, the weight of each attribute in the handoff criterion is determined by LS. The



(a) System throughput

(b) System dropping probability

(c) Average cost to user traffic of the algorithms

(d) The number of vertical handoffs

Figure 2: Performance of each algorithm.

simulation results display that the performance of the algorithm is affected by the allocated weight vector. Consequently in practice, we should consider both the characteristics of the traffic and the preference of the user and weigh the advantages and disadvantages before making the decision. According to the analysis and simulation results, the PROMSIS algorithm can achieve the satisfactory performance for the network and the user.

For future work, more comparisons with other vertical handoff methods will be further discussed and other techniques to solve the decision problem, such as the game theory, will also be taken into account.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Contract no. 61271235 and Natural Science Foundation of Education Committee of Jiangsu Province (no. 11KJB510014).

References

[1] A. Singhrova and N. Prakash, "Vertical handoff decision algorithm for improved quality of service in heterogeneous wireless networks," *IET Communications*, vol. 6, no. 2, pp. 211–223, 2012.

[2] F. Kaleem and A. Mehbodniya, "Dynamic target wireless network selection technique using fuzzy linguistic variables," *China Communications*, vol. 10, no. 1, pp. 1–16, 2013.

[3] K. Yang, I. Gondal, B. Qiu, and L. S. Dooley, "Combined SINR based vertical handoff algorithm for next generation heterogeneous wireless networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 4483–4487, Washington, DC, USA, November 2007.

[4] K. Yang, I. Gondal, and B. Qiu, "Multi-dimensional adaptive SINR based vertical handoff for heterogeneous wireless networks," *IEEE Communications Letters*, vol. 12, no. 6, pp. 438–440, 2008.

[5] L. Bin and L. Shengmei, "Vertical handoff algorithm based on mobility prediction," *Application of Electronic Technique*, vol. 39, no. 1, pp. 93–95, 2013.

[6] R. O. Parreiras, J. H. R. D. Maciel, and J. A. Vasconcelos, "The a posteriori decision in multiobjective optimization problems with smarts, promethee II, and a fuzzy algorithm," *IEEE Transactions on Magnetics*, vol. 42, no. 4, pp. 1139–1142, 2006.

[7] J. P. Brans, P. Vincke, and B. Mareschal, "How to select and how to rank projects: the Promethee method," *European Journal of Operational Research*, vol. 24, no. 2, pp. 228–238, 1986.

[8] E. Stevens-Navarro and V. W. S. Wong, "Comparison between vertical handoff decision algorithms for heterogeneous wireless networks," in *Proceedings of the IEEE 63rd Vehicular Technology Conference (VTC-Sping '06)*, pp. 947–951, Melbourne, Australia, May 2006.

[9] A. T. W. Chu, R. E. Kalaba, and K. Spingarn, "A comparison of two methods for determining the weights of belonging to fuzzy sets," *Journal of Optimization Theory and Applications*, vol. 27, no. 4, pp. 531–538, 1979.