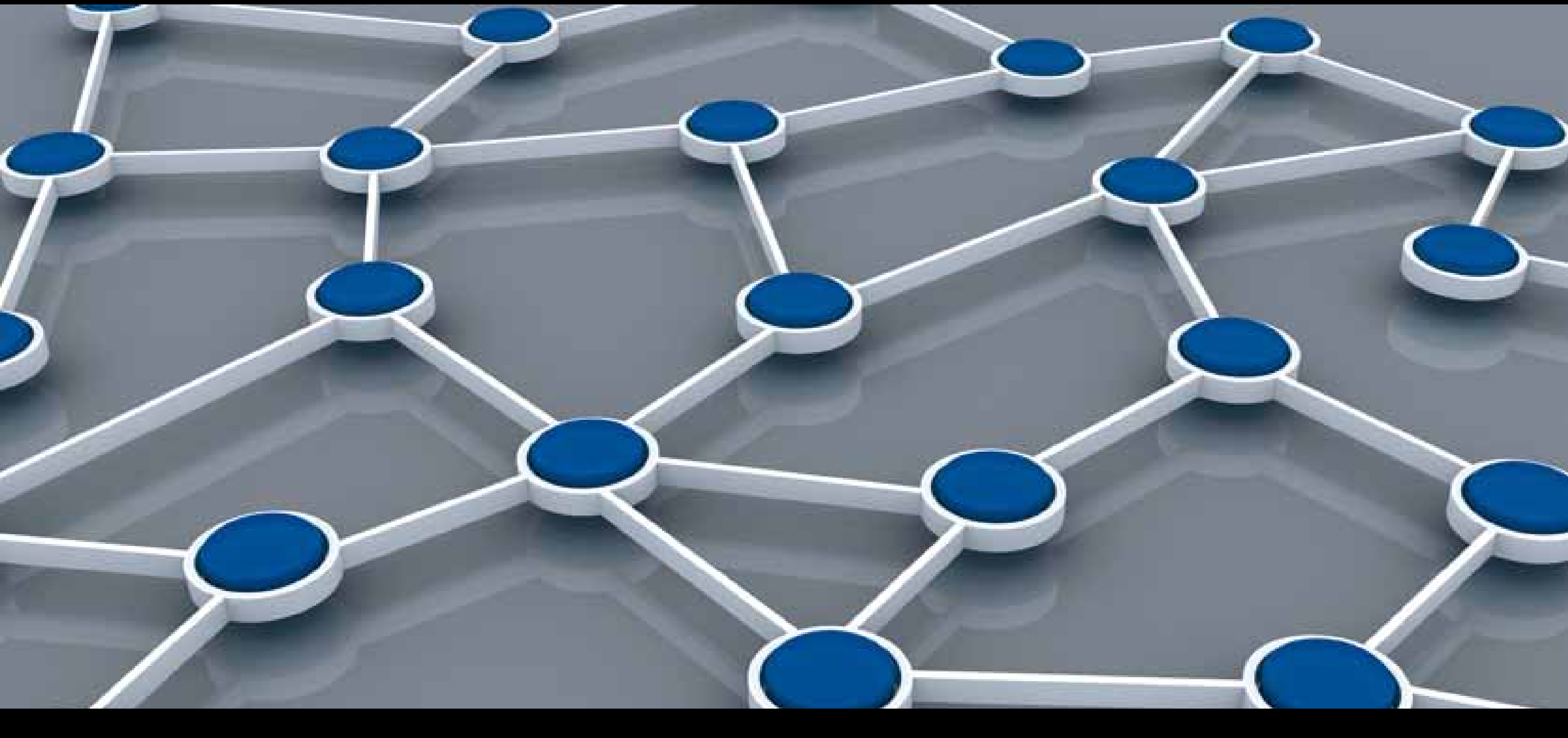
International Journal of Distributed Sensor Networks

Wireless Sensor Networks for the Internet of Things

Guest Editors: Yongzhao Zhan, Lu Liu, Liangmin Wang, and Yulong Shen



Wireless Sensor Networks for the Internet of Things

International Journal of Distributed Sensor Networks

Wireless Sensor Networks for the Internet of Things

Guest Editors: Yongzhao Zhan, Lu Liu, Liangmin Wang, and Yulong Shen

Editorial Board

Contents

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 717125, 2 pages
http://dx.doi.org/10.1155/2013/717125

Editorial

Wireless Sensor Networks for the Internet of Things

Yongzhao Zhan,[1] Lu Liu,[1,2] Liangmin Wang,[3] and Yulong Shen[4]

[1] *School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China*
[2] *School of Computing and Mathematics, University of Derby, Derby DE22 4DR, UK*
[3] *School of Computer Science and Engineering, Anhui University, Hefei, Anhui 230039, China*
[4] *School of Computer Science, Xidian University, Xi'an, Shaanxi 710071, China*

Correspondence should be addressed to Yongzhao Zhan; yzzhan@ujs.edu.cn

Received 1 August 2013; Accepted 1 August 2013

Wireless sensor networks (WSNs) have been increasingly popular which attract attention worldwide. WSN is now one of the key enablers for the Internet of Things (iThings) where WSNs will play an important role in future internet by collecting surrounding context and environment information. The innovations of integration of WSNs into iThings offer many interesting avenues of research for scientific communities. The research into WSNs for iThings is extremely important which could possibly change our day-to-day lives.

The purpose of the special issue is to collate efforts and achievements on the research of WSNs for iThings. This special issue will focus on fundamental issues in research and development of WSNs for iThings as well as new challenges in modelling and simulation of novel WSNs for iThings. This special issue contains thirteen papers which were selected from twenty-seven submissions after a peer review process.

Security is one of the most important research issues on WSNs for iThings. In an effort to address this challenge, S. Jiang et al. design a lightweight mobile reauthentication protocol for individual mobile network nodes which consumes less communication and computation recourse and, in the meantime, protects the privacy of mobile sensor nodes. The security and performance evaluation results show that the reauthentication scheme satisfies the security requirements for mobile WSNs, which is also suitable for the WSN environment with limited resources. L. Chen et al. focus on security solutions for the vehicular ad hoc networks (VANETs) in the emergency communication cases, in which the communication infrastructures are not always available. The security, feasibility, and efficiency of the proposed EPEC approach have been verified by the theoretical and experimental analyses.

Y. Guo et al. develop a selection mechanism for threshold key management in MANETs with selfish nodes. This mechanism can dynamically select a coalition of nodes to carry out the threshold key management service in runtime. The evaluation results show that proposed mechanism not only improves both the success ratio of key management service and lifetime of the network and but also reduces both the cost of participating nodes and compromising probability of MANETs, when compared with existing work in the presence of selfish nodes. X.-y. Chen et al. address the issues of node replication attacks on the area of wireless sensor networks. A location-binding symmetric key scheme and a detecting scheme have been proposed in the paper to address this issue. Extensive simulations have been conducted, and the simulation results indicate that the detection overheads are low and evenly distributed among all the sensor nodes.

Data integrity is one of major concerns in security. L. Zhu et al. focus on the issue of data integrity in WSNs and propose an efficient integrity-preserving data aggregation protocol (EIPDAP) to guarantee the integrity of aggregation results. EIPDAP can promptly verify the integrity of aggregation results once receiving the corresponding authentication information, thus significantly reducing energy consumption and communication delay.

Routing plays an important role in wireless sensor network communication. W. Xin-sheng et al. propose a load-balanced secure routing protocol for wireless sensor networks. The simulation results show that LSRP has achieved better routing performance and security. Z. Li et al. propose a traffic prediction-based fast rerouting algorithm (TPFR) to enable effective communication between the cluster heads and a sink node in WMSNs (TPFR) by using autoregressive moving average (ARMA) model. According to their evaluation results, TPFR can achieve the traffic load balancing and reduce energy consumption among network nodes by using fast rerouting scheme based on traffic prediction.

J. Gao et al. present a novel anticollision algorithm named query splitting-based anticollision (QSA) for mobile RFID-based Internet of Things. QSA reduces the number of collisions efficiently and makes it possible to identify multiple mobile tags without rollback. According to the performance evaluation results published in this paper, the proposed QSA algorithm takes fewer timeslots and achieves better performance in identifying mobile tags. Pei et al. propose a combing cogitative radio method in WSNs, named CWSNs, which integrates the idea of "the last diminisher" in the field of fairness allocation, the demand of secondary users (SUs), and channel characteristics. The research shows that CWSNs can achieve a higher transmission range, better use of the spectrum, lower energy consumption, and better communication quality.

The underwater acoustic sensor network is a special type of wireless sensor networks. Unlike the terrestrial wireless sensor networks that mainly rely on radio waves for communications, underwater sensor networks utilise acoustic waves. S. Xiong et al. present the so-called (response to the earliest transmitter of RTS MAC) RET-MAC protocol for underwater acoustic sensor networks. The simulation results reveal that RET-MAC can not only achieve higher fairness and balanced throughput but also reduce energy consumption and delay.

New research results on service provision on wireless sensor networks for Internet of Things have also been included in this special issue. N. Xi et al. identify the security constraints for each service participant to secure the information flow in a service chain based on the lattice model. A distributed verification framework is then presented in the paper which enables different service participants to verify their information flow policies. The evaluation results show a significant decrease in verification cost and a better load balance between sensor nodes. J. Zhu et al. propose a novel adaptive multihypothesis (MH) prediction algorithm for distributed compressive video sensing. The simulation results show that the proposed framework can provide better reconstruction quality than the framework using original MH prediction algorithm.

Mobile social network is an important application in future Internet of Things. Zhu et al. explore the delay tolerance in the message delivery from the source to the destination in mobile social networks. The simulation results indicate that the proposed theoretical models match very well with the simulation trace statistics.

Acknowledgment

Finally, we would like to express our sincere thanks to the colleagues who have served as reviewers for the submissions.

Yongzhao Zhan
Lu Liu
Liangmin Wang
Yulong Shen

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 674698, 9 pages
http://dx.doi.org/10.1155/2013/674698

Research Article

QSA: Query Splitting-Based Anticollision for Mobile RFID-Based Internet-of-Things

Jianliang Gao, Jianxin Wang, Jianbiao He, and Weiping Wang

School of Information Science and Engineering, Central South University, Changsha 410083, China

Correspondence should be addressed to Weiping Wang; wpwang@csu.edu.cn

Received 28 March 2013; Accepted 2 July 2013

Academic Editor: Lu Liu

Mobility is a common character of the emerging RFID-based internet-of-things. However, most of prior RFID anticollision algorithms ignore the movement of tags, which can degrade the identification performance seriously and even result in tag starvation problem. This paper presents a novel anticollision algorithm named Query Splitting-based Anticollision (QSA) for mobile RFID-based internet-of-things. By designing adaptive query, QSA reduces the number of collisions efficiently and makes it possible to identify multiple mobile tags without rollback. In QSA, we propose a query stack technology to avoid the rollback operation caused by new arriving tags, which solves the tag starvation problem under mobile environments. The performance evaluation shows that the proposed algorithm takes fewer timeslots and has better performance in identifying mobile tags.

1. Introduction

Radio frequency identification (RFID) is a contactless wireless communication technology. The scope for using this technology boosts as RFID tag becomes a low-cost device due to mass production [1]. As the rapid proliferation of RFID tags, it has given rise to various concepts that integrate the physical world with the virtual one. One of the most popular concepts is the Internet-of-Things (IoT), a vision in which the Internet extends into physical entities. In the IoT, RFID is the foundation to connect the things together [2].

RFID systems consist of a reading device called reader, and multiple tags which are attached to physical entities in the IoT. The reader is typically a powerful device with ample memory and computational resources. On the other hand, tags vary significantly in their computational capabilities. Among tag types, passive ones are emerging to be a popular choice for large scale deployments due to their low cost [3]. For passive tags, they respond only at reader commands with the energy provided by the reader [2, 4].

In RFID systems, the reader usually needs to communicate with multiple tags. If there are multiple tags in a reader's interrogation zone, the reader receives the responses from these tags simultaneously. For a reader, it is not able to distinguish exact information from the interfered wireless signals, which is called collision. Collision is a serious problem in RFID systems since the reader will not receive the messages rightly once collision occurs [5]. An example is shown in Figure 1. Simultaneous responses transmitted by multiple tags collide, resulting in an increase of identification delay, even failure of reading the tags. Therefore, an efficient anticollision algorithm is required to reduce collisions and to achieve fast identification.

The tag collision problem becomes more serious in mobile RFID-based IoT. The physical entities with tags are often mobile, which facilitates competitive advantage through benefits such as improved efficiency, increased visibility, reduced cost, and many others [6]. The mobile devices can be part of numerous products, gadgets, and vehicle parts [7]. These devices close the gap between the real word and its virtual representation via, for example, seamless identification and integration with other wirelessly-embedded devices and their surroundings.

However, the mobile devices result in new challenges for anticollision problem. One of the challenges is tag starvation. For example, as shown in Figure 1, the reader is processing the collision caused by Tag 1 to Tag N. Assuming the reader has partitioned the tags for several rounds and will identify one tag out right now. If Tag $N+1$ enters the reader's interrogation zone at this time, the new arrived tag might cause the rollback

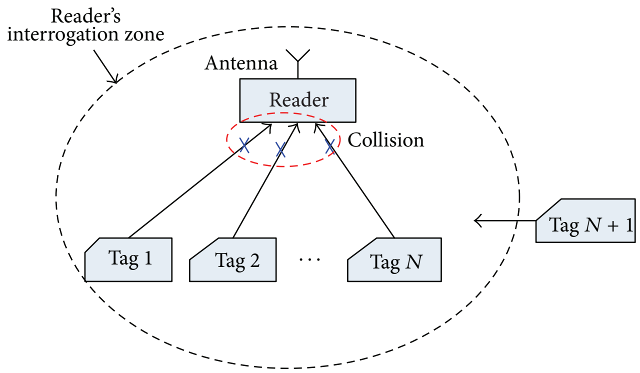


Figure 1: The tag collision with mobile tag.

of the partition operations. Therefore, the done operations have to be repeated and the delay of identifying the tags increases. If the new tags enter the reader's interrogation zone at an interval that causes rollback repeatedly, some tags might depart the reader's interrogation zone before being identified, which is called "tag starvation" in mobile environments. Therefore, it is necessary to provide efficient anticollision for mobile RFID-based IoT.

In this paper, we will deal with the "tag starvation" and performance problem for anticollision in mobile IoT. By the design of query stack and query rules, the proposed anticollision algorithm avoids repetition operations and decreases the collisions consequently. In summary, our contributions include the following.

(i) We characterize the problem of anticollision in mobile environment, which is more practical for the emerging IoT technology.

(ii) We propose a novel anticollision algorithm named Query Splitting-based Anticollision (QSA) for mobile RFID-based IoT, which solves the problem of tag starvation and performance degradation resulted from the tag mobility.

(iii) Simulation and analysis evaluate the efficiency of the proposed algorithm. The results show the better performance of the proposed algorithm to the prior methods.

The rest of this paper is organized as follows. Section 2 outlines preliminary, includes background, related work. Section 3 presents the problems for mobile tags. Section 4 gives the detailed design of our algorithm. The simulation results are provided in Section 5. Finally, Section 6 concludes the paper.

2. Preliminary

2.1. Collision Detection. Code technologies are widely used for collision detection. Manchester code is one of the most popular technologies for RFID systems [5]. In Manchester code, the value of one bit is defined as the voltage transition within a bit window. A bit "0" is coded by a positive transition, while a bit "1" is coded by a negative transition. In RFID systems, if two (or more) tags transmit different values simultaneously, the positive and negative transitions of the received bits cancel each other out. This state is not permissible in Manchester code during data transmission and is recognized as an error. Therefore, Manchester code makes it possible to "trace a collision to an individual bit" and "find where the collided bit is" [8].

Figure 2 shows an example of Manchester code for collision detection. The IDs of tag 1 and tag 2 are "10101100" and "10001001," respectively. When tags 1 and 2 send their IDs simultaneously using Manchester code, the decoded data from the interfered signal received by the reader is "$10x01x0x$," where "x" represents a collided bit. In this example, the locations of the collided bits are the 3rd, 6th, and 8th bits. This information helps the reader separate the collided tags into subsets more smartly and identify the tags more quickly.

Manchester code can be utilized to detect the collision bits, but the tags should be strictly synchronized. Fortunately, the tags in passive RFID systems are all driven by the reader with both energy and the same clock frequency.

2.2. Anticollision Algorithms. Many researches have focused on the issue of anticollision including tag-driven and reader driven procedures [9]. Tag-driven anticollision protocols function asynchronously [5]. For example, in Aloha-based protocols [10], time is divided into slots and each tag randomly transmits its ID in each timeslot. The tags continuously retransmit their IDs until the reader acknowledges their transmission. However, the Aloha-based protocols have several serious problems. For example, a specific tag may not be identified for a long time, leading to the so-called "tag starvation" problem. The performance of Aloha-based protocols is sensitive to the number of tags. Furthermore, it is very difficult to predict the number of tags in mobile environments, if not impossible.

For reader-driven anticollision protocols, they function synchronously, since all tags are controlled and checked by the reader simultaneously. Therefore, the reader can avoid tag starvation under static environments. Furthermore, they can be categorized into Binary Tree algorithm (BT) [11–13] and Query Tree algorithm (QT) [14–16].

2.2.1. Binary Tree Algorithm (BT). BT performs collision resolution by splitting collided tags into disjoint subsets. These subsets become increasingly smaller until they contain one tag. Each tag has a random binary number generator. For example, in Figure 3, tags with a counter value of zero are considered to be in the transmit state; otherwise, tags are in the sleep state. After each timeslot, the reader informs tags whether there is a collision or not. If there was a collision, each tag in the transmit state generates a random binary number. Tags will become in sleep state after being identified.

As can be seen that the average number of timeslots to identify the first tag is

$$T(N) = \log_2 N, \tag{1}$$

where N is the number of tags. Note that the reader needs to broadcast the universal condition (all bits are "1") before the

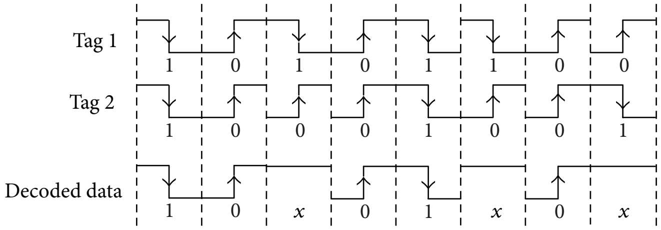


Figure 2: Collision detection with Manchester code.

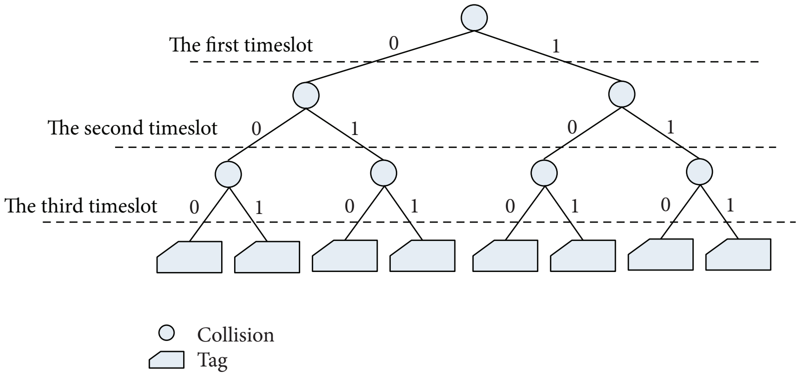


Figure 3: Binary tree based anticollision procedure.

first timeslots to collect all tags lying in its interrogation zero. Furthermore, the total number of timeslots for identifying all N tags is

$$T(N) = \sum_{k=1}^{N} \log_2 k. \tag{2}$$

The time consumption is the most serious defect since it has not recorded the history information, which results in a large amount of repeat operations.

2.2.2. Query Tree (QT). Query tree (QT) algorithms store tree construction at the reader, and tags only need to have a prefix matching circuit. The reader transmits a serial number to tags, which they then compare against their IDs. The tags whose IDs equal to or lower than the serial number respond to the reader's command. The reader then monitors tags reply bit by bit using Manchester code, and once a collision occurs, the reader splits tags into subsets based on collided bits. The reader then transmits another query by replacing the most significant collided bit with "0" and sets the other bits to "1." This procedure stops until a single tag has been selected out.

Rollback Query Tree (RQT) is proposed to reduce the average number of timeslots for identifying the tags. During the partition operation, these records are saved at the reader. Thus, the anticollision can avoid the repeating operations based on the saved information when the reader tries to select out the next tag. The timeslot number of RQT is

$$T(N) = \log_2 N + N - 1, \tag{3}$$

where N is the number of tags.

3. Problem for Identifying Mobile Tags

In static environments, Rollback Query Tree reduces the number of timeslots to deal with collision in RFID systems. However, both tag starvation and delay problems occurred in mobile environments. If one tag enters the reader's interrogation zone when the reader is processing the entered tags, some obtained results might be destroyed by the new arrived tag. The total number of identifying tags is

$$T(N+M) = T(N) + \sum_{i=1}^{M} \left(T_i + R_i\right), \tag{4}$$

where N, M are the numbers of the tags in the reader's interrogation zone already and the new arrived tags, respectively; $T(N)$ is the timeslots number as is defined in formula (3). T_i, R_i are the timeslots for processing the new arrived tag i and the timeslots that caused by the repeat operations. The increased timeslot for a new arrived tag i is $T_i + R_i$.

If the increased time is greater than the interval time between the arriving tags, the existing tags might not be processed in a long time since the reader has to reexecute the rollback operations for the new arrived tags. Therefore, tags starvation problem will happen if the following condition is met:

$$T_{i \to i+1} \leq T_i + R_i, \tag{5}$$

where $T_{i \to i+1}$ is the interval time between the new arrived tag i to $i+1$. If each tag i meets the requirement in formula (5), no tags will be selected out successfully. This requirement is considerably strong, but to a moving tag, it will fail to be identified only if the reader has not processed before it leaves the reader's interrogation zone.

In the following, we take an example to illustrate the rollback caused by a new arriving tag. As shown in Figure 4, there are already four tags (IDs: 10100011, 11100010, 11100011, and 11110010) in the reader's interrogation zone. After two rounds of query (named q_1 and q_2), the tag 10100011 is the first one to be identified. Then, the reader should broadcast the queries q_3, q_4, and q_5 to select the next smallest tag 11100010 in the reader's interrogation zone. If a new tag 01100011 enters the interrogation zone between broadcasting q_4 and q_5, it will respond to the query q_5 "≤11100010." Then, the reader receives the response "$x110001x$" (combined by 11100010 and 01100011) instead of selecting out tag 11100010. Therefore, the reader should broadcast the new request command "≤11111111" according to the highest uncertain "x" bit. Unfortunately, this query command includes the responses from all tags, and the queries "q_3" and "q_4" have to be executed again in the following. In an extreme situation, the operation might be repeated many times for the consecutive arriving tags as shown in formula (5). Thus, tag starvation is possible for query tree schemes in mobile environments.

4. QSA: Query Splitting-Based Anticollision for Mobile IoT

4.1. Overview of the Anticollision Algorithm. Firstly, a set of queries $Q = (q_1, q_2, \ldots, q_l)$ is defined for the reader, where q_1 is initialized as "$\leq\{1\}^n$". The reader executes the following steps to identify tags.

(1) Broadcast the current query in Q to all tags.

(2) When receiving the responses from tags:

(2.1) if the reply is string w without "x" bits, then select and process the tag with ID w;

(2.2) if a collision is detected; that is, the reply is string w with "x" bits, then set the next query in Q;

(2.3) if there is no reply from tags, do nothing.

(3) Update Q according to the received responses.

Repeat the above procedure until Q is empty. In this procedure, four commands (REQUEST, SELECT, READ-DATA, and DESELECT) are adopted as defined in Table 1.

For tags, let $w = w_1 w_2, \ldots, w_n$ be the tag's ID. The query is defined as follows: if $w \leq q$, then the tag sends string w to the reader, where q is the query string received from the reader.

As can be seen from the above procedure that the key problem is how to design the query scheme, our main idea is to keep the history records which can avoid rollback of the done operations.

4.2. Design of Adaptive Query

4.2.1. Query Stack Design for Rollback Operation. As shown in Table 1, REQUEST is the query command for the reader to split tags into subsets. To reduce the total timeslots for identifying all tags, we design a query stack structure to implement the rollback operations during query process. "Push Query" and "Pop Query" are the two basic operations of query stack. Each query response is pushed into query stack if and only if the response includes "x."

Table 1: Definition of commands.

REQUEST (ID_Condition)	Reader sends REQUEST command with parameter ID_Condition to tags. Tags compare their IDs against the received ID_Condition and reply their IDs to the reader if ID ≤ ID_Condition.
SELECT (ID)	Reader selects the tag ID.
READ-DATA	Reader reads data from the selected tag.
DESELECT	Reader cancels the selected tag, and this tag thus enters silent state.

We also take the example in Figure 4 to explain the stack operation. The response is "$111x001x$" for the query q_3: "≤11111111." The response "$111x001x$" includes "x"; thus, it is pushed into the query stack as shown in Figure 5. In the same way, the response "$1110001x$" is also pushed into the query stack with the query q_4: "≤11101111." When the reader broadcasts the query q_5: "≤11100010," a new tag "01100011" enters the reader's interrogation zone. Thus the response will be "$x110001x$" and it is also pushed into the stack as shown in Figure 5.

Once a response includes no "x," the reader will select this tag (using the command SELECT as shown in Table 1) and read the selected tag (using the command READ-DATA as shown in Table 1). After that, the reader will pop one element from the query stack. For example, when the new arrived tag "01100011" has been processed, the element "$x110001$" will then be popped out. The following design is how to set the "x" bits and broadcast new query commands.

4.2.2. Rule Design for Query Condition. The rule for query command is one of the most important designs in QSA. As shown in Table 2, the rules can be presented as two kinds. One is for the obtained response, and the other is for the pop stack.

(1) Rule for Obtained Response. When the reader receives response including "x" bits, the next query condition should be set according to the collision bits in the response. The second line in Table 2 shows the rule for this kind of obtained response.

In the first iteration, all bits are set as "1," that is, Max(ID), to collect the responses from all tags. For example, the parameter is "11111111" for eight bits ID. In the following iteration, the highest bit "x" is set as "0"; the bits which are lower than the highest "x" are all set as "1." According to the setting of "x" bits, we can draw the following theorem.

Theorem 1. *Rule for obtained response implements binary splitting for multiple tags.*

Proof. The response is a string $S = \{0, 1, x\}^n$, where n is the bit number of tag ID. The highest "x" bit in S is denoted as k, and it must meet the following requirement:

$$S[k] = \{x\}, \qquad S[n \cdots k+1] = \{0,1\}^{n-k}. \tag{6}$$

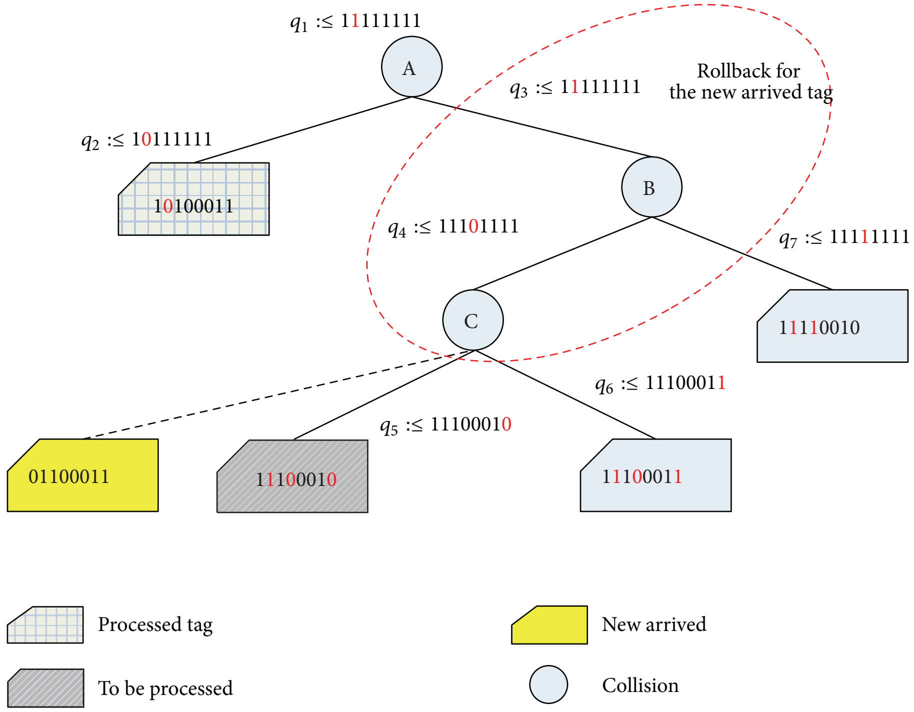


Figure 4: Rollback problem for the new arrived tag.

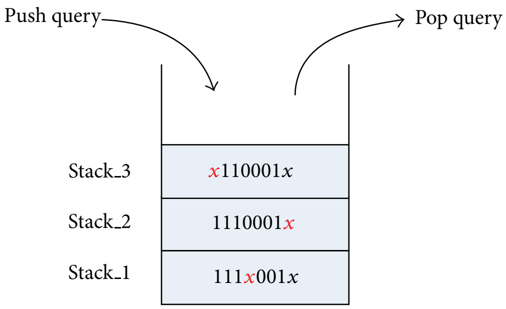


Figure 5: Query stack for new arrived tags.

Table 2: Rule design for query condition.

	The first iteration	The nth iteration
Rule for obtained response	Max(ID)	$\text{bit}(k) = 0, \text{bit}(k-1\cdots 1) = \{1\}^{k-1}$, (where k is the highest "x" bit in the response)
Rule for pop stack	Cannot happen	if there is new arrived tag $\text{bit}(k) = 1$; $\text{bit}(k-1\cdots 1)= \{0\}^{k-1}$ else $\text{bit}(k\cdots 1) = \{1\}^{k}$ end

Then, there is at least one tag whose kth bit is "1" and at least one tag whose kth bit is "0." Otherwise, the kth bit cannot be "0." Therefore, REQUEST "$\leq \text{bit}(k) = 0, \text{bit}(k-1\cdots 0) = 1$" splits the tags into two catalogs as binary division: one catalog is with $S[k] = \{0\}$, and the other is with $S[k] = \{1\}$. □

This rule can reduce the query scope by dividing the responding tags into two catalogs until only one tag respond.

(2) Rule for Pop Stack. After one tag is selected and processed, the top element is popped out. This element includes "x" bit. The rule for pop stack deals with the problem of setting these bits, which is shown in the third line of Table 2. Note that it cannot be the first iteration in anticollision process ("cannot happen" in Table 2) since the stack is empty at the beginning.

To the popped element, there are two situations: without new arrived tag and with arrived tag when it pushed into the stack. For the situation without new arrived tag, all bits are set as "1" in order to include all unprocessed tags; that is, $\text{bit}(k\cdots 1) = \{1\}^{k}$, where k is the highest "x" bit. On the other hand, if there are new arrived tags, the highest "x" bit is set as "1," and the lower bits are all set as "0"; that is, $\text{bit}(k-1\cdots 1) = \{0\}^{k-1}$.

Whether there is new arrived tag or not, it can be judged according to the following formula:

$$\text{Highest}\,x\ (\text{Top}) > \text{Highest}\,x\ (\text{Top-1}), \qquad (7)$$

where Highest x is the function to obtain the highest "x" bit in its parameter, Top means the first element in the query stack, and Top-1 is the second one.

Theorem 2. *If the highest "x" bit of the first stack element is higher than that of the second stack element, that is, the condition of formula* (7) *is met, there must be new arrived tag.*

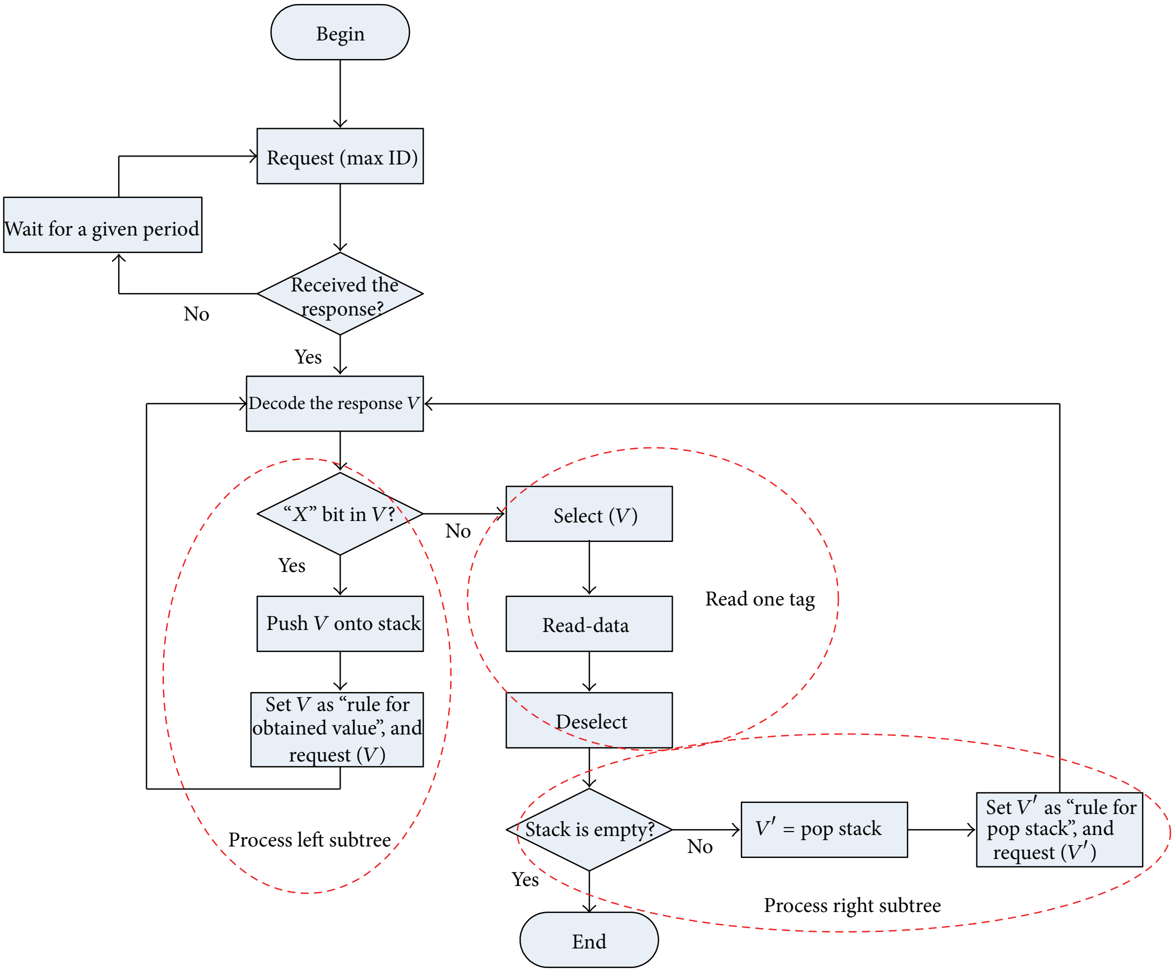


Figure 6: Flow chart of the proposed QSA algorithm.

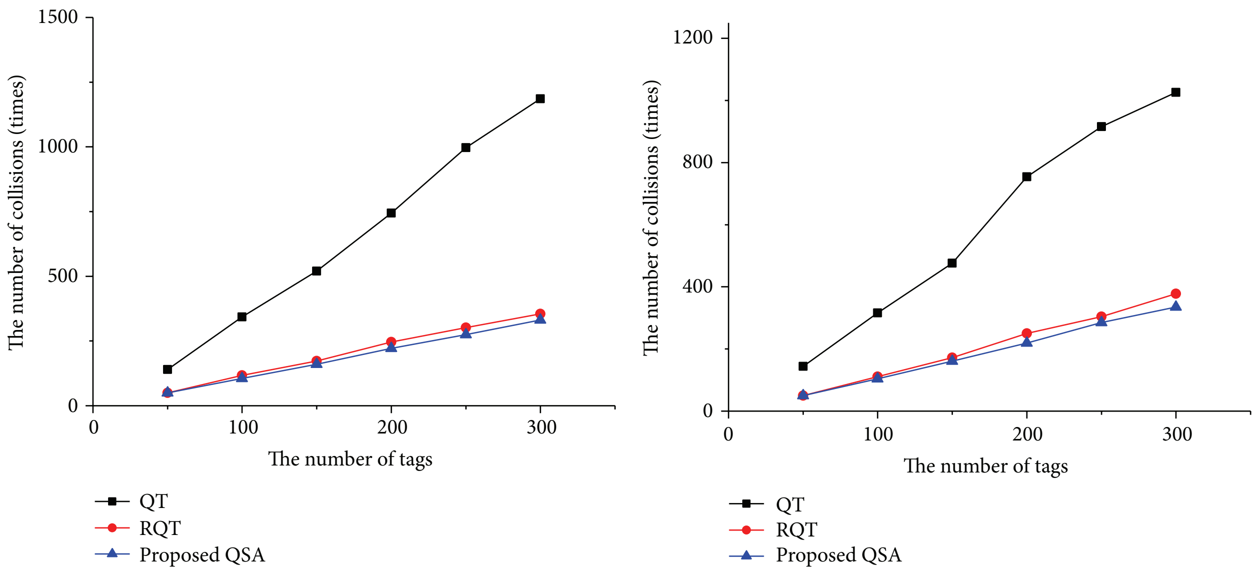
Proof. According to rule for obtained response, the reader first processes the tags whose highest "x" bit is "0." Therefore, only when the higher "x" bit is processed, it is possible for lower "x" to be processed. Thus, the highest "x" always decreases as the distance to the top of query stack. However, it is possible that the "x" can be at any bit if new tags arrive. That is to say, formula (7) is a result caused by new arrived tags. □

For example, as shown in Figure 5, the highest "x" bit in the top stack element Stack_3 is 8, and that in Stack_2 is 1. Formula (7) meets and there is new arrived tag. In the example, the query for Stack_2 element "1110001x" is q_5 "≤11100010" in the example of Figure 4. While the highest "x" bit in top element Stack_1 is the 8th bit. There is new arrived tag whose 8th bit is "0." Otherwise, the new arrived tag is impossible to be included by the query "≤11100010" and the top element cannot be "x" at the 8th bit. In fact, there is a new arrived tag whose ID is 01100011 in the example.

4.3. The Procedure of the Proposed QSA. Based on the design of query stack and query rule, we describe the procedure of the proposed QSA algorithm in the following. In QSA, the reader first broadcasts the query with the parameter (max_ID) and receives the replies from all tags in its interrogation zone. If the received response of this query includes "x," the tree can be divided into the left subtree and the right subtree according to the highest bit "x." In the left subtree, the highest "x" bit is "0," and that of the right tree is "1." The procedure of the anticollision algorithm is shown in Figure 6. It mainly includes three parts: process the left subtree, process the right subtree, and read one tag.

4.3.1. Process the Left Subtree. If there is "x" in the received response, the reader begins to process the left subtree. Firstly, the reader pushes the response which has "x" bits into the query stack. We design this query stack which can record the entrance to the right subtree. Then, it sets the query condition according to the rule for obtained response in Table 2. Finally, the reader broadcasts the query with the set condition, which enables a new round communication.

4.3.2. Read One Tag. If the received response has no "x" bit, there is no collision and a single tag is identified. The reader will select this tag and read data from it. After the tag is



(a)

(b)

Figure 7: The number of collisions versus various numbers of tags. (a) Tag ID is 64 bits; (b) tag ID is 128 bits.

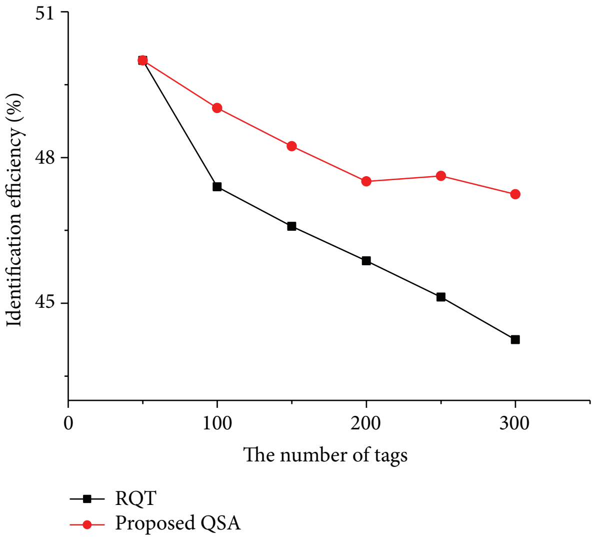


Figure 8: Identification efficiency.

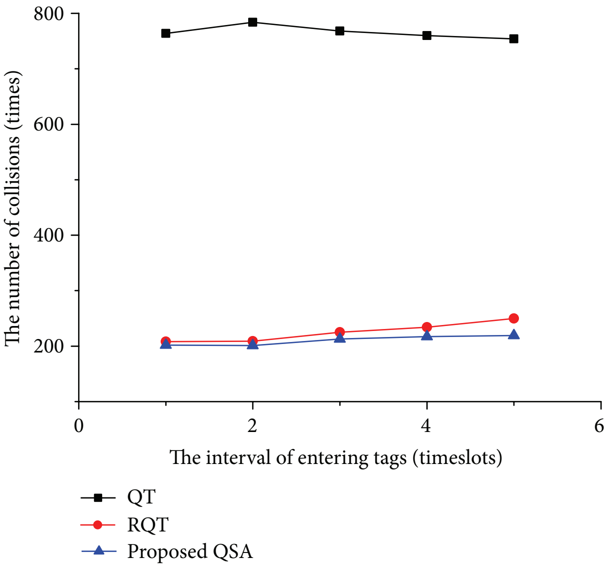


Figure 9: The number of collisions versus various mobile velocities.

processed, command "DESELECT" will make this tag keep silent state in the following.

4.3.3. Process the Right Subtree. After one tag is read, the algorithm will begin to process the right subtree. Firstly, the top element of the query stack is popped out, and it is set according to the rule for pop stack. As described in previous Section, there are two possibilities, that is, with new arrived tag and without new arrived tag when setting the query condition. If there is new tag which causes higher "x" bit, the query will enter the left tree process after receiving the response. Otherwise, the process will further deal with the right subtree. Note that the process will end until the query stack is empty, which means that all tags have been identified already.

5. Simulation Results

To evaluate anticollision performance, we compare the time consumption and identification efficiency of the proposed QSA algorithm with that of two representative schemes Query tree (QT) and Rollback Query Tree (RQT). Under mobile environments, we measure the number of timeslots used to read out all tags. Firstly, we fix mobile velocity of the tags and evaluate the results. The mobile velocity of tags is set as the following rule: a group of five tags enter the reader's interrogation zero at the interval of five timeslots.

Figure 7 shows the results of collision numbers under various numbers of tags. As can be seen that QT meets the most number of collisions, since it has not recorded the query results during the procedure of finding the least tag ID. RQT always takes more collisions than the proposed QSA algorithm, which is consistent with formula (4). Both 64 bits and 128 bits tag ID in Figures 7(a) and 7(b), the proposed QSA takes the least number of collisions, which illustrates the efficiency of the QSA since it overcomes the rollback operations when new tags arrive.

To identify the tags as soon as possible, the algorithm should keep high identification efficiency. Identification efficiency is defined as the ratio of the number of success timeslots to the total timeslot number. Higher identification efficiency means less wasted timeslots for collisions. Figure 8 describes the comparison of identification efficiency between RQT and the proposed QSA. The identification efficiency of QSA is higher than that of RQT. It validates the analysis of number of consumed timeslots in Figure 7(a). The identification efficiency of the proposed QSA especially decreases slowly as the tag number increases. While the RQT is of lower identification efficiency when the tag number increases. For example, the identification efficiency of the RQT is below 45% when the tag number is 300.

From the theoretical analysis, it can be known that the mobile velocity will affect the results of collisions. Finally, we evaluate the impact of tag mobile velocity. In this experiment, total 200 tags enter the reader's interrogation zone at different intervals, which varies from 1 to 5 timeslots. The same with the above experiments, five tags come into the reader's interrogation zone at a pointed timeslot. As can be seen from Figure 9, the proposed QSA achieves the least number of collisions under all intervals of entering tags. An interesting change happens in QT scheme: the collision number increases firstly and then decreases. It is because the tag number increases quickly at the smaller interval of entering tags, which causes more collisions. The gap between RQT and the proposed QSA increases as the interval increases. It indicates that the velocity of mobile tags will cause different rollback operations in RQT, while it is eliminated in our QSA algorithm.

6. Conclusions

Collision is considered as one of the most important issues that affect the identification process in RFID systems. Mobile tags of IoT especially induce new problems such as performance degrade and tag starvation. In this paper, we propose a novel anticollision algorithm named QSA to solve these problems. The proposed QSA can overcome the problems resulted from mobile tags efficiently. Compared to prior schemes, the QSA presents a better performance at different mobile velocity and various tag numbers.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant nos. 61106036, 61173169, and 61272147, in part by the science and technology projects of Hunan province, and in part by State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences.

References

[1] A.-H. Mohsenian-Rad, V. Shah-Mansouri, V. W. S. Wong, and R. Schober, "Distributed channel selection and randomized interrogation algorithms for large-scale and dense RFID systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1402–1413, 2010.

[2] F. Villanueva, V. David, and M. Francisco, "Internet of Things architecture for a RFID-based product tracking business model," in *Proceedings of the 6th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pp. 811–816, 2012.

[3] Y.-H. Chen, S.-J. Horng, R.-S. Run et al., "A novel anti-collision algorithm in RFID systems for identifying passive tags," *IEEE Transactions on Industrial Informatics*, vol. 6, no. 1, pp. 105–121, 2010.

[4] J. Panneerselvam, L. Liu, R. Hill, Y. Zhan, and W. Liu, "An investigation of the effect of cloud computing on network management," in *International Conference on High Performance Computing and Communications*, pp. 1794–1799, 2012.

[5] D. K. Klair, K.-W. Chin, and R. Raad, "A survey and tutorial of RFID anti-collision protocols," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 3, pp. 400–421, 2010.

[6] Y. Jiang and R. Zhang, "An adaptive combination query tree protocol for tag identification in RFID systems," *IEEE Communications Letters*, vol. 16, no. 8, pp. 1192–1194, 2012.

[7] S. Jiang, J. Miao, and L. Wang, "Mobile node authentication protocol for crossing cluster in heterogeneous wireless sensor network," in *Proceedings of the 3rd International Conference on Communication Software and Networks (ICCSN '11)*, pp. 205–209, May 2011.

[8] X. Jia, Q. Feng, and C. Ma, "An efficient anti-collision protocol for RFID tag identification," *IEEE Communications Letters*, vol. 14, no. 11, pp. 1014–1016, 2010.

[9] L. Kang, J. Zhang, K. Wu, D. Zhang, and L. Ni, "RCSMA: receiver-based carrier sense multiple access in UHF RFID systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 735–743, 2012.

[10] M. Al-Medhwahi, A. Alkholidi, and H. Hamam, "A new hybrid frame ALOHA and binary splitting algorithm for anti-collision in RFID systems," in *Proceedings of the 18th International Conference on Software, Telecommunications and Computer Networks (SoftCOM '10)*, pp. 219–224, September 2010.

[11] Y.-C. Lai and C.-C. Lin, "Two blocking algorithms on adaptive binary splitting: single and pair resolutions for RFID tag identification," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 962–975, 2009.

[12] Y.-C. Lai and C.-C. Lin, "A pair-resolution blocking algorithm on adaptive binary splitting for RFID tag identification," *IEEE Communications Letters*, vol. 12, no. 6, pp. 432–434, 2008.

[13] Y.-C. Lai and L.-Y. Hsiao, "General binary tree protocol for coping with the capture effect in RFID tag identification," *IEEE Communications Letters*, vol. 14, no. 3, pp. 208–210, 2010.

[14] Y. Lai, L. Hsiao, H. Chen, C. Lai, and J. Lin, "A novel query tree protocol with bit tracking in RFID tag identification," *IEEE Transactions on Mobile Computing*, pp. 1–13, 2012.

[15] J. H. Choi, D. Lee, and H. Lee, "Query tree-based reservation for efficient RFID tag anti-collision," *IEEE Communications Letters*, vol. 11, no. 1, pp. 85–87, 2007.

[16] J. Shin, B. Jeon, and D. Yang, "Multiple RFID tags identification with M-ary query tree scheme," *IEEE Communications Letters*, vol. 17, no. 3, pp. 604–607, 2013.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 596352, 13 pages
http://dx.doi.org/10.1155/2013/596352

Research Article

Load-Balanced Secure Routing Protocol for Wireless Sensor Networks

Wang Xin-sheng,[1] Zhan Yong-zhao,[1] and Wang Liang-min[2]

[1] *School of Computer Science and Telecommunication Engineering, Jiangsu University, China*
[2] *Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, Anhui University, China*

Correspondence should be addressed to Wang Xin-sheng; wxs@ujs.edu.cn

Received 28 February 2013; Accepted 28 May 2013

Academic Editor: Lu Liu

To solve the problems of limited energy of the nodes and security of routing in wireless sensor networks, load-balanced secure routing protocol (LSRP), a load-balanced secure routing protocol for wireless sensor networks, is proposed. Based on structured topology of hexagonal mesh, hops at different directions are calculated on the optimal route for transmitting data packets in LSRP. Depending on characters of hops, the nodes can rapidly find a route among multiple optimal routes by the policy of the twice probability routing selection. Data breach is prevented by data encryption, and data security is realized by one-way hash key chain and symmetric key authentication. LSRP offers preventions against usual attacks, and it also takes into account traffic load balance. Analysis and simulation results show that LSRP has better performance on traffic load balance and security.

1. Introduction

As a convenient tool to capture information, wireless sensor networks can access information in fields that are beyond the arm of flesh. Special fields of application such as military and antiterrorism require security of sensitive data, which arouses scholars' attention on the security of wireless sensor networks [1, 2]. However, complex security measures based on cryptography are inapplicable owing to the calculation and storage capability of the nodes of wireless sensor networks. Open wireless communications means with limited band width facilitate attacks such as eavesdropping and DoS. The multihop transmission and self-organization approach causes deficiency of key infrastructure and possibility of malicious nodes to mix in the network to implement insider attack. All of these problems pose a greater security challenge to wireless sensor networks than traditional network [3].

The discovery of self-organizing routing, the approach of multihop data forwarding, and the mode of open wireless communication pose two threats to routing security in wireless sensor networks [4]: on one hand, there might be potential threats to security in the course of packet transmission, such as eavesdropping, altering, and discarding, which will result in breach, inauthenticity or loss of the content; on the other hand, the attackers might manipulate the packets on communication links to attack the network through routing and cause performance deterioration or even breakdown of the network. This makes routing security an important subject in studies about the security of wireless sensor networks. A series of secure routing protocols have been proposed against various kinds of routing attacks. For example, GPSR [5] can detect black hole regions through periodic broadcast probe request and effectively detect and counteract sinkhole attack and wormholes attack. SRWA [6] uses mobile agent to reduce false positive to defense wormholes attack. SeRWA [7] protocol uses symmetric key cryptography to defense wormhole attack and can find a secure route against a wormhole attack. SPINS [8] can realize authentication, encryption and refreshing of data and authentication of broadcast packets under the condition of limited resources, and effectively detect and counteract data eavesdropping, altering, and replay attacks. EENDMRP [9] uses the multiple paths and digital signature crypto system to transmitted data packets and effectively prevent selective forwarding, sinkhole, and altering attacks. By importing tokens, SRD [10] can detect and prevent acknowledgement spoofing and false routing information attack. SDDR [11] uses the μTESLA (microtimed, efficient, streaming, loss-tolerant authentication) algorithm in order to prevent black hole and acknowledgement spoofing attacks. INSENS [12] and TRANS [13] adopt measures like link-layer encryption and authentication, multipath routing, identity authentication, two-way connection authentication, and authentication broadcast to effectively prevent false routing information, Sybil attack, and HELLO FLOOD attack. ATSR [14] uses accurate location information to implement a distributed trust model to prevent selective forwarding and Sybil attacks. Multipath and multibase station routing [15] can effectively prevent HELLO FLOOD attack and replay attack through the key and one-way hash key chain assigned by multitree key protocol. Multipath routing [16, 17] can effectively prevent particular attacks with the feature of attracting all traffic to pass the malicious nodes, such as wormhole, sinkhole, and selective forwarding attacks. By checking the credit of the nodes, ARRIVE [18] can effectively prevent selective forwarding attack. However, these algorithms and protocols are mainly targeted at one or several types of attacks and have disadvantages in excessively large load of calculation and communication.

Taking both security and energy saving into consideration so as to extend the service life of the network is still a burning problem. By combining topology generation and routing discovery, this paper reduces the complexity of routing discovery by combine topology generation and routing discovery, based on this, puts forward a secure routing protocol based on the twice probability routing selection, LSRP (load-balanced secure routing protocol). LSRP realizes routing security by one-way hash key chain and symmetric key cryptography and balances network load through optimizing routing to extend the service life of the network. Section 2 elaborates the routing protocol LSRP. Section 3 analyzes the security of LSRP and makes comparison with relevant tasks. Section 4 gives demonstration through simulation.

2. Load-Balanced Secure Routing Protocol (LSRP)

Topology control can effectively reduce energy consumption of wireless sensor network [19]. The literatures [20–23] reach the conclusion after analysis and comparison that hexagonal mesh structure can use redundant nodes to store energy and as a result has prominent advantage in effectively lengthening the service life of the network. Meanwhile, regular-shaped topology also provides applicable rules for route discovery and positioning of malicious nodes. On the basis of the approach stated in the literature [15], this section adds security design and puts into effect a secure way to generate hexagonal mesh topology, and then secure routing protocol LSRP is set up on this topology. LSRP realizes routing discovery and selection, data packet transmission and security authentication, and defense against routing attack.

2.1. Generation of Network Topology. Sensor nodes are deployed to the detected region by scattering. Before that, symmetric key K_a corresponding to the base station and temporary shared symmetric key K_{temp} are saved at each node. The latter one will be deleted after the node completes topology construction. Upon being scattered to the target region, nodes form a structured network topology made up of the same hexagonal cell in logic by broadcast communication. Formation of the topology includes four phases as follows.

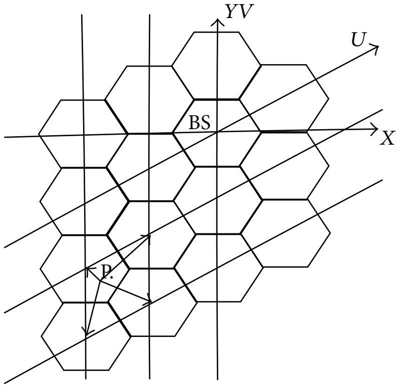


Figure 1: Calculate the ID of the RC of node P.

(1) Node initialization. In this phase, nodes acquire own and neighbors' locations. The node obtains the respective position (x, y) through GPS, then broadcasts Hello packet $((x, y), \text{MAC}_{K_{\text{temp}}}((x, y)))$ at a distance of $2a$. a is the side length of the regular hexagonal cell (shortened as RC) while $\text{MAC}_{K_{\text{temp}}}((x, y))$ is the message authentication code generated by using K_{temp} for verifying the authenticity of (x, y). The node receives Hello packet and verifies the authenticity of (x, y) through K_{temp} and $\text{MAC}_{K_{\text{temp}}}((x, y))$. After Hello packet passes authentication, the node saves the position information of neighboring nodes.

(2) Cell partition. In this phase, nodes determine which RC they affiliate to. BS broadcasts partitioning message which contains the location of BS and a. To facilitate easier notations, we introduce set of coordinates (u, v) where the V-axis coincides with the Y-axis, and U-axis is 30 degrees tilted counterclockwise from the X-axis. The (u, v) coordinates of RC center are referred to as the ID of RC. Once node P receives the partitioning message, it calculates the IDs of the four adjacent RCs, as shown in Figure 1. P then calculates the distances between itself and these RCs' centers and adapts the ID of the RC whose center is closest to it.

(3) Active node election. In this phase, active node is picked out according to the following rules: assuming G is a node coordinates set in an RC, one node whose coordinate is $\text{Min}_G\ (x, y)$ is picked out. This is the active node of the RC it belongs to. $\text{Min}_G\ (x, y)$ is defined as below: set G as the node's coordinate set; $(x, y) \in G$; any (x_0, y_0) meets the criteria of $(x_0, y_0) \in G$ and is different from (x, y); if (x, y) meets the criteria: $x < x_0$ or $x = x_0$ and $y < y_0$, then (x, y) is the minimum coordinate of G, written as $\text{Min}_G\ (x, y)$. All other nodes then enter into sleep state. Sleep node periodically sends an inquiring message to the active node, and the active node either keeps it asleep or lets it become new active node to continue its work.

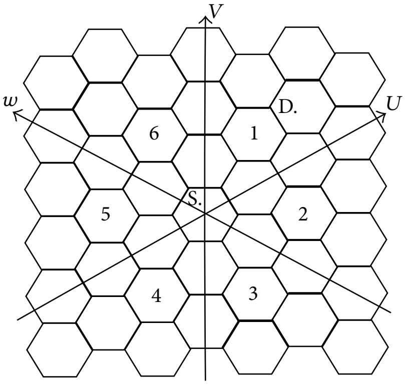


Figure 2: U, V coordinate system.

(4) Secure architecture construction. In this phase, secure architecture is constructed, that is, the communication relations are set up between RCs. Each RC's active node broadcasts request packet $((u, v), \mathrm{MAC}_{K_{\mathrm{temp}}}((u, v)))$. (u, v) is the coordinate of the node sending request packet. The active node receives request packet and verifies its facticity. Then, according to (u, v) and the node's own coordinate, the active node will be able to determine whether the node sending packet is the active node of the neighboring RC. If it is, add it to the table of neighboring RCs.

2.2. Routing Discovery and Selection. According to the fact that LSRP is based on hexagonal mesh topology, a routing discovery and selection method is designed.

The main idea of the method is as follows. First of all, calculate the number of hops in U, V, and W directions from the source node to the destination node. Then, choose the transmission routing according to the policy of the twice probability routing selection, that is, according to certain probability, choose a direction R_1 among U, V, and W, and randomly generate the number of continuous hops, T_1, in R_1 direction according to certain probability rules by referring to the total number of hops in R_1 direction and that of other directions. The packet will take T_1 hops continuously along R_1 direction. If there are unfinished hops along other directions, choose another direction R_2 according to certain rules. If there are T_2 hops along R_2 direction, take T_2 hops continuously along R_2 direction. And the like, until it comes to the destination node.

The detail of routing discovery and selection is as follows.

2.2.1. Routing Discovery. As shown in Figure 2, routing discovery is to calculate the number of hops along the shortest path from S to D, that is, u, v, w (u denotes u hops in direction U. v denotes v hops in direction V. w denotes w hops in direction W). Among u, v, w, at least one is equal to 0.

According to the above result, we designed Algorithm 1 to calculate the initial values of u, v, and w. In OPA_*uvw* algorithm, the case that two out of three directions of hops are zero is considered, which states that only one optimal path between source node and destination node. For the case, u, v, and w are updated, and two new paths whose hops are one more than that of the optimal path are added for improving performance on traffic load balance.

2.2.2. Routing Selection. According to the type of node, routing selection is divided into the following two types.

(1) Source Node Routing Selection. After source node S monitor one event, it needs to select one path in advance to transmit the event message to destination node D, that is, it needs to determine routing information (u, v, w, t, s) and routing direction. t denotes direction of packet forwarding for next hop node. s denotes hops in direction t. According to characters of u, v, and w, OPA_*uvwts* algorithm for the twice probability routing selection to calculate the values of routing information is designed as shown in Algorithm 2.

(2) Intermediate Node Routing Selection. After intermediate node has received the data packet, it needs to determine next hop routing information (u, v, w, t, s) and direction. According to routing information (u, v, w, t, s) in the data packet, update_*uvwts* algorithm is designed for computing next hop routing information and direction as shown in Algorithm 3.

2.3. Data Packet Transmission. Data packet is forwarded according to the routing computed by above routing algorithm. In order to strengthen security, acknowledgement packet, alert packet, and notice packet are additionally introduced in LSRP. Acknowledgement packet is for detect selective forwarding attack. Alert packet is for transmitting alert message containing the position of the attackers to the source node. Notice packet is for transmitting message of attack existence in the path to the source node.

The routing transmission of different types of packets is shown in Figure 3.

The realization process of the data packet transmission is as below.

Set S as the source node, B as the intermediate node, and D as the destination node.

Step 1. S: generate ($\mathrm{encry}_{\mathrm{data}}$, $\mathrm{MAC}_{\mathrm{SD}}(\mathrm{encry}_{\mathrm{data}})$, $\mathrm{MAC}_{\mathrm{OHK}}(\mathrm{counter})$, $\mathrm{routing}_{uvwts}$, direction).

Step 2. S → B: {$\mathrm{routing}_{uvwts}$, $\mathrm{encry}_{\mathrm{data}}$, counter, $\mathrm{MAC}_{\mathrm{SD}}(\mathrm{encry}_{\mathrm{data}})$, $\mathrm{MAC}_{\mathrm{OHK}}(\mathrm{counter})$}.

Step 3. B: verify_cout(counter), update_*uvwts* ($\mathrm{routing}_{uvwts}$).

Step 4. B → B: {$\mathrm{routing}_{uvwts}$, $\mathrm{encry}_{\mathrm{data}}$, counter, $\mathrm{MAC}_{\mathrm{SD}}(\mathrm{encry}_{\mathrm{data}})$, $\mathrm{MAC}_{\mathrm{OHK}}(\mathrm{counter})$}.

Step 5. B → D: {$\mathrm{routing}_{uvwts}$, $\mathrm{encry}_{\mathrm{data}}$, counter, $\mathrm{MAC}_{\mathrm{SD}}(\mathrm{encry}_{\mathrm{data}})$, $\mathrm{MAC}_{\mathrm{OHK}}(\mathrm{counter})$}.

Step 6. D: verify_cout(counter), verify_MAC ($\mathrm{encry}_{\mathrm{data}}$), decrypt($\mathrm{encry}_{\mathrm{data}}$), judge_attack().

Detailed descriptions about the data packet transmission are given below.

Step 1. The source node S constructs the data packet shown in Figure 4. The routing information (u, v, w, t, s) and

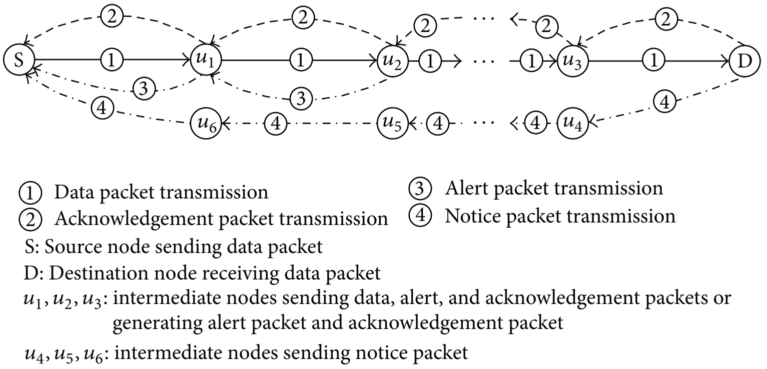

S
u_1
u_2
u_3
D
u_4
u_5
u_6
1
2
3
4
⋯
① Data packet transmission
② Acknowledgement packet transmission
③ Alert packet transmission
④ Notice packet transmission
S: Source node sending data packet
D: Destination node receiving data packet
u_1, u_2, u_3: intermediate nodes sending data, alert, and acknowledgement packets or generating alert packet and acknowledgement packet
u_4, u_5, u_6: intermediate nodes sending notice packet

Figure 3: Routing transmission of different types of packets.

```
Input: source node S $(u_s, v_s)$, destination node $D$ $(U_d, V_d)$
  $u_D = u_d - u_s$; $v_D = v_d - v_s$;
   neither of $u_D$ and $v_D$ is 0
      {if $u_D$ and $v_D$ have opposite signs,
            {if $|u_D| > |v_D|$, then $(u, v, w) = (u_D + v_D, 0, v_D)$
                if $|u_D| < |v_D|$, then $(u, v, w) = (0, v_D + u_D, u_D)$
                 if $|u_D| = |v_D|$
                  {$w = v_D$
           //convert one-way path to multi-way path; sign (x) means the sign of x
                  if $|w| > 1$, then $(u, v, w) = (\text{sign}(w)(-1), \text{sign}(w)(1), \text{sign}(w)(|w| - 1))$
                  if $|w| = 1$, then $(u, v, w) = (0, 0, v_D)$}}
       if $u_D$ and $v_D$ have the same sign, then $(u, v, w) = (u_D, v_D, 0)$}
  if one between $u_D$ and $v_D$ is 0 and the one not equal to 0 is larger than 1
    {      //convert one-way path with number of hops over 1 to multi-way path
            if $|v_D| > 1$, then $(u, v, w) = (\text{sign}(v_D)(1), \text{sign}(v_D)(|v_D| - 1), \text{sign}(v_D)(1))$
            if $|u_D| > 1$, then $(u, v, w) = (\text{sign}(u_D)(|u_D| - 1), \text{sign}(u_D)(1), \text{sign}(u_D)(-1))$}
  if between $u_D$ and $v_D$, one is 0 and the other is equal to 1, then $(u,v,w) = (u_D, v_D, 0)$
Output: $(u, v, w)$
```

Algorithm 1: OPA_*uvw*.

```
Input: $(u, v, w)$
  if one among $u$, $v$, $w$ is 0, set $w = 0$ (similar treatment in the case of $u = 0$ or $v = 0$)
        {the active node randomly chooses one direction in $U$ and $V$ by the principle of equal probability,
        supposing $U$ direction is chosen (similar treatment for $V$ direction)
        direction = sign($u$)
        //$t$ takes value 1, 2, 3, standing for $U$, $V$, $W$ direction respectively
        if $|v| = 1$, then $(u, v, w, t, s) = (\text{sign}(u)(|u| - 1), v, w, \text{sign}(u)(1), (|u| - 1))$
        if $|v| \neq 1$
              {   randomly select a figure $u1$ among $1 \sim |u|$ by the principle of equal probability,
                 if $u1 > 1$, then $(u, v, w, t, s) = (\text{sign}(u)(|u| - 1), v, w, \text{sign}(u)(1), (u1 - 1))$
                 if $u1 = 1$, then $(u, v, w, t, s) = (\text{sign}(u)(|u| - 1), v, w, \text{sign}(v)(2), |v| - 1)$}}
  if none among $u$, $v$, $w$ is 0
        {choose one direction randomly in $U$, $V$ and $W$ by the principle of equal probability
         supposing $U$ direction is chosen (similar treatment for $V$ or $W$ direction)
              direction = sign($u$)
              if $u \cdot v < 0$ and $v \cdot w > 0$, then $(u, v, w, t, s) = (0, v, w, \text{sign}(w)(3), |w|)$
              if $u \cdot v > 0$ and $v \cdot w > 0$, then $(u, v, w, t, s) = (0, v, w, \text{sign}(v)(2), |v|)$
              if $u \cdot v > 0$ and $v \cdot w < 0$, then $(u, v, w, t, s) = (u, 0, 0, \text{sign}(u)(1), |u|)$}
Output: $(u, v, w, t, s)$ and direction
```

Algorithm 2: OPA_*uvwts*.

```
Input: $(u, v, w, t, s)$
  Assuming $t$ is $U$ direction (For $V$ or $W$ direction similarly processing)
  if next hop   node in t direction is normal node
    {     direction = sign($u$)
          If $s > 1$, then $(u, v, w, t, s) = (\mathrm{sign}(u)(|u| - 1), v, w, t, (s - 1))$
          If $s = 1, v \neq 0$, then $(u, v, w, t, s) = (\mathrm{sign}(u)(|u| - 1), v, w, \mathrm{sign}(v)(2), |v|)$
          If $s = 1, w \neq 0$, then $(u, v, w, t, s) = (\mathrm{sign}(u)(|u| - 1), v, w, \mathrm{sign}(w)(3), |w|)$
          If $s = 1, v = 0, w = 0$, then $(u, v, w, t, s) = (0, 0, 0, 0, 0)$;
          If $s = 0$, it means that event packet reaches destination node }
  if next hop   node in $t$ direction has been marked with failure node
    {     if $v \neq 0$, then $\{(u, v, w, t, s) = (\mathrm{sign}(u)(|u| + 1), \mathrm{sign}(v)(|v| - 1), w, t, s + 1)$,
                      direction = sign($v$)(2)}
          if $w \neq 0$, then $\{(u, v, w, t, s) = (\mathrm{sign}(u)(|u| + 1), v, \mathrm{sign}(w)(|w| - 1), t, s + 1)$,
                      direction = sign($w$)(3)}
          if $v = 0, w = 0$, then $\{(u, v, w, t, s) = (u, 0, \mathrm{sign}(u)(-1), t, s)$,
                      direction = sign($u$)(2)}}
Output: $(u, v, w, t, s)$ and direction
```

Algorithm 3: update_*uvwts*.

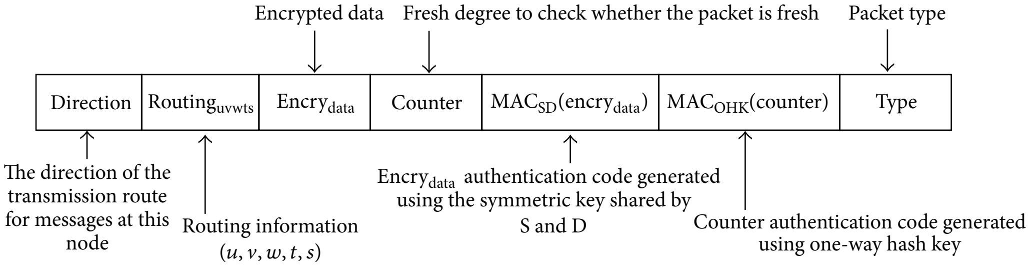

Encrypted data
Fresh degree to check whether the packet is fresh
Packet type
Direction
Routing$_{uvwts}$
Encry$_{data}$
Counter
MAC$_{SD}$(encry$_{data}$)
MAC$_{OHK}$(counter)
Type
The direction of the transmission route for messages at this node
Routing information (u, v, w, t, s)
Encry$_{data}$ authentication code generated using the symmetric key shared by S and D
Counter authentication code generated using one-way hash key

Figure 4: Format of data packet.

transmission direction can be got by OPA_*uvw* algorithm and OPA_*uvwts* algorithm (see Algorithms 1 and 2).

Step 2. S sends the data packet to the intermediate node at next hop.

Step 3. The intermediate node receives the data packet, firstly verifies the fresh degree of the packet via counter, and then according to the received (u, v, w, t, s) and above update_*uvwts* algorithm it calculates the routing$_{uvwts}$ and direction from the intermediate node to the destination node D.

Step 4. The intermediate node sends the data packet to the neighboring downstream intermediate node along the transmission direction outputted in Step 3. Next, it does the following operations.

(1) If it is the node that generates acknowledgement packet, it constructs the acknowledgement packet shown in Figure 5.

$K_{i,j,}$ the current one-way hash key used in MAC$_{OHK}$ (ACK) is calculated by

$$K_{i,j} = F^{N_t - \mathrm{int}((T_c - T_b)/L)}\left(K_{i,n}\right). \tag{1}$$

N_t is the total number of one-way hash keys that can be used by the node; T_b is the time of the key used by the node at the beginning; T_c is the current time of the node; L is the life cycle of one-way hash key; $K_{i,n}$ stands for the last key in the node's one-way hash key chain; int denotes rounding function.

Direction comes from the data packet saved in the buffer. The acknowledgement packet is sent towards upstream in the opposite direction to data packet transmission.

TTL is determined by the policy preset in the protocol, that is, the number of hops required to arrive at the previous node generating acknowledgement packet.

(2) Waiting acknowledgement packet from its downstream node. If receiving acknowledgement packet in prescribed time, it does the following operations.

(i) Check whether the one-way hash key, $K_{i,j,}$ used by the packet is valid.

If $K_{i,j,}$ meets the below criterion:

$$\frac{(T_{c2} - T_{c1})}{L} - m > 1; \tag{2}$$

T_{c2} is the time of the acknowledgement packet received this time; T_{c1} is the time of the acknowledgement packet received last time generated by the same node; L is the life cycle of one-way hash key; m meets the criterion: $K_{i,p} = F^m(K_{i,j})$;

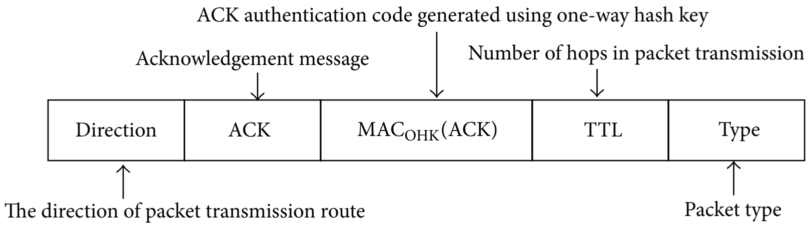

ACK authentication code generated using one-way hash key
Acknowledgement message
Number of hops in packet transmission
Direction
ACK
MAC$_{OHK}$(ACK)
TTL
Type
The direction of packet transmission route
Packet type

Figure 5: Format of acknowledgement packet.

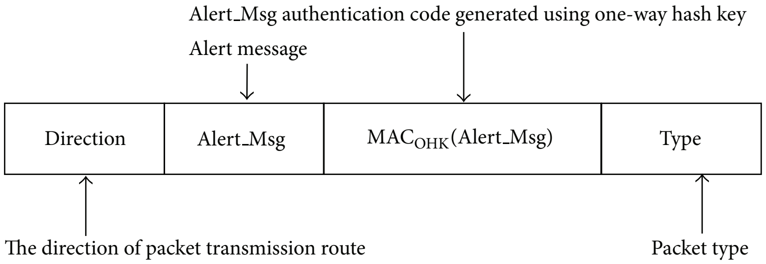

Alert_Msg authentication code generated using one-way hash key
Alert message
Direction
Alert_Msg
MAC$_{OHK}$(Alert_Msg)
Type
The direction of packet transmission route
Packet type

Figure 6: Format of alert packet.

$K_{i,j}$ and $K_{i,p}$ are the one-way hash keys received this time and last time from the node generating the acknowledge packet; then $K_{i,j}$ is considered invalid, send the alert packet illustrated in Figure 6 to S.

Alert_Msg contains information identifying its downstream neighboring node as a malicious node. Direction comes from the data packet saved in the buffer. The alert packet is sent towards upstream in the opposite direction to data packet transmission.

(ii) If the time to receive the acknowledge packet overruns the expected time, send the alert packet to S.

(iii) Check whether ACK is authentic via MAC$_{OHK}$(ACK). If not, discard the packet.

(iv) Add 1 to the number of acknowledgement packets received. If it is under the expected value and overruns the stipulated time limit, send the alert packet to S; if it is up to the expected value, delete the data packet temporarily saved in the buffer.

(v) If TTL > 0, deduct one from TTL value and send the acknowledgement packet to the upstream nodes.

Step 5. The intermediate node sends the data packet to D according to the routing information.

Step 6. The destination node receiving the data packet makes the following four operations.

(1) Via MAC$_{SD}$ (encry$_{data}$), check whether encry$_{data}$ is authentic. If not, discard the packet.

(2) Check the authenticity of counter via MAC$_{SD}$ (counter), and then check whether the packet is fresh by comparing it with relevant values of the current node. If not fresh, discard it.

(3) Decipher the data content of the packet.

(4) Check whether there is attack.

Set the number of packets received by D from S as S_r, and the number of packets already sent by S as S_s. S_r needs to be initialized and recalculated at certain interval or after attack is repaired. S_l is the number of packets sent by S contained in the last packet received at last calculation cycle or in the first packet received after attack is repaired, while σ is an adjustable parameter related to the network's packet loss rate. If

$$\frac{\left|S_s - S_l - S_r - 1\right|}{\left|S_s - S_l\right|} < \sigma, \tag{3}$$

it means that the packet is normally received and update S_r with $S_r + 1$; otherwise, it means that there is attack. D will set S to alert state and use other secondary routes (referring to the routes comprised of nodes surrounding the optimal route with one hop more than the optimal route) to send the notice packet illustrated in Figure 7 to inform S about intruding node likely to make selective forwarding attack. When S receives the notice packet, it changes the data packet transmission mode, from regular mode to detection mode.

3. LSRP Performance Analysis

We evaluate LSRP comprehensively both in theory and by simulation, with focus on analyzing its security and traffic load balance.

3.1. LSRP Security Analysis. LSRP safeguards network security from the below aspects.

(1) Defense against eavesdropping attack. In order to capture high-sensitive data transmitted between the nodes, the attacker tries to get relevant information by eavesdropping the communication link.

To make sure the packet content is breach-proof, before transmission, LSRP encrypts the packet content, as described in Step 1 of above data packet transmission, and generates encrypted message encry$_{data}$. This can prevent outsider

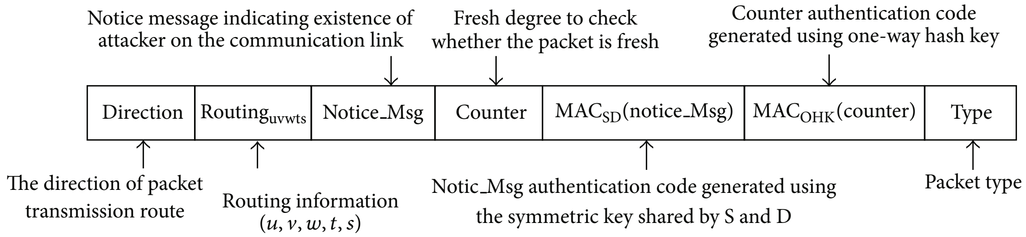

Notice message indicating existence of attacker on the communication link
Fresh degree to check whether the packet is fresh
Counter authentication code generated using one-way hash key
Direction
Routing$_{uvwts}$
Notice_Msg
Counter
MAC$_{SD}$(notice_Msg)
MAC$_{OHK}$(counter)
Type
The direction of packet transmission route
Routing information (u, v, w, t, s)
Notic_Msg authentication code generated using the symmetric key shared by S and D
Packet type

Figure 7: Format of notice packet.

attackers from eavesdropping the communication link to intercept the packet and steal its content.

(2) Defense against altering attack. If these exists insider attacker in the communication link, the insider_attacker can send a counterfeit packet to the receiver by altering the data packet and result in the receiver's making incorrect judgment or operation.

LSRP uses symmetric key and one-way hash key to generate authentication code to prevent the packet from being altered. For example, in Step 1 of above data packet transmission, symmetric key shared by S and D is used to generate MAC$_{SD}$ (encry$_{data}$) for encry$_{data}$, while in Step 4 of data packet transmission, one-way hash key is used to generate MAC$_{OHK}$ (ACK) and MAC$_{OHK}$ (Alert_Msg) for the acknowledgement packet and alert packet, respectively. After receiving the packet, the receiver verifies the authentication code. If the packet information is inconsistent, it is ascertained that the packet content has been altered and hence there exists altering attack in the communication link.

(3) Defense against replay attack. The attacker intends to drain network energy and interfere in normal packet transmission by continuously replaying the old packet.

LSRP prevents the packet from being replayed by outsider attackers by inserting counter tag, which indicating fresh degree of the packet, and its authentication code into the packet. For instance, in Step 1 of data packet transmission, counter and MAC$_{OHK}$ (counter) are used. As each receiver has a corresponding counter in itself, by comparing it with counter in the packet, it can determine whether the packet is fresh or not. If not, discard the packet. MAC$_{OHK}$ (counter) guarantees the authenticity of counter. In this way, replay attack can be prevented. Moreover, thanks to the application of counter, that is, packet fresh degree, cycling attack [24] is also counterchecked.

(4) Defense against Wormholes and Sinkhole attacks. In Wormholes attack, the attacker receives the information at one end of the network through low-latency link and at the same time by virtue of its high performance sends the information to the cahoot at the other end to replay it, so as to produce high-performance communication link, attract the nodes to use the link where the attacker lurks, and then carry out larger sabotage by combining selective forwarding attack. In Sinkhole attack, a compromise node is produced to attract almost all traffic within certain region to pass through it, creating a sinkhole centering on the attacker, and then to carry out larger destruction by combining selective forwarding attack.

From the perspective of security, one important advantage of routing protocols based on geographical position is that it makes it difficult for the attackers to make Wormholes and Sinkhole attacks [24]. LSRP belongs to this category and can well defend against Wormholes and Sinkhole attacks. Routing protocols constructing topology initiated by base station, such as REAR [25], are prone to Wormholes and Sinkhole attacks. In the construction of the topology used by LSRP, the geographical positions of the base station and local nodes, the side length of RC and localized interaction are adopted, which make Wormholes unable to come into being. As the transmission route of data packet is realized by the policy for the twice probability routing selection proposed in this paper, the traffic is naturally routed to the physical position of the base station and is hardly attracted to other places to form sinkhole. Consequently, LSRP is almost immune to Wormholes and Sinkhole attacks.

(5) Defense against Sybil attack. A feature of Sybil attack is that the attacker keeps changing identity to attract as many packets as possible to go through it in the disguise of nodes at different positions and then carries out larger sabotage by combining selective forwarding attack. Sybil attack poses huge threat to multipath routing and geographical position based routing. Routing protocols mentioned in the literatures [5, 16, 17, 21, 26] are prone to Sybil attack.

LSRP is a routing protocol based on geographical position and therefore prone to Sybil attack. LSRP defends against Sybil attack by using symmetric key. In order to make Sybil attack, the attacker needs to put the disguised node in the transmission route of data packet. According to LSRP, to become a transmitting node in the route, the node needs to save its information in the neighboring nodes. One node accepts another node as its neighboring node in the course of topology construction. In the topology construction process given in Section 2, message authentication code generated using K_{temp} is used for identity authentication between the nodes. Without K_{temp}, the attacker cannot pass packet authentication, accordingly cannot disguise a node to become the neighboring node of other legal nodes and hence incapable of making attack.

Even if the attacker captures the node and gets the symmetric key, in LSRP, it is difficult to disguise itself as other nodes and make Sybil attack, for the below reasons: in LSRP, as each node and the base station share a unique symmetric key and the ID of each node is verified via the symmetric key, the attacker can hardly get the symmetric key of several nodes by capturing one node to disguise itself as several nodes. Hence, it is hard to make Sybil attack in this way.

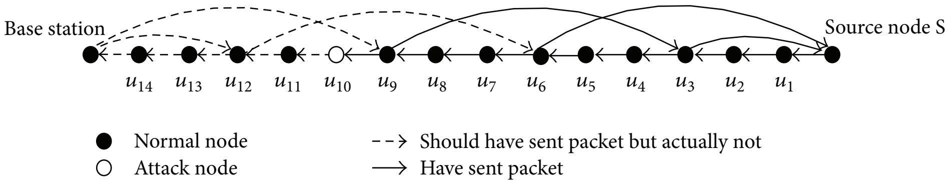


Figure 8: The case that the attacker discards data packet and does not return acknowledgement packet.

(6) Defense against HELLO FLOOD attack. In HELLO FLOOD attack, by right of high-power transmission, the attacker makes many nodes believe that it is their neighbor and causes those nodes send packets to an unknown place. As a result, the network is plunged into a mess.

Similar to the defensive measures against Sybil attack, LSRP also uses symmetric key to defend against HELLO FLOOD attack. HELLO FLOOD implements attack by making many legal nodes believe it is their neighbor, while the key of Sybil attack also lies in turning the attacker into the neighboring node of the legal nodes. These two types of attacks differ in the radiated power and the destruction target of the attackers. The approach for verification of legal neighboring nodes adopted in the defense against Sybil attack is also applicable in the defense against HELLO FLOOD attack. With it, the attacker is unable to win the legal nodes' trust and is rejected from adding to the neighbor table of the legal nodes. Hence, HELLO FLOOD attack is effectively prevented in the same way.

(7) Defense against selective forwarding attack. In selective forwarding attack, the attacker gains its end to sabotage network information by forwarding some information only and discarding the other. For some other attacks aimed at routing, such as Wormholes, Sinkhole, Sybil, they usually unite with selective forwarding attack to exert huge destructive force. Therefore, defending against selective forwarding attack is of great importance. Moreover, as this attack discards packet selectively and is more concealed, defense is even more difficult and the countermeasures are more complicated.

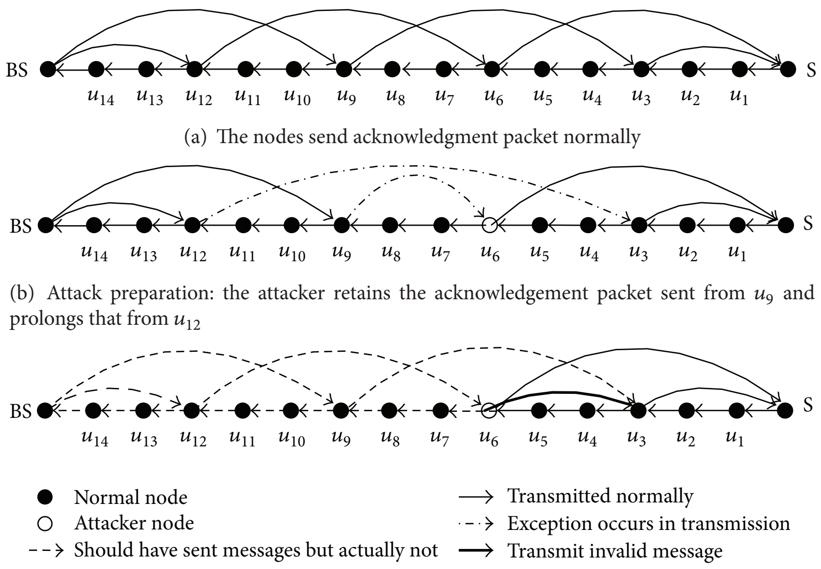
In LSRP, selective forwarding attack is detected by checking the number of packets sent by the source node and the number of packets already received from the source node accord with formula (3). When an attack is detected, in Step 4 of data packet transmission, a measure for positioning and detecting selective forwarding attack is provided to search for the position of the intruding node. This measure can detect the position of the attacker in the case of the following three attacks with time- and acknowledgement-based multihop detection technology:

(i) The attacker randomly discards packets and does not return acknowledging packets. LSRP chooses some nodes from the route to return acknowledgement packet to its upstream nodes, who then decide whether the neighboring downstream node is an attacker according to the number of received acknowledgement packets. For example, in the case of Figure 8, the attacker u_{10} discards the packet from u_9; therefore u_{10}'s downstream nodes are unable to send acknowledgement packet, which causes u_9, u_8, and u_7 to receive one acknowledgement packet only (if there is no attack, two acknowledgement packets should be received by each node). Then, u_9 generates an alert packet, reporting that u_{10} is an attacker, and sends it to S. u_8 and u_7 might also generate alert packet, but S can fix on the position of the attacker according to the last "time node seeing the previous data packet." Therefore, the judgment can be formed that the attacker specified by u_9 is the real attacker.

(ii) When the attacker finds that there is attack detection action, it does not discard the packet but intentionally prolongs the time to return acknowledgement packet. Delayed reply of acknowledgement packet causes upstream nodes far away from the attacker unable to receive the acknowledgement packet and consequently generates an alert packet by mistake, which causes legal nodes to be mistaken for the attacker. In Step 4 of data packet transmission, LSRP validates whether the downstream neighboring node is an attacker by checking the interval between sending the data packet and receiving the acknowledgement packet. If the interval overruns certain threshold value, it is affirmed that the downstream neighboring node is an attacker.

(iii) The case as illustrated in Figure 9 occurs. It is divided into two stages: attack preparation and attack implementation. At the former stage, the attacker intercept the acknowledgement packet, so as to intercept the one-way hash key K_x needed for fabricating an acknowledgement packet at next stage. At the later stage, the attacker discard new receiving data packet, fabricates a new acknowledgement packet with key K_x and sends it to the upstream. In this case, despite the acknowledgement packet is used for detection, it is hard to find the attacker's position though it has discarded the data packet. Regarding this problem, LSRP realizes prevention of malicious altering of packets by stipulating the Time to Live of each key in one-way hash key chain. When a node receives an acknowledgement packet, formula (2) is used to check if the key is within the valid time. If yes, keep upward transmission; otherwise, generated an alert packet and send it to S.

(8) Defense against acknowledgement spoofing attack. In acknowledgement spoofing attack, the attacker eavesdrops the packet sent to other neighboring nodes, sends acknowledgement spoofing packet to the source node that sends the



(a) The nodes send acknowledgment packet normally

(b) Attack preparation: the attacker retains the acknowledgement packet sent from u_9 and prolongs that from u_{12}

(c) Attack implementation: the attacker sends counterfeit acknowledgement packet of u_9 to upstream

FIGURE 9: One case of selective forwarding attack.

packet and makes it believe that a weak link is robust or an expired link is "alive"; hence packet loss is incurred.

This kind of attack can be regarded as a particular case of selective forwarding attack, because the destination node cannot receive packets sent by the source node as the attacker sends false acknowledgement packet and leads to data packet loss. LSRP can find out the position of the invalid RC by the approach of detecting the position of the intruding node in selective forwarding attack and treats the invalid RC as the attacker of selective forwarding. In this way, though the real attacker sending the false acknowledgement packet is not dealt with, it is not capable of acknowledgement spoofing attack anymore, because a better communication link is chosen to realize secure packet transmission. Hence, acknowledgement spoofing attack is effectively prevented.

3.2. LSRP Traffic Load Balance Analysis. As a secure routing protocol, LSRP features routing selection based on hexagonal mesh topology, one prominent advantage of which is that the route is determined only in relation with the node's coordinate, dispensing with generation of aroute leading to the destination node by flooding or searching for other destination nodes in other directions. It can save the energy consumed in routing searching. DPRA [15] is also a routing protocol based on hexagonal mesh topology, but it has only realized routing selection, and hasn't taken routing security into account. In addition, though DPRA is intended to pick out a suitable routing via the probability formula $P = (P_u, P_v, P_w) = (|u| + |v| + |w|)^{-1}(|u|, |v|, |w|)$ to balance network traffic load, it is still inferior to LSRP in traffic load balance. This section analyzes traffic load balance of LSRP in comparison with DPRA.

We analyze the load of RCs passed by packets when packets are sent from the source node S (u_s, v_s) to the destination node D (u_d, v_d). Suppose $w = 0$ (analysis is the same in the case of $u = 0$ or $v = 0$). $P_{i,j}$ is set to denote the probability of packet's passing through intermediate RC node $M(u_k, v_k)$, $i = |u_k - u_s|$, $j = |v_k - v_s|$.

(1) When M falls into the middle RCs as shown in the shaded part of Figure 10, the $P_{i,j}$ of LSRP and DPRA accords with formula (4) and (5), respectively,

$$P_{i,j} = \frac{1}{2} \cdot \frac{1}{|u|} + \frac{1}{2} \cdot \frac{1}{|v|}, \tag{4}$$

$$\begin{aligned} P_{i,j} = {} & P_{i-1,j}\frac{|u| - i + 1}{|u| + |v| - i - j + 1} \\ & + P_{i,j-1}\frac{|v| - j + 1}{|u| + |v| - i - j + 1}. \end{aligned} \tag{5}$$

From formula (4) and (5), we know that in LSRP the probability of packet's passing through M is $(1/2)(1/|u| + 1/|v|)$, while that in DPRA is related both to the values of u and v and to the node's position; therefore load balance in LSRP is superior to that in DPRA.

(2) When M falls into the surrounding RCs in 1, 2, 3, and 4 parts of Figure 10, traffic load balance is analyzed as follows.

(i) When M is a node of part 1, the $P_{i,j}$ of LSRP and DPRA accords with formula (6) and (7), respectively, with value falling into $[1/2|u|, 1/2]$ and $[\prod_{t=0}^{|u|-1}((|u| - t)/(|u| + |v| - t)), |u|/(|u| + |v|)]$, respectively,

$$P_{i,j} = \frac{1}{2} \cdot \frac{|u| - i + 1}{|u|}, \tag{6}$$

$$P_{i,j} = \prod_{t=0}^{i-1} \frac{|u| - t}{|u| + |v| - t}. \tag{7}$$

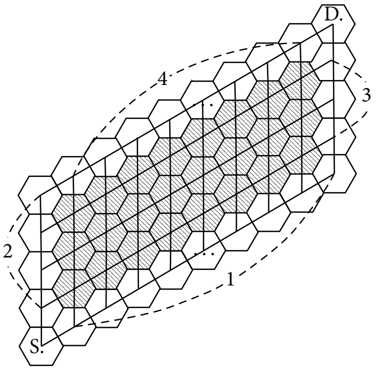


Figure 10: RCs traversed by the data packet.

(ii) When M is a node of part 2, the $P_{i,j}$ of LSRP and DPRA accords with formula (8) and (9), respectively, with value falling into $[1/2|v|, 1/2]$ and $[\prod_{t=0}^{|v|-1}((|v| - t)/(|u| + |v| - t)), |v|/(|u| + |v|)]$, respectively,

$$P_{i,j} = \frac{1}{2} \cdot \frac{|v| - j + 1}{|v|}, \tag{8}$$

$$P_{i,j} = \prod_{t=0}^{j-1} \frac{|v| - t}{|u| + |v| - t}. \tag{9}$$

(iii) When M is a node of part 3, the $P_{i,j}$ of LSRP and DPRA accords with formula (10) and (5), respectively, with value falling into $[(1/\ 2)((1/|v|) + (1/|u|)), (1/2)((1/|u|) - (1/|v|) + 1)]$ and $[\prod_{t=0}^{|v|-2}((|v| - t)/(|u| + |v| - t)), |v|/(|u| + |v|)]$,

$$P_{i,j} = \frac{1}{2} \cdot \frac{1}{|u|} + \frac{1}{2} \cdot \frac{j}{|v|}. \tag{10}$$

(iv) When M is a node of part 4, the $P_{i,j}$ of LSRP and DPRA accords with formula (11) and (5), respectively, with value falling into $[(1/2)((1/|v|) + (1/|u|)), (1/2)((1/|v|) - (1/|u|) + 1)]$ and $[\prod_{t=0}^{|u|-2}((|u| - t)/(|u| + |v| - t)), |u|/(|u| + |v|)]$,

$$P_{i,j} = \frac{1}{2} \cdot \frac{i}{|u|} + \frac{1}{2} \cdot \frac{1}{|v|}. \tag{11}$$

The above analysis shows that the probability of packets' passing through M node in 1, 2, 3, and 4 regions in LSRP and DPRA falls into interval $A = [\min(1/2|u|, 1/2|v|), \max((1/2)((1/|v|)-(1/|u|)+1), (1/2)((1/|u|)-(1/|v|)+1))]$ and interval $B = [\min(\prod_{t=0}^{|u|-1}((|u| - t)/(|u| + |v| - t)), \prod_{t=0}^{|v|-1}((|v| - t)/(|u|+|v|-t)), \max(|u|/(|u|+|v|)), |v|/(|u|+|v|)]$. As $A \subset B$, it can be deduced that node load balance in 1, 2, 3, and 4 regions in LSRP is superior to that in DPRA.

4. Simulation Experiment

We evaluated LSRP in depth through simulation in NS2. As a security mechanism has been added and the protocol itself is secure, experimental evaluation mainly focuses on load balance of network traffic and energy consumption of the network. As for the scenario of experiment, assuming that 3000 nodes are randomly generated and distributed over 632 RCs on an 800 × 800 m^2 site, each RC only has one active node, the sensitive radius of the nodes in the RC is 20 m and the communication radius is 40 m. The experiment uses the same energy consumption model as described in literature [27], with the initial energy of each node set to 100 J.

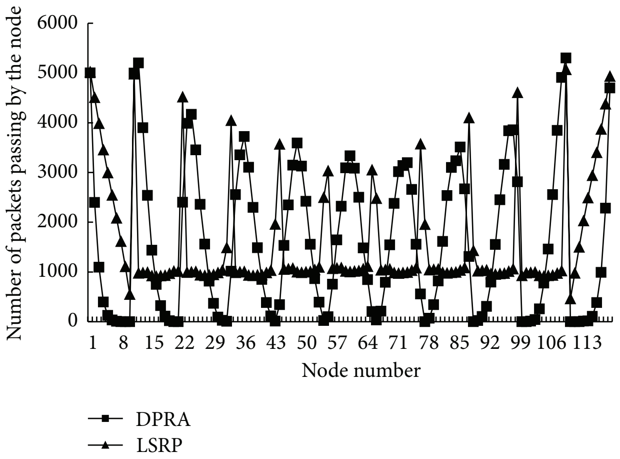


Figure 11: Load from the node (−10, 10) to the node (0, 0).

4.1. Network Traffic Load Balance. Figures 11 and 12 map the simulation results of sending 10000 data packets from the source nodes RC(−10, 10), RC(−10, 6) to the destination node RC(0, 0), respectively. In the chart, Node No. refers to the sequential number of the nodes ordered in u direction on multiple optimal routes. From Figures 11 and 12, it can be seen that LSRP features better load balance than DPRA under the condition that hops in U, V, and W directions are balanced or not balanced. They also show that the more unbalanced the hops in U, V, and W directions, the better load balance is realized by LSRP than by DPRA. The experiment figures reflect that LSRP has indeed improved traffic load balance.

4.2. Network Energy Consumption. In order to defend against various kinds of attacks, in addition to symmetric key and one-way hash key chain, LSRP also adds acknowledgement packet, alert packet and notice packet. Transmitting these three kinds of packets intended to defend against selective forwarding attack increases energy consumption. By comparing the solution to defend against selective forwarding attack in LSRP with that proposed in the literature [28] by Xiao et al., we illustrate the issues of energy consumption of the nodes and delay of data packets.

Figures 13 and 14 examine the energy consumption situation of the nodes when the transmission interval is 0.02 seconds (i.e., conditions with packet loss; Figure 15 shows a situation of packet loss when 1000 data packets are sent at different intervals and without attack), under the condition of without attacker or with attacker. From the charts, we can

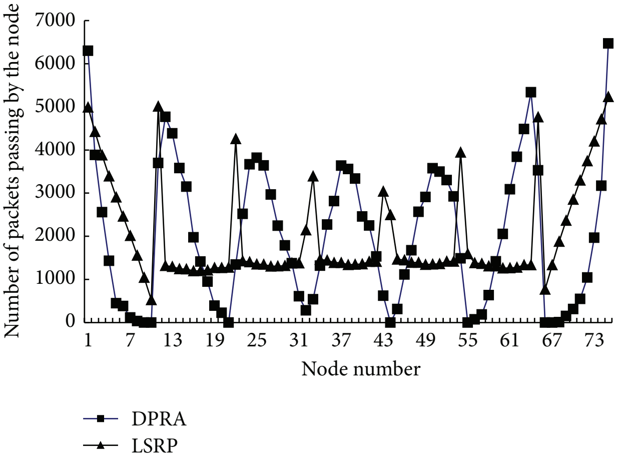


FIGURE 12: Load from the node (−10, 6) to the node (0, 0).

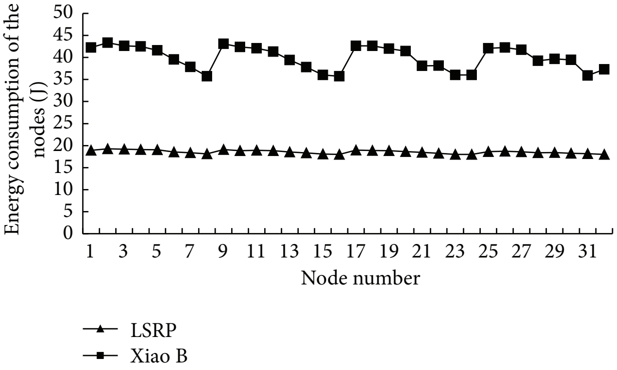


FIGURE 13: Energy consumption of each node on the optimal route after 1000 packets are sent from the source node (−7, −3) to the destination node (0, 0), without attack.

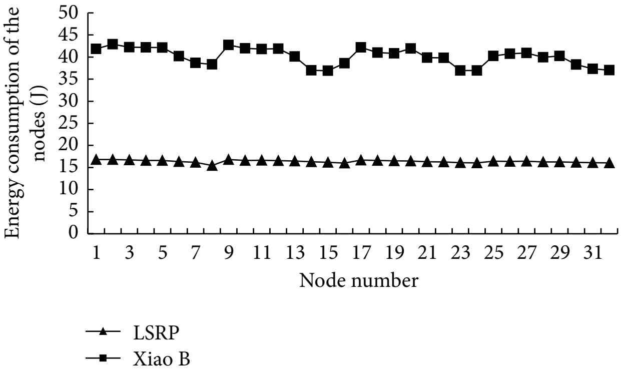


FIGURE 14: Energy consumption of each node on the optimal route after 1000 packets are sent from the source node (−7, −3) to the destination node (0, 0), with one node making attack.

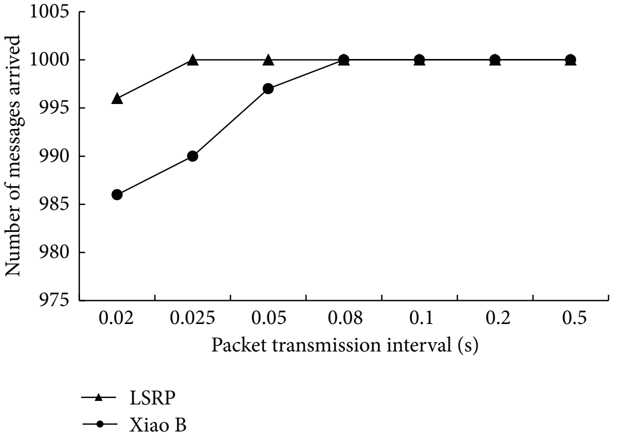


FIGURE 15: The number of messages received by the destination node after 1000 packets are sent from the source node (−7, −3) to the destination node (0, 0), without attack.

see that in LSRP energy consumption at each node is lower than that in Xiao's solution. This is because LSRP only invokes attack detection solution when attack is spotted and returns to the status of no attack detection after the attacker is located and dealt with; while in Xiao's solution, attack detection is done every time a data packet is sent, therefore extra energy is consumed.

Figure 16 examines the average energy consumption of the nodes under different transmission intervals. The chart shows that in the case of the same transmission interval, the average energy consumption in LSRP is less than that in Xiao's solution. This is because LSRP only invokes attack detection solution when attack is spotted and returns to the status of no attack detection after the attacker is located and dealt with; while in Xiao's solution, attack detection is done every time a data packet is sent, therefore extra energy is consumed. Meanwhile, Figure 16 also reveals the trend of descending at first and then ascending gradually of the average energy consumption, that is network energy consumption is closely related to the frequency of packet transmission. Descending at the beginning is because the network becomes less busy and less crowded, packet loss is reduced, and accordingly the number of alert packets decreases; consequently, energy consumption of the nodes is reduced. Later, the average energy consumptions mounts up because under no network congestion and no packet loss, as the packet transmission interval lengthens, so does the node's idle time. However, the node still consumes energy at idle time, so more and more energy is consumed.

Figures 17 and 18 examine delay of data packets under without attacker and with one attacker. The charts show that delay of data packets is relevant to the frequency of packet transmission. In the case of short transmission interval, the packet arrival time in LSRP is much shorter than that in Xiao's solution. This is also because LSRP only invokes attack detection solution when attack is spotted; therefore the number of acknowledgement packets is less than that of Xiao's solution, so is packet delay or congestion. When packet transmission interval is larger than 0.08 seconds and lengthens gradually, the arrival time of LSRP is a little shorter than and very close to that of Xiao's solution. This is because as the transmission interval lengthens, so does the node's idle time. The percentage of energy consumed at idle time

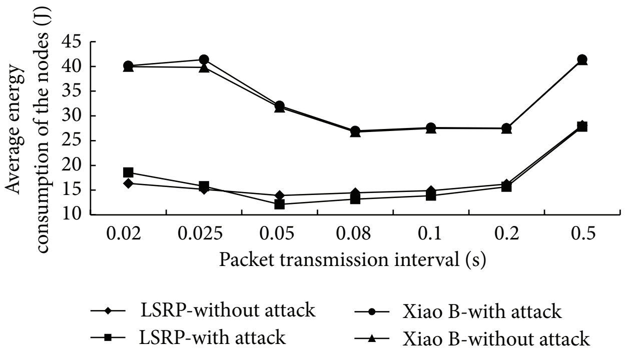


Figure 16: Average energy consumption of the nodes on the optimal route after 1000 packets are sent from the source node (−7, −3) to the destination node (0, 0).

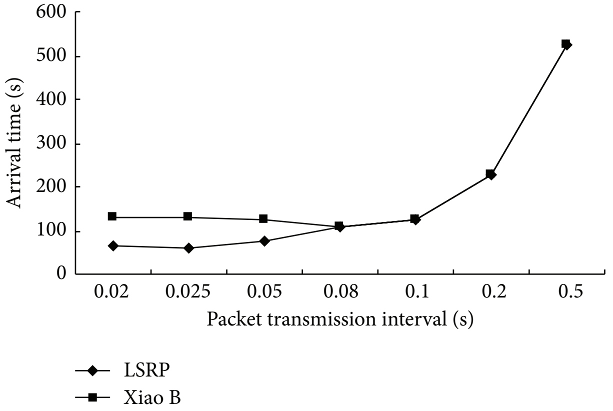


Figure 17: Time to arrive at the destination node after 1000 packets are sent from the source node (−7, −3) to the destination node (0, 0), without attacker.

increases and energy consumption of the node hinges on its idle time.

5. Conclusion

With rapid development of wireless sensor network applications, to guarantee routing reliability of the sensor network is a fundamental requirement to the security of the entire network and has become the major challenge in the research on wireless sensor security applications. This paper proposed an load-balanced WSN secure routing protocol, LSRP. Based on energy-saving hexagonal mesh topology, LSRP realizes security control over sensor network routing by making use of encryption technology, one-way hash key chain, and symmetric key technology and topology structure based on geographical position. In addition, through the policy of the twice probability optimized routing selection, it allows each RC to share data transmission more evenly, balances network traffic load, and effectively prevents some RCs from dying too quickly, and consequently lengthens the life cycle of WSN.

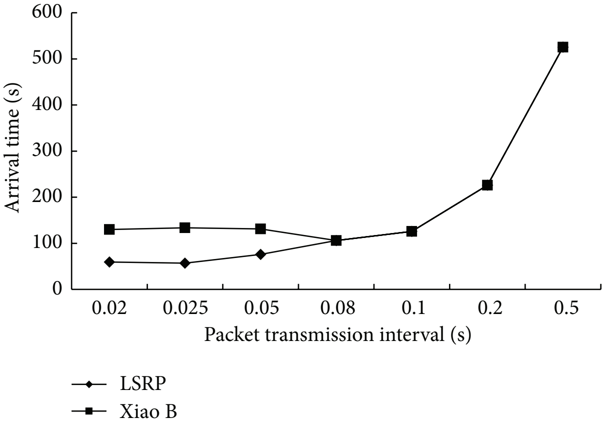


Figure 18: Time to arrive at the destination node after 1000 packets are sent from the source node (−7, −3) to the destination node (0, 0), with one attacker.

Acknowledgments

This work is supported by the Natural Science Foundation of China under Grant no. 61272074, the Natural Science Foundation of Jiangsu Province under Grant no. BK2011464, and the project of the Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education and the Foundation of Jiangsu University under Grant nos. 12JDG104 and 12JDG103. And the author Wang Liang-min is supported by the Disguised Researcher Program of Jiangsu Province of China (2012-wlw-020), and the academic leader is supported by Qinglan Project of Jiangsu Province of China.

References

[1] D. R. Raymond and S. F. Midkiff, "Denial-of-service in wireless sensor networks: attacks and defenses," *IEEE Pervasive Computing*, vol. 7, no. 1, pp. 74–81, 2008.

[2] M. Sadeghi, F. Khosravi, K. Atefi, and M. Barati, "Security analysis of routing protocols in wireless sensor networks," *International Journal of Computer Science Issues*, vol. 9, no. 1, pp. 465–472, 2012.

[3] X. Ren, "Security methods for wireless sensor networks," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '06)*, pp. 1925–1930, June 2006.

[4] Q.-Q. Pei, Y.-L. Shen, and J.-F. Ma, "Survey of wireless sensor network security techniques," *Journal of China Institute of Communications*, vol. 28, no. 8, pp. 113–122, 2007.

[5] B. Karp and H. T. Kung, "GPSR: greedy Perimeter Stateless Routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM '00)*, pp. 243–254, August 2000.

[6] T. J. Sebastian, "Secure route discovery against wormhole attacks in sensor networks using mobile agents," in *Proceedings of the 3rd International Conference on Trendz in Information Sciences and Computing (TISC '11)*, pp. 110–115, December 2011.

[7] S. Madria and J. Yin, "SeRWA: a secure routing protocol against wormhole attacks in sensor networks," *Ad Hoc Networks*, vol. 7, no. 6, pp. 1051–1063, 2009.

[8] A. Perrig, R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler, "SPINS: security protocols for sensor networks," *Wireless Networks*, vol. 8, no. 5, pp. 521–534, 2002.

[9] M. G. Shiva, R. J. D'Souza, and G. Varaprasad, "Digital signature-based secure node disjoint multipath routing protocol for wireless sensor networks Source," *IEEE Sensors Journal*, vol. 12, no. 10, pp. 2941–2949, 2012.

[10] C. Yin, S. Huang, P. Su, and C. Gao, "Secure routing for large-scale wireless sensor networks," in *Proceedings of the International Conference on Communication Technology (ICCT '03)*, pp. 1282–1286, Institute of Electrical and Electronics Engineers, April 2003.

[11] N. El-Bendary, O. S. Soliman, N. I. Ghali, A. E. Hassanien, V. Palade, and H. Liu, "A secure directed diffusion routing protocol for wireless sensor networks," in *Proceedings of the 2nd International Conference on Next Generation Information Technology (ICNIT '11)*, pp. 149–152, June 2011.

[12] J. Deng, R. Han, and S. Mishra, "INSENS: intrusion-tolerant routing for wireless sensor networks," *Computer Communications*, vol. 29, no. 2, pp. 216–230, 2006.

[13] S. Tanachaiwiwat, P. Dave, R. Bhindwale, and A. Helmy, "Secure locations: routing on trust and isolating compromised sensors in location-aware sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys' 03)*, pp. 324–325, Association for Computing Machinery, November 2003.

[14] M. García-Otero, T. Zahariadis, F. Álvarez et al., "Secure geographic routing in ad hoc and wireless sensor networks," *Eurasip Journal on Wireless Communications and Networking*, vol. 2010, Article ID 975607, 2010.

[15] M. A. Hamid, M. Mamun-Or-Rashid, and S. H. Choong, "Defense against lap-top class attacker in wireless sensor network," in *Proceedings of the 8th International Conference Advanced Communication Technology (ICACT '06)*, pp. 314–318, February 2006.

[16] N. Nasser and Y. Chen, "Secure multipath routing protocol for wireless sensor networks," in *Proceedings of the 27th International Conference on Distributed Computing Systems Workshops (ICDCSW '07)*, p. 12, Institute of Electrical and Electronics Engineers, June 2007.

[17] N. Nasser and Y. Chen, "SEEM: secure and energy-efficient multipath routing protocol for wireless sensor networks," *Computer Communications*, vol. 30, no. 11-12, pp. 2401–2412, 2007.

[18] C. Karlof, Y. Li, and J. Polastre, "ARRIVE: algorithm for robust routing in volatile environments," Tech. Rep. UCB/CSD-03-1233, Computer Science Department, University of California at Berkeley, 2002.

[19] P. Santi, "Topology control in wireless ad hoc and sensor networks," *ACM Computing Surveys*, vol. 37, no. 2, pp. 164–194, 2005.

[20] H. Zhang and A. Arora, "GS3: scalable self-configuration and self-healing in wireless sensor networks," *Computer Networks*, vol. 43, no. 4, pp. 459–480, 2003.

[21] X. Wang and T. Berger, "Topology control, resources allocation and routing in wireless sensor networks," in *Proceedings of the IEEE Computer Society's 12th Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS '04)*, pp. 391–399, IEEE Computer Society, October 2004.

[22] X. Wang and T. Berger, "Self-organizing redundancy-cellular architecture for wireless sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '05)*, pp. 1945–1951, Institute of Electrical and Electronics Engineers, March 2005.

[23] X.-S. Wang, Y.-Z. Zhan, and L.-M. Wang, "STCP: secure topology control protocol for wireless sensor networks based on hexagonal mesh," in *Proceedings of the4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, pp. 1–4, IEEE Computer Society, 2008.

[24] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 293–315, 2003.

[25] H. Hassanein and J. Luo, "Reliable energy aware routing in wireless sensor networks," in *Proceedings of the 2nd IEEE Workshop on Dependability and Security in Sensor Networks and Systems (DSSNS '06)*, pp. 54–62, Institute of Electrical and Electronics Engineers Computer Society, April 2006.

[26] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Personal Communications*, vol. 7, no. 5, pp. 16–27, 2000.

[27] 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.

[28] B. Xiao, B. Yu, and C. Gao, "CHEMAS: identify suspect nodes in selective forwarding attacks," *Journal of Parallel and Distributed Computing*, vol. 67, no. 11, pp. 1218–1230, 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 240230, 10 pages
http://dx.doi.org/10.1155/2013/240230

Research Article

Detecting and Defending against Replication Attacks in Wireless Sensor Networks

Xiang-yi Chen,[1] Li-xia Meng,[2] and Yong-zhao Zhan[1]

[1] *School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China*
[2] *Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, Anhui University, Hefei 230039, China*

Correspondence should be addressed to Yong-zhao Zhan; yzzhan@ujs.edu.cn

Received 27 February 2013; Accepted 18 May 2013

Academic Editor: Lu Liu

Wireless sensor networks are susceptible to node replication attacks due to their unattended nature. Existing replicas detection schemes can be further improved in regard of detection probabilities, detection overheads, and the balance of detection overheads among sensor nodes. In this paper, we make the following contributions: first, we point out the unrealistic assumption that the replica node would behave honestly as the benign sensor nodes; thus the existing detection schemes would fail if the replica nodes cheat or collude with the compromised node. Then, we propose a location-binding symmetric key scheme forcing the replica nodes to be inserted only in the vicinity of the compromised node. Later, a detecting scheme is presented to inspect the location claims within the neighborhood. Finally, analysis shows that our scheme helps to detect and defend against replication attacks effectively and efficiently. Extensive simulations are conducted and the results show that the detection overheads are low and evenly distributed among all the sensor nodes.

1. Introduction

Wireless sensor networks (WSNs) are generally deployed in the unattended environments for some missions, such as environment monitoring and enemy surveillance. The unattended nature and the lack of tamper-resistant hardware cause wireless sensor networks to be vulnerable to various insider attacks, threatening the operation of WSNs.

Replication attack is one of the insider threats. The attacker captures one or more sensor nodes, tampers with them and obtains the credential materials, such as the identity and keys, then clones some nodes as replica nodes, and surreptitiously inserts these replicas in the network. Subsequently, the attacker may launch a variety of insidious attacks, such as data injection, selecting forwarding, routing loop, or even topology partition. Just as shown in Figure 1, a network was formed by the normal nodes (without frame). The captured and compromised nodes are represented in the solid frame, and replica nodes are represented in the dashed frame.

Thus, detection of replica nodes becomes one research hotspot in WSN [1]. The first distributed replication detection schemes RM and LSM were proposed by Parno et al. [2]. In RM scheme, nodes broadcast to neighboring nodes the location claim message signed by ID-based public key scheme. Then the neighbors forward such received claim message with a specified probability to randomly selected network nodes, which act as witness. According to the birthday paradox [1], the nodes owning the same ID would select same witness nodes with a big probability. These witness nodes eventually detect replicas successfully. To further increase the detection probability, LSM scheme is also proposed. In this scheme, the nodes in the forwarding path of the claim messages also store and compare the messages. Thus witness line segment is formed from the source to the destination; then the witness line of the same ID will cross at some node with a large probability and the node at the cross point acts as the witness node. Compared with RM scheme, the detection probability is increased at the cost of memory storage. However, LSM scheme has the crowded-center problem because the witness line is prone to cross at the center of network with a big probability.

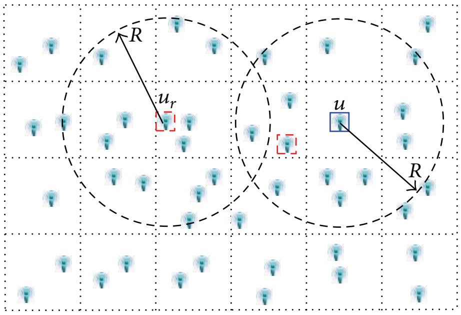


Figure 1: Replication attacks in wireless sensor network.

2. Related Work

The existing detection schemes can be classified as centralized approaches and distributed approaches.

2.1. Centralized Detection Approaches. The schemes in [3–6] assume a central base station to conduct the detection. Choi et al. [3] proposed to detect the replica nodes by set. The network is divided into disjoint subregions. A header node is enumerated to report the member list to the base station in each subregion. The reports from all of the header nodes are computed by set. The intersection of two sets is checked; any nonempty intersection implies the existence of the replica sensor node. Brooks et al. [4] proposed a centralized scheme to detect replication attacks by using random key predistribution. Every sensor node should report the usage of its keys. If the usage of some key exceeded the threshold, then the sensor node was identified to be suspicious. Ho et al. [5] presented a SPRT method for replica detection in mobile sensor networks, in which the base station checks whether the speeds of the mobile sensor nodes exceed the threshold. Based on a state-of-the-art signal processing technique, compressed sensing, Yu et al. [6] proposed CSI to detect replication attacks.

2.2. Distributed Detection Approaches. In distributed approaches [2, 7–14], the replication attacks detection is conducted by reporting the location claim messages to randomly chosen witness nodes in the network. Paradoxes of the location claims indicate the detection of replication attacks. To further improve the detection probability, Conti et al. [7] proposed RED scheme, in which a random seed was shared and upgraded in the network. The same random seed and the same pseudorandom function result in the same witness node chosen by replica nodes and the compromised node. But it is difficult to share and upgrade such random seed across the whole network. Zhu et al. [8] proposed another detection scheme by using localized multicast. Ho et al. [9] proposed to take advantage of the group deployment knowledge to further raise detection probability and lower detection overheads. Zhang et al. [10] proposed four detection schemes, B-MEM, BC-MEM, C-MEM, and CC-MEM, to address the cross-over problem and the crowded-center problem in the detection. Li and Gong [11] proposed RDE scheme to utilize the local neighborhood geographic information for replication attacks detection. Zeng et al. [12] proposed two detection schemes, RAWL and TRAWL, to distribute the witness sensor node to the network. Wang and Shi [13] introduced mobile patrollers to detect replica nodes; the result shows this solution is effective and also energy efficient to prolong the lifetime of network. Xing and Cheng [14] proposed two replication detection schemes from both the time domain and the space domain in MANETs (mobile ad hoc networks). The basic idea is to utilize a cryptographic one-way hash function to force the replica sensor nodes to keep on generating paradoxes.

It is assumed by existing schemes that compromised sensor node and the replica node carry out the detection procedures honestly. However, it is not always true; it is more likely that the program code in the compromised sensor node and the replica sensor node has been modified by the attacker for the purpose of escaping from being detected. To make things worse, the replica node may collude with the compromised node, which will lead to the failure of existing detection schemes.

In this work, we seek to detect and defend against the replication attacks with fewer communication, computation, and memory overheads than previous works. We propose a location-binding pairwise key scheme, forcing the attacker to insert the replica nodes to the vicinity of the compromised node. Then, the neighbor sensor nodes around the replica sensor nodes are the first possible witnesses to detect the replication attacks.

The remainder of this paper is organized as follows: the next section illustrates the network model and assumptions; Section 4 proposes our location-binding pairwise key management scheme which is used in our scheme to defend against the replication attacks; Section 5 presents our replicas detection scheme; Section 6 analyzes the security and efficiency of our scheme, and extensive simulations supporting the analytical findings are also shown. Discussion follows in Section 7. Section 8 concludes the proposed scheme.

3. Network Model and Assumptions

In this paper, we assume that there are only stationary sensor nodes in the wireless sensor network. We also assume that the communications between the stationary sensor nodes are bidirectional, which is also an assumption of most of previous detection schemes.

Stationary nodes can get their geographic location by using positioning device (e.g., GPS device) or positioning algorithms [15–18]. Also, we assume that all the sensor nodes are loosely time synchronized using time synchronization techniques, such as [19, 20]. Consider

$$f(\alpha, \beta) = \sum_{i,j=0}^{t-1} a_{ij} \alpha^{i} \beta^{j}. \qquad (1)$$

Prior to network deployment, we assume that a trusted authority (TA) chooses one t-degree bivariate symmetric polynomial in (1) with the coefficients a_{ij} $(0 \le i, j \le t-1)$ over the finite field GF(q), where q is a prime large enough to accommodate the keys. The symmetry of the polynomial assures the equation $f(\alpha, \beta) = f(\beta, \alpha)$, which also implies the equation of the coefficients $a_{ij} = a_{ji}$. Also, TA select a cryptographic one-way hash function $h(\cdot)$, which maps arbitrary-length input into fixed-length output. TA preloads every sensor node with the symmetric polynomial $f(\alpha, \beta)$ and the hash function $h(\cdot)$. Table 1 lists all the relevant symbols and the corresponding meaning.

4. Location-Binding Pair-Wise Key Scheme

Most of the existing detection schemes explore the public key algorithms for encryption and digital signature. Even if public key algorithms are feasible and available in wireless sensor networks, the large computation overheads will exhaust the limited battery power and pose great challenges to the resource-limited sensor nodes.

Due to the low cost, small factor, and limited resources in the sensor nodes, symmetric encryption algorithms are more suitable for wireless sensor networks. In [21], the key management in distributed sensor networks is studied. Fei et al. [22] proposed a time-space related symmetric key predistribution scheme. Wang et al. [23] proposed an updateable key management scheme with intrusion tolerance through symmetric polynomial and one-way hash chain.

We assume the sensor nodes adopt a location-binding pair-wise key scheme (LBK), which tightly binds the sensor node's identity with its geographical location to resist against node capture and replica nodes. In the following, we first propose our location-binding pair-wise key management scheme.

4.1. Sensor Deployment and Localization. After sensor nodes are deployed into the targeted region, each sensor node carries out the bootstrap procedure as described by Algorithm 1. The sensor node obtains its geographical location (x, y). Then, the sensor nodes use their location (x, y) and the collision-resistant one-way hash function $h(\cdot)$ to compute its identity as $\mathrm{ID} = h(x \| y)$, where $\|$ denotes message concatenation. In this way, we bind each sensor node's identity tightly with its geographic location to resist against replication attack.

Once the identity is generated, a sensor node, for example, sensor node u, generates its key ring polynomial $K_u(\beta) = f(\mathrm{ID}_u, \beta)$, then erases the symmetric polynomial $f(\alpha, \beta)$ permanently.

All the sensor nodes are required to conduct the previous process within short time duration $T1$ lest the attacker captures the symmetric polynomial.

4.2. Neighborhood Pairwise Key Establishment. Every sensor node broadcasts its identity within its radio range to establish its neighborhood. Then each sensor node uses its neighbor lists to calculate the relevant pair-wise key for encryption and decryption. For example, every sensor node establishes pair-wise keys with its neighboring sensor node according to Algorithm 2.

To defend against node attacks from compromising t sensor nodes and deriving the symmetric polynomial $f(\alpha, \beta)$, every sensor node, for example, sensor node u, is required to get its geographic location and computes its key ring polynomial $K_u(\beta)$ within a time limit $T1$ after bootstrap. Once $K_u(\beta)$ is generated or $T1$ expires every sensor node erases the relevant polynomial $f(\alpha, \beta)$ permanently. Algorithm 1 describes such a bootstrap procedure.

Table 1: Notation summary.

Notation	Meaning
n	Total number of sensor nodes
u, v, w	Sensor nodes u, v, and w
u_r	Replica sensor node of u
u_w	Witness sensor node of u and u_r
$f(\alpha, \beta)$	Bi-variate symmetric polynomial
$h(x)$	One-way hash function
$K_u(\beta)$	Key ring polynomial of sensor node u
ID_u, ID_v, ID_w	Respective identity of sensor nodes u, v, and w
K_{uv}, K_{uw}, K_{vw}	Pairwise key between sensor nodes u, v, and w
Δ	Tolerance of time synchronization
R	Sensor communication range
B	Average number of sensor nodes in the communication range
N	Number of witness sensor nodes
d	Distance between u and u_r
γ	The ratio of d to R
S	Area of the deployment region of the network
S_s, S_c, S_w	Area of sensor communication, blind zone, and witness zone

Also, every sensor node, for example, sensor node u, should establish neighborhood and compute the pair-wise keys with at most $t-1$ neighbor nodes within a time limit $T2$ after key ring polynomial $K_u(\beta)$ is generated. Once keys are computed or $T2$ expires every sensor node erases its key ring polynomial $K_u(\beta)$ permanently. In this way, even if the attacker compromises a sensor node, $t-1$ keys are not enough to deduce the key ring polynomial $K_u(\beta)$. Algorithm 2 details the whole process.

For those sensor nodes that have more than $t-1$ neighbor nodes, the pair-wise keys can be negotiated through the common neighbor nodes in their neighborhood. For example, in Figure 2, sensor node u has more than $t-1$ neighbor nodes, in which sensor node v has already established the pair key K_{uv} with node u according to Algorithm 1; also node v has established pair-wise key K_{vw} with node w, while sensor node u has to negotiate with node w to share a pair-wise key. The negotiation needs 4 steps as shown in Figure 2:

(1) node u sends a negotiation request message to node w, including the identity of u and w, the message type (*Nego_Req*), the neighbor list of node u, and a nonce. The nonce is used to defend against replay attack;

(2) node w receives node u's neighbor list, chooses one common node, for example, node v, generates the shared key K_{uw} randomly, and answers node u with

```
Start timeout timer $T1$
Get the location $l_u = (x, y)$
Compute sensor identity $\mathrm{ID}_u = h(x \| y)$
if $(T1)$ then
    Compute key ring polynomial $K_u(\beta) = f(\mathrm{ID}_u, \beta)$
    Erase the symmetric polynomial $f(\alpha, \beta)$
else
    Erase the symmetric polynomial $f(\alpha, \beta)$
    Halt
end if
```

Algorithm 1: Sensor node bootstrap procedure.

```
Broadcast($\mathrm{ID}_u, l_u$)
Initialize and start timeout timer $T2$
$c = 0$
while $(T2$ && $c < t - 1)$ do
    Receive $(\mathrm{ID}_v, l_v)$ from a neighbor node
    If $|l_u - l_v| \le R$ and $\mathrm{ID}_v$ not in the neighbor list
        Compute pair-wise key $K_{uv} = K_u(\mathrm{ID}_v)$
        Store $\mathrm{ID}_v$ and $K_{uv}$
        $c = c + 1$
    end if
end while
Erase the key ring polynomial $K_u(\beta)$
```

Algorithm 2: Direct neighborhood pair-wise key establishment.

the answer message, including the identity of u and w, the message type (*Nego_Anw*), and the secret part, which contains node v's identity, the nonces of nodes u and w, encrypted with the shared key K_{uw};

(3) node w sends a message to node v, which includes the identities of nodes w and v, the message type (*Nego_Ask*), and the secret part, containing the identity of nodes u and v, the nonce of w, and the shared key K_{uw}, encrypted with the shared key K_{vw};

(4) node v replies to node u with a message, including the identities of v and w, the message type (*Nego_Rly*), and the secret part, which contains the identities of w and u, the shared key K_{uw}, and the nonce of node w, encrypted with the shared key K_{uv}.

Node u decrypts to get K_{uw} and then decrypts the secret part of message from node w to verify the identity of node v and the nonce of node w. If the verifications pass, then node u stores ID_w and K_{uw}; otherwise node w is removed from the neighbor list of node u.

5. Replicas Detection

Replica nodes detection can be conducted right after the sensor nodes are initialized and location-based pair-wise keys are generated. Node location inspection is carried out to detect the forged nodes who claim to have the benign nodes' location. If the node passes location inspection, then the encrypted messages are deciphered using the pair-wise key according to the node's identity. If this decryption fails, then the node is identified as a replica node.

Because of the above pair-wise key scheme and the location-binding sensor nodes' identity, the attacker is restricted in placing the replica nodes after compromising benign nodes and making replicas. The best strategy for the attacker is to put the replica nodes in the vicinity of the compromised node. This way, the replica nodes can utilize the stolen credentials to communicate with the neighbor sensor nodes of the compromised node and hide in the network without being detected.

Thus, the replica nodes are restricted to a region around the sensor node. Based on this idea, we propose a replication detection scheme within neighborhood.

Every sensor node broadcasts location claim message to its neighborhood periodically:

$$\mathrm{Message}_u : (\mathrm{ID}_u, l_u, t_u, \mathrm{MAC}) . \qquad (2)$$

ID_u, l_u, t_u, and MAC denote the identity, location, time, and message authentication code, respectively. The sensor nodes in the neighborhood carry out Algorithm 3 to conduct replication attack detection.

When the sensor node v receives a message from its neighbor node u, it computes the distance to check whether or not the claim is issued from its neighbor sensor nodes. If this check fails, the sensor node who issued this location claim is identified as a replica node.

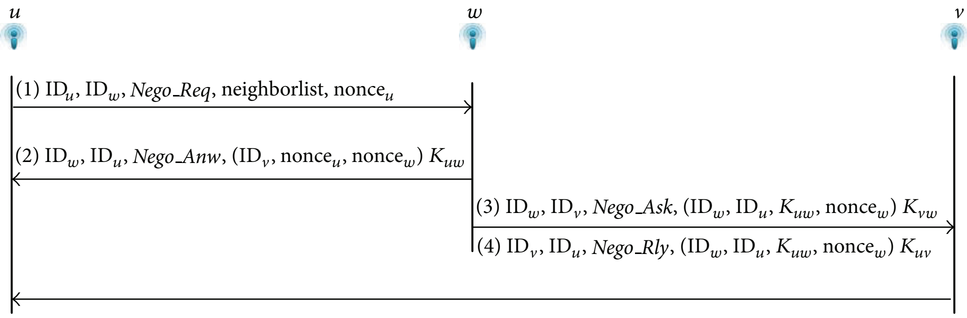

u
w
v
(1) ID_u, ID_w, *Nego_Req*, neighborlist, nonce_u
(2) ID_w, ID_u, *Nego_Anw*, (ID_v, nonce_u, nonce_w) K_{uw}
(3) ID_w, ID_v, *Nego_Ask*, (ID_w, ID_u, K_{uw}, nonce_w) K_{vw}
(4) ID_v, ID_u, *Nego_Rly*, (ID_w, ID_u, K_{uw}, nonce_w) K_{uv}

Figure 2: Pair-wise key negotiation.

```
Input: $ID_u, l_u, t_u$, MAC
Output: detection result: 0 = not a replica node, 1 = replica, 2 = N/A (not available)
function NDRA($ID_u, l_u, t_u$, MAC)
  if $h(ID_u \| l_u \| t_u) \neq MAC$ and $|t_u - t_v| > \Delta$ then
    return 2
  end if
  if $|l_u - l_v| > R$ and $h(x \| y) \neq ID_u$ then
    return 1
  end if
    return 0
end function
```

Algorithm 3: Neighborhood detection of replication attack.

6. Analysis and Simulations

To inspect our defending and detection scheme, both numerical and simulations results are shown in this section. First, we give the security analysis in Section 6.1, then the numerical results follow in Section 6.2, and simulation results are shown in Section 6.3.

6.1. Security Analysis

Definition 1. The communication coverage of a sensor node is a circle with center at the sensor node and radius equal to the communication range R. The area is written as S_s.

The sensor nodes in the communication coverage of sensor node u are neighbor sensor nodes of u.

Definition 2. Blind zone is defined as the common communication coverage of sensor node u and its replica sensor node u_r. Sensor nodes in the blind area cannot distinguish sensor nodes u and u_r. This zone and its area are written as S_c.

Definition 3. Witness zone is defined as the region in the communication coverage of sensor node u_r but not in the communication coverage of sensor node u. Sensor nodes in the witness zone, written as witness sensor nodes, can detect the replica sensor node u_r. This zone and its area are written as S_w.

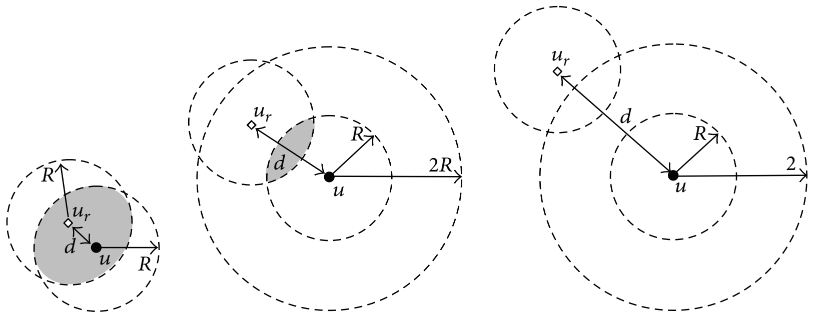
We also have the following facts.

> *Fact 1.* Replica node u_r can be detected if and only if at least one witness sensor node exists in witness zone.
>
> *Fact 2.* Sensor node u and its replica node u_r can collude if and only if there are only two sensor nodes in the blind zone: u and u_r.

Scenario 1: The Distance between u and u_r: $d \leq R$. If the attacker compromises sensor node u, then it makes one replica node u_r, deploys u_r in a distance $d \leq R$, just as shown in Figure 3(a). Then node u_r shares the same credentials as u. The common neighbor nodes in the blind zone cannot distinguish u from u_r. While sensor node in the witness zone of node u_r can detect the forged geographic location of node u_r. So, if the attacker wants to make u_r undetected, he must make sure that there are no sensor nodes in the witness zone S_w. Thus, replica node u_r can cheat the sensor nodes in the blind zone S_c using the legitimate security credentials.

In scenario 1, u and u_r cannot collude, because sensor nodes in the blind zone can detect the colluding message in which the identities are the same. So, if u wants to collude with u_r, there should be no sensor nodes other than u and u_r in the blind zone S_c.

Lemma 4. *Replica sensor node u_r stays undetected and colludes with sensor node u, if and only if $N = 1$; that is, replica sensor node u_r has only one neighbor sensor node u.*



(a) $d \leq R$ (b) $R \leq d \leq 2R$ (c) $d > 2R$

Figure 3: Replication node u_r in three scenarios.

Proof. This lemma holds according to Fact 1 and Fact 2. □

This lemma indicates that it's difficult for the attacker to insert replica sensor node u_r in the sensor network ensuring that u_r stays undetected and colludes with u.

Lemma 5. *If replica sensor node u_r stays undetected and does not collude with sensor node u, then replica sensor node u_r can cheat all the sensor nodes in the blind zone of sensor node u and replica sensor node u_r.*

Proof. This lemma holds trivially. □

This lemma means that we need some new schemes to detect the replica sensor node u_r.

Scenario 2: The Distance between u and u_r: $R < d \leq 2R$. In Figure 3(b), if replica node u_r is placed at a distance $d \in (R, 2R]$ away from sensor node u, then node u and node u_r cannot communicate directly, but there are still possibilities that u and u_r share common neighbor nodes. Even though compared with scenario 1, the common area between node u and node u_r is smaller, the threats of replication attacks still exist.

Lemma 6. *If replica sensor node u_r is inserted at a distance $R < d \leq 2R$ away from u, then u and u_r cannot collude.*

Proof. This lemma holds trivially. □

Scenario 3: The Distance between u and u_r: $d > 2R$. In Figure 3(c), if the replica node u_r is deployed away from node u at a distance larger than $2R$, then node u_r not only cannot communicate with node u directly, but also shares no common neighbor node with node u. So, node u_r will have to establish pair-wise keys with neighbor nodes around, the identity of u_r is the same with u; and it is easy to be detected. But if the attacker changes node u_r's identity, it is impossible to deduce the key ring polynomial without the symmetric polynomial. So, node u_r can only eavesdrop on the communications around itself, whereas it cannot decrypt any of the messages without relevant pair-wise key. Actually, node u_r is isolated to die. The attacker would not benefit too much from scenario 3.

Lemma 7. *In scenario 3 with $u - u_r$ distance $d > 2R$, u_r can stay undetected if and only if u_r is isolated without any neighbor sensor nodes.*

Proof. This lemma holds according to *Fact 1.* □

6.2. Numerical Discussion. In this section, we execute the mathematical analysis to get the numerical results of our proposed defending and detection scheme.

Given that there are n sensor nodes evenly distributed in the network, the area of network deployment region is S, the communication radius of the radio is R, and the detection probability in three scenarios is calculated as follows.

The blind zone in scenario 1 and scenario 2 is the common communication area between the benign node u and the replication node u_r, which can be computed as in (3). Then the area of the witness zone of u_r can be computed by (4). In Figure 4, the common area S_c can be computed as

$$\begin{aligned}
S_c &= 2\,(\text{area of sector AuB} - \text{area of triangle AuB}) \\
&= 2\left(\left(\frac{2\beta}{2\pi}\right)\times\pi R^2 - \left(\frac{1}{2}\right)\times\left(\frac{d}{2}\right)\times 2\left(\sqrt{R^2-\left(\frac{d}{2}\right)^2}\right)\right) \\
&= 2\left(\left(\arccos\left(\frac{d}{2R}\right)\div\pi\right)\times\pi R^2 - 0.25d\times\left(\sqrt{4R^2-d^2}\right)\right), \\
S_c &= 2R^2\arccos\left(\frac{d}{2R}\right) - 0.5d\sqrt{4R^2-d^2}.
\end{aligned} \tag{3}$$

Sensor nodes in the blind zone cannot distinguish sensor node u from replica node u_r; however, the sensor nodes in the witness zone can detect the replica node u_r by checking the distance between the claimed location of u_r. So, to detect u_r, the more nodes in the witness zone, the better. Once the sensor nodes are deployed in the region, the detection relies more on the area of the witness zone. The area of the witness zone can be calculated as

$$\begin{aligned}
S_w &= S_s - S_c \\
&= \pi R^2 - 2R^2\arccos\left(\frac{d}{2R}\right) + 0.5d\sqrt{4R^2-d^2}.
\end{aligned} \tag{4}$$

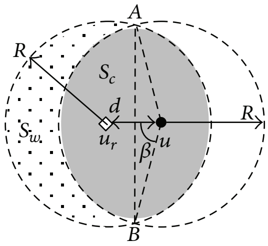


FIGURE 4: Blind zone S_c and witness zone S_w.

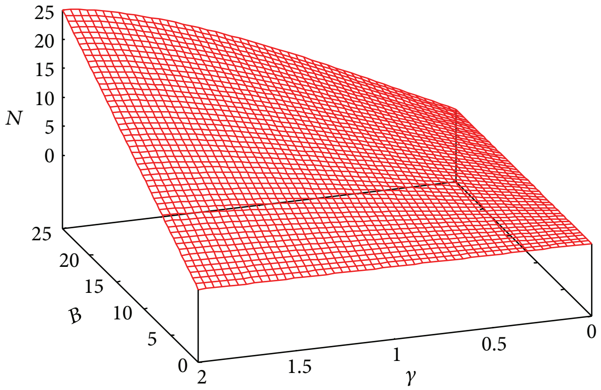


FIGURE 5: The number of witness nodes.

The number of sensor nodes depends on the distribution of the sensor nodes in the deployment area. Suppose that the sensor nodes are evenly distributed in the area. Then the density of sensor nodes is n/S and the sensor nodes in the witness zone can be computed as

$$N = \left(\pi R^2 - 2R^2 \arccos\left(\frac{d}{2R}\right) + 0.5d\sqrt{4R^2 - d^2}\right) \times \left(\frac{n}{S}\right). \tag{5}$$

To detect the replica node u_r, the number of sensor nodes in the witness zone must be greater than or equal to 1 as in the inequality

$$N \geq 1. \tag{6}$$

To facilitate our analysis, we let γ denote the ratio between the distance of $u - u_r$ and the communication radius R as in (7). In scenario 1 and scenario 2, γ is in the range of 0 to 2 $(0 \leq \gamma \leq 2)$, while in scenario 3, γ is greater than 2. Consider

$$\gamma = \frac{d}{R}. \tag{7}$$

Also, we let B denote the average number of sensor nodes in the communication range. B can be derived as in (8). The greater B is, the more neighbor sensor nodes exist:

$$B = \frac{\pi R^2}{S} \times n. \tag{8}$$

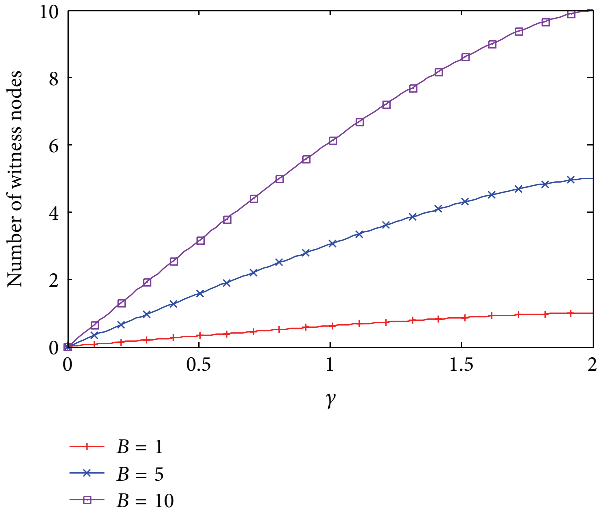


FIGURE 6: Number of witness nodes.

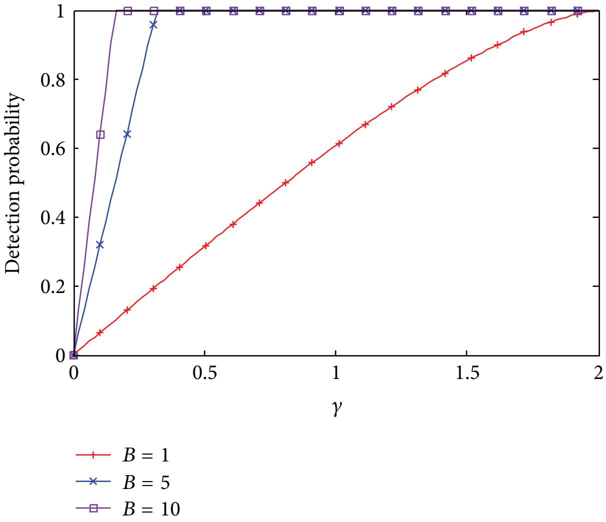


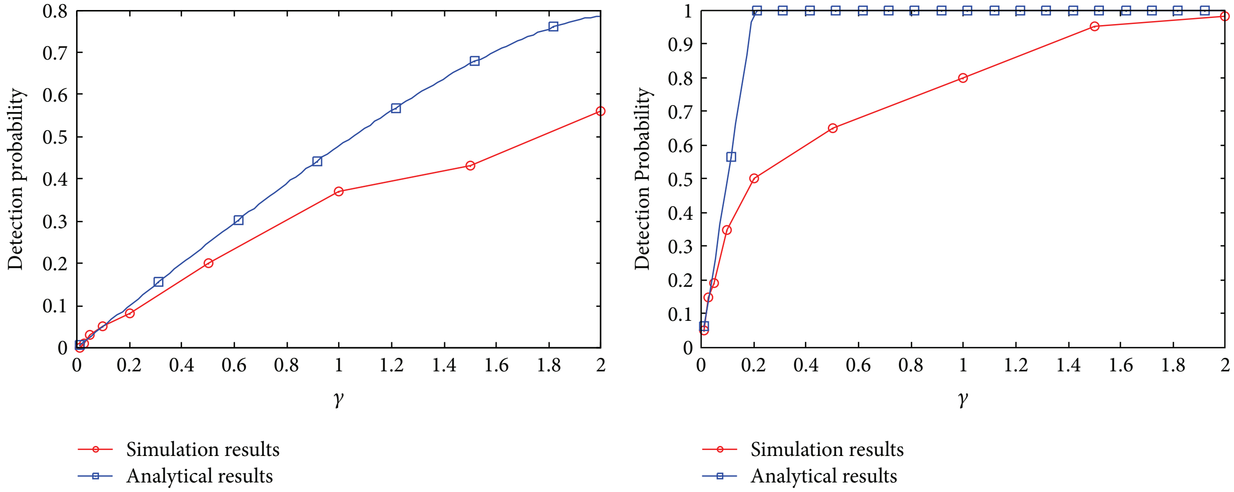
FIGURE 7: Detection probability.

Then (5) can be rewritten as function of γ and B, yielding

$$N = \left(\frac{1}{\pi}\right) \times B \times \left(\pi - 2\arccos(0.5\gamma) + 0.5\gamma\sqrt{4 - \gamma^2}\right). \tag{9}$$

We can see from (9) that the number of witness sensor nodes on the replica node u_r is only related with B and γ. Figure 5 shows that the number of witness sensor nodes changes with γ and B, where B is assumed to be in the range of 0 to 10. Figure 6 shows that the number of the witness nodes changes with the ratio γ when fixing B to 1, 5, and 10, respectively.

If N is greater than or equal to 1, it means that the replica node u_r would be detected definitely. Otherwise, there would be some false positives in the detection scheme. To ensure



(a) Simulations with $n = 100$

(b) Simulation with $n = 1000$

Figure 8: Simulation results and analytical results.

a 100% detection probability without false positives, B and γ should satisfy the inequality

$$N \geq 1,$$
$$\left(\frac{1}{\pi}\right) \times B \times \left(\pi - 2\arccos(0.5\gamma) + 0.5\gamma\sqrt{4-\gamma^2}\right) \geq 1,$$
$$B \geq \frac{\pi}{\left(\pi - 2\arccos(0.5\gamma) + 0.5\gamma\sqrt{4-\gamma^2}\right)}. \tag{10}$$

This means that we can get higher detection probability by increasing the density of sensor nodes. Meanwhile, it also means that the attacker has to insert the replica node close to the compromised sensor node u to stay undetected. The closer the replica node is to the compromised node, the less the potential threat of the replication attacks is

At last, we let P denote the detection probability as in (11). Figure 7 shows that the detection probability changes with the ratio γ when B is fixed to be 1, 5, and 10, respectively. Consider

$$P = \begin{cases} N, & B < \dfrac{\pi}{\left(\pi - 2\arccos(0.5\gamma) + 0.5\gamma\sqrt{4-\gamma^2}\right)}, \\ 1, & \text{otherwise}. \end{cases} \tag{11}$$

6.3. Detection Performance. To inspect the detection scheme, simulations in different scenarios are carried out in the NS2 network simulator [24].

6.3.1. Replication Detection. In the simulations, the WSN deployment area is 5000 meters both in the width and in the length. The wireless nodes' physical radio model is TwoWay-Ground and the MAC layer protocol is IEEE 802.11. In NS2, the default communication radius of the wireless node is 250 meters.

Simulations are conducted in two network deployments; there are 100 sensor nodes in one deployment, while there are 1000 sensor nodes in the other deployment. In both deployments, the sensor nodes are distributed randomly with a uniform distribution. There is only one replica sensor node in every round of the simulations; the replica sensor node claims to own a location-binding identity, but the identity is forged. The average count of neighbor nodes can be computed with (7), yielding the average sensor nodes count; $B = 0.7854$ and $B = 7.854$, respectively.

We carried out 100 rounds simulations in both deployments. The detection probability was computed as the ratio of the successful detection rounds to the total 100 rounds. We change the distance ratio γ defined in (6) from 0.0125, 0.025, 0.05, 0.1, 0.2, 0.5, 1.0, 1.5 to 2.0 in the simulations, in order to examine the detection probability in 3 scenarios.

The results are shown in Figure 8, in which the analytical results are also shown for comparison. It can be seen that the larger the ratio γ is, the larger the detection probability is, which fits the analytical results.

6.3.2. Detection Overhead. At last, simulations are carried out to analyze the detection costs of our scheme. In the simulations, the total number of sensor nodes is fixed to $n = 1000$; only one replica node is randomly selected from all the sensor nodes.

Figure 9 shows the results. Every stationary sensor node has to send about 1 data packet to its neighbor sensor nodes, while it has to receive about at most 7 data packets from its neighbor nodes on average. That is, in total, at most 8 data packets have to be transmitted on average.

7. Comparison with Related Work

In our scheme, every sensor node asynchronously reports its location-based identity to its neighbor nodes and also has to receive claim messages from all its neighbor nodes. The total communication and computation overheads incurred by the detection are $O(B)$. Memory requirement to store the

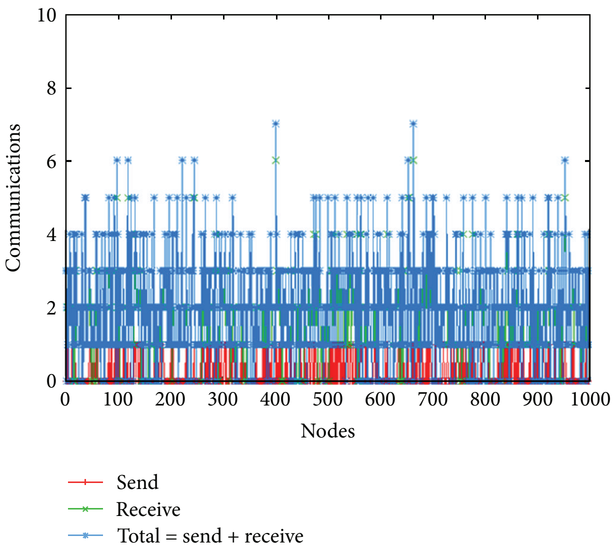


Figure 9: Detection overheads.

Table 2: Comparison with other schemes.

	Communications	Memory
RM [2]	$O(n)$	$O(\sqrt{n})$
LSM [2]	$O(\sqrt{n})$	$O(\sqrt{n})$
RED [7]	$O(\sqrt{n})$	$O(1)$
CSI [6]	$O(logn)$	—
Our scheme	$O(B)$	$O(B)$

pair-wise keys with neighbor nodes in every sensor node is $O(B)$.

The detection algorithm is executed by every node in the network in a distributed manner, eliminating the packets forwarding in the network. Thus, the detection overheads are well balanced over the network.

Table 2 compares the detection overheads with RM [2], LSM [2], RED [7], and CSI [6] schemes, in which B represents the average number of sensor nodes in the communication range.

As distributed approaches, RM [2], LSM [2], and RED [7] all assume public key schemes, which poses great challenges to the resource-limited sensor nodes. Yet, as a centralized scheme, CSI assumes the base station to conduct detection, which has the common problems of centralized approaches, such as single-point failure and unbalanced overheads among the sensor networks [1].

8. Conclusions

We present a detecting and defending scheme against replication attacks within neighborhood by using location-binding symmetric key scheme. The replica nodes are restricted to the vicinity of the compromised node. The location claim messages would not be forwarded, eliminating the overheads of other sensor nodes out of the neighborhood.

The analysis results show that the average number of sensor nodes and the distance between the compromised sensor node and the replica node determine the detection probability. The greater sensor nodes is, the higher the number of the detection probability; the farther apart the compromised node is from the replica node, the higher the detection probability is. Simulation results confirm the analytical findings. Compared with previous replica detection schemes, our scheme greatly lowers relevant detection overheads.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 61202474, the Natural Science Foundation of Jiangsu Province under Grant no. BK2011464, and the Project of the Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, Anhui University.

References

[1] L. M. Wang, F. Li, S. M. Xiong et al., "Research on detection methods for insidious attack of wireless sensor networks," *Computer Science*, vol. 38, no. 4, pp. 97–129, 2011.

[2] B. Parno, A. Perrig, and V. Gligor, "Distributed detection of node replication attacks in sensor networks," in *Proceedings of the IEEE Symposium on Security and Privacy (IEEE S and P '05)*, pp. 49–63, Oakland, Calif, USA, May 2005.

[3] H. Choi, S. Zhu, and T. F. L. Porta, "SET: detecting node clones in sensor networks," in *Proceedings of the 3rd International Conference on Security and Privacy in Communication Networks (SecureComm '07)*, pp. 341–350, Nice, France, September 2007.

[4] R. Brooks, P. Y. Govindaraju, M. Pirretti, N. Vijaykrishnan, and M. T. Kandemir, "On the detection of clones in sensor networks using random key predistribution," *IEEE Transactions on Systems, Man and Cybernetics Part C*, vol. 37, no. 6, pp. 1246–1258, 2007.

[5] J. Ho, M. Wright, and S. K. Das, "Fast detection of replica node attacks in mobile sensor networks using sequential analysis," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 1773–1781, Rio de Janeiro, Brazil, April 2009.

[6] C. M. Yu, C. S. Lu, and S. Y. Kuo, "CSI: compressed sensing-based clone identification in sensor networks," in *Proceedings of the 8th IEEE International Workshop on Sensor Networks and Systems for Pervasive Computing (PERCOM '12)*, pp. 290–295, Lugano, Switzerland, 2012.

[7] M. Conti, R. Di Pietro, L. V. Mancini, and A. Mei, "A randomized, efficient, and distributed protocol for the detection of node replication attacks in wireless sensor networks," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 80–89, Quebec, Canada, September 2007.

[8] B. Zhu, V. G. K. Addada, S. Setia, S. Jajodia, and S. Roy, "Efficient distributed detection of node replication attacks in sensor networks," in *Proceedings of the 23rd Annual Computer Security Applications Conference (ACSAC '07)*, pp. 257–266, Miami Beach, Fla, USA, December 2007.

[9] J. Ho, D. Liu, M. Wright, and S. K. Das, "Distributed detection of replica node attacks with group deployment knowledge in

wireless sensor networks," *Ad Hoc Networks*, vol. 7, no. 8, pp. 1476–1488, 2009.

[10] M. Zhang, V. Khanapure, S. Chen, and X. Xiao, "Memory efficient protocols for detecting node replication attacks in wireless sensor networks," in *Proceedings of the 17th IEEE International Conference on Network Protocols (ICNP '09)*, pp. 284–293, Princeton, NJ, USA, October 2009.

[11] Z. Li and G. Gong, "Randomly directed exploration: an efficient node clone detection protocol in wireless sensor networks," in *Proceedings of the IEEE 6th International Conference on Mobile Adhoc and Sensor Systems (MASS '09)*, pp. 1030–1035, Macao, China, October 2009.

[12] Y. Zeng, J. Cao, S. Zhang, S. Guo, and L. Xie, "Random-walk based approach to detect clone attacks in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 5, pp. 677–691, 2010.

[13] L. Wang and Y. Shi, "Patrol detection for replica attacks on wireless sensor networks," *Sensors*, vol. 11, no. 3, pp. 2496–2504, 2011.

[14] K. Xing and X. Cheng, "From time domain to space domain: detecting replica attacks in mobile ad hoc networks," in *Proceedings of the IEEE (INFOCOM '10)*, San Diego, Calif, USA, March 2010.

[15] S. Čapkun and J. Hubaux, "Secure positioning of wireless devices with application to sensor networks," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, pp. 1917–1928, Miami, Fla, USA, March 2005.

[16] A. Caruso, S. Chessa, S. De, and A. Urpi, "GPS free coordinate assignment and routing in wireless sensor networks," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, pp. 150–160, Miami, Fla, USA, March 2005.

[17] D. Dubhashi, O. Häggström, L. Orecchia, A. Panconesi, C. Petrioli, and A. Vitaletti, "Localized techniques for broadcasting in wireless sensor networks," *Algorithmica*, vol. 49, no. 4, pp. 412–446, 2007.

[18] J. Yi, J. Koo, and H. Cha, "A localization technique for mobile sensor networks using archived anchor information," in *Proceedings of the 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '08)*, pp. 64–72, San Francisco, Calif, USA, June 2008.

[19] J. Elson and D. Estrin, "Time synchronization for wireless sensor networks," in *Proceedings of the 15th International Parallel and Distributed Processing Symposium (IPDPS '01)*, pp. 1965–1970, San Francisco, Calif, USA, 2001.

[20] J. Elson, L. Girod, and D. Estrin, "Fine-grained network time synchronization using reference broadcasts," in *Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI '02)*, vol. 36, pp. 147–163, Boston, Mass, USA, 2002.

[21] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in *Proceedings of the 9th ACM Conference on Computer and Communications Security*, pp. 41–47, Washington, DC, USA, November 2002.

[22] F. Fei, L. Jing, and Y. Xianglan, "Space-time related pairwise key predistribution scheme for wireless seneor networks," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '07)*, pp. 2692–2696, Shanghai, China, September 2007.

[23] L. Wang, T. Jiang, and X. Zhu, "Updatable key management scheme with intrusion tolerance for unattended wireless sensor network," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference: "Energizing Global Communications" (GLOBECOM '11)*, pp. 1–5, Houston, Tex, USA, December 2011.

[24] The Network Simulator—ns-2, http://www.isi.edu/nsnam/ns/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 358120, 8 pages
http://dx.doi.org/10.1155/2013/358120

Research Article

How Much Delay Has to Be Tolerated in a Mobile Social Network?

Yanqin Zhu, Haojun Zhang, and Qijin Ji

School of Computer Science and Technology, Soochow University, Suzhou 215006, China

Correspondence should be addressed to Yanqin Zhu; yqzhu@suda.edu.cn

Received 28 February 2013; Accepted 19 May 2013

Academic Editor: Lu Liu

Message delivery in a mobile social network (MSN) is difficult due to the fact that the topology of such network is sparse and unstable. Various routing schemes for MSNs were proposed to make the message delivery robust and efficient. However, little research has been conducted to explore how much delay has to be tolerated for the message delivery from the source to the destination. Since the social relationships among nodes are stable during a certain period of time, it is expected that the delay of message delivery in MSNs could be modeled with a probability model. In this paper, we take the first step to address this issue. We firstly extract three routing models from the existing routing schemes for MSNs and then develop the probability models of the message transmission delay for each abstract routing model. The simulation results show that the theoretical models match very well the simulation trace statistics.

1. Introduction

Mobile social network is a type of delay-tolerant networks (DTN [1]) which take into consideration the sociality of the participating nodes of the networks [2]. If the terminals of MSN are smartphones facilitated with rich sensing capability, it can function as a typical distributed sensor network [3]. Similar to the traditional mobile ad hoc networks (MANETs), no fixed infrastructures are built in such networks, and communications between the nodes depend on short-distance wireless links, like Wi-Fi or even Bluetooth. Since equipment in an MSN is usually carried by people moving freely, those nodes may organize themselves into highly dynamic topology, and the links between the nodes are usually intermittent and connected by opportunity due to the mobility of the nodes. As a result, large transmission delay is allowed in MSN. So in some extent, MSN is considered as a type of delay-tolerant networks. Although routing in an MSN might adopt the similar store-and-forward strategies [4] that dominate DTNs, it is essentially different from the DTN routing because the sociality of the user behavior can be cooperated explicitly to improve the efficiency of networking and communication in MSNs. Previous works [5–7] have shown that the performance of routing in MSN may heavily depend on the users' social behavior like that in a human social network, which is also the key characteristic for the design and analysis of the other issues in MSNs; see, for example [8, 9].

Traditional MANET routing protocols such as AODV [10], DSR [11], DSDV [12], and LAR [13] make assumption that the topologies of the networks are fully connected. These protocols will fail to route any message if there is no precomputed route from source to destination at the time of message being sent. In an opportunistic MSN, nodes carry the data to be forwarded and also are ready to forward data for other nodes. What is more, the mobility of nodes can be exploited to forward data opportunistically upon the encounter with each other. The key problem here is hence how to design appropriate relay selection strategies to improve the opportunity or probability of data forwarding. This improvement will make the data forwarding more efficient with less delay. Relay selection strategies depending on the history of contacts among nodes were proposed to find the right forwarding relays [14, 15].

While the end-to-end connections cannot be preset when the message is ready for transmission, social networks often demonstrate unique social characteristic like the so-called small-world phenomenon shown in the Milgram's 1967 mail transfer experiment [16]. In a social network, two-people contacts frequently usually have social ties. Recently, several sociality-aware routing strategies have been proposed to improve the routing efficiency for the MSNs [17–19].

Both contact history-based and social network metrics-based routing schemes designed for MSNs are best-effort, and there is no delay guarantee for the message delivery from the source to the destination. While MSNs are transmission delay-tolerant, it is still curious that how much delay must be tolerated for the message to be delivered from source to the destination or the delay bounds for message delivery in certain scenarios. These bounds are particularly interesting for the future multimedia application over the MSNs. While most of the work for data delivery in MSNs focuses on the design of new routing schemes, researches about how to evaluate the message delay remain rare and we try to make a first step to fill this gap.

In our opinion, the difficulty of the delay estimation for MSNs comes from two aspects: the dynamics of network topology and the uncertainty of the routing path. Though the mobility of nodes makes the networks' physical topology change dynamically, the nodes' social behavior tries to maintain regular contacts among nodes with social ties. We believe, therefore, it is possible to predict the routing behavior, at least the bounds of end-to-end delay of message delivery in an MSN.

In this paper, we illuminate a unified framework for evaluating the delay of message delivery in MSNs. Firstly, we propose general routing models extracted from the existing routing schemes. Then, we evaluate the delay with stochastic process theory based on the extracted routing model. The contributions of our paper are in threefold. First, we extract characteristics from the existing MSN routing strategies and classify them into three general models. Second, we propose a general method to evaluate the end-to-end delay in an MSN. And last, we design simulation experiments to validate the theoretical model. To our knowledge, this is the first effort to estimate the end-to-end delay of message delivery in an MSN via a probability framework.

The rest of this paper is organized as follows. Section 2 reviews the existing work. Section 3 presents the abstract routing model of the MSNs. Section 4 analyzes the delay bounds of the three types of general routing models with stochastic process theory and gives out the estimation results. Section 5 validates the performance of the proposed probability delay model by comparing the theoretical result with the simulation, and Section 6 concludes the paper.

2. Related Work

There is much work on the routing schemes for MANETs [20], from which the DTNs were evolved. The researches about routing schemes in a DTN might originate from the work of Jain et al. [4], in which messages are to be moved end-to-end across a time-varying connected graph but the topology dynamics may be known in advance. After that, there is much progress on this topic [21, 22].

Epidemic routing [23] which is originally designed for MANETs could deliver data fast and robustly by forwarding data to any encounters. To satisfy the limitation of energy and memory of nodes, a variety of relay selection strategies specially for DTNs are proposed [14, 15, 17–19]. Lindgren et al. proposed the ProPhet [14], making use of contact history to predict the probability of the meeting of the two nodes. Only nodes that can reach the destination with a higher probability will get the data from the relay.

Data dissemination in the delay tolerant MSNs is a critical component of many applications, for example, content update in an online social networks [24]. Sociality-based routing schemes for the DTNs have also been studied in recent years. SimBet routing [17] uses two social metrics (centrality and similarity) to estimate or predict the probability that potential relay nodes may reach the destination. Unlike the ProPhet and Epidemic, only one copy of data exists in the networks by SimBet routing. After that, Daly and Haahr tried to improve SimBet by using multiple copy of data forwarding [18]. Recently, Hui et al. [19] proposed a novel sociality-based forwarding algorithm, BUBBLE, which employed two social and structural metrics, namely, centrality and community.

Though there have been many research works on the design of routing schemes for the delay tolerant MSNs, model-based performance evaluation of these routing schemes is relatively scarce. Boldrini et al. [25] considered a utility-based cooperative data dissemination system in which the utility of data was defined based on the social relationships between users. Specifically, they designed a Markov model to characterize the data dissemination process in both its stationary and transient regimes. The main result of their analysis is that the data distribution process always converges to one of two possible stationary regimes.

Our study is obviously different from above work. In this paper, we concentrate on how to evaluate the delay from the source to the destination in a delay tolerant MSN by a probabilistic way. Based on the abstracted routing models, the data disseminating process is modeled as a stochastic process to estimate the delay to be experienced in such routing process.

3. Models

3.1. Network Model. In an MSN, nodes may be in a moving status and links between some nodes do not always exist. These characteristics make the topology of an MSN dynamic. However, during a certain time, the social relationships among the nodes are fixed. Being similar with [26, 27], we assume the pairwise node intercontact time is exponentially distributed. We consider the links among nodes as the social relationships of their holders, and the network model can be described as follows.

Consider an MSN with n mobile nodes, which can be denoted as a graph $G = (V, E)$, where V is the set of nodes and E is the set of links between nodes. If there is social relationship between node i and node j, let $e_{ij} \in E$ denote the edge e_{ij} between them. Letting λ_{ij} denote the weight of the link e_{ij}, then the intercontact time between nodes i and j obeys exponential distribution with parameter λ_{ij}. And further, the contacts between nodes i and j form a homogeneous Poisson process with the contact rate λ_{ij}.

3.2. The Model of MSN Routing. An MSN may experience frequent, persistent link partitioning and may never have a stable end-to-end path. Like in DTNs, routing in MSN employs the store-and-forward strategies over the opportunistic links. From various routing schemes existing currently, we extracted the following essential attributes which play the key roles for message delivery in an MSN.

(a) The relay selection strategy: this is without doubt the most important issue in the design of MSN routing, which directly affects the routing performance. Generally, we divided the strategies into two categories: nonstrategic that forwards data to any nodes it meets, and strategic that only forwarding data to those nodes which have better forwards quality.

(b) The data copying strategy: to improve the successful ratio of message delivery, the relays usually forward multiple copies of the data to their neighbors concurrently. However, this strategy significantly increases the consumption of the network resource. The data copying strategy here can be classified as single-copy and multicopy in general. The multicopy strategies can be further divided into finite-copy and infinite-copy strategies. While the routing scheme with finite-copy strategy usually forward fixed number of copies to the neighbors, the routing schemes with infinite-copy strategy simply forward the data to all the nodes it encounters.

(c) Clustering: while a good relay selection strategy can restrain the number of data copy to be forwarded, clustering is an effective approach to limit the over-consumption of the resource in a large scale MSN. According to whether the nodes will be clustered, the routing schemes for MSNs can be classified into layered routing and plain routing.

In this paper, only plain routing is considered. We extract three types of routing models from existing instances of routing schemes designed for MSNs based on the first two attributes described above.

3.2.1. Single-Copy Strategic Routing (SCSR). The routing strategy refers that node carrying data will not indiscriminately forward data to whichever it encounters but only chooses those nodes that can forward the data to destination with one or more hops. The usual routing strategies include what we have mentioned before, for example, the contacts history-based and social metric-based relay selection. A typical routing scheme belonging to this type is SimBet.

3.2.2. Multicopy Routing without Strategies (MCR-WS). This is the simplest routing schemes. Mobile node will forward its carried data to any nodes it encounters and keep the data at the same time. An infinite-copy strategy is more popular than the finite multiple copy strategy. Epidemic routing belongs to this type of routing schemes.

3.2.3. Multicopy Strategic Routing (MCSR). In this type of multicopy routing schemes, strategies are considered which makes routing decisions more complex. Since it combines the simplicity of multicopy routing and efficiency of strategic routing, it attracts more attention and most of the routing schemes belong to this type. Compared with MCR-WS routing, if the relay selection strategy is designed to be effective enough, MCSR could achieve similar performance while consuming less system resources. ProPhet is a typical routing scheme of this type.

3.3. Some Assumptions of the Routing Model. We make the following assumptions listed from weak to strong.

(1) There is enough time for the nodes to exchange their data when they contact.

(2) For the intercontact time is much longer than the data transmission time over the link, the latter would be ignored when evaluating the transmission delay.

(3) In the case of multicopy routing strategy, nodes' buffer capacity is large enough thus no packet will be discarded due to the lack of memory.

(4) TTL (time-to-live) field is set as time limitation and the data will be discarded actively if it does not arrive at the destination after *TTL* time's forwarding.

(5) All nodes will not receive the same data for two times by numbering the message with a global ID created by a Hash function.

4. Analysis

We will analyze the delay of message delivery in MSNs based on the three routing models in Section 3. Before the detailed analysis, we firstly present an important result in Lemma 1. Assume that the source S intends to send data to the destination D through the intermediate node set $\{N_1, N_2, \ldots, N_r\}$ as shown in Figure 1. Denote $\lambda_1, \lambda_2, \ldots, \lambda_{r+1}$ as the weight of intermediate links in sequence, and the corresponding intercontact time is $X_1, X_2, \ldots, X_{r+1}$. As described in Section 3.1, the intercontact time between nodes follows exponential distribution, so the probability density function of X_k is $f_{X_k}(x) = \lambda_k e^{-\lambda_k x}$. Then the maximal aggregated time needed to forward data from S to D is

$$Y = \sum_{i=1}^{r+1} X_i, \tag{1}$$

which is hypoexponentially distributed [28] according to the following lemma.

Lemma 1 (section 5.2.4 of [28]). *For an opportunistic path with r hops, the corresponding edges weight as* $\lambda_1, \lambda_2, \ldots, \lambda_{r+1}$, *then the probability density function of* Y *as in* (1) *is*

$$f_Y(x) = \sum_{k=1}^{r+1} C_k^{(r+1)} f_{X_k}(x), \tag{2}$$

where $C_k^{(r+1)} = \prod_{s=1,s\neq k}^{r+1} (\lambda_s / (\lambda_s - \lambda_k))$.

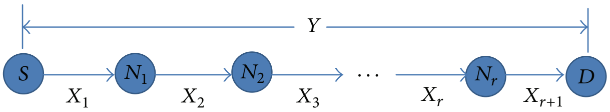


Figure 1: Opportunistic path connecting nodes S and D.

By formula (2) *and assumption 5, the expected delay of the path in Figure 1 is*

$$E[Y] = \int_0^T x p_Y(x)\,dx = \int_0^T x \sum_{k=1}^{r+1} C_k^{(r+1)} \lambda_k e^{-\lambda_k T} dx$$
$$= \sum_{k=1}^{r+1} C_k^{(r+1)} \left(\frac{1}{\lambda_k} - \left(T + \frac{1}{\lambda_k} \right) e^{-\lambda_k T} \right). \tag{3}$$

4.1. Delay of Multicopy Routing without Strategies. We firstly analyze the case of MCR-WS and the typical routing of this type is Epidemic [23]. Messages generated at source node will be forwarded to encountering nodes that have not received the data. Data will disseminate over the whole network quickly by multicopying in the network. As we know, MCR-WS usually enjoys the minimum delay but consumes much more system resources.

According to assumptions 1 and 2 in Section 3, the time needed to send data from the source to the destination is the sum of the intercontact time between intermediate nodes pairs. There might be several possible routing paths connecting the source to the destination. But due to the multicopy strategy, the transmission delay is always the minimum of delays along all the possible paths as the following lemma.

Lemma 2. *There are m paths $\{P_1, P_2, \ldots, P_m\}$ between the source node S and destination node D. The corresponding transmission delay of each path is elements of set $\{Y_1, Y_2, \ldots, Y_m\}$. Then with MCR-WS, the delay between S and D is $T(S, D) = \min\{Y_1, Y_2, \ldots, Y_m\}$.*

Proof. The proof is obvious and omitted. □

Because the weight of every edge constructing a path can be got according to formula (2), the cumulative distribution function $F_{Y_i}(x)$ of each path's delay can be derived by Lemma 1. By Lemma 2, the cumulative distribution function of $T(S, D)$ is given out in Proposition 3.

Proposition 3. *In the MCR-WS scheme, if $\{Y_1, Y_2, \ldots, Y_m\}$ are independent with each other, then the transmission delay from S to $DT(S, D) = \min\{Y_1, Y_2, \ldots, Y_m\}$ follows the distribution*

$$F_T(x) = 1 - \prod_{i=1}^{m} \left(1 - F_{Y_i}(x)\right). \tag{4}$$

Proof. According to Lemma 2, any one of the elements in $\{Y_1, Y_2, \ldots, Y_m\}$ satisfying $Y_i \leq x$ makes sure $T \leq x$. The probability of $Y_i > x$ for all i is $\prod_{i=1}^{m}(1 - F_{Y_i}(x))$, thus $F_T(x) = 1 - \prod_{i=1}^{m}(1 - F_{Y_i}(x))$. □

Proposition 3 is provable on conditions that all the possible paths are independent. If two paths do not share any node, these two paths are independent with each other. However, all paths in $\{Y_1, Y_2, \ldots, Y_m\}$ from S to D usually are not independent, and some correlated paths should be fixed approximately. Assuming that $P_i = \{N_1^{(i)}, N_2^{(i)}, \ldots, N_m^{(i)}\}$ and $P_j = \{N_1^{(j)}, N_2^{(j)}, \ldots, N_n^{(j)}\}$ are two paths sharing intermediate node, the shared node $N_{k_1}^{(i)}$ or $N_{k_2}^{(j)}$ is called broken node if $N_{k_1}^{(i)} = N_{k_2}^{(j)}$ and $N_{k_1+1}^{(i)} \neq N_{k_2+1}^{(j)}$. In contrast, $N_{k_1}^{(i)}$ or $N_{k_2}^{(j)}$ is called backing node if $N_{k_1}^{(i)} = N_{k_2}^{(j)}$ and $N_{k_1-1}^{(i)} \neq N_{k_2-1}^{(j)}$. By the position of the broken node and the next backing node, two subpaths can be found which have the same source S' (broken node) and the same destination D' (backing node). Because the two subpaths share no common intermediate nodes, they are independent. As a result, through Lemma 1 and Proposition 3, the expectation of transmission delay from S' to D' can be calculated as T', building a dummy connection between S' and D' instead of the paths which connecting them. The distribution of intercontact time Y' between S' and D' is approximated as exponential distribution with parameter $\lambda_{S'D'} = 1/T'$. Through the above approximate treatment, two related paths could be merged into one path. What is more, if the same merging process is done on all the paths $\{Y_1, Y_2, \ldots, Y_m\}$, a group of independent paths will be constructed. Finally, transmission delay evaluation of MCR-WS will be done using Proposition 3 directly.

4.2. Delay of Single-Copy Strategic Routing. Now we consider the second case, SCSR. The strategy here refers to that relaying nodes will not forward data to any encounters but select those that could forward the data to the destination with better quality. There are several different relay selection strategies. Some maintain a local encountering probability by contact history and others calculate the utility value using social metrics. Despite of the distinct of the strategies, no strategy can guarantee that the relay selection will always reach the destination. It is assumed that the probability of selecting a right relay is a constant $P_{\text{Eval-}W}$. We will show how to estimate the value of $P_{\text{Eval-}W}$ later.

Let V be the node forwarding data and $N(V)$ its neighbor set. Because not all neighbors can lead to the destination, let $W(V) \subseteq N(V)$ denote nodes that can forward data to the destination and tend to be selected as the next relay by the strategy. Due to the single-copy strategy, only one of the nodes in $N(V)$ will receive data from V. To make sure the data could be forwarded to the destination with highest probability, the intercontact history between node V and the candidates of relays is used to decide the probability of relay selection. Let $P_s(N_i)$ denote the probability that node $N_i \in W(V)$ is selected by V as the next relay. For node $N_i \in W(V)$, it is obvious that the more frequently contacting with V, the higher probability of $P_s(N_i)$. So we have

$$\sum_{i \in W(V)} P_s(N_i) = P_{\text{Eval-}W}$$
$$P_s(N_i) : P_s(N_j) = w(N_i, V) : w(N_j : V) \quad \forall i, j \in W(V). \tag{5}$$

In most cases, there are many possible paths connecting the source S and the destination D. All the possible path set can be found and it is denoted as PathSet $(S, D) = \{(P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m)\}$, in which Prob_i denotes the probability of selecting path P_i for message delivery. Assuming the path P_i includes the following nodes N_1 (source), $N_2, \ldots, N_{m-1}, N_m$ (destination), then

$$\text{Prob}_i = \prod_{i=2}^{m} p_s(N_i). \tag{6}$$

Because there is only one data copy in the network, with the assumption of independent probability of possible path selection, we can calculate the probability of the successful message delivery:

$$P_{\text{Eval}} = \sum_{i=1}^{m} \text{Prob}_i. \tag{7}$$

What is more, it is easy to get the probability of successful message delivery of the real scenario and we denote it as P_{Real}. Then letting $P_{\text{Eval}} = P_{\text{Real}}$, with (5), (6), and (7), the probability of selecting the proper relay $P_{\text{Eval-}W}$ can be calculated.

Through above discussion, the preparation for the evaluation of SCSR has been done and then Proposition 4 is given as follows.

Proposition 4. *In the type of SCSR, assume the possible routing path set is* $\text{Path}(S, D) = \{(P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m)\}$, *in which* Prob_i *denotes the probability of selecting path* P_i. *And let* Y_i *denote the transmission delay over path* P_i. *Then the delay* T *needed to send data from source to destination successfully has the following probability density function:*

$$f_T(x) = \sum_{i=1}^{m} f_{Y_i}(x) \times \frac{\text{Prob}_i}{P_{\text{Real}}}, \tag{8}$$

where $P_{\text{Real}} = \sum_{i=1}^{m} \text{Prob}_i$ *and* $f_{Y_i}(x)$ *is the probability distribution function of the variable* Y_i.

Proof. Firstly, it is proved that $f_T(x)$ satisfies the property of probability density function. Obviously, for all $x \in R$, $f_T(x) \geq 0$ and it has

$$\begin{aligned}
\int_{-\infty}^{+\infty} f_T(x) &= \int_{0}^{+\infty} f_T(x)\,dx \\
&= \int_{0}^{+\infty} \sum_{i=1}^{m} f_{Y_i}(x) \times \frac{\text{Prob}_i}{P_{\text{Real}}} dx \\
&= \sum_{i=1}^{m} \frac{\text{Prob}_i}{P_{\text{Real}}} \times \int_{0}^{+\infty} f_{Y_i}(x)\,dx \\
&= \sum_{i=1}^{m} \frac{\text{Prob}_i}{P_{\text{Real}}} = 1.
\end{aligned} \tag{9}$$

So $f_T(x)$ can be probability density function of variable T. Then we prove that the relationship in (8) makes sense. Because of the single-copy strategy, there is only one path that is selected to transmit data. The probability Prob_i of each path being chosen can be calculated according to (5), (6), and (7), and the distribution functions of transmission delay Y_i on each path are known; we have

$$p(T = x) = \sum_{i=1}^{m} p(Y_i = x) \times \text{Prob}_i. \tag{10}$$

By the influence of policy, the data may not be sent to the destination for selecting the wrong path, which makes the transmission delay infinite. If the case of choosing the broken path is not taken into consideration, the sum of chosen probability of all the possible left paths $\sum_{i=1}^{m} \text{Prob}_i$ is less than 1, which makes the component of T uncompleted. But if we normalize the probability of selecting the proper path, that is to say, letting the probability Prob_i be normalized as $\text{Prob}_i / P_{\text{Real}}$, which would keep the component of T completed, then we can get the probability density function $f_T(x) = \sum_{i=1}^{m} f_{Y_i}(x) \times \text{Prob}_i / P_{\text{Real}}$ under the condition of successful message delivery. □

By Proposition 4, the expected time delay from the source node S to the destination D in this type of routing schemes can be easily got as

$$E[T(S, D)] = \sum_{i \in \text{Path}(S,D)} E[Y_i] \times \frac{\text{Prob}_i}{P_{\text{Real}}}. \tag{11}$$

4.3. Delay of Multicopy Strategic Routing. The multicopy strategy improves the probability of successful message delivery comparing with the SCSR. And the relay selection strategy reduces the overhead of MCR-WS. Spyropoulos et al. [15] proved that the performance of MCSR could achieve the performance of Epidemic routing if the strategy was designed properly.

The method of dealing with relay selection strategy is similar to that in SCSR. We still assume the probability of selecting the proper relay is a constant $P_{\text{Eval-}W}$. However, due to the multicopy strategy, every node in $W(S)$ is selected with the same probability $P_{\text{Eval-}W}$. All paths connecting the source S and destination D will be denoted as $\text{Path}(S, D) = \{(P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m)\}$. Then the probability of transferring data on path P_i is

$$\text{Prob}_i = (P_{\text{Eval-}W})^{|P_i|-1}, \tag{12}$$

where $|P_i| - 1$ is the number of relays selected. Unless all the possible paths fail, the data will fail to be transmitted to the destination. So the probability of successfully sending data is $1 - \prod_{i=1}^{m}(1 - \text{Prob}_i)$. Letting

$$1 - \prod_{i=1}^{m} (1 - \text{Prob}_i) = P_{\text{Real}}, \tag{13}$$

then $P_{\text{Eval-}W}$ and each probability Prob_i can be calculated. To evaluate the transmission delay in the MCSR, Proposition 5 is given as follows.

Proposition 5. *In the type of MCSR, node S sends message to node D and the all possible routing path set is* $\text{Path}(S, D) = \{(P_1, \text{Prob}_1), (P_2, \text{Prob}_2), \ldots, (P_m, \text{Prob}_m)\}$. *If the routing paths are independent with each other, the transmission delay T follows:*

$$F_T(x) = 1 - \prod_{i=1}^{m} \left(1 - F_{Y_i}(x)\right) \times \text{Prob}_i. \tag{14}$$

Proof. We consider all the cases of $T > x$. Firstly, if the data is sent successfully to the destination, it requires all transmission delay on every possible routing path that has $Y_i > x$. Because all Y_i are independent with each other, through the distribution function of Y_i, the probability of $Y_i > x$ for $i = 1, 2, \ldots, m$ is $\prod_{i=1}^{m}(1 - F_{Y_i}(x))$. For every possible routing path is selected at the probability of Prob_i, the probability of $Y_i > x$ should be multiplied by Prob_i. Then we have $F_T(x) = 1 - \prod_{i=1}^{m}(1 - F_{Y_i}(x)) \times \text{Prob}_i$. Secondly, we verify whether $F_T(x)$ fulfills the definition of distribution function. Obviously, it is true that $\lim_{x\to\infty} F_T(x) = 1$, for $\lim_{x\to\infty} F_{Y_i}(x) = 1$, for all $i = 1, 2, \ldots, m$. □

Proposition 5 requires that all possible correlated routing paths are independent with each other. Some fixing similar to MCR-WS should be done. Two related routing paths that share intermediate nodes will be merged. The merging process is as same as the process in MCR-WS, but the probability of the new merged routing path should be re-calculated. Assume, in the merging process, two subpaths which have m_1 and m_2 nodes, respectively, include their common head node S' and tail node D'. After they have been merged, the probability of having data flow from S' to D' is $1 - (1 - (P_{\text{Eval-}W})^{m_1}) \times (1 - (P_{\text{Eval-}W})^{m_2})$. Through the path merging and probability modifying process, a new independent routing path will be got. Then using Proposition 5, the transmission delay in MCSR can be calculated.

5. Simulation

In this section, we validate our delay model by comparing the theoretical delay model with the simulation results. All of the three routing schemes corresponding to each routing model are included: Epidemic [23] for MCR-WS, SimBet [17] for SCSR, and ProPhet [14] for MCSR.

5.1. Simulation Setup. The simulation experiment is developed based on the general discrete event simulation platform OMNET++ [29]. Though the scale of network in our simulation could be set as any size only if it is within the resource constraints of the simulator, the theoretical delay model has no relevance to the scale. In the simulation, 30 nodes are randomly placed inside an area of 500 × 500 m^2. IEEE 802.11b DCF is used as the MAC layer protocol. The radio propagation range of all nodes is set as 100 meters. The nodes move according to the Gauss-Markov mobility models [30], with a minimum speed of 0 m/s and a maximum speed of 5 m/s. When two nodes are in the radio propagation range, it is considered that they meet and can communicate with each other. As models omit the transmission delay of the channel, the transmission delay of all channels is set as 0.

Table 1: Pearson Correlation Coefficient.

Routing model	Pearson Correlation Coefficient
MCR-WS	0.8591
SCSR	0.8021
MCSR	0.9065

During the simulation, the source and destination nodes are selected randomly from the node set. Data packets are generated by the source node every five seconds and 3000 packets will be sent during this interval. The time-to-live fields of these packets are set as 15 s, which means the message will be discarded if it has not arrived at the destination after 15 seconds. The intercontact time of nodes meeting each other is recorded to provide the social ties for the model-based numerical experiment. We approximate the parameter of the exponential distribution as the reciprocal of the average intercontact time. The parameter setting of ProPhet is as same as in [14], that is, $P_{\text{init}} = 0.75$, $\beta = 0.25$, and $\gamma = 0.98$. And the parameters of SimBet are just like in [17], that is, $\alpha = \beta = 0.5$. To simulate the packet routing process, every data packet is forwarded on the meeting of the expected nodes according to the routing schemes and the transmission delay is traced.

5.2. Model Evaluation. We try to evaluate the accuracy of the theoretical model by comparing the estimated delay with the corresponding statistics of the simulation result. We have already derived the probability density function or cumulative distribution function of the transmission delay of each routing model, respectively, as in (4), (8) and (14). Denote all the delays traced in the simulation as $\{d_1, d_2, \ldots, d_m\}$ and divide the time interval $[0, TTL]$ into constant k parts equally. The constant k should be carefully set. Then counting the number c_i of d_i belongs to each subinterval and let c_i/m be the probability of the subinterval. In one simulation, the comparison results are shown in Figures 2, 3, and 4. It is obvious that the delay in simulation matches very well the delay calculated by the proposed theory model.

Then we run the simulation for 20 times of each routing model, each of which includes sending 3000 data packets. Each time the network scale is also $N = 30$ but with different network topology. Through the simulation, the average transmission delay is obtained and the expected delay by the model could also be calculated. Pearson Correlation Coefficient [31] is a measure of the correlation between two variables X and Y, giving a value between +1 and −1 inclusive. The more close to +1 or −1, the more related they are. It is commonly denoted with r and can be obtained from the following formula:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}. \tag{15}$$

We use Pearson Correlation Coefficient to show further how much the average transmission delay in simulation matches the theoretical result. The result presented in Table 1

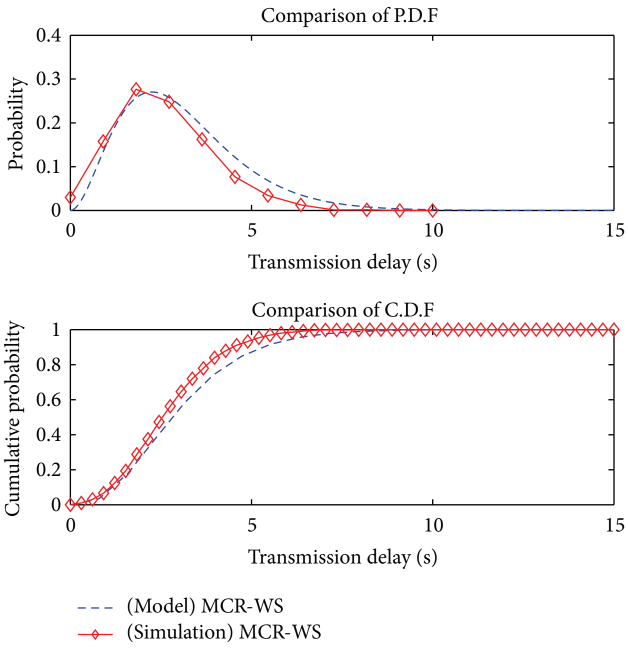


Figure 2: Comparison of MCR-WS.

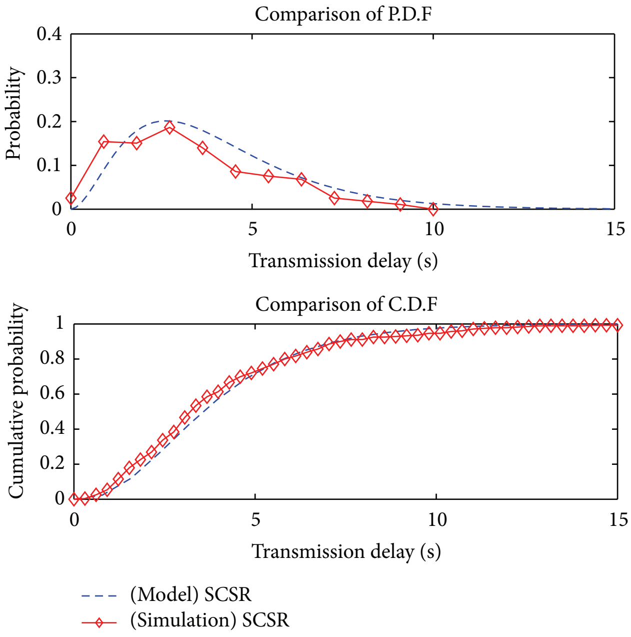


Figure 3: Comparison of SCSR.

shows that the simulation delays and theory-calculated delays are highly related.

6. Conclusions

In this paper, we proposed a probability model to estimate the delay of message delivery in the delay tolerant mobile social networks. We firstly extracted three general routing models from the existing various routing schemes for MSNs. According to an elegant result in probability theory, we constructed probability delay models for each of the three routing models. Then the simulation experiments were designed to validate the accuracy of the theoretical delay model. It was found that the delay statistics from the simulation trace matched very well the theoretical results, which means that proposed model is quite accurate for the prediction of delay of message delivery in an MSN.

During the model construction, we omitted the storage limitation of the nodes in an MSN. In the future work we will enhance the proposed delay model considering the message loss from the storage limit. For only the plain routing schemes for MSNs being considered in the current work, delay estimation for the layered routing schemes is also a valuable extension to this work.

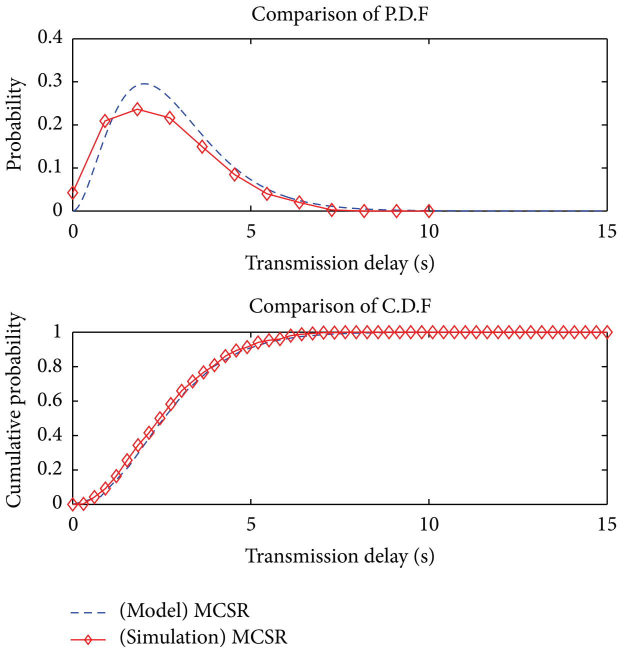


Figure 4: Comparison of MSCR.

Acknowledgments

The work was supported by National Science Foundation of China under Grant no. 61070170, Science Research Program of Universities in Jiangsu Province under Grant no. 11KJB520017, and Application Foundation Research of Suzhou (Jiangsu, China) under Grant no. SYG201238.

References

[1] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 27–34, August 2003.

[2] H. Falk, "Applications, architectures, and protocol design issues for mobile social networks: a survey," *Proceedings of the IEEE*, vol. 99, no. 12, pp. 2125–2129, 2011.

[3] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.

[4] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 145–158, New York, NY, USA, September 2004.

[5] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft, "A Socio-Aware Overlay for publish/subscribe communication in delay tolerant networks," in *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, pp. 225–234, New York, NY, USA, October 2007.

[6] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: a social network perspective," in *Proceedings of the 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'09)*, pp. 299–308, New York, NY, USA, May 2009.

[7] T. Henderson, D. Kotz, and L. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, pp. 187–201, New York, NY, USA, October 2004.

[8] A. Beach, M. Gartrell, S. Akkala et al., "WhozThat? Evolving an ecosystem for context-aware mobile social networks," *IEEE Network*, vol. 22, no. 4, pp. 50–55, 2008.

[9] W. D. Yu and A. Siddiqui, "Towards a wireless mobile social network system design in healthcare," in *Proceedings of the 3rd International Conference on Multimedia and Ubiquitous Engineering*, pp. 429–436, June 2009.

[10] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, pp. 90–100, February 1999.

[11] D. Johnson and D. Maltz, "Dynamic source routing in ad-hoc wireless networks," *Mobile Computing*, vol. 353, pp. 153–181, 1996.

[12] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proceedings of the Conference on Communications Architectures, Protocols and Applications*, vol. 24, no. 4, pp. 234–244, New York, NY, USA, 1994.

[13] Y.-B. Ko and N. H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," *Wireless Networks*, vol. 6, no. 4, pp. 307–321, 2000.

[14] A. Lindgren, A. Doria, and O. Schelen, "Probabilistic routing in intermittently connected networks," *Mobile Computing and Communications Review*, vol. 7, no. 3, pp. 19–20, 2003.

[15] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and wait: an efficient routing scheme for intermittently connected mobile networks," in *Proceedings of the ACM SIGCOMM Workshop on Delay-Tolerant Networking*, pp. 252–259, New York, NY, USA, August 2005.

[16] S. Milgram, "The small world problem," *Psychology Today*, vol. 1, no. 1, pp. 60–67, 1967.

[17] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANETs," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 32–40, New York, NY, USA, September 2007.

[18] E. M. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant MANETs," *IEEE Transactions on Mobile Computing*, vol. 8, no. 5, pp. 606–621, 2009.

[19] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble Rap: social-based forwarding in delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1576–1589, 2011.

[20] E. M. Royer and C.-K. Toh, "A review of current routing protocols for ad hoc mobile wireless networks," *IEEE Personal Communications*, vol. 6, no. 2, pp. 46–55, 1999.

[21] J. Ott, D. Kutscher, and C. Dwertmann, "Integrating DTN and MANET routing," in *Proceedings of the ACM SIGCOMM Workshop on Challenged Networks*, pp. 221–228, September 2006.

[22] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: positive and negative social effects," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 387–401, 2013.

[23] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Tech. Rep. CS-200006, Duke University, 2000.

[24] S. Ioannidis, A. Chaintreau, and L. Massoulié, "Optimal and scalable distribution of content updates over a mobile social network," in *Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM '09)*, pp. 1422–1430, April 2009.

[25] C. Boldrini, M. Conti, and A. Passarella, "Modelling data dissemination in opportunistic networks," in *Proceedings of the 3rd ACM Workshop on Challenged Networks*, pp. 89–96, New York, NY, USA, September 2008.

[26] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[27] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.

[28] S. M. Ross, *Introduction To Probability Models*, Academic Press, 2006.

[29] OMNeT++ Community Site, http://www.omnetpp.org/.

[30] D. B. Johnson and D. A. Maltz, "Dynamic source routing in Ad Hoc wireless networks," in *Mobile Computing*, vol. 353 of *The Kluwer International Series in Engineering and Computer Science*, pp. 153–181, 1996.

[31] http://en.wikipedia.org/wiki/Pearson_correlation_coefficient.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 385471, 12 pages
http://dx.doi.org/10.1155/2013/385471

Research Article

RET-MAC: A New Fair MAC Protocol for Underwater Acoustic Sensor Network

Shuming Xiong,[1,2] Cheng Yuan,[1] Lixin Tian,[3] and Yongzhao Zhan[1]

[1] *School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China*
[2] *Faculty of Science, Jiangsu University, Zhenjiang 212013, China*
[3] *Nanjing Normal University, Nanjing 210023, China*

Correspondence should be addressed to Shuming Xiong; xsm@ujs.edu.cn

Received 28 February 2013; Accepted 28 April 2013

Academic Editor: Lu Liu

In underwater acoustic sensor networks (UASNs), the propagation delay of acoustic signals is much longer than Radio Frequency (RF). There is spatial unfairness problem caused by space-time uncertainty. Hence, the design of Medium Access Control (MAC) is a challenging issue. In underwater, it not only considers transmission time but also takes location into account. In this paper, we propose Response to the Earliest Transmitter of RTS MAC (RET-MAC) protocol to solve the problem. RET-MAC adopts adaptive RTS Contention Phase (RTS CP) to determine the earliest transmitter of RTS. And CTS Delay Phase (CTS DP) is added to postpone sending CTS in order to avoid collision. In addition, we propose CTS back-off mechanism to adjust the length of CTS DP as needed. Contention back-off mechanism is used to reduce network congestion and increase fairness further. The simulation results show that our scheme can achieve higher fairness and throughput; at the same time it also guarantees lower energy consumption and delay.

1. Introduction

Recently, with the development of the computer technology, microelectronics technology, and communication technology, underwater acoustic sensor networks (UASNs) have also got great progress. There are a wide range of applications of UASNs, such as oceanographic data collection, pollution monitoring, offshore exploration, disaster prevention, assisted navigation, and tactical surveillance applications [1]. A Media Access Control (MAC) protocol is used to coordinate the access to the shared acoustic channel among multiple nodes, which is very critical to the network performance [2]. However, unlike the terrestrial wireless sensor networks that mainly rely on radio waves for communications, underwater sensor networks utilize acoustic waves, which present a much harsher environment for both the physical and the data-link layers [3]. In fact, underwater acoustic communications are characterized by three major differences with respect to terrestrial radio: the very low propagation speed; the strongly anisotropic nature, whereby horizontal channels are usually harsher than vertical channels; the significant difference between the power required to operate acoustic transducers and the power required to receive or listen to an acoustic signal [4].

In designing resource-sharing schemes for underwater networks, one needs to keep in mind the peculiar characteristics of the acoustic channel [5]. In [6], Syed et al. point out that the long propagation delay of acoustic media leads to spatial unfairness problem. The nodes nearer the receiver occupy the channel quickly. On the contrast, other nodes away from receiver may be in “starvation” state for a long time. And sensor nodes are battery-powered, which makes the energy of nodes very limited, especially the energy cost required by communications [7]. In addition, compared with terrestrial wireless sensor networks, replacement of these low cost batteries in underwater is a challenging task and uneconomical too [8]. Therefore, there is an important significance to design the fairness and low energy consumption of UASN MAC protocol.

MAC protocols decide how multiple nodes share the underlying acoustic channel, which is critical to the overall network performance [9]. In this paper, we propose Response

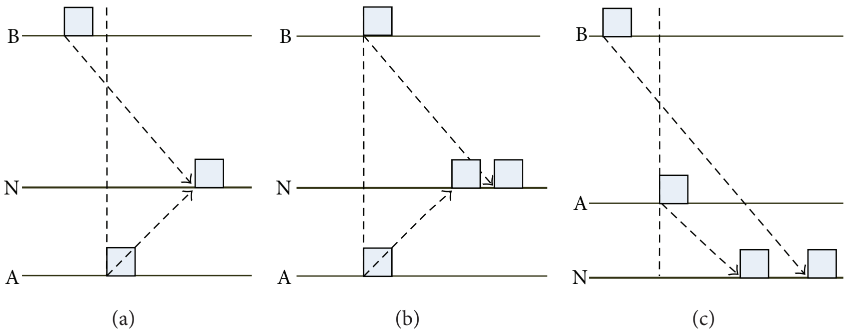


FIGURE 1: (a) Different transmission time but collision at N. (b) Same transmission time, no collision at N. (c) Spatial fairness problem.

to the Earliest Transmitter of RTS MAC (RET-MAC) protocol, which can ensure node to contend the channel fairness. According to the sending time of RTS packet, the protocol uses RTS Contention Phase (RTS CP) to determine the earliest transmitter of RTS in order to satisfy spatial fairness. On the other hand, our protocol postpones the transmission of CTS packet by CTS Delay Phase (CTS DP) to avoid collision. Above all, we adopt adaptive RTS CP and CTS DP to reduce delay. In addition, contention back-off mechanism is added to increase fairness further. The simulation results show that RET-MAC not only achieves higher fairness and lower delay, but also has better throughput and energy efficiency.

The rest of this paper is organized as follows. In Section 2, the related work is introduced. In Section 3, we discuss the spatial unfairness issue in UASNs and present our RET-MAC protocol. Then we analyze the protocol parameters in Section 4. After that, we evaluate the performance of RET-MAC in Section 5. Finally, we conclude this paper and discuss some future research in Section 6.

2. Related Work

Currently, there are many MAC protocols for terrestrial RF-based sensor networks, which are devoted to conserve energy and increase system performance by avoiding collision, such as S-MAC [10], CC-MAC [11], and Z-MAC [12]. However, they are not applicable to UASN due to long propagation delay and the characteristics of underwater environment [13]. The MAC protocol design of UASNs is facing many greater challenges, gradually attracting researchers' attention.

In [14], Peng et al. introduce a contention-based MAC protocol with parallel reservation (COPE-MAC) for UASN. In order to establish communication with less rounds of handshakes, they propose parallel reservation, while using cyber carrier sensing to detect and avoid collisions with computation. COPE-MAC can improve MAC performance in both network throughput and energy efficiency. In [3], Chirdchoo et al. study Aloha-based variant protocols, proposing two Aloha-based random access MAC protocols, namely, Aloha-CA and Aloha-AN, for UASN. The two protocols combine Aloha with carrier sensing, use short reservation frame to contend channel, and add some relevant information of data to reservation frame. Other nodes sense the channel to achieve the information which will be sent from neighbor nodes and calculate the busy time of the channel. In the busy time, they take into sleep state in order to reduce energy consumption. Both schemes can boost the throughput by reducing the number of collisions. In [15], Park and Rodoplu propose UWAN-MAC protocol suitable for UASN.

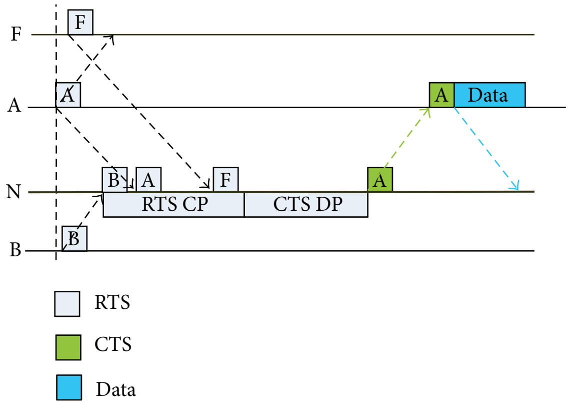


FIGURE 2: Basic idea of the RET-MAC protocol.

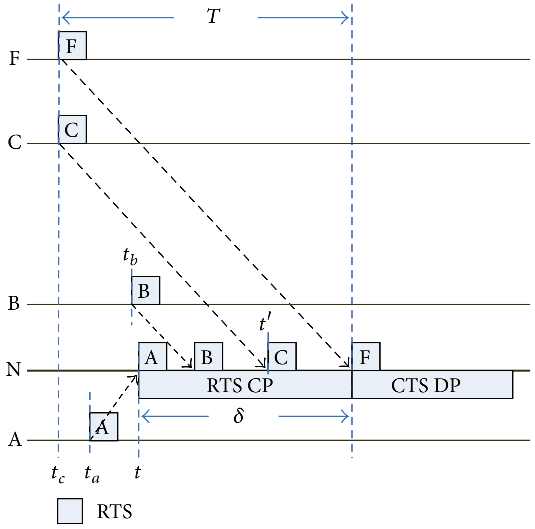


FIGURE 3: Analysis of RTS CP.

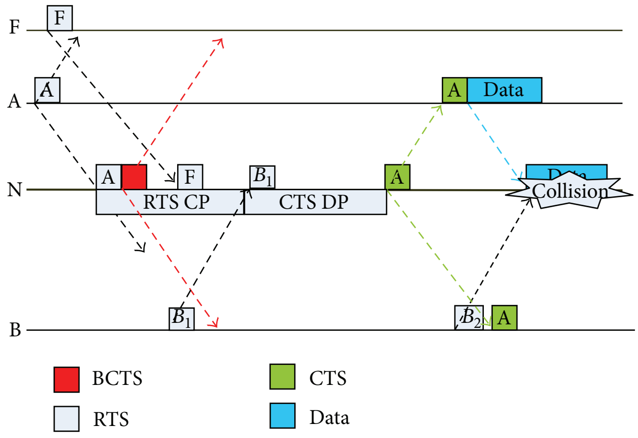


Figure 4: Add BCTS to avoid collision.

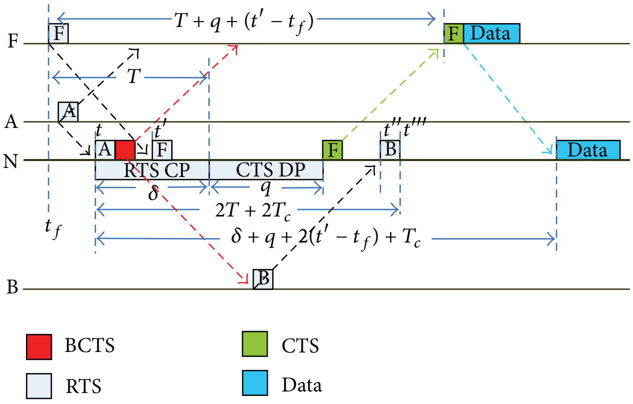


Figure 5: Analysis of CTS DP.

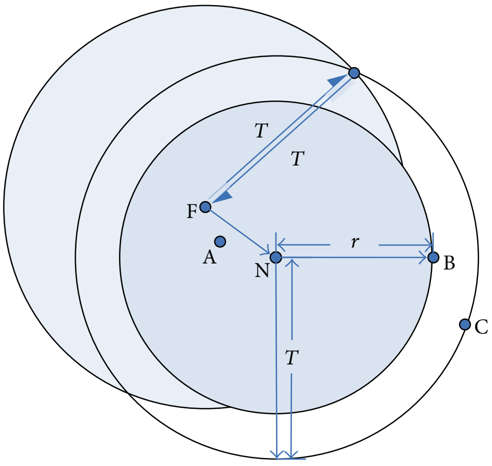


Figure 6: Example of network configuration.

It uses relative time stamps, not only in the transmission of data but also in the establishment of communication with newcomers, ingeniously solving synchronization problem between nodes sending and receiving data. And at the same time, it effectively saves energy and reduces the data collision rate. In [16], Hsu et al. propose a Spatial-Temporal MAC (ST-MAC) protocol, which is designed to overcome spatial-temporal uncertainty based on TDMA-based MAC scheduling for energy saving and throughput improvement. They construct the Spatial-Temporal Conflict Graph (ST-CG) to describe the conflict delays and propose the Traffic-based One-step Trial Approach (TOTA) to solve the coloring problem. Through a comprehensive study, ST-MAC has better network throughput and energy cost compared with existing schemes. In [17], Azar and Manzuri propose a MAC protocol based on reserved time slot. It employs a synchronization algorithm to synchronize all nodes and uses the listen/sleep periodic operation for saving energy. In addition, the protocol uses short ranges of underwater acoustic communication links to achieve higher throughput. In [18], Cho et al. propose a PR-MAC protocol, in which nodes exchange information between 2hop neighbor nodes in random accesses period. According to predetermined priority, each node is assigned transmission opportunities with contention free method by TDMA scheme. The reservation period is divided into several slots to transfer data reducing collision. PR-MAC can also reduce energy consumption by reducing period with active mode. In [19], Hong et al. propose an efficient continuous time scheduling TDMA protocol (ECS), including the continuous time based and sender oriented conflict analysis model, the transmission moment allocation algorithm, and the distributed topology maintenance algorithm. By using continuous time based transmission moment allocation scheme, differences of link delays are further utilized and channel utilization of receiver node is improved. At the same time, ECS has higher network throughput and better efficiency.

However, these research works of underwater acoustic sensor networks MAC protocols mainly focused on collision avoidance, throughput, and energy efficiency, but most of them rarely consider fairness. In contention protocols, the long propagation delay of acoustic media leads to spatial unfairness problem. Xie and Cui propose a reservation-based MAC protocol (R-MAC) in [20]. R-MAC has three phases to allocate the channel resources including latency detection, period announcement, and periodic operation. Nodes transfer data at its own slots. At the beginning, node in latency detection phase detects the propagation latency to all its neighbors. In the period announcement phase, each node randomly selects its own listen/sleep schedule and broadcasts this schedule. The data are transmitted in the periodic operation phase. In R-MAC, an intended receiver randomly selects one reservation from the reservations it collects to support fair access of the channel. However, it has low throughput and is not suitable for intensive network. Later, in [6], Syed et al. point out the problem that is long propagation delay of acoustic channel leading to spatial unfairness. They propose the T-Lohi MAC protocol to solve the problem. T-Lohi uses the random back-off method to ensure the fair access of the channel, to overcome the problem of “the nodes nearer the receiver occupy the channel quickly.” However, T-Lohi does not consider hidden terminal problem, resulting in that network throughput becomes lower. In [13], Liao and Huang propose SF-MAC protocol to solve the spatial

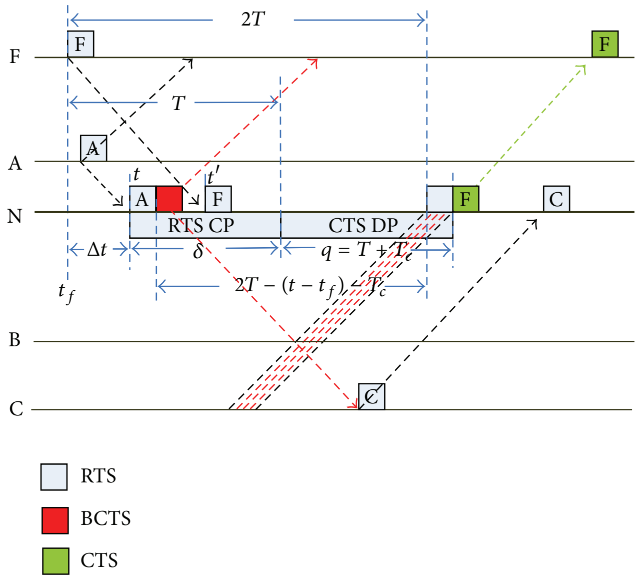


Figure 7: Analysis of MP-CD.

unfairness. SF-MAC adopts the RTS/CTS handshake method to solve the hidden terminal problem, and during contention period determines the earliest transmitter of RTS packet by Probability Rule. However, in SF-MAC, the contention period is fixed, and the length of the handshake time is longer. It seriously impacts on network throughput. In this paper, RET-MAC protocol considers the hidden terminal problem and solves spatial unfairness. At the same time, it uses adaptive handshake time and some mechanisms to reduce delay and energy cost and improve network throughput.

3. RET-MAC Protocol Design

In this section, we first discuss space-time uncertainly and spatial unfairness problem in UASNs. Secondly, we introduce network model and the basic idea of RET-MAC protocol. Then we analyze RTS CP and CTS DP. Finally, CTS back-off mechanism is used to adjust CTS DP. And we propose contention back-off mechanism to increase fairness further and reduce network congestion.

3.1. Spatial Unfairness of UASN. In terrestrial short-range RF network, propagation delay is negligible. While in UASNs it is essential to consider the location of nodes and transmission time due to the long propagation delay of acoustic media. In [21], Syed et al. first propose that due to long propagation delay of acoustic media, a collision at receiver has space-time uncertainty. In Figure 1(a), when nodes A and B transmit packets at different time, owing to the low propagation delay between A and N, the signals of A and B arrive N at the same time. There happens collision at N. In Figure 1(b), both nodes A and B transmit packets at the same time, owing to different propagation delay node N may successively receive the signals of A and B without collision. These examples show that in UASNs a collision in receiver not only depends on packets transmission time, but also depends on the location of nodes.

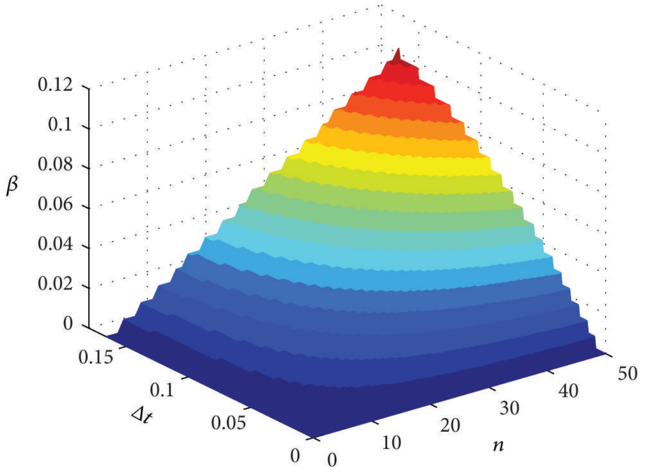


Figure 8: The relationship of n, Δt, and β.

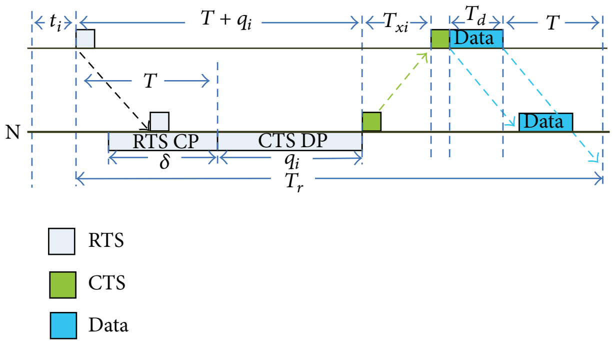


Figure 9: Data communication cycle.

Propagation delay of signal is proportional to the distance between nodes, in competition protocol node nearer the receiver easily obtain channel leading to spatial unfairness problem. As shown in Figure 1(c), nodes A and B request the channel, B sends request earlier, and A sends later, while the request packet of A arrives N earlier, node A may first obtain the channel. Therefore, the nodes nearer the receiver occupy the channel quickly, which make relatively distant nodes that cannot fairly use channel. That is spatial unfairness problem.

According to the above discussion of spatial unfairness problem, especially when all nodes frequently require for sending data, nodes away from the receiver cannot effectively obtain the channel. So, this paper proposes RET-MAC protocol to satisfy the spatial fairness. And it adopts adaptive RTS CP and CTS DP to achieve lower energy consumption and delay.

3.2. UASN Model. RET-MAC adopts receiver-based protocol and mainly handles the spatial unfairness. Therefore, in this paper network model consists of a single receiver and n contenders; contenders send RTS to receiver in order to obtain channel and then send data later. This paper assumes that network has these following properties.

(1) All nodes have a unique ID and are relatively fixed.

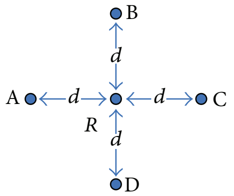


Figure 10: Simulation topology of delay.

(2) All nodes have similar capabilities (processing/communication), and clock is a weak synchronization.

(3) All contenders contend channel and send their data to the receiver.

(4) Contenders are randomly distributed in the largest transmission range of the receiver.

The third and fourth properties of network mainly proceed from the receiver. The protocol allocates the channel resources according to the sending time of RTS packet, satisfying that the channel is allocated to the earliest request node.

3.3. RTS Contention Phase (RTS CP). Based on RTS/CTS method RET-MAC responses to the earliest transmitter of RTS packet to allocate the channel, and all the control packets contain sending time of this packet. During RTS CP the receiver determines the contender which first sent RTS. Then, in order to avoid collision the receiver postpones CTS DP to send CTS. Other contenders which failed competition continue to contend channel after this communication is over. Figure 2 describes the basic idea of the RET-MAC protocol. Assuming that node N is receiver, nodes A, B, and F are contenders. Node N received RTS from A, B, and F in RTS CP and determined A that first sent RTS. Then, receiver N postponed CTS DP to send CTS to A. Finally, node A sent data to N.

Due to long propagation delay of acoustic media, receiver captures the first RTS which may not be the earliest sent, so we add RTS CP to determine the earliest transmitter of RTS packet.

Properties. The length of RTS CP is determined by both the currently earliest sending time of RTS and the time of the first captured RTS. And it is changing dynamically.

Proof. Assuming that the length of RTS CP is δ, the maximum transmission range of acoustic signal is R. The velocity of underwater acoustic is V. So, the maximum propagation delay is $T = R \times V$. In Figure 3, node N is receiver. Nodes A, B, C, and F are contenders. The distance between N and F is R.

Firstly, node N captures the first RTS from A at time t. At this time, N regards A as the earliest transmitter of RTS. Because the maximum length of propagation time is T, if the sending time of RTS is earlier than t_a, it certainly reaches N in the period of $T - (t - t_a)$. So, δ is equal to $T - (t - t_a)$. Then, node N captures RTS from B, while $t_b > t_a$, this RTS is ignored. At time t', N captures RTS from C, and $t_c < t_a$, so C is regarded as the earliest transmitter of RTS. If there is earlier RTS, it certainly arrives at N in the period of $T - (t' - t_c)$ after time t'. At this time, δ is equal to $T-(t-t_c)$ less than $T-(t-t_a)$. As you see, the length of RTS CP is contraction. And so on, we can achieve the earliest transmitter of RTS packet.

Therefore, the length of RTS CP is determined by both the earliest send time of RTS and the time of the first captured RTS. And it is changing dynamically according to currently earliest transmitter. □

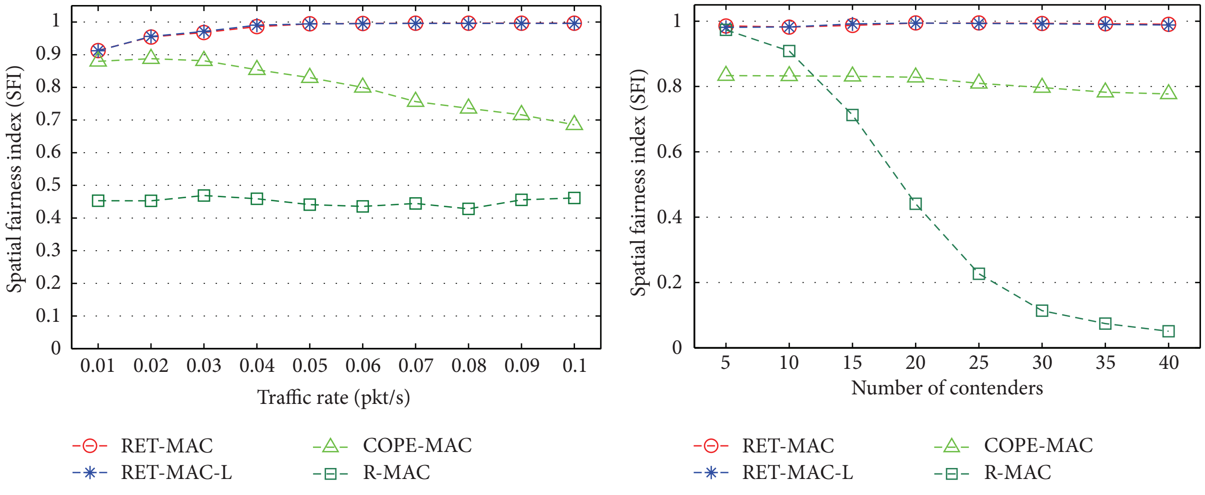
The pseudo-code of RTS CP algorithm determining the earliest transmitter of RTS and the end time of RTS CP is shown in Algorithm 1. The input of the algorithm is current captured RTS packet (r_frame).

3.4. Beforehand CTS (BCTS). In RET-MAC, contenders listen to any control packets from other nodes before sending RTS. In this case the channels have been occupied, and the contenders keep silent and do not participate in this competition. In order to avoid collision, RET-MAC postpones CTS DP to send CTS. However, the hidden contenders that is out of the maximum propagation range of some contenders keep silent until hearing CTS. If these contenders have data to send before receiving CTS, there may cause collision at receiver. Therefore, we add a Beforehand CTS (BCTS) to prevent collision caused by hidden nodes. When the receiver captures the first RTS, it immediately broadcasts BCTS to notify its neighbor that the channel is occupied.

In Figure 4, node N is receiver. Nodes A, B, and F are contenders. Node A achieved the channel and sent data, while node B was out of the maximum propagation range of A, and node B sent RTS (B_2) before hearing CTS. In this case, there may cause collision. Thus, when receiver captures the first RTS from A, it immediately broadcasts BCTS. Node B keeps silent after hearing BCTS. When the length of CTS DP is longer, it can effectively avoid collision caused by hidden nodes.

3.5. CTS Delay Phase (CTS DP). Due to adding BCTS in RET-MAC, and when the length of CTS DP is longer, it can effectively avoid collision. However, the longer handshake time increases energy consumption and seriously affects network throughput. So, in order to reduce delay and energy consumption and improve network throughput, the protocol should consider how to shorten the length of CTS DP without collision. Assuming that the length of CTS DP is q, all control packets (RTS, CTS, and BCTS) have equal size and their transmission times are T_c.

In Figure 5, node N is receiver. Nodes A, B, and F are contenders. The instance between N and B is the maximum transmission range, and the propagation delay between them is T. During RTS CP N determines the earliest transmitter F, and the length of RTS CP (δ) is equal to $T - (t - t_f)$. The propagation delay between F and N is $t' - t_f$; we can conclude that the time interval from t_f to the time of receiving CTS of F is $T + q + (t' - t_f)$, where the processing time of node is



(a)

(b)

FIGURE 11: (a) Spatial fairness index with traffic rate. (b) Spatial fairness index with number of contenders.

```
Function RTS_COMPETE
Begin function
  //Achieve send time of RTS
  r_time = r_frame → send_time
  if (r_time ≥ first_sendRTS_time)
          free(r_frame)              //Free this packet
  else
  //Update the earliest transmitter
          first_sendRTS_node = r_frame → sender
  //Update the earliest transmit time of RTS
          first_sendRTS_time = r_time
  //Update the end time of RTS CP
          RTS_compete_endtime = $T$+r_time
  end if
End Function
```

ALGORITHM 1: Pseudo-code of the RTS CP algorithm.

negligible. In order to avoid CTS collision, it must meet the constraints of

$$T + q + \left(t' - t_f\right) \geq 2T + T_c \Longrightarrow q \geq T + T_c - \left(t' - t_f\right). \quad (1)$$

In addition, the time interval from t to the time of receiving data of N is $\delta + q + 2(t' - t_f) + T_c$, and the last RTS may arrive N at t''. In order to avoid data collision, it must meet the constraints of

$$\delta + q + 2\left(t' - t_f\right) + T_c \Longrightarrow q \geq 2T + T_c - 2\left(t' - t_f\right) - \delta. \quad (2)$$

We have $0 \leq t' - t_f \leq T$, $\delta = T - (t - t_f)$, and $t' - t_f \geq t - t_f$. Thus, when we take $q \geq T + T_c$, it can fully satisfy (1) and (2). However, when the length of CTS DP is longer, the handshake time is larger. In this case the network throughput becomes very poor. Therefore, in RET-MAC, in order to reduce delay and improve network throughput we take the initial value of q equal to $T + T_c$.

BCTS is added, and node N may capture the last RTS at time t''. After time t''' there will not be any RTS which arrive at N. However, we take the initial value of q is $T + T_c$ and have

$$\delta + q = 2T + T_c - \left(t - t_f\right) < 2T + 2T_c. \quad (3)$$

From (3), at the end time of CTS DP, there may be RTS which arrive at N. Therefore, if N is in receiving state at the end time of CTS DP, we need to dynamically extend the CTS DP. This paper proposes CTS back-off mechanism to extend CTS DP.

3.6. CTS Back-off Mechanism. When receiver sends CTS at the end time of CTS DP, it may be receiving RTS. This paper proposes CTS back-off mechanism to extend CTS DP in order to be ready for sending CTS again. As the above discussion, the initial value of q is $T + T_c$, and when $q \geq 2T + 2T_c - \delta$, not any control packets arrive at receiver. Therefore, the length of CTS DP q is shown in

$$q = k_i = \begin{cases} T + T_c, & i = 0, \\ T + T_c + \dfrac{k_{i-1}}{2} - \dfrac{\delta}{2}, & i > 0,\ k_{i-1} + \delta < 2T, \\ 2T + 2T_c - \delta, & i > 0,\ k_{i-1} + \delta \geq 2T, \end{cases} \quad (4)$$

where i is the number of extending CTS DP. In fact, at the end time of CTS DP, the probability of receiving RTS is low. And with the length of CTS DP increasing, the probability decreases. In the part of analysis, this paper analyzes the maximum probability of extending CTS DP when q takes the initial value.

3.7. Contention Back-off Mechanism. Receiver only communicates with one contender in each communication. In other words, other failed contenders continue to contend the

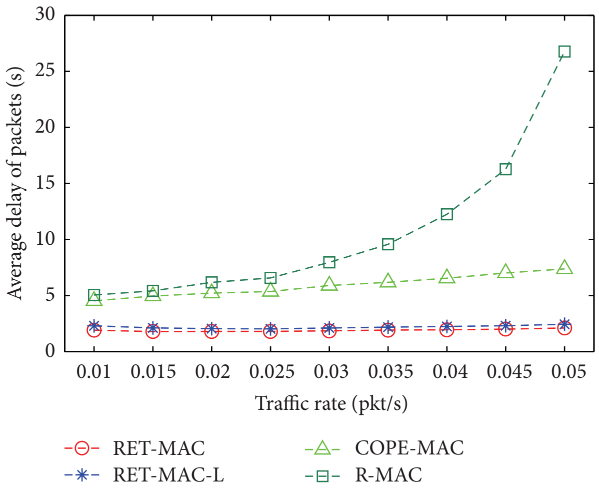


Figure 12: Average delay of packets with traffic rate.

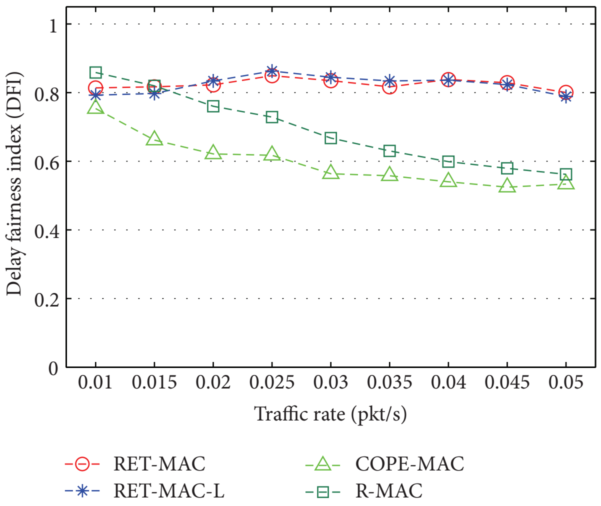


Figure 13: Delay fairness index with traffic rate.

channel after this communication ends. The failed contenders calculate the end time of this communication according to the send time of CTS and propagation delay. And in order to save energy, the failed contenders enter into sleep state until this communication ends. In order to satisfy the fairness of contending channel and avoid network congestion caused by all nodes sending RTS in short time, we propose contention back-off mechanism which makes all contenders postpone $\alpha(x)$ time to send RTS. In contention back-off mechanism, we take the number of contending channel of current data packet into $\alpha(x)$ that can increase the fairness of access channel further. The value of $\alpha(x)$ is given by

$$\alpha(x) = \left(\text{random }\ [0,1] + \left(1 - \frac{\mathrm{CN}(x)}{n}\right)\right) \times \mathrm{CW}, \qquad (5)$$

where n is the number of receiver's neighbor nodes, which is included in BCTS to notify neighbors. CN(x) is the number of what the current data packet contend channel. If data packet is sent successfully, CN(x) is 0, and the maximum value of CN(x) is n. CW is the size of contention window, we take it as T in simulation. From (5) it can be seen that when CN(x) is larger, $\alpha(x)$ is smaller. Contenders send RTS earlier and more easily achieve channel.

4. Analysis of Protocol Parameters

In this section, we first discuss the maximum probability of extending CTS DP in order to explain that the probability of extending CTS DP is very lower under the initial case. In addition, we also analyze the maximum throughput of network in detail under the above network model.

4.1. The Maximum Probability of Extending CTS DP. The Maximum Probability of extending CTS DP (MP-CD) is defined as the maximum probability of what CTS DP is extended to be ready again for sending CTS, because receiver may be receiving control packet at the end time of CTS DP. In Figure 6, node N is receiver. Nodes A, B, C, and F are contenders. The distance between C and N is the maximum propagation range. Nodes B and C are out of the max propagation range of F. We analyze MP-CD under the condition of $q = T + T_c$.

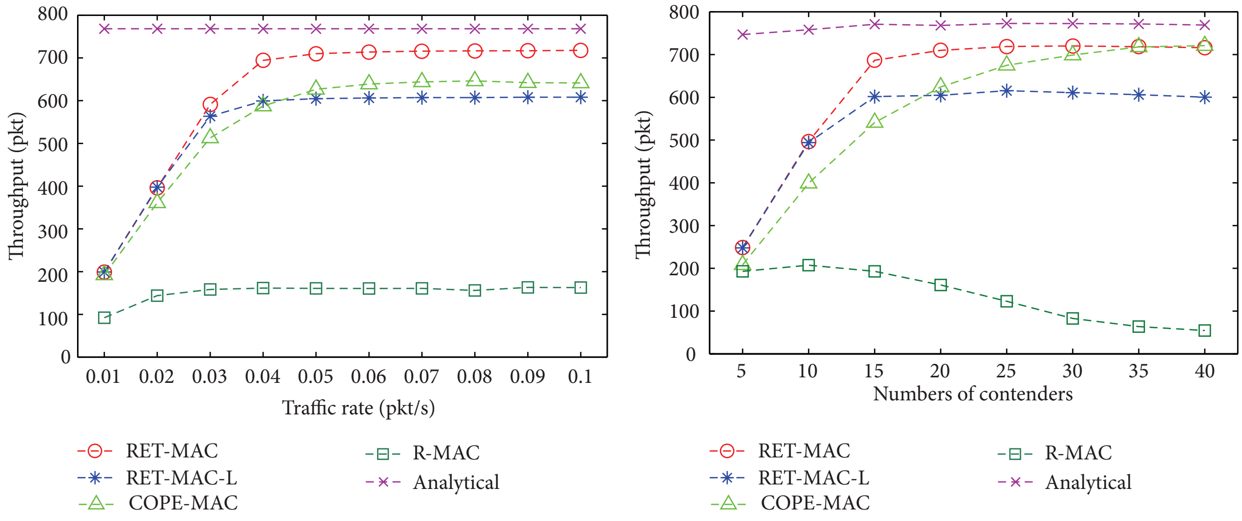
As shown in Figure 7, node N in RTS CP determined F which first sent RTS. The time interval from the sending time of F's RTS to the end time of CTS DP is $2T+T_c$. And at the end time of CTS DP, BCTS packet have been sent out for about $2T+\Delta t$, where $\Delta t = t - t_f$, and we define $r = (2T - \Delta t - T_c)/2$. Therefore, the control packets of all contenders in shadow area of Figure 6 cannot arrive at N at the end time of CTS delay phase. The contenders in the white areas of Figure 6 only in certain time send RTS that may arrive at N at the end time of CTS DP.

We assume that all contenders have data packets to send at the end of each communication and send RTS to contend channel. According to contention back-off mechanism, the maximum length of contention windows is $2T$. And some contenders send RTS in the time interval of the red areas of Figure 7 that may lead to extend CTS DP. Therefore, one contender may lead to the probability of extending CTS DP is

$$\varepsilon = \frac{T_c}{2T}. \qquad (6)$$

In addition, we assume that the distance between receiver and the earliest transmitter of RTS is d. We can conclude $0 < \Delta t \le d/C$, and the blank areas of Figure 6 are given by

$$\begin{aligned} s = {} & \pi R^2 - 2R^2 \arcsin\sqrt{1 - \frac{d^2}{4R^2}} + d\sqrt{R^2 - \frac{d^2}{4}} \\ & + 2\left(R^2 - r^2\right)\arccos\frac{R^2 + d^2 - r^2}{2dR} - 2r^2 \arccos\frac{R^2 + r^2 - d^2}{2Rr} \\ & - \frac{R^2 + d^2 - r^2}{2d^2}\sqrt{4R^2 d^2 - \left(R^2 + d^2 - r^2\right)^2}. \end{aligned} \qquad (7)$$



(a)

(b)

Figure 14: (a) Throughput with traffic rate. (b) Throughput with number of contenders.

The number of nodes in the blank areas of Figure 6 satisfies $N = \lceil n \times s/\pi R^2 \rceil$; n is the number of the receiver's neighbor. Thus, the maximum probability of extending CTS DP is given by

$$\beta = 1 - (1 - \varepsilon)^N. \tag{8}$$

Assuming that the maximum propagation range of nodes R is 500 meters, acoustic velocity V is 1500 meters per second, d set $R/2$. The size of control packets is 40 bit. Each modem transmits data at a speed of 10 kb/s, and n increases from 1 to 50. We can calculate that T_c is equal to 0.004 seconds, T is equal to 1/3 seconds, and the value of ε is 0.006, $0 < \Delta t \leq 1/6$. At this time, the relationship of n, Δt, and β is shown in Figure 8.

As shown in Figure 8, when n is constant, Δt is larger, and the value of β is larger. For example, we take $n = 20$, $\Delta t = 1/6$ seconds, so $\beta = 0.0413$. We can see that the probability of extending CTS DP is very low. So, the length of CTS DP mostly is equal to $T + T_c$. And according to (3), RET-MAC shortens the length of the sum of RTS CP and CTS DP. That is said that RET-MAC reduces delay and then improves network throughput. In fact, the value of β is far small than theoretical value. First of all, not all nodes have data packets to send at the same time. Secondly, contenders keep silent after hearing other RTS or BCTS. Finally, we conclude the theoretical value is larger than the fact value, because our assuming is the highest possible.

4.2. Maximum Throughput. This section estimates the maximum throughput of RET-MAC in the above network model. In this network model, the maximum throughput can be shown as the maximum packet which the receiver has received. We assume that the network is in ideal state without any collision. The processing delay of node is negligible. Every node has data packets to send and fairly transfer packets.

Figure 9 describes a data communication cycle. t_i is the idle time interval between the $(i - 1)$th communication and the ith communication. T_{xi} is the propagation delay between receiver and the earliest transmitter of RTS in ith communication. T_d is transmission delay of data packet. q_i is the length of CTS DP in ith transmission. So, the ith data packet is received successfully that need time T_{ri}, which is given by

$$T_{ri} = t_i + 2T + q_i + T_{xi} + T_c + T_d. \tag{9}$$

Assuming that the network operation time is T_t, η is throughput of receiver, so we can conclude that

$$\begin{gathered} T_t = \sum_{i=1}^{\eta} t_i + 2\eta T + \sum_{i=1}^{\eta} q_i + \sum_{i=1}^{\eta} T_{xi} + \eta T_c + \eta T_d \\ \Longrightarrow \eta = \frac{T_t}{(1/\eta)\sum_{i=1}^{\eta}(t_i + q_i + T_{xi}) + 2T + T_c + T_d}. \end{gathered} \tag{10}$$

From (10), the maximum throughput is relative to q and T. However, in specific operating conditions the value of T is fixed. There is no idle time with short CTS DP that can achieve high throughput. So, we take $t_i = 0, q = T + T_c$, and the theoretical maximum throughput of network is

$$\eta_{\max} = \frac{T_t}{3T + 2T_c + \overline{T_x} + T_d}, \tag{11}$$

where $\overline{T_x}$ is the average propagation delay between receiver and contenders. In simulation section, the difference between simulation result of RET-MAC throughput and the value of analysis is small. It is said that RET-MAC can effectively use channel resource and have lower collision rate.

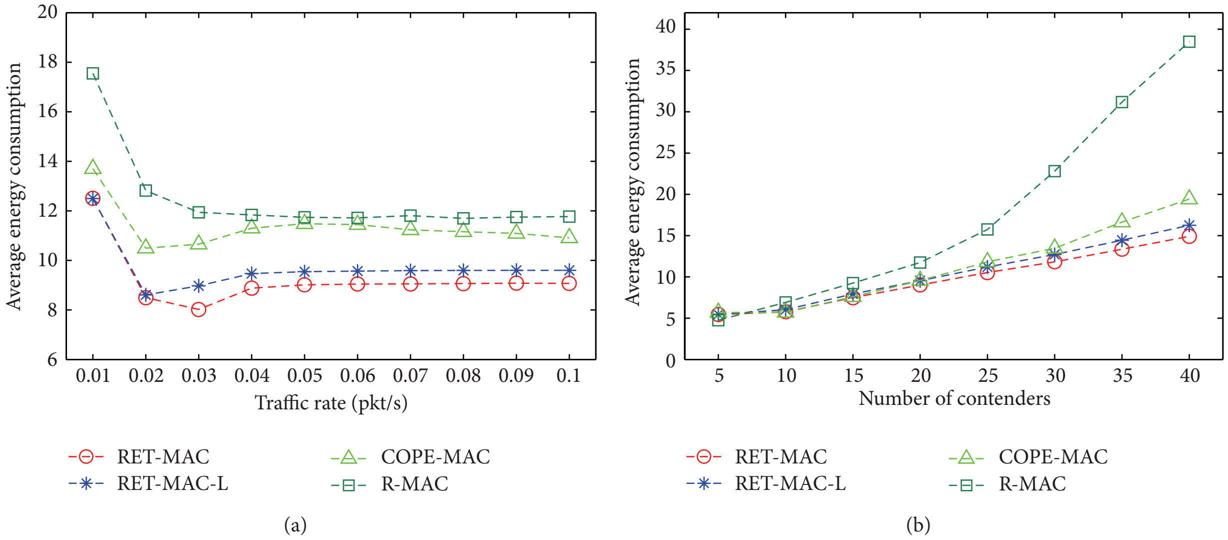


Figure 15: (a) Average energy consumption with traffic rate. (b) Average energy consumption with number of contenders.

5. Protocol Performance Evaluation

In this section, we first present main performance evaluation index. According to fairness index [18], we define Spatial Fairness Index and Delay Fairness Index. Secondly, simulation environment and working parameters are present. Finally, we analyze and compare RET-MAC with other MAC protocol by evaluating some important medium access metrics such as fairness, delay, throughput, and energy consumption.

5.1. Main Performance Evaluation Index. Fairness is an important aspect of underwater network protocol. RET-MAC is to solve spatial unfairness problem. This paper analyzes spatial fairness from throughput. If contenders have very good spatial fairness, when they have equal traffic rate, all nodes have equal data packets to send not because of their distance from the receiver. Therefore, this paper uses throughput to react spatial fairness. In addition, the transmit delay of data packets is also an important aspect of underwater network protocol; lower and fairness transmit delay can effectively improve network performance. The fairness index [22] has been widely used as a standard to deal with fairness problem. Thus, we also define the spatial fairness index and the delay fairness index based on it.

(1) *Spatial Fairness Index.* In this section, we analyze spatial fairness from throughput. We assume that in aforementioned network model there are n contenders. According to this paper needs, we define Spatial Fairness Index (SFI) as follows:

$$\mathrm{SFI} = \frac{\left(\sum_{i=1}^{n} S_i\right)^2}{n \times \left(\sum_{i=1}^{n} S_i^2\right)}, \qquad (12)$$

where n is the number of contenders, and S_i is throughput of ith node. If all contenders can fairly access channel, their throughput is equal, at this time SFI set 1. If one node always occupies channel, SFI set $1/n$. In short, if the value of SFI is close to 1, the spatial fairness of network is better.

(2) *Delay Fairness Index.* The transmit delay of packet includes waiting delay, transmission delay, and propagation delay. In UWSN, propagation delay is longer. And it is proportional to the distance between nodes. Therefore, under network model of Figure 10 we analyze delay fairness. In this way, transmission delay and propagation delay of packet are equal, so delay fairness of packet depends on fairness of waiting time.

We still use the fairness index [22] to define Delay Fairness Index (DFI) as follows:

$$\mathrm{DFI} = \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n \times \left(\sum_{i=1}^{n} d_i^2\right)}, \qquad (13)$$

where n is the number of packets which is received successfully. If delay is equal, DFI set 1. If delay jitter is larger, DFI is small. In short, if the value of DFI is close to 1, fairness of waiting delay is better.

5.2. Simulation Environment and Working Parameters. We implement RET-MAC protocol in Aqua-Sim [23], an NS-2 based simulator for UWSN, developed at the Underwater Sensor Network (UWSN) Lab at the University of Connecticut. The simulation environment consists of a single receiver and n transmitters. All transmitters are randomly dispersed in the sensing region of receiver. Each transmitter will transmit data, which follows the Poisson arrival process with average traffic generation rate λ. For power consumption, we also use the numbers from a practical acoustic modem as follows: transmitting power 50 W, receiving power 3 W, and idle power 80 mW [24]. All simulations last for 1000 seconds and all the results are obtained from the average of 100 runs. The setting of key simulation parameters are listed in Table 1.

Table 1: Simulation parameters.

Simulation parameters	Values
Propagation range R	500 meters
Underwater Acoustic Velocity V	1500 meters per second
Transmitting Power	50 W
Receiving Power	3 W
Idle Power	80 mW
Data transmission rate	10 kbps
Size of Data packet	60 Bytes
Size of Control packet	40 bits
Simulation Cycle	1000 seconds
Contention Windows CW	1/3 seconds

We compare RET-MAC with another MAC protocol by evaluating the important medium access metrics such as fairness, throughput, energy consumption, and delay. For first three aspects, we consider two different scenarios where Scenario-1 has 20 transmitters, and the data generation rate λ of each transmitter increases from 0.01 to 0.1 packets per second. In Scenario-2, we fix traffic rate λ to 0.05, and the number of contenders increases from 5 to 40. For delay simulation environment, we adopt 4 contenders and one receiver based on Figure 10. The data generation rate λ in this scenario increases from 0.01 to 0.05 packets per second.

5.3. Analysis of Simulation Results. In this section, through simulations, we would like to study the performance of RET-MAC and compare it with COPE-MAC [14], R-MAC [20], and RET-MAC-L in UWSN. COPE-MAC protocol is a new MAC protocol based on RTS/CTS. It adopts parallel reservation to improve communication efficiency and uses cyber carrier sensing to detect and avoid collisions. In the long propagation delay of UASNs especially, COPE-MAC can improve MAC performance in both network throughput and energy efficiency. R-MAC is a reservation-based MAC protocol. It carefully schedules the transmissions of control and data packets avoiding data packet collision completely. In R-MAC, each node adopts periodic listen/sleep to reduce energy waste in idle state and overhearing. And intended receiver randomly selects one reservation from the reservations that it collects, so it supports fairness. RET-MAC-L is a special case of RET-MAC protocol. Due to adding BCTS, we take $\delta + q = 2T + 2T_c$. In this case when receiver sends CTS, there will not be any node occupying the channel. So we called it RET-MAC-L protocol.

(1) *Spatial Fairness*. According to Section 5.1, we analyze spatial fairness by (12) SFI. Figure 11 shows that both RET-MAC and RET-MAC-L have very good spatial fairness. No matter what data rate is high or the number of nodes is large, SFI is approximately equal to 1. Thus, both ways of response-to-the-earliest-transmitter and contention back-off mechanism can provide effectively fair access channel of contenders. However, SFI of COPE-MAC decreases with traffic rate and the number of nodes increasing. Although in COPE-MAC the intended receiver will randomly pick the senders to establish connection and by one reservation it can transmit more packets at different time, it does not weigh the fairness access of all nodes. Also, despite the fact that R-MAC supports fairness, in Figure 11(b) when the number of nodes becomes larger, collision rate is very high. So, its fairness shows very bad compared with other protocols. While just the number of contenders is small, R-MAC shows better fairness. For example, in Figure 11(b) the traffic rate is fixed, when the number of contenders is 5, and SFI is equal to 1. But when the number of contenders increases to 40, SFI is less than 0.05.

(2) *Delay*. Figure 12 shows the average transmit delay of packets under the structure of Figure 10. As we can see from Figure 12, RET-MAC has lower average transmit delay of packet, because using dynamic adaptive RTS CP and CTS DP makes both the handshake time and makes the transmit delay of packet shorter. While RET-MAC-L uses fixed time, so the average delay of packets is slightly higher than RET-MAC. For R-MAC protocol, collision rate increased with traffic rate. So, when traffic rate is higher, average delay of packets is larger. Figure 13 shows delay fairness of packets. We analyze delay fairness of packets by (13) DFI in Section 5.1. The delay fairness of packets of both COPE-MAC and R-MAC reduces with traffic rate increasing. However, RET-MAC still has much more stable delay fairness; its DFI maintain at 0.8 or more, even remaining good delay fairness in case of lower average delay.

(3) *Throughput*. In the above simulation model, channel utilization can be replaced by throughput. As shown from Figure 14, both two changes of traffic rate and contenders, RET-MAC has better throughput. In Figure 14 especially, when traffic rate or contenders are lower, the packet delivery ratio of RET-MAC is equal to 100%, because RET-MAC used dynamically adaptive RTS CP and CTS DP, under avoiding collision which can effectively shorten the length of handshake time to improve network throughput. Furthermore, in Figure 14 data line of “Analytical” is drawn according to (11) and simulation environment parameters. There is only a small gap between RET-MAC throughput and analytical thought in Figure 14. There are two main reasons: first, owing to adopt contention back-off mechanism, there is a small idle time between twice data communications; second, the data line of “Analytical” is calculated under the ideal state.

COPE-MAC can achieve higher network throughput with using parallel reservations and cyber carrier sensing mechanism, but it is still lower than RET-MAC. Due to using a fixed length of sum of RTS CP and CTS DP, the handshake time of RET-MAC-L becomes longer than RET-MAC. Therefore, its throughput becomes lower. R-MAC uses listen/sleep mechanism which seriously impact on the network throughput. And R-MAC does not apply to intensive networks; otherwise there is lower throughput.

(4) *Energy Consumption*. Energy consumption is another important aspect of UASNs, because lower energy consumption can effectively extend lifetime of network. From Figure 15(a), we can see that for all protocols, average energy

consumption of packet decreases at first with traffic rate increasing. When traffic rate is equal to 0.01 pkt/s, in idle state energy consumption accounts for most of all energy consumption. So, average energy consumption of packet is higher. After traffic rate arrived at 0.04 pkt/s, the data lines of average energy consumption leveled off. Figure 15(a) also shows that RET-MAC can achieve much higher energy efficiency than other protocols. From Figure 15(b), we can see that average energy consumption increases with number of contenders increasing. Compared with other protocols, RET-MAC has lower energy consumption. Although R-MAC adopts listen/sleep mechanism to save energy, network collision increases with contenders. Therefore, when the number of contenders arrives at 40, average energy consumption of packet is higher. In addition, because RET-MAC shortens handshake time and reduces waiting time of contenders, average energy consumption of RET-MAC is lower than RET-MAC-L.

6. Conclusion

This paper discussed spatial unfairness problem and proposed Response to the Earliest Transmitter of RTS MAC (RET-MAC) protocol to achieve fairness. RET-MAC adopts adaptive RTS CP to determine the earliest transmitter of RTS. And CTS DP is added to postpone sending CTS in order to avoid collision. We also proposed CTS back-off mechanism to adjust the length of CTS DP as needed. In addition, contention back-off mechanism is used to reduce network congestion and increase fairness further. The simulation results show that our scheme can achieve higher fairness and throughput. At the same time, it also guarantees lower energy consumption and delay. As future work, we will still explore fairness problem when time is not synchronization, and study fairness control is based on network load and amount of sensed information in order to enhance fair using of network bandwidth.

Acknowledgments

This work is supported by the Foundation of Jiangsu University under Grant no. 12JDG103, National Natural Science Foundation of China under Grant no. 61272074, Science and Research Projects for Graduate Students of Jiangsu University under Grant no. 11A135, the Foundation of Jiangsu University under Grant no. 12JDG104 and College Students' Practice Innovation Training Program of Jiangsu Province under Grant no. 2012JSSPITP1256.

References

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, 2005.

[2] Z. Peng, Y. Zhu, Z. Zhou, Z. Guo, and J. H. Cui, "COPE-MAC: a Contention-based medium access control protocol with Parallel Reservation for underwater acoustic networks," in *Proceedings of the IEEE Oceanic Engineering Society (OCEANS '10)*, pp. 1–10, May 2010.

[3] N. Chirdchoo, W. S. Soh, and K. C. Chua, "Aloha-based MAC protocols with collision avoidance for underwater acoustic networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 2271–2275, May 2007.

[4] P. Casari and M. Zorzi, "Protocol design issues in underwater acoustic networks," *Computer Communications*, vol. 34, no. 17, pp. 2013–2025, 2011.

[5] J. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater senor networks: applications, advances, and challenges," *Philosophical Transactions of the Royal Society A*, vol. 370, pp. 158–175, 1958.

[6] A. A. Syed, W. Ye, and J. Heidemann, "Comparison and evaluation of the T-Lohi MAC for underwater acoustic sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 9, pp. 1731–1743, 2008.

[7] L. Chen, X.-J. Li, Y.-B. Guo, and L.-M. Wang, "RUF-MAC: related and urgent first MAC for wireless sensor networks," *Applied Mathematics & Information Sciences*, vol. 7, no. 3, pp. 1237–1244, 2013.

[8] G. A. Shah, "A survey on medium access control in underwater acoustic sensor networks," in *Proceedings of the International Conference on Advanced Information Networking and Applications Workshops (WAINA '09)*, pp. 1178–1183, May 2009.

[9] Z. Zhou, Z. Peng, P. Xie, J.-H. Cui, and Z. Jiang, "Exploring random access and handshaking techniques in underwater wireless acoustic networks," *EURASIP Journal on Wireless Communications and Networking*, 2013.

[10] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 3, pp. 493–506, 2004.

[11] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp. 316–329, 2006.

[12] I. Rhee, A. Warrier, M. Aia, J. Min, and M. L. Sichitiu, "Z-MAC: a hybrid MAC for wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 511–524, 2008.

[13] W.-H. Liao and C. Huang, "SF-MAC: a spatially fair MAC protocol for underwater acoustic sensor networks," *IEEE Sensors Journal*, vol. 12, no. 6, pp. 1686–1694, 2012.

[14] Z. Peng, Y. Zhu, Z. Zhou, Z. Guo, and J. H. Cui, "COPE-MAC: a Contention-based medium access control protocol with Parallel Reservation for underwater acoustic networks," in *IEEE Oceanic Engineering Society (OCEANS' 10)*, May 2010.

[15] M. K. Park and V. Rodoplu, "UWAN-MAC: an energy-efficient MAC protocol for underwater acoustic wireless sensor networks," *IEEE Journal of Oceanic Engineering*, vol. 32, no. 3, pp. 710–720, 2007.

[16] C. C. Hsu, K. F. Lai, C. F. Chou, and K. C. J. Lin, "ST-MAC: spatial-temporal MAC scheduling for underwater sensor networks," in *Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM '09)*, pp. 1827–1835, April 2009.

[17] Z. Azar and M. T. Manzuri, "A latency-tolerant MAC protocol for underwater acoustic sensor networks," in *Proceedings of the International Conference on Control, Automation and Systems (ICCAS '10)*, pp. 849–854, October 2010.

[18] H.-J. Cho, J.-I. Namgung, N.-Y. Yun, S.-H. Park, C.-H. Kim, and Y.-S. Ryuh, "Contention free MAC protocol based on priority in underwater acoustic communication," in *Proceedings of the*

IEEE Oceanic Engineering Society (OCEANS '11), pp. 1–7, June 2011.

[19] L. Hong, F. Hong, Z. Guo, and Z. Li, "ECS: efficient communication scheduling for underwater sensor networks," *Sensors*, vol. 11, no. 3, pp. 2920–2938, 2011.

[20] P. Xie and J. H. Cui, "R-MAC: an energy-efficient MAC protocol for underwater sensor networks," in *Proceedings of the 2nd Annual International Conference on Wireless Algorithms, Systems, and Applications (WASA '07)*, pp. 187–195, August 2007.

[21] A. A. Syed, W. Ye, J. Heidemann, and B. Krishnamachari, "Understanding spatio-temporal uncertainty in medium access with ALOHA protocols," in *Proceedings of the ACM the 2nd Workshop on Underwater Networks*, pp. 41–48, ACM, September 2007.

[22] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Tech. Rep. TR-301, Digital Equipment Corporation, Hudson, Mass, USA, 1984.

[23] P. Xie, Z. Zhou, Z. Peng et al., "Aqua-sim: an NS-2 based simulator for underwater sensor networks," in *Proceedings of IEEE/MTS Oceans (IEEE OCEANS '09)*, pp. 1–7, October 2009.

[24] L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, and K. Ball, "The WHOI Micro-Modem: an acoustic communications and navigation system for multiple platforms," in *Proceedings of the IEEE Oceans Conference*, vol. 2, pp. 1086–1092, September 2005.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 176293, 11 pages
http://dx.doi.org/10.1155/2013/176293

Research Article

Traffic Prediction-Based Fast Rerouting Algorithm for Wireless Multimedia Sensor Networks

Zhiyuan Li,[1] Junlei Bi,[1] and Siguang Chen[2]

[1] *School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China*
[2] *College of Science and Engineering, City University of Hong Kong, Hong Kong*

Correspondence should be addressed to Zhiyuan Li; lizhiyuan81@126.com

Received 23 February 2013; Accepted 28 April 2013

Academic Editor: Lu Liu

Rerouting has become an important challenge to Wireless Multimedia Sensor Networks (WMSNs) due to the constraints on energy, bandwidth, and computational capabilities of sensor nodes and frequent node and link failures. In this paper, we propose a traffic prediction-based fast rerouting algorithm for use between the cluster heads and a sink node in WMSNs (TPFR). The proposed algorithm uses the autoregressive moving average (ARMA) model to predict a cluster head's network traffic. When the predicted value is greater than the predefined network traffic threshold, both adaptive retransmission trigger (ART) that contributes to switch to a better alternate path in time and trigger efficient retransmission behaviors are enabled. Performance comparison of TPFR with ant-based multi-QoS routing (AntSensNet) and power efficient multimedia routing (PEMuR) shows that they: (a) maximize the overall network lifespan by load balancing and not draining energy from some specific nodes, (b) provide high quality of service delivery for multimedia streams by switching to a better path towards a sink node in time, (c) reduce useless data retransmissions when node failures or link breaks occur, and (d) maintain lower routing overhead.

1. Introduction

Efficiently transmitting multimedia streams in wireless multimedia sensor networks (WMSNs) is a significant challenging issue, due to the limited transmission bandwidth and power resource of sensor nodes. Three recent surveys [1–3] on current trends and future directions in WMSNs show that to overcome various failures, such as node failures, link breaks, network congestion, and dynamic holes, routing has the responsibility of choosing an alternate path that is not optimal to continually deliver the multimedia streams which can cause huge rerouting overhead. These three surveys also expatiate that there is no solution focusing on addressing the rerouting problem of multimedia streaming in WMSNs. Thus, more rerouting algorithm explorations are required to adapt to topology changes caused by various failures and guarantee the quality of service of multimedia streaming delivery.

Rerouting over WMSNs is different from the existing routing protocols for scalar wireless sensor networks [2, 3]. It is a very critical and challenging issue due to the stringent quality of service (QoS) requirements of multimedia (video streaming, still images, and audio) transmission, such as (1) the end-to-end delay, (2) the packet delivery rate, and (3) the PSNR (peak signal-to-noise ratio) level. Hence, a fast rerouting mechanism is required in order to avoid various failures resulting in service interruption.

This paper proposes a traffic prediction-based fast rerouting (TPFR) algorithm for use among the cluster heads in WMSNs. TPFR runs on the top of the uneven cluster network topology, because the uneven clustering network model may provide a valuable solution to balance the network loads and prolong the lifetime of WMSNs [4]. According to the literature [5], the intercluster multipath routing is discovered. And then, we use autoregressive moving average (ARMA) model to predict the cluster head's network traffic. When the predicted value is greater than the predefined network traffic threshold, both adaptive retransmission trigger (ART) that contributes to switch to the better alternate path and trigger efficient retransmission behaviors are enabled. In consequence, this failure area is smoothly bypassed, and multimedia streams are continually forwarded to the destination

Table 1: Multipath routing and transmission protocols for WMSNs.

Protocol	Network architecture	Geographic routing	Operational layer	Fault-tolerant mechanism	Performance metric
MMSPEED [6]	Flat	Yes	Routing/MAC	Weak	Delay/overhead
TPGF [7]	Flat	Yes	Routing	Medium	Delay/hop count
MPMPS [8]	Flat	Yes	Routing	Medium	Distance/delay/data type
AntSensNet [5]	Hierarchical	No	Transport/routing	Medium	Delay/packet delivery ratio/overhead
PEMuR [9]	Hierarchical	No	Routing/MAC	Medium	Delay/energy consumption/PSNR

node. Finally, TPFR is implemented on the NS-2 platform. Compared with similar algorithms, TPFR can significantly improve the quality of data transmission services. Moreover, TPFR has lower energy consumption and routing overhead and can prolong the network lifetime.

The rest of the paper is organized as follows. Section 2 introduces an overview of existing related works. Section 3 provides the network architecture, the system model, and the assumptions. Section 4 presents the traffic prediction-based fast rerouting algorithm. Section 5 presents the theoretical analysis and the performance evaluation. Finally, Section 6 concludes the paper.

2. Related Work

In this section, we focus on multipath routing protocols for WMSNs that include routing and scheduling functionalities, and we summarize them in Table 1.

Multipath and Multi-SPEED (MMSPEED) routing protocol [6] supports probabilistic QoS guarantee by provisioning QoS in two domains, timeliness and reliability. MMSPEED adopts a differentiated priority packet delivery mechanism in which QoS differentiation in timeliness is achieved by providing multiple network-wide packet delivery speed guarantees. MMSPEED needs the support of IEEE 802.11e at the MAC layer with its inherent prioritization mechanism based on the differentiated interframe spacing (DIFS). Each speed level is mapped onto a MAC layer priority class. For supporting service reliability, probabilistic multipath forwarding is used to control the number of delivery paths based on the required end-to-end reaching probability. In the scheme, each node in the network calculates the possible reliable forwarding probability value of each of its neighbors to a destination by using the packet loss rate at the MAC layer. According to the required reliable probability of a packet, each node can forward multiple copies of packets to a group of selected neighbors in the forwarding neighbor set to achieve the desired level of reliability. MMSPEED could use its redundant path selection scheme for load balancing, which is not only for reliability enhancement, but also to improve the overall network lifetime. However, the drawback of the protocol is that it shows degraded performance in handling various holes and the sudden network congestion.

The two-phase geographical greedy forwarding (TPGF) routing protocol [7] focuses on exploring and establishing the maximum number of disjoint paths to the destination in terms of the minimization of path length, the end-to-end transmission delay, and the energy consumption of nodes. The first phase of TPGF algorithm explores the possible paths to the destination. During this phase, a step back and mark is used to bypass voids and loops until successfully a sensor node finds a next-hop node which has a routing path to the base station. The second phase is responsible for optimizing the discovered routing paths with the shortest transmission distance (i.e., choosing a path with least number of hops to reach the destination).

The MPMPS (multipriority multipath selection) routing protocol [8] is an extension of TPGF. MPMPS highlights the fact that not every path found by TPGF can be used for transmitting video because a long routing path with long end-to-end transmission delay may not be suitable for audio/video streaming. Furthermore, because in different applications, audio and video streams play different roles and the importance level may be different, it is better to split the video stream into two streams (video/image and audio). For example, video stream is more important than audio stream in fire detection because the image reflects the event; while audio stream is more important in deep ocean monitoring. Therefore, we can give more priority to the important stream depending on the final application to guarantee the using of the suitable paths.

It is worth to note that both TPGF and MPMPS are offline multipath routing protocols. However, these "offline multipath" protocols have to explore the multiple routes that may exist between the source and the destination before the actual data delivery phase. They may not be well adapted for large-scale highly dense unattended network deployments and for networks with frequent node mobility.

The literature [5] proposes a QoS routing algorithm for WMSNs based on an improved ant colony algorithm (AntSensNet). The AntSensNet algorithm introduces routing modeling with four QoS metrics associated with nodes or links. The algorithm can find a route in a WMSNs that satisfies the QoS requirements of an application, while simultaneously reducing the consumption of constrained resources as much as possible. Moreover, by using clustering, it can avoid congestion after quickly judging the average queue length and solve convergence problems, which are typical in ant colony optimization. In addition, AntSensNet is able to use an efficient multipath video packet scheduling in order to get minimum video distortion transmission.

Power efficient multimedia routing (PEMuR) [9] aims at both energy efficiency and high QoS attainment. To achieve its objectives, PEMuR proposes the combined use of an

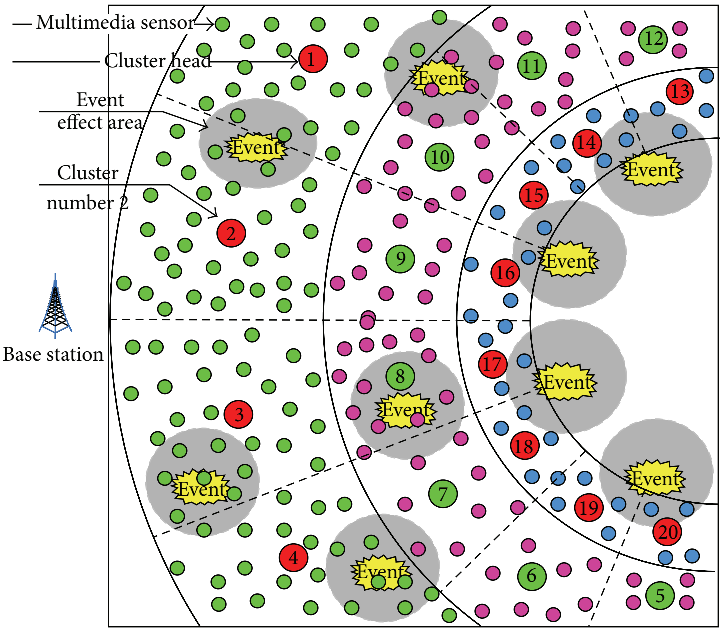


Figure 1: Graphical depiction of the nonuniform clustering architecture adopted by TPFR.

energy aware hierarchical routing protocol with an intelligent video packet scheduling algorithm. The routing protocol enables the selection of the most energy efficient routing paths and manages the network load according to the energy residues of the nodes. In this way, an outstanding level of energy efficiency is achieved. Additionally, the proposed packet scheduling algorithm enables the reduction of the video transmission rate with the minimum possible increase of distortion. In order to do so, it makes use of an analytical distortion prediction model that can accurately predict the resulted video distortion due to any error pattern. Thus, the algorithm may cope with limited available channel bandwidth by selectively dropping less significant packets prior to their transmission.

Both AntSensNet and PEMuR are "online" energy efficient hierarchical multipath routing protocols. However, these "online multipath" protocols lack fast rerouting mechanism when the various failures of nodes or links happen. Thus, the drawback of the two protocols is that the QoS of the video stream transmission rapidly degrades in handling various holes and sudden failures. In consequence, they may not be well adapted for the resource-constrained WMSNs and the stringent quality of service (QoS) requirements of multimedia transmission.

Hence, we propose a novel online fast rerouting algorithm called TPFR that (1) reduces video distortion using multi-priority-level multipath transmission model, (2) predicts network traffic through cluster heads using autoregressive moving average (ARMA) model, and (3) fast routes packets through better alternate paths using the traffic prediction strategy for avoiding various failures.

3. System Model

The many-to-one traffic pattern results in the hot spot problem when the multihop forwarding mode is adopted in intercluster communication for WMSNs. Because the cluster heads closer to the base station are burdened with heavier relay traffic, the area near the base station becomes a hot spot. Nodes in the hot spot drain their energy and die much faster than other nodes in the network, reducing sensing coverage and causing network partitions. Although many protocols proposed in the literature reduce energy consumption on forwarding paths to increase energy efficiency, they do not necessarily extend the network lifetime due to the unbalanced energy consumption.

We divide the network into uneven clusters using our proposed protocol, called UCBCPNS [10], where each cluster is deployed with heterogeneous sensors (camera, audio, and scalar sensors) that communicate directly in a certain schedule with a cluster head and relay their sensed data to it. Moreover, these heterogeneous sensor nodes have the same radio interface and propagation range. A cluster head has more resources, and it is able to perform intensive data processing. These powerful nodes and cluster heads are deployed nonuniformly in the network, and they are wirelessly connected with the sink either directly (in case of first-level cluster heads) or through other cluster heads in multihop mode. The graphical depiction of the nonuniform clustering architecture is shown in Figure 1. Our algorithm runs on the top of the nonuniform clustering network topology.

Then, the queuing model on a multimedia sensor node is designed, which is shown in Figure 2. According to the urgency and importance of the data streams, the model sets the different priorities for the different types of data streams and allows the high priority data stream to firstly transmit on a better path. For example, there are three types of data streams to be transmitted, such as video stream, sound stream, and scalar data stream. According to different application scenarios and demands, the system may

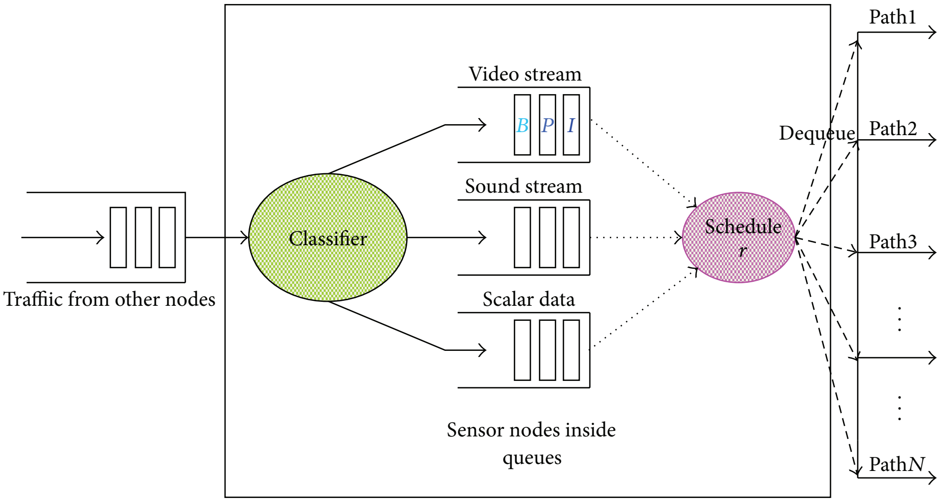


FIGURE 2: Queuing model on a multimedia sensor node.

automatically set different priority levels for different types of traffic. When a cluster head receives different types of traffic from other cluster heads, the received traffic is divided into three types, namely, video stream, sound stream, and scalar data by using the classifier model in the node's inside queues. Then the system makes a decision on the forwarding sequence of different types of traffic reference to the priorities set by itself. It is worth to note that the video sequence begins with an I-frame and is followed by P-frames and B-frames. I-frame in the video streams is a key frame, and P-frames and B-frames are nonkey frames. In a group of pictures, the decoding of P-frames and B-frames depends on the I-frame. If the I-frame is lost, the P-frames and the B-frames become useless data, which not only affects the quality of the video decoding, but also will result in the waste of network resources [11]. In our scheme, I-frame is firstly delivered on a better path.

The scheduler in Figure 2 has two functions which are similar to the function of routing. One is responsible for delivering the higher priority data streams to the optimal primary routing, and the other is responsible for fast rerouting the data streams to another better alternate route when various failures happen. The first function has been achieved using the AntSensNet [5], and the second function will be achieved using the TPFR proposed in this paper.

4. Traffic Prediction-Based Fast Rerouting

Internet traffic prediction plays a fundamental role in network design, management, control, and optimization [12]. Essentially, the statistics of network traffic itself determines the predictability of network traffic. Network traffic prediction for WMSNs is the process of mapping past (and present) traffic values onto future traffic values through linear or nonlinear mapping functions as shown in

$$\widehat{X}(t+k) = F\left[X(t), X(t-1), \ldots, X(t-p+1)\right], \qquad (1)$$

where the function F maps the past p traffic values $X(t)$, $X(t-1), \ldots, X(t-p+1)$ onto the k step-ahead traffic value $\widehat{X}(t+k)$.

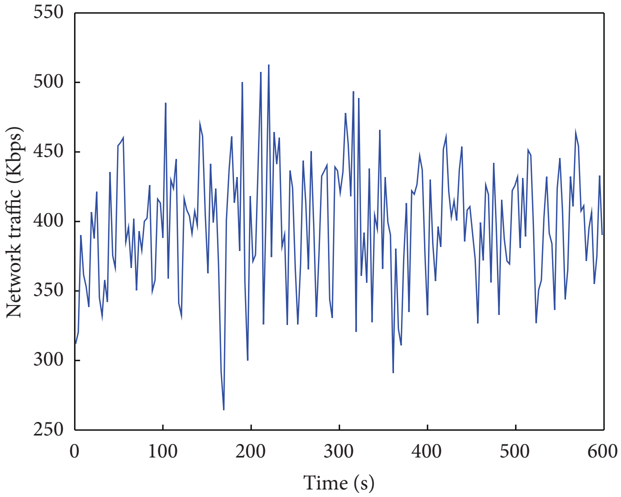


FIGURE 3: Original network traffic.

The design of a traffic prediction scheme mainly concerns constructing or devising the proper mapping functions.

4.1. Traffic Prediction Model Using Autoregressive Moving Average. We firstly gather enough network traffic from a gateway. The hybrid network traffic includes the multimedia data generated by the MeshEye nodes and the scalar data generated by the Mica2 nodes, which is shown in Figure 3. Assume that the time series of the collected traffic is $\{X_i\}$. Then the time series $\{X_i\}$ is analyzed using the famous commercial statistical software named SAS, and we find that the time series $\{X_i\}$ is a stationary and non-Gaussian white noise sequence. The modeling is described as follows.

(1) Sample autocorrelation coefficients and partial correlation coefficients: we obtain the sample autocorrelation coefficient of X_i using

$$\widehat{\rho}_k = \frac{\sum_{t=1}^{n-k}\left(x_t - \overline{x}\right)\left(x_{t+k} - \overline{x}\right)}{\sum_{t=1}^{n}\left(x_t - \overline{x}\right)^2}, \quad \forall\, 0 < k < n. \qquad (2)$$

The sample partial correlation coefficient of X_i is obtained by using

$$\widehat{\phi}_{kk} = \frac{\widehat{D}_k}{\widehat{D}}, \quad \forall\, 0 < k < n, \tag{3}$$

where

$$\begin{gathered} \widehat{D} = \begin{vmatrix} 1 & \widehat{\rho}_1 & \cdots & \widehat{\rho}_{k-1} \\ \widehat{\rho}_1 & 1 & \cdots & \widehat{\rho}_{k-2} \\ \vdots & \vdots & & \vdots \\ \widehat{\rho}_{k-1} & \widehat{\rho}_{k-2} & \cdots & 1 \end{vmatrix}, \\ \widehat{D}_k = \begin{vmatrix} 1 & \widehat{\rho}_1 & \cdots & \widehat{\rho}_1 \\ \widehat{\rho}_1 & 1 & \cdots & \widehat{\rho}_2 \\ \vdots & \vdots & & \vdots \\ \widehat{\rho}_{k-1} & \widehat{\rho}_{k-2} & \cdots & \widehat{\rho}_k \end{vmatrix}. \end{gathered} \tag{4}$$

After that, we find that the two correlation coefficients $\widehat{\rho}_k$ and $\widehat{\phi}_{kk}$ show significant tailing. Thus, we use the ARMA (p, q) model to fit the time series $\{X_i\}$.

(2) The order p and the order q of the ARMA process: the Akaike information criterion (AIC) [13] is used to select the order p and the order q, which is shown in formula

$$\mathrm{AIC} = -2 \ln \left(\widehat{\beta}\right) + 2M, \tag{5}$$

where M denotes the number of unknown parameters in the model and $\widehat{\beta}$ denotes the maximum likelihood estimates of β.

The logarithm likelihood function in (5) is denoted by formula

$$\ln \left(\widehat{\beta}; x_1, \ldots, x_n\right) = -\left[\frac{n}{2} \ln \widehat{\sigma}_{\varepsilon}^2 + \frac{n}{2} + \frac{n}{2} \ln 2\pi\right]. \tag{6}$$

Combine formula (5) and formula (6), and then we can get formula

$$\mathrm{AIC}(p, q) = n \ln \widehat{\sigma}_{\varepsilon}^2 + 2(p + q + 1). \tag{7}$$

We solve the minimum value of the function AIC (p, q) and obtain that the minimum of AIC occurs at the order p equals 2 and the order q equals 1. As a consequence, we use the ARMA $(2, 1)$ model to fit the time series $\{X_i\}$, which is shown in

$$\begin{gathered} \Phi(B) X_i = \Theta(B) a_i, \\ \Phi(B) = 1 - \phi_1 B - \phi_2 B^2, \\ \Theta(B) = 1 - \theta_1 B, \end{gathered} \tag{8}$$

where B is a backward shift operator, $\{a_i\}$ is a Gaussian white noise with mean zero and variance σ^2, and ϕ_1, ϕ_2, θ_1, and σ_a^2 (white noise variance) are parameter estimation. We use the least squares estimation method to estimate the parameters ϕ_1, ϕ_2, θ_1, and σ_a^2 in time series $\{X_i\}$ due to the resource-constrained sensor nodes. We can obtain the estimated parameters $\widehat{\phi}_1 = -0.63703$, $\widehat{\phi}_2 = 0.33314$, $\widehat{\theta}_1 = -0.93656$, and $\sigma_a^2 = 0.00172$. These parameters satisfy the stationarity condition of the time series $\{X_i\}$, namely, $\widehat{\phi}_1 + \widehat{\phi}_2 < 1$, $\widehat{\phi}_2 - \widehat{\phi}_1 < 1$, and $|\widehat{\phi}_2| < 1$. Hence, we can get the ARMA fitted model which is shown in

$$X_t + 0.63703 X_{t-1} - 0.33314 X_{t-2} = a_t + 0.93656 a_{t-1}. \tag{9}$$

According to the stationarity and invertibility conditions of the ARMA model, we also get

$$\begin{gathered} x_t = \sum_{i=0}^{\infty} G_i \varepsilon_{t-i}, \\ \varepsilon_t = \sum_{j=0}^{\infty} I_j x_{t-j}, \end{gathered} \tag{10}$$

where $\{G_i\}$ denotes the Green function values and $\{I_j\}$ denotes the values of the inverse functions

$$x_t = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} G_i I_j x_{t-i-j}. \tag{11}$$

We refer to x_{t+l}, for all $l \geq 1$, as the l-step ahead forecast of $\{X_i\}$. x_{t+l} can be expressed as a linear function of the past p traffic values, which is shown in

$$\widehat{x}_t(l) = \sum_{i=0}^{\infty} D_i x_{t-i}, \tag{12}$$

where D_i is the coefficient matrix of the past traffic values and $\widehat{x}_t(l)$ denotes the l-step ahead forecast of the time series $\{x_t\}$.

The forecasting error of the time series $\{x_t\}$ is shown in

$$e_t(l) = x_{t+l} - \widehat{x}_t(l). \tag{13}$$

The minimum variance of the forecasting error above is denoted by

$$\mathrm{Var}_{\widehat{x}_t(l)}\left[e_t(l)\right] = \min\left\{\mathrm{Var}\left[e_t(l)\right]\right\}. \tag{14}$$

Hence, we also get the explicit expressions of $\widehat{x}_t(l)$ and $e_t(l)$, which are shown in

$$\begin{gathered} \widehat{x}_t(l) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} G_{l+i} I_j x_{t-i-j}, \quad \forall\, l \geq 1, \\ e_t(l) = \sum_{i=0}^{l-1} \sum_{j=0}^{\infty} G_i I_j x_{t+l-i-j}. \end{gathered} \tag{15}$$

Theoretical and experimental results show that multistep prediction may bring about much greater forecast error and complexity [13], and hence we only give 1-step ahead forecast model for the resource-constrained wireless multimedia sensor networks.

For the ARMA $(2, 1)$ model, the 1-step ahead forecast model and its associated forecast error are shown in

$$\widehat{x}_t(1) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} G_{1+i} I_j x_{t-i-j}, \qquad e_t(1) = \sum_{j=0}^{\infty} G_0 I_j x_{t+1-j}. \tag{16}$$

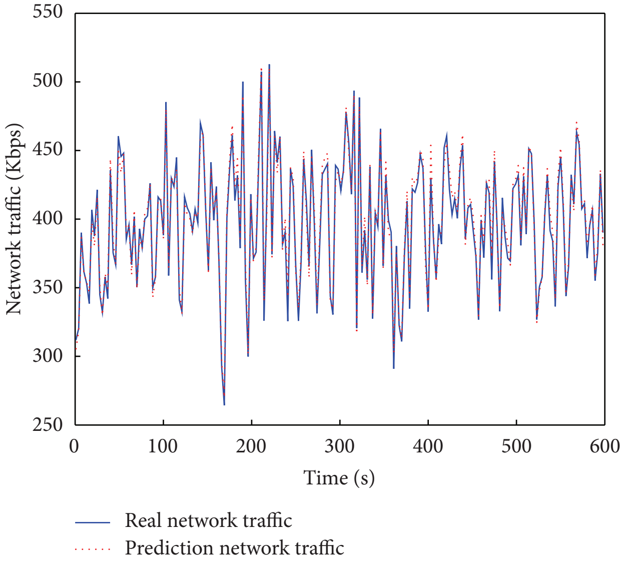


Figure 4: 1-step ahead forecast of X_i.

The model is implemented on Matlab 7.0. The comparison between the real network traffic and the prediction network traffic is shown in Figure 4. The results show that the model can accurately predict the WMSNs traffic. Furthermore, the model has some benefits, such as linear computing and low complexity.

4.2. Traffic Prediction-Based Fast Rerouting Strategy. Firstly, set a traffic threshold value denoted by Max based on the processing capability of a sensor node. Denote the network traffic at time i by X_i. We refer to $\widehat{X}_{t+1}$ as the 1-step ahead forecast of $\{X_i\}$ at the forecast origin t, and we refer to $P_t(1)$ as the probability that $\widehat{X}_{t+1}$ is greater than Max:

$$P_t(1) = P\left(\widehat{X}_{t+1} > \text{Max} \mid X_t, X_{t-1}, X_{t-2}, \ldots, X_{t-i}\right). \quad (17)$$

The probability distribution of X_i is subjected to the probability distribution of a_i. Here a_i obeys the normal distribution; therefore X_i also obeys the normal distribution. According to the statistical analysis of the collected multimedia sensor traffic, we can find that X_i obeys the $N(\mu_X, \sigma_X^2)$ distribution, where μ_X and σ_X denote the mean and variance of the time series $\{X_i\}$. Then we can obtain

$$\begin{aligned} P\left(\widehat{X}_{t+1} \le \text{Max}\right) &= P\left(\frac{\widehat{X}_{t+1} - \mu_X}{\sigma_X} \le \frac{\text{Max} - \mu_X}{\sigma_X}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\text{Max} - \mu_X)/\sigma_X} e^{t^2/2} dt, \\ P_t(1) &= 1 - P\left(\widehat{X}_{t+1} \le \text{Max}\right) \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\text{Max} - \mu_X)/\sigma_X} e^{t^2/2} dt. \end{aligned} \quad (18)$$

As a consequence, we can obtain a traffic prediction-based fast rerouting strategy which is shown in Theorem 1.

Theorem 1. *A sufficient condition for the adaptive path switching is that the probability of the $\widehat{X}_{t+1}$ value greater than the preset traffic threshold* Max *is equal to* $1 - (1/\sqrt{2\pi}) \int_{-\infty}^{(\text{Max} - \mu_X)/\sigma_X} e^{t^2/2} dt$.

Let us illustrate the results of the theorem using an example. The graphical depiction of the example is shown in Figure 5.

In the Figure, route 1 is a primary path from a source node to the base station. Both route 2 and route 3 are alternate paths for route 1.

Case I is that we do not use traffic prediction-based fast rerouting strategy. When node B is unable to process packets from other nodes, it takes the initiative to discard the packets. However, the node A continues to transmit the rest data packets until it finds the failure of node B. Node A will send the invalid message of node B to the source node. After that, the source node will forward the rest traffic through the alternate route 2 or route 3.

Case II is that we use traffic prediction-based fast rerouting strategy. When node B discovers that it satisfies the sufficient condition of Theorem 1, it will forward the urgent message to the source node via multihop wireless links at once. When the source node receives the urgent message, the efficient retransmission behavior is triggered. Obviously, the forwarding packets may bypass the fault area in advance and are smoothly rerouted through the alternate path 2, which can greatly improve the reliability of the data transmission and reduce the transmission delay. In addition, the other advantage of the fast rerouting strategy is that the data retransmission times, the energy consumption, and the routing control overhead may also be greatly reduced.

5. Theoretical Analysis and Performance Evaluation

5.1. Theoretical Analysis

5.1.1. Performance Analysis. Retransmission is one of the greatest impact factors on network performance due to various failures, such as network congestion, coverage hole, and routing hole. Some backgrounds and the symbol definitions are introduced as follows.

We firstly introduce the first-order radio model adopted in [14]. By using this approach, an energy loss of d^2 due to channel transmission is assumed. The energy $E_{Tx}(k, d)$ that a node dissipates for the radio transmission of a message of k bits over a distance d is due to running both the transmitter circuitry and the transmitter amplifier and is given by

$$E_{Tx}(k, d) = E_{\text{elec}} * k + \varepsilon_{\text{amp}} * k * d^2, \quad (19)$$

where E_{elec} is the transmitter circuitry dissipation per bit, which is supposed to be equal to the corresponding receiver circuitry dissipation per bit, and ε_{amp} is the transmit amplifier dissipation per bit per square meter.

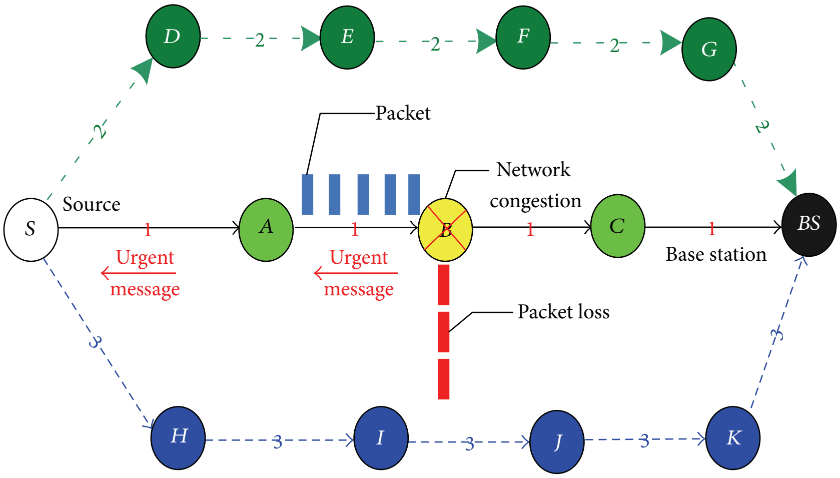


FIGURE 5: Graphical depiction of fast rerouting.

Similarly, the energy $E_{Rx}(k)$ dissipated by a node for the reception of a k-bit message is due to running the receiver circuitry. It is given by

$$E_{Rx}(k) = E_{Rx-\text{elc}}(k) = E_{\text{elec}} * k. \tag{20}$$

Secondly, we refer to E_c as the communication energy consumption of a node, λ denotes the number of links from a source node to the sink, l is an average path length of a link, T_{ave} is the average transmission delay from a source node to the sink node, and ρ is the packet loss rate.

The probability of data successfully retransmitted at the first time is $(1-\rho)^{\lambda-1}$. Let α be equal to $(1-\rho)^{\lambda-1}$, and the probability of data successfully retransmitted at the second time is $(1-\alpha)\alpha$. Similarly, we can obtain that the probability of data successfully retransmitted at the third time is $(1-\alpha)^2\alpha$. The average retransmission time T is given by

$$\begin{aligned} T &= \alpha + 2\alpha(1-\alpha) + 3\alpha(1-\alpha)^2 + \cdots \\ &= \left[\frac{\alpha}{1-\alpha}\right] \times \frac{(1-\alpha)}{[1-(1-\alpha)]^2} = \frac{1}{\alpha} = \frac{1}{(1-\rho)^{\lambda-1}}. \end{aligned} \tag{21}$$

We combine (19) and (20) with (21) to get

$$E_c = E_{Tx} + E_{Rx} = \frac{k}{(1-\rho)^{\lambda-1}}\left[2E_{\text{elec}} + \varepsilon_{\text{amp}} d^2\right]. \tag{22}$$

Obviously, E_c is proportional to $kd^2/(1-\rho)^{\lambda}$, in which it must be noted that the communication energy consumption E_c decreases proportional to every decrease in the ρ and λ values.

The average transmission delay from a source node to the sink node is given by

$$T_{\text{ave}} = \frac{DlT}{B_w}, \tag{23}$$

where B_w denotes the average available bandwidth of a sensor and D denotes the data to be transmitted.

Obviously, T_{ave} is proportional to lT/B_w, in which it must be noted that the T_{ave} value increases proportional to every increase in the l and T values, and it increases proportional to every decrease in B_w.

Our algorithm uses the fast rerouting strategy based on traffic prediction to bypass the fault area in advance and is smoothly rerouted through better alternate path. Compared with similar routing algorithms for WMSNs [5, 9], TPFR has lower T value, λ value, and ρ value and higher B_w value. These parameters play an important role in improving the network performance, such as reducing the transmission delay, network energy consumption, and prolonging the network lifetime.

5.1.2. Control Overhead Analysis. We refer to M_d as the size of a datagram, M_i denotes the size of a control message, and M_d is greater than M_i. Then we let m denote the number of the retransmitted datagrams and $T_{\max}$ ($T_{\max} < T$) denote the maximum number of retransmissions. Additionally, ω denotes the path length from a source node to a failure node and E_ε denotes the energy consumption of a processor on executing the traffic prediction algorithm.

In this paper, the routing algorithms without the rerouting mechanism are named non-TPFR.

For the non-TPFR algorithms, from the failure to the fault recovery, the data packets D_1 and the energy consumption E_1 generated by the algorithms are given by the following, respectively:

$$D_1 = mM_dT_{\max} + \omega M_i, \tag{24}$$

$$E_1 = \left(E_{\text{elec}} + \varepsilon_{\text{amp}} * d^2\right) \times \left[mM_dT_{\max} + \omega M_i\right] + E_{\text{elec}}\omega M_i. \tag{25}$$

For our proposed algorithm, from the failure to the fault recovery, the data packets D_2 and the energy consumption E_2 generated by TPFR are given by the following, respectively:

$$D_2 = \omega M_i, \tag{26}$$

$$E_2 = \left(2E_{\text{elec}} + \varepsilon_{\text{amp}} \times d^2\right) \times \omega M_i + E_\varepsilon. \tag{27}$$

Table 2: Simulation environment and used parameters.

Simulation parameter	Value
Network size	400 × 400 m^2
Node number	400
Link layer	LL
MAC layer	IEEE802.11
IFQ type	Queue/DropTail/PriQueue
IFQ length	10
Antenna type	Antenna/OmniAntenna
Physical type	Phy/WirelessPhy
Channel type	Channel/WirelessChannel
Energy model	EnergyModel
Initial energy of a node	0.5 J
Bandwidth	2 Mbit/s
Traffic threshold	1.7 Mbit/s

Obviously, from the failure to the fault recovery in networks, D_1 generated by the non-TPFR algorithms is greater than D_2 generated by our algorithm. In addition, the energy consumption of the processing module and the sensing module is far less than that of the wireless communication module. For example, the energy consumption of the 1 bit information transmitted 100 meters is roughly equivalent to that of the execution of the 3000 computer instructions. Thus, E_1 is greater than E_2.

In summary, from the failure to the fault recovery, the efficiency of our algorithm with more local computations is better than the non-TPFR algorithms with more communications and retransmissions. Moreover, our algorithm has very strong practicality, and it has important implications for improving the survivability of WMSNs.

5.2. Performance Evaluation

5.2.1. Simulation Parameters Settings. In this part, we simulate our proposal using NS-2 version 2.29 which is a discrete event network simulator for over 100 experiments with various random topologies. The network size is 400 m × 400 m deployed with 400 nodes for duration of 1200 time rounds. The traffic is CBR of 600 packet/second, and the packet size is 316 bytes. The video traces come from MDC Foreman video test sequences [15] provided by a study group for the video tracking in Arizona State University. In the current video traces, there are 300 frames, and the frame rate is 30 frames/s, corresponding to a frame period equivalent to 36 ms. Additionally, we assume that the frame period is equal to the size of a transmission window. We adopt IEEE802.11 for the MAC layer as shown in Table 2 which lists the parameters we used in our simulation.

In the simulations, we focus on measuring the performance metrics after the network has set up to include the average end-to-end delay, the average packet delivery ratio, the peak signal to noise ratio, the energy consumption, the remaining alive nodes, and the communication overhead. To prove the effectiveness of TPFR, we have also implemented the AntSensNet algorithm (ant-based multi-QoS routing) [5] and PEMuR (power efficient multimedia routing) [9], and we compared the simulation results.

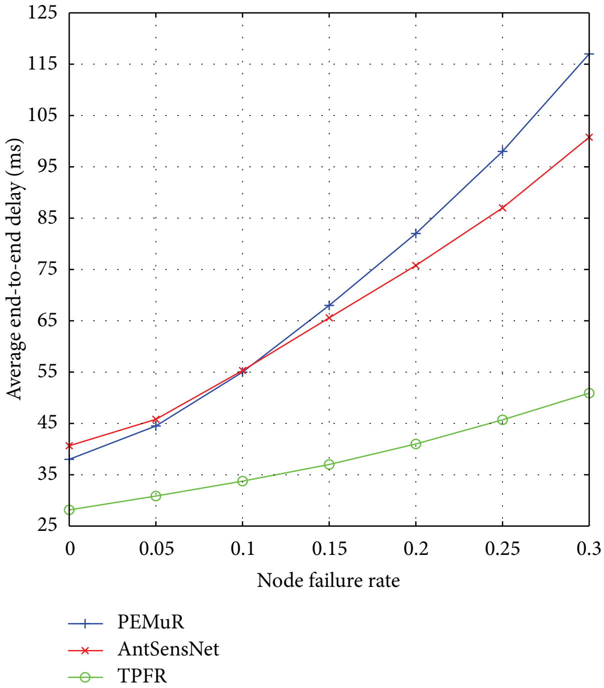


Figure 6: End-to-end delay performance.

5.2.2. Simulation Results Evaluation. Figure 6 shows the end-to-end delay, which is one of the important QoS parameters as the real-time multimedia packets have strict playout deadlines. We compare the average end-to-end delay of our algorithm combined with the AntSensNet routing discovery technique with the other routing protocols (PEMuR and AntSensNet). As shown in the figure, the TPFR design methodology outperforms the two classical multimedia routing protocols.

It is shown clearly that with the increase of node failure rate, our fast rerouting design has the minimum end-to-end delay and outperforms the other routing protocols because it depends on selecting a better alternate path in terms of the bandwidth, the minimum hop count, and the remaining energy before a node failure through the proposed traffic prediction mechanism. It is worth to note that PEMuR and AntSensNet only perform well at low node failure rate or link breaks, but with higher node failures and link breaks, the end-to-end delay increases exponentially due to the various failures of cluster heads which cause lost packets retransmission frequently. More importantly, the two protocols lack the rerouting mechanism.

The packet delivery ratio involves the ratio of successfully delivered data packets to the total data packets sent from the source to their destination. The average packet delivery ratio (PDR) is shown in Figure 7 where our rerouting algorithm outperforms the other algorithms, which confirms the previous theoretical analysis. We obtain this result due to the use of the traffic prediction technology that bypasses various failure

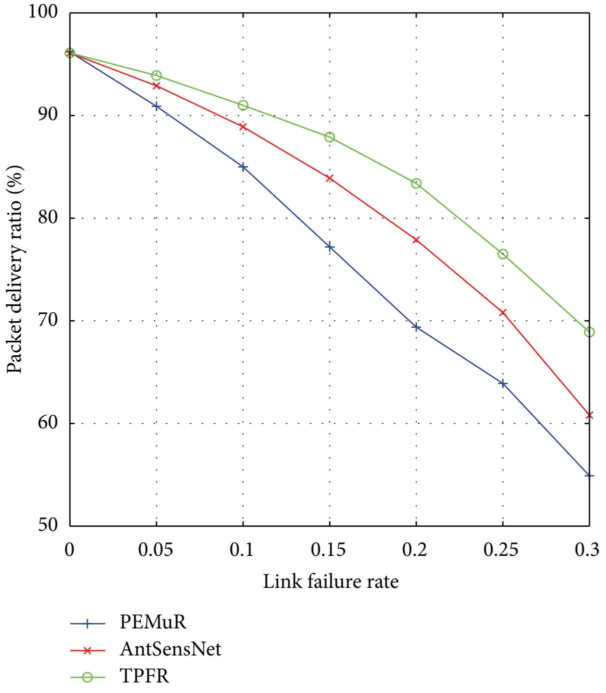


Figure 7: Packet delivery ratio.

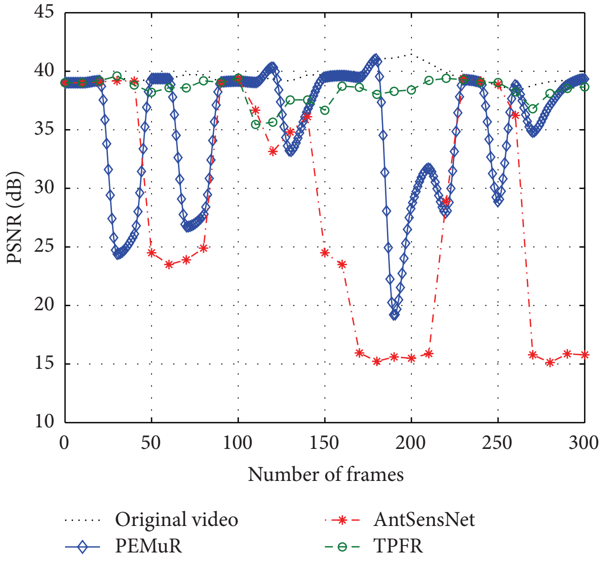


Figure 8: Received video quality of Foreman video.

areas such as network congestion, any node failure or link break, besides the selection of paths with better link quality based on the bandwidth and the remaining energy. Thus the number of lost packets significantly decreases. Such results were expected, and this investigation confirms the authors' hypotheses.

Figure 8 shows the average PSNR of the Foreman video when a node failure rate ranges from 0 to 0.3. We can see that the perceived video quality (PSNR) was higher for the simulations using TPFR when compared to the other protocols under the nonuniform node distribution. And the simulation curve of TPFR is consistent with the original video sequence. This is because the protocols PEMuR and AntSensNet are not able to efficiently handle the retransmission of video streams when node failures or link breaks occur. They are only specialized in minimizing the video distortion under an errorless transmission environment.

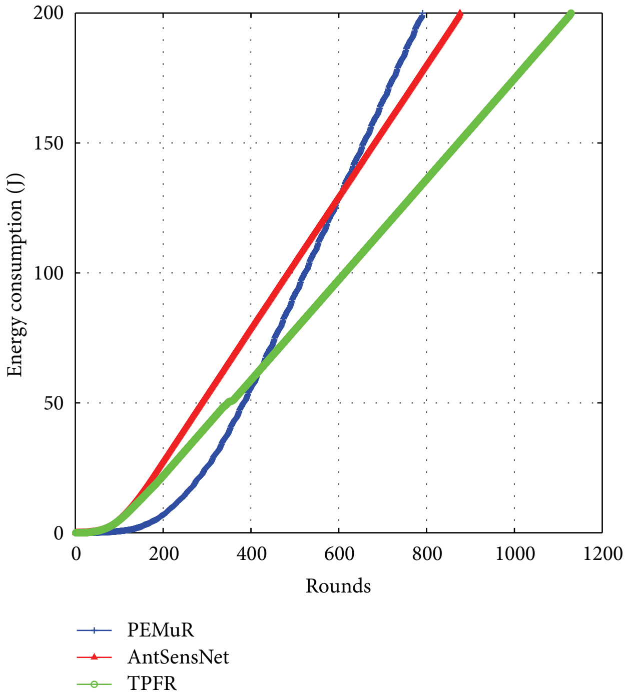


Figure 9: Energy consumption of network.

With respect to the average energy consumption, our proposed algorithm has less energy consumption than the AntSensNet algorithm as shown in Figure 9 with different time rounds because of the many benefits that they get from the traffic prediction-based fast rerouting. However, the PEMuR algorithm has better energy efficiency before 400 time rounds because both AntSensNet and TPFR algorithms lack sufficient information to find appropriate routes during this period. After this period, when the algorithms converge and the ants have gathered enough node and route information, the quality of routes discovered for our algorithm is superior to that found by PEMuR. In a word, with increasing time and failures, we notice that PEMuR and AntSensNet algorithms suffer from packet collisions and interferences and consume more energy for retransmitting lost packets, while TPFR exploits the benefits from the adaptive fast rerouting scheduling to prevent such problems and hence has less energy consumption.

The depletion of nodes over time is a typical metric of the energy efficiency of a routing protocol. Figure 10 shows the number of alive nodes in networks has changed over time, and the TPFR protocol is significantly better than the other routing protocols in retarding the time of node depletion. For the PEMuR protocol, the first node depletion time is at 311 rounds and the last node depletion time is at 791 rounds. For the AntSensNet protocol, the first node depletion time

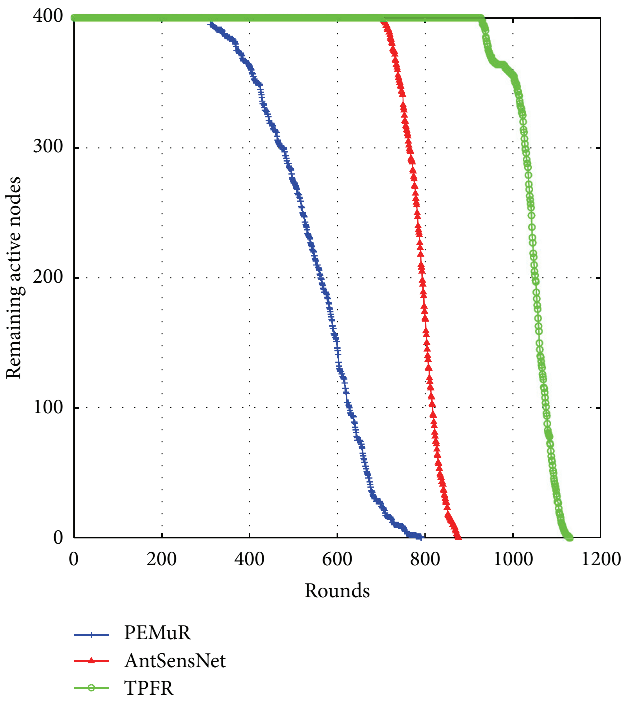


Figure 10: Remaining alive nodes in network.

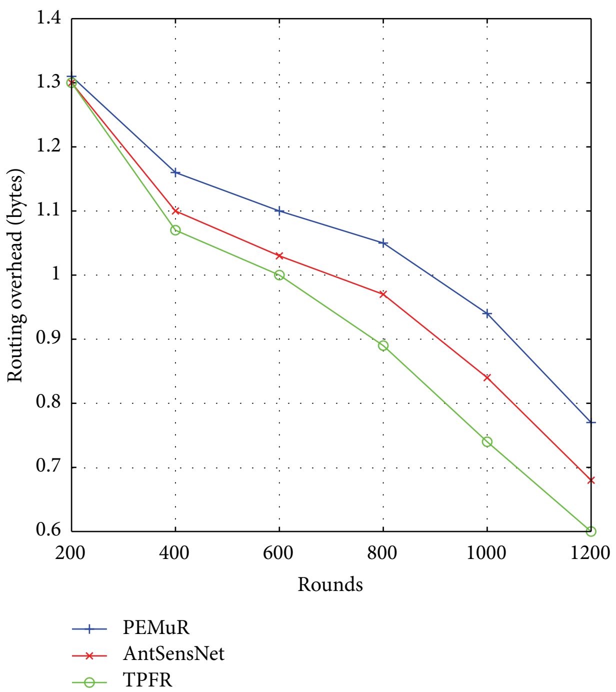


Figure 11: Routing overhead.

is at 702 rounds and the last node depletion time is at 876 rounds. For our proposed scheme, the first node depletion time is at 927 rounds and the last node depletion time is at 1129 rounds. The communication module consumes more energy than other modules in a wireless multimedia sensor. In our scheme, we use more traffic prediction computations instead of communications. Hence, our protocol has lower communicational energy consumption and can prolong the network lifetime.

The extra control packets are required in order to periodically monitor and maintain path conditions. And the routing overhead is shown in Figure 11. With increasing time, the mean routing overhead is reduced for the three algorithms; however, TPFR has a lower reduction of routing overhead than other algorithms. Due to such periodic updates, they constantly require a certain amount of routing overhead. The overhead of PEMuR can be reduced by piggybacking the control information on data packets if there is traffic between a sink and cluster heads. And that of AntSensNet can be reduced by embedding data into forward ants (a specimen of data ants) and piggybacking the pheromone information on data packets. In fact, TPFR is an improved AntSensNet scheme. TPFR uses computational overhead instead of communicational overhead, and hence it has lower routing overhead. Additionally, the simulation result remains consistent in the theoretical analysis of Section 5.1.2.

6. Conclusions

This paper presented TPFR, a novel fast rerouting algorithm over WMSNs, which aims at both energy savings and high QoS. The innovation of our proposed algorithm lies in the combined use of ant-based hierarchical routing protocol using multiple QoS metrics along with a traffic prediction-based fast rerouting algorithm. The adopted rerouting algorithm not only proposes an energy efficient rerouting policy, but also manages the network load according to the energy residues of the nodes and prevents useless data retransmissions through the proposed use of the intelligent rerouting algorithm. In this way, an outstanding level of energy efficiency and high QoS under the node failures or link breaks network environments is achieved.

Extended simulation tests performed showed that the utilization of TPFR enables the considerable retardation of the energy depletion of the video nodes. The enhancement in energy performance metrics provided by TPFR becomes even greater in the case of a nonuniform node energy distribution. Additionally, it was shown that TPFR succeeds in maintaining high levels of the average end-to-end delay, the packet delivery ratio (PDR), the perceived video quality (PSNR), and routing overhead for a nonuniform energy distribution. These advantages of TPFR enhance the belief that this scheme is indeed capable of achieving efficient multimedia stream communication in real-life applications.

The authors of this paper have already started to study this research work under the network invasion. We plan to apply the intrusion tolerance approach to solve a new challenging problem. According to this approach, even if the network is under DDoS attack, WMSNs is still able to provide available quality of service.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (61202474, 61201160, and

61272074) and by the Senior Professional Scientific Research Foundation of Jiangsu University (12JDG049).

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.

[2] I. T. Almalkawi, M. G. Zapata, J. N. Al-Karaki, and J. Morillo-Pozo, "Wireless multimedia sensor networks: current trends and future directions," *Sensors*, vol. 10, no. 7, pp. 6662–6717, 2010.

[3] S. Ehsan and B. Hamdaoui, "A survey on energy-efficient routing techniques with QoS assurances for wireless multimedia sensor networks," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 2, pp. 265–278, 2012.

[4] X. Fan, W. Shaw, and I. Lee, "Layered clustering for solar powered wireless visual sensor networks," in *Proceedings of the 9th IEEE International Symposium on Multimedia (ISM '07)*, pp. 237–244, Taichung, Taiwan, December 2007.

[5] L. Cobo, A. Quintero, and S. Pierre, "Ant-based routing for wireless multimedia sensor networks using multiple QoS metrics," *Computer Networks*, vol. 54, no. 17, pp. 2991–3010, 2010.

[6] E. Felemban, C. G. Lee, and E. Ekici, "MMSPEED: multipath Multi-SPEED Protocol for QoS guarantee of reliability and timeliness in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 6, pp. 738–753, 2006.

[7] L. Shu, Y. Zhang, and L. T. Yang, "TPGF: geographic routing in wireless multimedia sensor networks," *Telecommunication Systems*, vol. 44, no. 1-2, pp. 79–95, 2010.

[8] L. Zhang, M. Hauswirth, and L. Shu, "Multi-priority multi-path selection for video streaming in wireless multimedia sensor networks," in *Ubiquitous Intelligence and Computing*, vol. 5061 of *Lecture Notes in Computer Science*, pp. 439–452, 2010.

[9] D. Kandris, M. Tsagkaropoulos, I. Politis, A. Tzes, and S. Kotsopoulos, "Energy efficient and perceived QoS aware video routing over Wireless Multimedia Sensor Networks," *Ad Hoc Networks*, vol. 9, no. 4, pp. 591–607, 2011.

[10] Z. Y. Li and R. C. Wang, "Secure coverage-preserving node scheduling scheme using energy prediction for wireless sensor networks," *The Journal of China Universities of Posts and Telecommunications*, vol. 17, no. 5, pp. 100–108, 2010.

[11] X. Cao, R. C. Wang, and H. P. Huang, "Multi-path routing algorithm for video stream in wireless multimedia sensor networks," *Journal of Software*, vol. 23, no. 1, pp. 108–121, 2012.

[12] S. A. M. Östring and H. Sirisena, "The influence of long-range dependence on traffic prediction," in *Proceedings of International Conference on Communications (ICC '01)*, pp. 1000–1005, June 2000.

[13] R. Tsay, *Analysis of Financial Time Series-3rd Edition*, John Wiley & Sons, Hoboken, NJ, USA, 2010.

[14] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Siences (HICSS '00)*, pp. 1–10, Maui, Hawaii, USA, January 2000.

[15] "Video trace library," 2010, http://trace.eas.asu.edu/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 247931, 9 pages
http://dx.doi.org/10.1155/2013/247931

Research Article

Adaptive Multihypothesis Prediction Algorithm for Distributed Compressive Video Sensing

Jinxiu Zhu,[1,2] Ning Cao,[3] and Yu Meng[1,2]

[1] *College of Computer and Information, Hohai University, Changzhou 213022, China*
[2] *Changzhou Key Laboratory of Sensor Networks and Environmental Perception, Changzhou 213022, China*
[3] *College of Computer and Information, Hohai University, Nanjing 210098, China*

Correspondence should be addressed to Jinxiu Zhu; zhujinxiu1972@163.com

Received 20 February 2013; Accepted 23 April 2013

Academic Editor: Liangmin Wang

A novel adaptive multihypothesis (MH) prediction algorithm for distributed compressive video sensing (DCVS) is proposed in this paper. In the proposed framework, consistent block-based random measurement for each video frame is adopted at the encoder independently. Meanwhile, a mode decision algorithm is applied in CS-blocks via block-based correlation measurements at the decoder. The inter-frame MH mode is selected for the current block wherein the interframe correlation coefficient value exceeds a predetermined threshold. Otherwise, the intraframe MH mode is worthwhile to be selected. Moreover, the adaptive search window and cross-diamond search algorithms on measurement domain are also incorporated to form the dictionary for MH prediction. Both the temporal and spatial correlations in video signals are exploited to enhance CS recovery to satisfy the best linear combination of hypotheses. The simulation results show that the proposed framework can provide better reconstruction quality than the framework using original MH prediction algorithm, and for sequences with slow motion and relatively simple scene composition, the proposed method shows significant performance gains at low measurement subrate.

1. Introduction

In the wireless video sensor network, wireless video cameras are widely used to timely operate and store the video data which is waiting for the later upload to a fixed network base station. However, in such a network, the resources, for example, the computing and the storage capacity, are more limited at the encoder than the decoder. To address this issue, an advanced video coding scheme, referred to distributed video coding (DVC) [1], is presented to satisfy the low-complexity capability at the encoder. Its main ideal is founded on the principle of distributed source coding (DSC) with a number of separate encoders and one joint decoder. However, it still suffers from the disaster where huge amounts of raw image data are captured at the decoder with the memory-intensive consumption, even for still image. Fortunately, with the development of the single-pixel camera architectures [2], the underlying compressed sensing (CS) theory seems to be a key approach to mitigate such a problem [3]. As an effective data compression method, CS enables to directly achieve the compressed data via a random projection on the raw image data. Recently, distributed compressive video sensing (DCVS) [4] is proposed to flexibly utilize the CS in the framework of the DVC. The compressed video data is firstly obtained at a low complexity encoder. Then the CS reconstruction is performed with the correlation exploitation among successive frames at a high-complexity decoder [5–8].

Motivated by recent progress in the distributed video coding, we propose a novel distributed compressive video sensing framework, wherein several stopping criteria are used to improve and speed up the multihypothesis prediction algorithm for the CS frame reconstruction by the dual-discrete wavelet transform (DDWT) [9] basis. The remainder of the paper is organized as follows. In Section 2, the DCVS and MH prediction are reviewed for the block-based CS. In Section 3, the DVCS framework based on adaptive MH prediction is introduced. The experimental results are presented in Section 4. Finally, relevant conclusions and some remarks are given in Section 5.

2. Related Works

2.1. Distributed Compressed Video Sensing. To the best of our knowledge, the DCVS framework mainly can be divided into two categories according to the original video data reception at the encoder. At the encoder of DCVS, video frames are grouped into group of pictures (GOP) consisting of a key frame (also called K-frame) and a number of nonkey frames (also called CS-frames). In the scheme proposed in [4], the encoder obtains the original information, key frames are encoded using traditional MPEG/H.264 encoding, while for CS frames, measurements are taken via random measurement matrix. Side information is generated from the neighboring reconstructed K-frames. The disadvantage of this framework is that the complex MPEG/H.264 encoding is still required. On the other hand, in [10] where the encoder gets the measurements directly but not the original information, their approach is different from the scheme proposed in [4] in which CS measurements are applied to both key and nonkey frames. K-frames are reconstructed using GPSR [11] while stopping criteria based on side information generated from the K-frames are used during the reconstruction process of the CS-frames.

However, there still exit a few disadvantages in the aforementioned methods. For example, the generation rule of side information (SI) is usually simple due to releasing the computation burden of the coders. The CS reconstruction process also cannot perform effectively with the rough prediction. As a result, the performance of the DCVS cannot achieve the best. Thereby, we focus on a novel framework where several criterions are proposed and also use adaptive search window to improve the traditional MH prediction algorithm for the better CS-frame reconstruction.

2.2. Multihypothesis Prediction for Block-Based CS of Video. To alleviate the huge computation and memory burden for both the sensing and reconstructing processes, images are usually decomposed into small blocks for the further process. An approach for block-based CS (BCS) for 2D images is proposed with the assumption of the independence among blocks in [12]. Thereby, from the perspective of the incoherence principle in CS theory, block-based measurements seem to be less efficient than frame-based measurements due to that the former sensing matrix is block diagonal and the latter one is dense. Nonetheless, by the lower request of incoherence, BCS can preserve local information that helps the decoder construct more accurate SI based on the interframe sparsity model with the sparsity-constraint block prediction. Thereby, the work in [13] explores the sparsity of small interframe difference to remove the temporal redundancy, though it is not suited for video sequences with large interframe difference and fast motion. Later, the CS reconstruction of video is presented with an alternative way [14]. An explicit prediction using motion estimation (ME) and motion compensation (MC) is utilized to overcome the above problem. Another ME/MC-based reconstruction is also proposed with the block-based measurement of a CS-frame to form a block-by-block MH motion-compensated prediction [15]. Recently, an improved strategy for incorporating MH prediction into the block-based compressed sensing with smooth projected Landweber (MH-BCS-SPL) reconstruction of video is proposed [16] which could get a more accurate prediction by finding a linear combination of all the blocks/hypotheses in the search window.

2.3. Our Contributions. Different from current MH prediction approaches which only focus on these modified MH regularization algorithms, our approach combines both MH prediction mode and dictionary acquisition in the aforementioned DCVS schemes. Our main contribution in this paper is as follows.

(i) A novel block mode decision at the decoder is performed for blocks in CS-frame. In particular, the inter- and intramodes are adaptive used based on the correlation of CS measurements in order to obtain better reconstruction quality.

(ii) Cross-diamond search algorithm on measurement domain is adopted to build the dictionary in Inter-MH prediction mode, which provides much sparser representation for the corresponding blocks.

(iii) A practical and real-time system is designed for the DCVS via the above adaptive MH prediction.

3. Adaptive Multihypothesis Prediction Algorithm of Our DCVS

3.1. Proposed DCVS Scheme. The proposed DVCS framework based on the adaptive MH prediction (AMH_DVCS) is described in this section. As illustrated in Figure 1, at the encoder, the frames of a video sequence are divided into two categories: K-frames and CS-frames. To simplify the encode framework, in both K-frames and CS-frames the consistent block-based random measurements are adopted. Sequentially, the measurements are transmitted to the decoder with their corresponding nearby integers. Hereby, the sparse basis matrix Ψ is used as the DDWT basis. Without loss of generality, the subrate (S_K) is set higher in K-frames than in CS-frames (S_{CS}), due to the K-frames usually reconstructed with more advanced quality. In the section of the decoder, a mode decision algorithm is formed based on the block correlation of CS measurements. The blocks in CS-frame (CS-blocks) are classified into two types, that is, inter-frame MH prediction modes (Inter_MH mode) and intra-frame MH prediction modes (Intra_MH mode). Then the bitstream of measurements is processed with the selected decoding mode. On one hand, the K-frame is reconstructed by intra-frame reconstruction. On the other hand, the SI frame is generated by the adaptive reconstruction for the CS frame. Thus, residual frames are sparse recovered with SI frames, and the further decoded CS-frames are produced. The more detailed process is given in Section 3.2.

3.2. Adaptive Multihypothesis Prediction for CS-Frame Reconstruction

3.2.1. Decoding Mode Decision Based on the Correlation of CS Measurements. In our work, the adaptive decoding mode

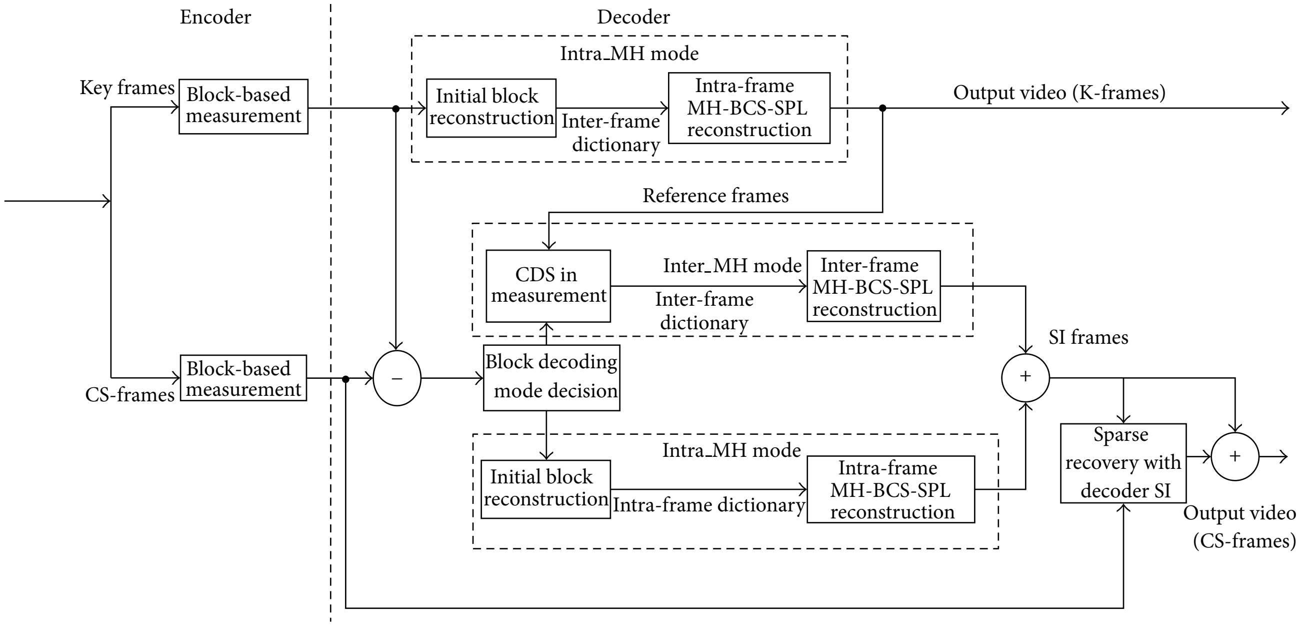


Figure 1: Proposed DCVS framework based on adaptive MH prediction.

is adopted to exploit the temporal and spatial correlation of video sequences. For convenience, some attributions of parameters in the DCVS are first discussed. Generally, CS measurements can be modeled as random Gaussian sources. And the dependence between two random variable quantities is indicated by Pearson's correlation coefficient [17]. The work in [18] explores that the frames in various video sequences have the high correlation among CS measurements with the corresponding value even above 0.9. Therefore, we define the correlation coefficient (CC) function of CS measurements as

$$r(y_1, y_2) = \frac{\sum_{i=1}^{N}[y_1(i) - \overline{y_1}] \cdot [y_2(i) - \overline{y_2}]}{\sqrt{\sum_{i=1}^{N}[y_1(i) - \overline{y_1}]^2}\sqrt{\sum_{i=1}^{N}[y_2(i) - \overline{y_2}]^2}}, \quad (1)$$

where y_1 and y_2 are the different measurement vectors of block, N is the length of a measurement vector. Then, the mode decision procedure is described as follows.

Step 1. Calculate the inter-frame CC function $r(y_B^{\mathrm{cs}}, y_B^K)$, where y_B^{cs} and y_B^K are the measurement vectors of current coding CS-block and nearby key frames, respectively.

Step 2. Given a predetermined threshold T_1, the Inter_MH mode is selected for the current block, if $r(y_B^{\mathrm{cs}}, y_B^K) > T_1$, and the Intra_MH mode is done otherwise. That is,

$$B\_\mathrm{mode} = \begin{cases} \text{Inter\_MH} & r\left(y_B^{\mathrm{cs}}, y_B^K\right) > T_1, \\ \text{Intra\_MH} & r\left(y_B^{\mathrm{cs}}, y_B^K\right) \le T_1. \end{cases} \quad (2)$$

3.2.2. Multihypothesis Prediction Algorithm. MH prediction has a major distinctive advantage wherein multiple predictions are used to yield a composite prediction, compared with the traditional simple single-hypothesis (SH) prediction. For this reason, various improved MH prediction methods are developed and widely incorporated in the recent video coding to enhance the video-coding quality. These methods impose specific structures on the hypotheses to form the more strict ultimate prediction with the target of the additional motion-vector rate constraint by multiple predictions of a block. Moreover, because the MH predictions are all performed at the decoder, without the corresponding rate burden, the more complex but better forms of MH prediction can be considered. The major work of ME/MC in residual reconstruction is to create an MH predictive block with the windows, whose distance is as close as possible to original block, in given reference frames. The optimal sparse coefficient $w_{t,i}$ can be cast as

$$w_{t,i} = \arg\min_{w} \|x_{t,i} - H_{t,i}w\|_2^2, \quad (3)$$

where $x_{t,i}$ is the ith block for the tth frame and $H_{t,i}$ is a matrix whose columns are the rasterizations of the possible blocks within the search space of the reference frames within the search window. And the MH predictive block $\tilde{x}_{t,i}$ is constructed as

$$\tilde{x}_{t,i} = H_{t,i}w_{t,i}. \quad (4)$$

However, in the encoder, we can only receive the measurements $y_{t,i} = \Phi x_{t,i}$. Thereby, in the CS framework, the estimated sparse coefficient $\widehat{w}_{t,i}$ can be obtained as

$$\widehat{w}_{t,i} = \arg\min_{w} \|\Phi(x_{t,i} - H_{t,i}w)\|_2^2 = \arg\min_{w} \|y_{t,i} - \Phi H_{t,i}w\|_2^2. \quad (5)$$

The most common method to solve the least-squares (LSQ) problem is Tikhonov regularization [19] which imposes a l_2 penalty on the norm of $\widehat{w}_{t,i}$ as

$$\widehat{w}_{t,i} = \arg\min_{w} \|y_{t,i} - \Phi H_{t,i}w\|_2^2 + \lambda^2\|\Gamma w\|_2^2, \quad (6)$$

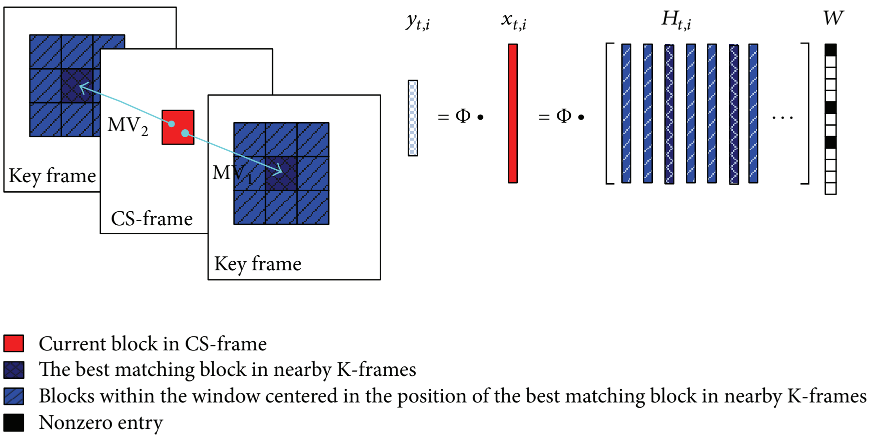

$y_{t,i}$
$x_{t,i}$
$H_{t,i}$
W
$= \Phi \bullet$
$= \Phi \bullet$
$\cdots$
Key frame
MV_2
CS-frame
MV_1
Key frame
🟥 Current block in CS-frame
■ The best matching block in nearby K-frames
▨ Blocks within the window centered in the position of the best matching block in nearby K-frames
■ Nonzero entry

Figure 2: The Inter_MH mode assumes a (vectorized) block in a CS-frame can be represented weightily by a linear combination of (vectorized) temporal blocks within the window centered in the position of the best matching block in preceding and following key frames.

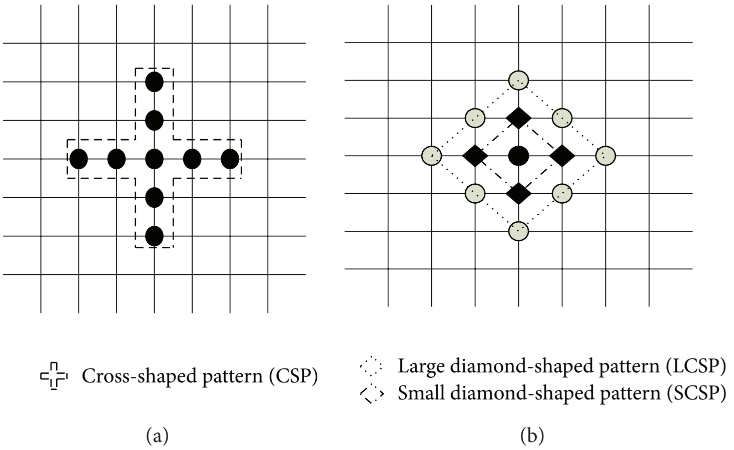

Cross-shaped pattern (CSP)
Large diamond-shaped pattern (LCSP)
Small diamond-shaped pattern (SCSP)
(a)
(b)

Figure 3: Search patterns used in the CDS algorithm. (a) CSP. (b) LDSP and SDSP.

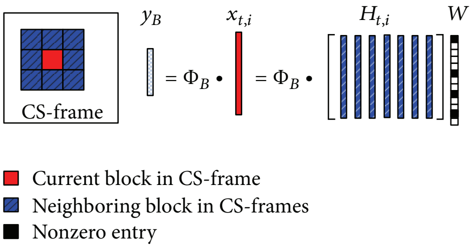


Figure 4: The Intra_MH mode assumes a (vectorized) block in a CS-frame can be represented weightily by a linear combination of (vectorized) spatial neighboring blocks within the selected window in the CS-frame itself.

where Γ is known as the Tikhonov matrix, and λ is a scale factor. For each block, $\widehat{w}_{t,i}$ can be calculated directly by the Tikhonov solution as

$$\widehat{w}_{t,i} = \left((\Phi H_{t,i})^T (\Phi H_{t,i}) + \lambda^2 \Gamma^T \Gamma\right)^{-1} (\Phi H_{t,i})^T y_{t,i}. \quad (7)$$

By taking (7) into (4), the prediction $\widetilde{x}_{t,i}$ can be finally achieved as

$$\widetilde{x}_{t,i} = H_{t,i} \widehat{w}_{t,i}. \quad (8)$$

Finally, all the predicted blocks $\widetilde{x}_{t,i}$ are put together providing the SI frame.

3.2.3. Inter_MH Mode Principles. The Inter_MH mode is shown in Figure 2. Block-based measurements, along with preceding and following key frames, are used to generate MH block prediction. As for Inter_MH mode, we use a dictionary of temporal blocks in the adjacent key frames as the matrix $H_{t,i}$. Different from the original MH-BCS-SPL reconstruction [16] with the fixed search window size, $W = 6$, and the window centered in the position of current CS-block, our proposed method uses the cross-diamond search algorithm (CDS) in [20] for the fast block motion estimation to calculate the maximum correlation coefficient (MCC) of CS measurements. In the CDS algorithm, an initial cross-shaped pattern (CSP), as shown in Figure 3(a), is proposed as the initial step, and then a large diamond-shaped pattern (LDSP) and small diamond-shaped pattern (SDSP) are

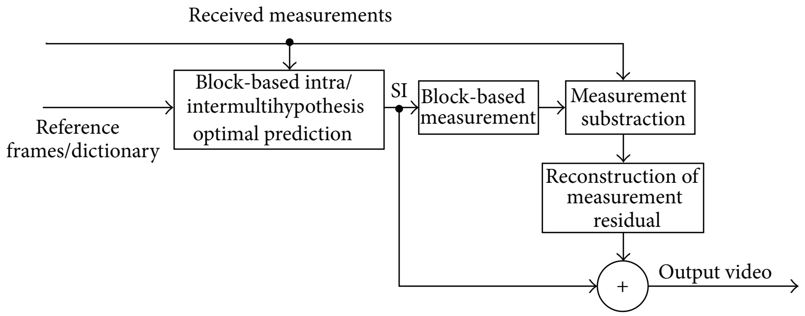
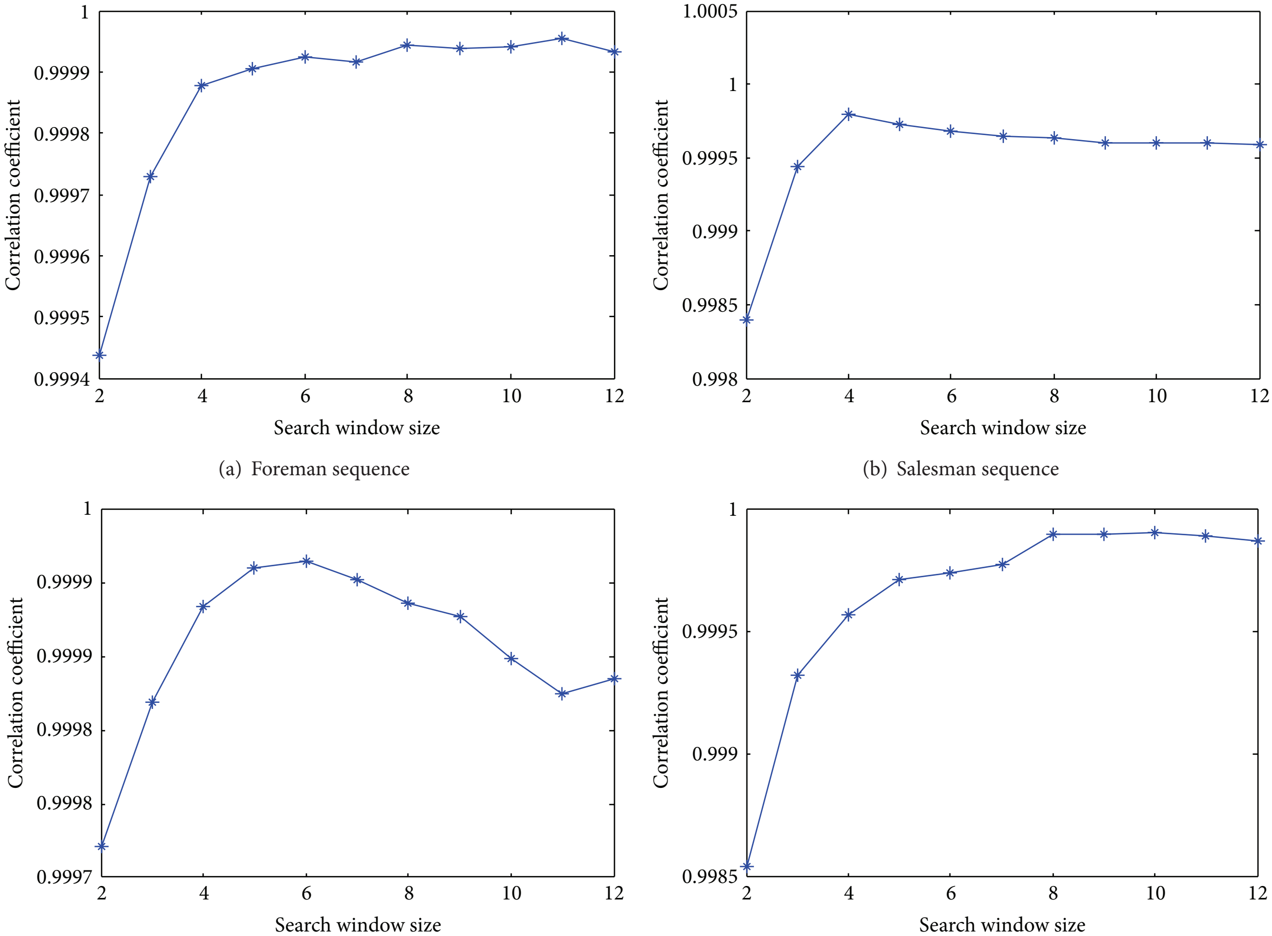

Received measurements
Reference frames/dictionary
Block-based intra/ intermultihypothesis optimal prediction
SI
Block-based measurement
Measurement substraction
Reconstruction of measurement residual
+
Output video

Figure 5: The process of sparse reconstruction with decoder SI.



(a) Foreman sequence

(b) Salesman sequence

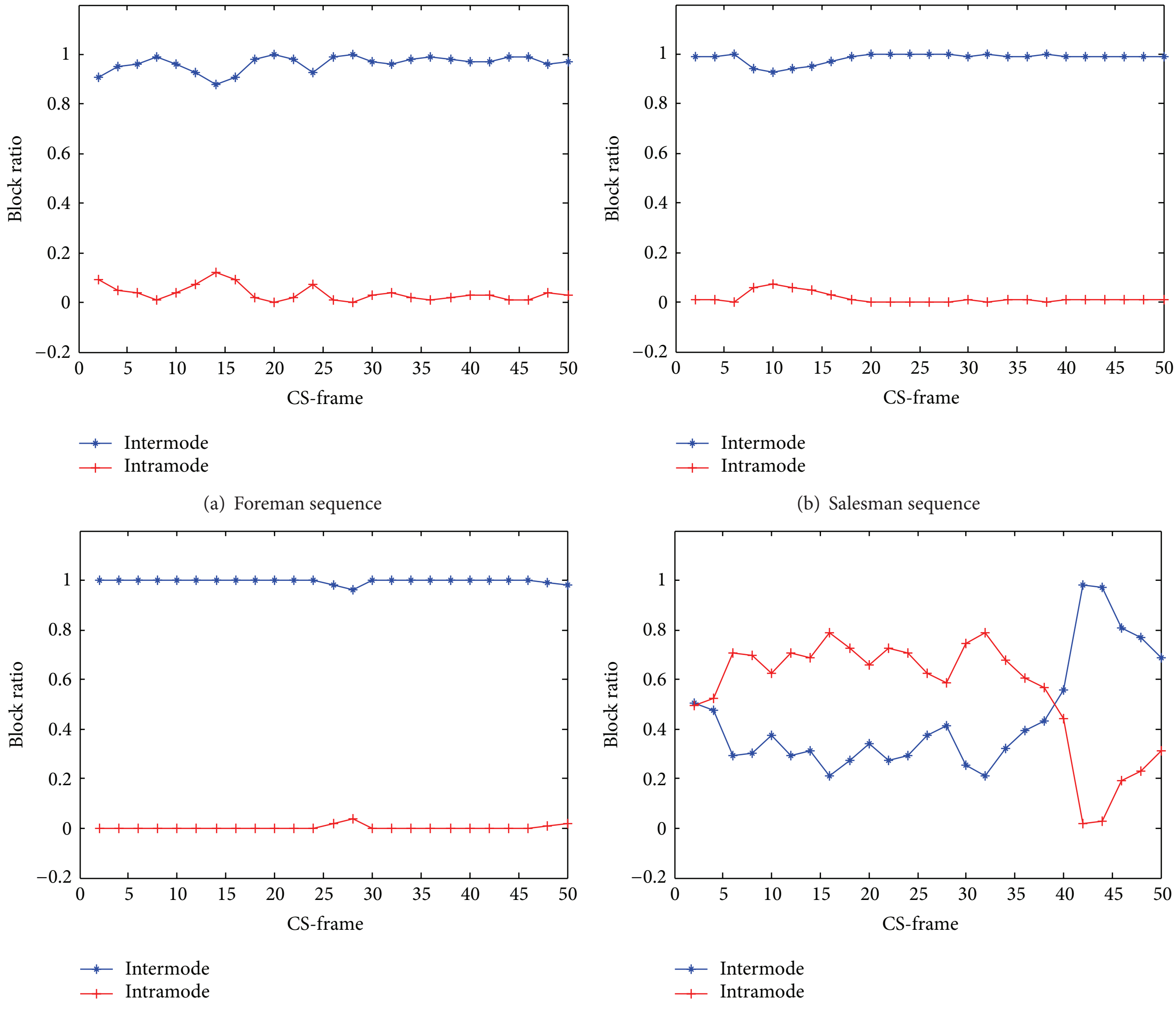
(c) Mother-daughter sequence

(d) Football sequence

Figure 6: Relationship between the search window size W and the correlation of original CS-frame and its SI predicted by MH-BCS-SPL algorithm on the 50th frame of four video sequences.

used, as depicted in Figure 3(b). The best matching block is obtained by computing MCC between the current CS block and the blocks within a search window while the window size increases from 2 to 14. The process of the window match is performed until the incremental MCC is not larger than a threshold, for example, 0.005. Then the block at the current search window is viewed as the best matching block. The search window is a square region which centered on the position of the best matching block. Because the best matching block has the maximum correlation with the current block in CS frame, the generated prediction is sufficiently accurate for small search window size, so we set the window size W to be 3 for lower computation.

In the Inter_MH mode, it assumes the atoms in the dictionary are composed by a set of linear combinations of temporal blocks which are also called MH predictions



(a) Foreman sequence

(b) Salesman sequence

(c) Mother-daughter sequence

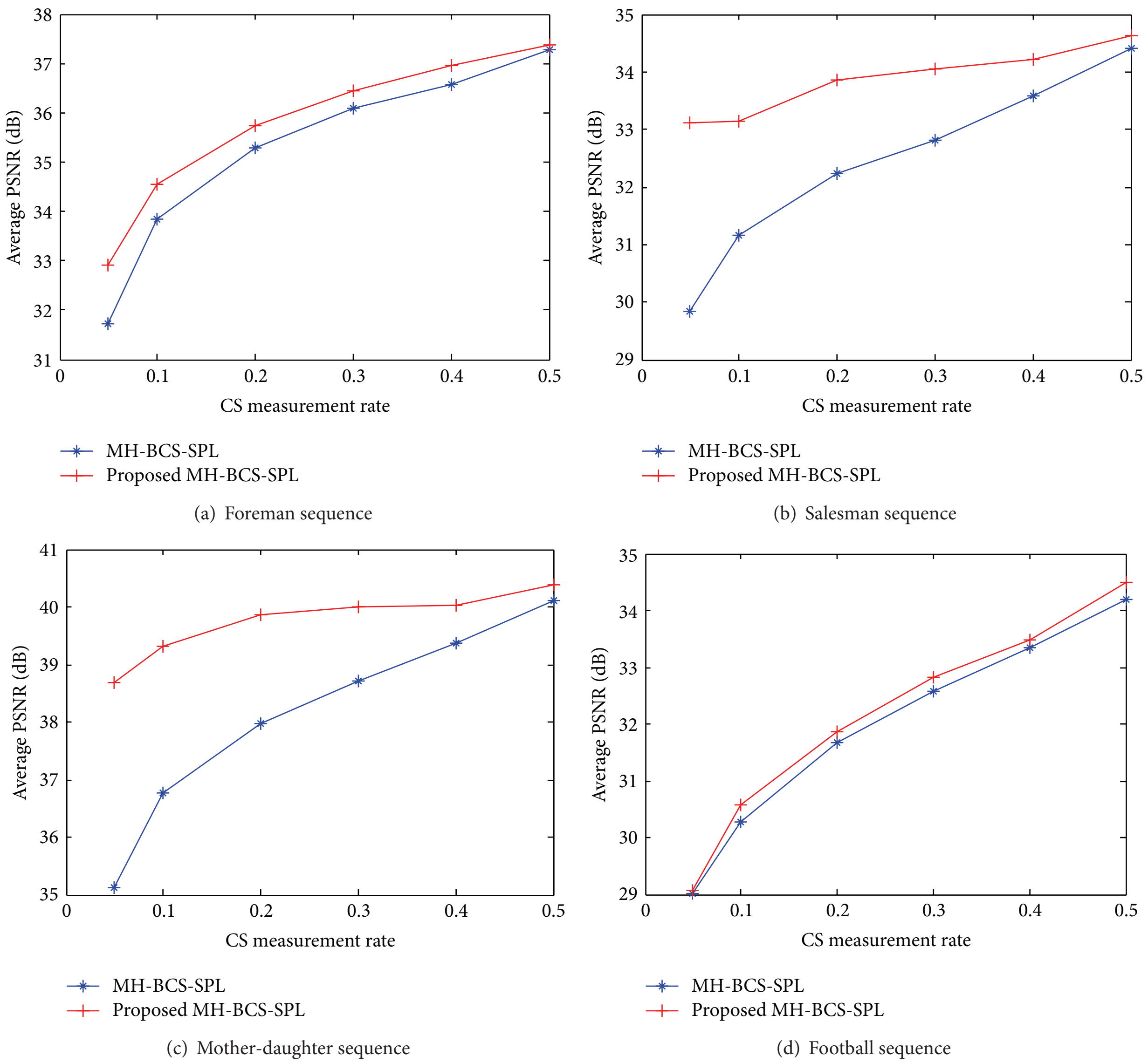
(d) Football sequence

Figure 7: Block ratio between Inter_MH mode and the Intra_MH mode of blocks for the first 50 frames of the four sequences.

within the window centered on the position of the best matching block in the adjacent key frames available at the decoder. Given its corresponding compressed measurements, Our method can predict the block optimally to generate the SI for a CS block by using (7) and (8).

3.2.4. Intra_MH Mode Principles. The Intra_MH mode is shown in Figure 4. Just as in the Inter_MH mode, the block match is also used. The only difference is that the dictionary is composed by spatial blocks in CS-frame. The best matching block is obtained by computing MCC between the current CS block and the blocks within a search window in CS-frame itself with the size W increasing from 2 to 11. If the difference of MCC between the current and next sizes is not larger than 0.00001, the current size is selected as the size of window where blocks are used as the atoms of the dictionary. Similarly, given its compressed measurements, this mode enables the block to be optimally predicted to generate the SI for a CS block.

3.2.5. Sparse Reconstruction with Decoder SI. With the aforementioned adaptive decoding modes for each CS block, all the recovered CS-blocks are put together which generates the SI frame. AMH_DVCS employs a very simple but effective algorithm to realize the sparse reconstruction with the SI frame; that is, subtract the measurement vector of the current frame from the measurement vector of the SI frame to form a new measurement vector of the prediction error. When the prediction is sufficiently precise, it can be faithfully recovered from its compressed measurements. The approximation of an input frame is then simply recovered by adding the prediction error to the SI frame. The process of sparse representation with decoder SI is summarized in Figure 5. The BCS-SPL algorithm is used in the algorithm of sparse representation with the decoder SI.



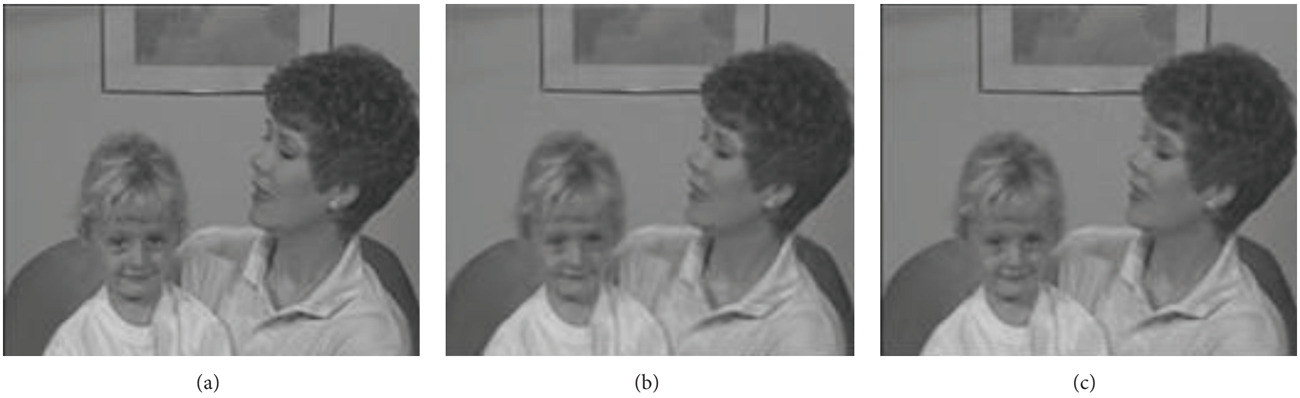
(a) Foreman sequence

(b) Salesman sequence

(c) Mother-daughter sequence

(d) Football sequence

Figure 8: Performance comparison between the proposed MH-BCS-SPL and the original BCS-SPL: average reconstruction quality of the first 50 frames of the four sequences.



(a)

(b)

(c)

Figure 9: Reconstruction frame 50 of the Mother-daughter sequence with $S_{CS} = 0.3$ and $S_K = 0.7$. (a) Original image. (b) Frame reconstructed using the original MH-BCS-SPL with PSNR = 38.1821 dB. (c) Frame reconstructed using the proposed MH-BCS-SPL with PSNR = 39.9817 dB.

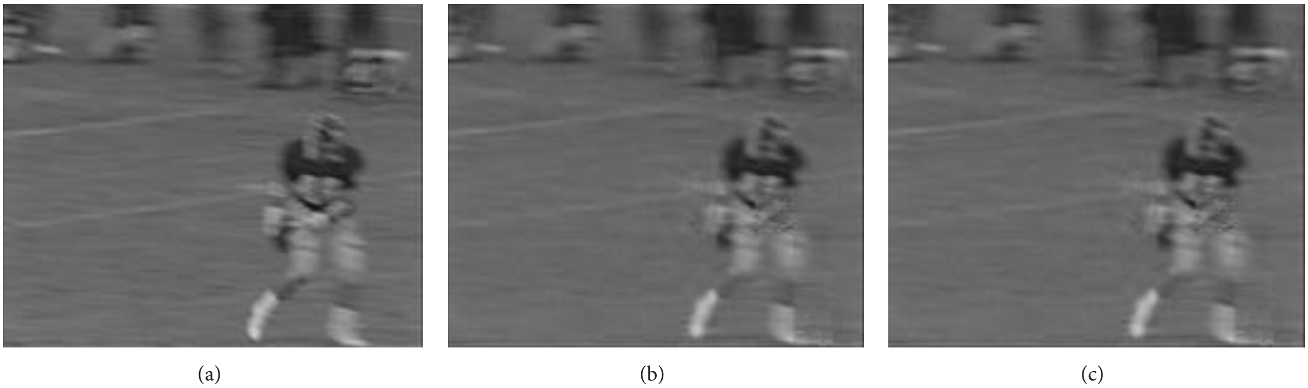


FIGURE 10: Reconstruction frame 50 of the Football sequence with $S_{CS} = 0.3$ and $S_K = 0.7$ (a) Original image. (b) Frame reconstructed using the original MH-BCS-SPL with PSNR = 31.3012 dB. (c) Frame reconstructed using the proposed MH-BCS-SPL with PSNR = 33.9063 dB.

4. Simulation Results

The performance of our proposed framework is test on four QCIF video sequences, that is, Foreman, Salesman, Mother-daughter, and Football, with GOP = 2 and the block size $B = 16$. Several subrates are also employed to evaluate the proposed method with the baseline of the original MH-BCS-SPL algorithm.

4.1. Results of the Correlation with Different Search Window Size. Figure 6 shows the relationship between the search window size W and the average correlation of the blocks between original CS-frame and its SI frame predicted by the original MH-BCS-SPL algorithm for 50th frame of four video sequences. From Figure 6, we can see that at first the correlation increases with the search window size growing, which means a greater W can provide a better reconstruction quality. But when the W exceeds a suitable size, the correlation increases slower or decreases as the W growing. We know that the lager the W is, the more complex the reconstruction is. Thus, we make use of the adaptive search window size in order to trade off between the reconstruction quality and complexity.

4.2. Results for the Decoding Mode Selection. As long as the decoding mode is selected, the reconstruction quality is relatively insensitive to T_1. From our simulations, $T_1 = 0.999$ achieves the maximum reconstruction quality and then is used in our paper. The ratio between the Inter-MH mode and the Intra-MH mode of blocks in the first 50 frames for four sequences is provided in Figure 7. Figure 7 shows that for the sequence with little motion, most blocks select the Inter-MH mode, while most blocks select Intra-MH mode for the sequence with fast or complex motion. In the Football sequence, the motion of the first 38 frames is intense while in the rest frames the motion is little and simple, so from Figure 7(d) we can see that the curve is intersecting in the 39th frame.

4.3. Performance Comparison with Original MH-BCS-SPL Algorithm. The average PSNR performances with different subrates for the four sequences using the proposed algorithm and the original MH-BCS-SPL algorithm are shown in Figure 8. The numerical values on the x-axis denote the subrates of the CS-frames with a fixed K-frame subrate $S_K = 0.7$ while those on the y-axis represent the average reconstruction quality (PSNR in dB) of CS-frames.

As can be seen in Figure 8, the proposed adaptive MH-BCS-SPL provides superior reconstruction quality over the original MH-BCS-SPL across the range of tested subrates. It also can be seen that for sequences with slow motion and relatively simple scene composition, such as the Mother-daughter sequence, the proposed method shows significant performance gains, while the gap between adaptive MH-BCS-SPL and MH-BCS-SPL narrows as the subrate increases. However, for the Football sequence with fast or complex motion, the performance gains are not substantial, while the gap between adaptive MH-BCS-SPL and MH-BCS-SPL broadens as the subrate increases.

We also compare the visual quality of the reconstruction results. Figures 9 and 10 show examples for the reconstructed frame 50 of the Mother-daughter sequence and Football sequence with $S_{CS} = 0.3$ and $S_K = 0.7$. These also demonstrate better performance of our adaptive framework. Thus, the proposed method gives better subjective and objective reconstruction quality. Compared to the original MH-BCS-SPL scheme, our method explores both the temporal and spatial correlation not only between neighboring frames but also between neighboring blocks in CS-frame itself to build adaptive dictionary for more spare description and, therefore, is able to achieve higher quality.

5. Conclusion

In this paper, a new distributed compressive video sensing framework based on adaptive MH prediction is proposed to simultaneously capture and compress videos at the low-complexity encoder and efficiently reconstruct videos at the decoder. The proposed framework can estimate the inter-frame correlation between colocated blocks in neighbor frames based on CS measurements to further divide these blocks into two types. We exploit both the temporal and spatial correlation between neighbor frames and blocks in CS-frame. To enhance CS recovery, the adaptive MH predictions are developed to find the best linear combination of hypotheses. Our simulation results demonstrate that the proposed framework can provide better reconstruction quality than the original MH-BCP-SPL algorithm. Additional considerable gain, approximately 0.5–3.6 dB, in the average PSNR can be achieved compared with the prior works.

References

[1] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[2] M. F. Duarte, M. A. Davenport, D. Takbar et al., "Single-pixel imaging via compressive sampling: building simpler, smaller, and less-expensive digital cameras," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, 2006.

[4] P. N. Josep, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proceedings of the Picture Coding Symposium (PCS '09)*, May 2009.

[5] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 1169–1172, April 2009.

[6] V. Stanković, L. Stanković, and S. Cheng, "Compressive image sampling with side information," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 3037–3040, November 2009.

[7] H. W. Chen, L. W. Kang, and C. S. Lu, "Dynamic measurement rate allocation for distributed compressive video sensing," in *Visual Communications and Image Processing*, Proceedings of the SPIE, July 2010.

[8] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 1393–1396, November 2009.

[9] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, 2001.

[10] L.-W. Kang and C. S. Lu, "Distributed compressive video sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1169–1172, Taipei, Taiwan, April 2009.

[11] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[12] L. Gan, "Block compressed sensing of natural images," in *Proceedings of the International Conference on Digital Signal Processing*, pp. 403–406, Cardiff, UK, July 2007.

[13] J. Zheng and E. L. Jacobs, "Video compressive sensing using spatial domain sparsity," *Optical Engineering*, vol. 48, no. 8, pp. 1–10, 2009.

[14] H. Jung and J. C. Ye, "Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: what we can learn from video compression techniques," *International Journal of Imaging Systems and Technology*, vol. 20, no. 2, pp. 81–98, 2010.

[15] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 1393–1396, Cairo, Egypt, November 2009.

[16] E. W. Tramel and J. E. Fowler, "Video compressed sensing with multihypothesis," in *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., pp. 193–202, Snowbird, Utah, USA, March 2011.

[17] S. Hoggar, *Mathematics of Digital Images*, Cambridge, UK, 2006.

[18] Y. Baig, E. M. K. Lai, and A. Punchihewa, "Distributed Video Coding Based on Compressed Sensing," *ICME Workshops*, pp. 325–330, 2012.

[19] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V.H. Winston & Sons, Washington, DC, USA, 1977.

[20] S. Zhu and K. K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 287–290, 2000.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 913782, 8 pages
http://dx.doi.org/10.1155/2013/913782

Research Article

A Privacy-Preserving Reauthentication Scheme for Mobile Wireless Sensor Networks

Shunrong Jiang, Jiapeng Zhang, JingJun Miao, and Conghua Zhou

The Department of Internet of Things, Jiangsu University, Zhenjiang 212013, China

Correspondence should be addressed to Shunrong Jiang; jsywow@gmail.com and Conghua Zhou; chzhou@ujs.edu.cn

Received 1 March 2013; Revised 23 April 2013; Accepted 23 April 2013

Academic Editor: Liangmin Wang

The mobile wireless sensor network (MWSN) is a new style WSN with mobile sinks or sensors in the network. MWSN has advantages over static WSN in the aspect of better energy efficiency, improved coverage, and superior channel capacity. However, mobile nodes also bring some security problems. For example, it is difficult to ensure secure communications among the mobile nodes and static nodes. In this paper, we design a lightweight mobile reauthentication protocol for mobile nodes. The designed protocol provides forward secure pairwise key for the mobile node when it moves from one cluster to another. Thus, the mobile sensor node can be authenticated by the new cluster head, and the privacy of his origin area is protected. In addition, the security and performance analysis shows that our scheme meets the need of lower communication and computation overhead, while achieving security requirement for mobile sensor node application in MWSN.

1. Introduction

WSN has become more and more prospective in human life with the development of hardware and communication technologies. However, due to the static network style, there are some natural limitations of WSN, such as network connectivity and network lifetime [1–4]. Furthermore, more and more researches find that the mobility in WSN not only improves the overall network lifetime and the data capacity of the network, but also addresses delay and latency problems [5–9]. There are many researches on how to realize better energy efficiency, improve coverage, enhance target tracking, and cause superior channel capacity for MWSN. However, limited researches consider the issue caused by the mobile sensor nodes, such as credibility with low consumption overhead and secure communication in MWSN. While more and more application scenarios require mobile sensors in WSN, such as traffic detection, animal observation E-Health, and battlefield. Furthermore, some present researches begin to consider the mobile adversary [10], which brings new security problems. Therefore, we should pay attention to realize the mutual authentication between the mobile node and the cluster efficiently, generate the new pairwise key, and make sure of the security of data transmission.

The framework of MWNS is given as in Figure 1. The network considers four types of entities:

(1) base station—as usual, the base station is assumed to be absolutely secure, which has plenty bandwidth, energy, storage space, and computation capability;

(2) cluster head—cluster head is assumed to have more storage space, energy, communication range, and computation capability than sensor node, and notice that, in general, the communication range of cluster head is also larger than the sensor node;

(3) static sensor node—we consider static sensor nodes in our network model, since they can work for the cluster head, and relay for mobile sensor node which has smaller communication range than cluster head, and in general, we assume that it has limited storage space, energy, and communication range;

(4) mobile sensor node—the mobility is the only difference between the mobile sensor node and the static sensor node, and the mobile sensor node roams from one cluster to another cluster and communicates with the nodes in the cluster.

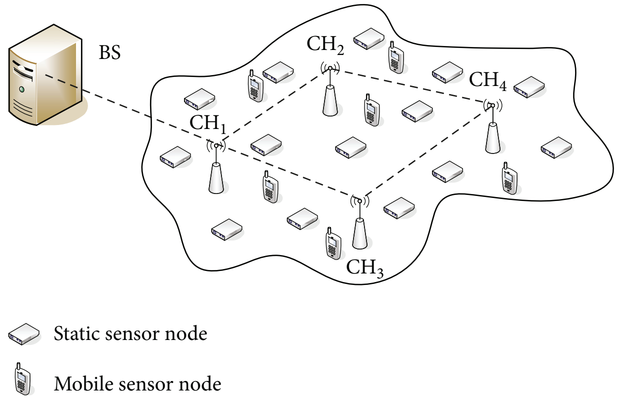


Figure 1: The framework of MWSN.

Based on the framework of MWSN, some security problems are brought by the mobility of mobile sensor node. One problem is the identity authentication of mobile sensor node in the new cluster. The other is the new session key generated between mobile sensor node and new cluster to ensure security communication. Moreover, we should protect the privacy of the mobile sensor node which means attackers cannot track it. Therefore, the security requirements of MWSN are given as follows:

(1) the authentication of identity—making sure that the identities of both parties who generate the key are trusted;

(2) data integrity—ensuring that only authorized person can modify the transmission of information;

(3) message privacy—transferring data can only be read by authorized parties;

(4) key freshness: ensuring that the existing session key is only used at current stage;

(5) node Resiliency—the network recovers when some nodes are captured by malicious attackers;

(6) privacy-preserving—since the mobile sensor node roams in WSN, attackers cannot track the mobile sensor node;

(7) scalability—allowing revoking and joining nodes. With the expanding of the network scale, it has little effect on the storage space of nodes and communication consumption.

In this paper, we focus on the security requirements caused by mobile sensor nodes. For the mobile sensor node in MWSN, we present an efficient node reauthentication and key generation scheme for mobile sensor nodes which consumes less communication and computation overhead and protects the privacy of the mobile sensor node. The security and performance analysis shows that our re-authentication scheme for mobile sensor node cannot only efficiently realize the secure requirements for MWSN, but also suit for the limited resource WSN.

The rest of this paper is organized as follows. In Section 2, we introduce the related work. We present our protocol in Section 3. Section 4 gives the security analysis of re-authentication protocol. Section 5 gives the performance analysis and simulation. Finally, we conclude the paper in Section 6.

2. Related Work

We introduce our related work from three aspects: the lightweight authentication schemes for WSN, the research of mobile sink in MWSN, and the re-authentication schemes for the mobile sensor node in MWSN.

The demand of lightweight is mostly considered in sensor network. All nodes in sensor network are considered to be static initially. For example, Perrig et al. [11] proposed a typical authentication scheme named μTESLA (Timed Efficient Stream Loss-Tolerant Authentication) by using the one-way hash chain. The protocol publishes the authentication key K_{mac} through delay to ensure that before the K_{mac} is published, the attacker cannot forge the correct broadcast packets. Du et al. [12] constructed an authentication path based on the public key mechanism by using Merkle Tree to reduce the computation and communication overheads. He also proposed dividing the entire WSN network into region Merkle Tree which can reduce the height of the Tree and the hops of the authentication. Ibriq and Mahgoub [13] proposed an efficient authentication program in which BS (Base Station) acts as the role of Certificate Authentication (CA) and assigns part of its functions to CH (Cluster Head). A sink can generate a key from "Partial Key Escrow Tab" [13] in all nodes and can be elected as Cluster Head. After the data integrated, messages are exchanged among cluster heads and finally transmitted to BS. However, since the partial key escrow Tab should be stored in every node, this scheme needs additional storage space. All these authentication protocols are for static nodes without considering roaming issue.

The advantages of mobile sink in MWSN have attracted much attention. Zhang et al. [14] proposed several efficient schemes to restrict the privilege of a mobile sink without impeding its capability of carrying out authorized operations for an assigned task. To prevent the authenticator from revealing information due to mobile sink compromises, the privileges of the authenticator are restricted by adding parameters, such as the starting time and the ending time of a task, the type of a task, and ID of the mobile sink. Vieira et al. [15] proposed a bioinspired location service named Phero-Trail location service protocol. In Phero-Trail, location information is stored in a 2D upper hull of a Sensor Equipped Aquatic Swarm, and a mobile sink uses its trajectory projected to the 2D hull to maintain location information. This enables mobile sensors to efficiently locate a mobile sink. The results show that Phero-Trail performs better than existing approaches. Agrawal et al. [16] proposed a key update protocol which securely updates the session key between a pair of nodes with the help of random inputs in mobile sensor networks. The security analysis shows that the proposed protocol resists known-key, impersonation, replay, worm, and sink hole attacks, while also provides forward secrecy, key freshness, and key control.

Table 1: Notation.

Notation	Description
T	Timestamp
M_A	The mobile node
CH_A, CH_B	The cluster head
$K_{A,B}$	The pair-wise key for A and B
$\{M\}_K$	Encrypt message M by K
$\mathrm{MAC}(k, M)$	The message authentication code of M using K
$H()$	Hash function
$\|\|$	Message connecting
$\oplus$	Xor

Recently the security of mobile sensor nodes in WSN has been paid more and more attention. Han and Kim [17] proposed the re-authentication issue concerning mobile nodes moving among sink nodes. The scheme considers the sink in the home cluster as a trusted third party. It prestores authentication information in all surrounding neighbor clusters and transfers the credible information to the new sink. The communication and computation overhead of re-authentication is reduced through credible trust. Qiu et al. [18] considered a sensor node roaming within a very large and distributed wireless sensor network, such as the application of healthcare field, in which the sensor nodes are deployed in the patient's body. When a dynamic sensor node moves to new area and wants to attack a router or a cluster head in this area, it first sends a request message to the base station. After verifying validity of the request message, the base station generates the session key for mobile node and the router and sends it to the router, and then the router sends the material of session key to the mobile node to generate the session key. Qiu also improves the E-G scheme to guarantee that two sensor nodes share at least one key with probability 1 with less storage and energy overhead. The disadvantages of Han's scheme are as follows. First, it only takes the mobile node, sink node, and base station into consideration. Then, the communication overhead of the program mostly concentrates on the mobile node, so it has influence on the lifetime of the mobile node. Lastly, the re-authentication material is prestored in the neighbor clusters, which exists unnecessary communication overhead and information leak. In Qiu's scheme, the base-station is always online and provides the full utilities. The re-authentication also depends on the base station which incurs large communication overhead.

3. The Proposed Protocol

With the mobility of MWSN, the mobile sensor nodes may move from one cluster to another. If we repeat the new nodes addition process proposed in [4], the scheme will degrade to the E-G [19] scheme. Besides, some predistribution schemes need to interrupt the operation of network and implemented by man, which is unrealistic for the running wireless network. Therefore, the roaming behavior of mobile sensor nodes must consider how to get trust from the new cluster and generate the pair-wise key to achieve security communication. Considering the security requirements caused by mobile sensor nodes in the MWSN, we design our re-authentication protocol as follows. For convenience, we assume that after the static WSN generated the communication keys for the entities, the mobile sensor nodes join the network from some clusters.

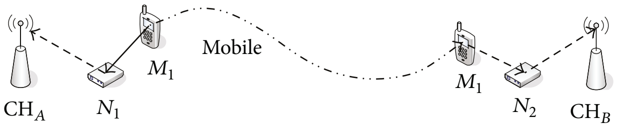
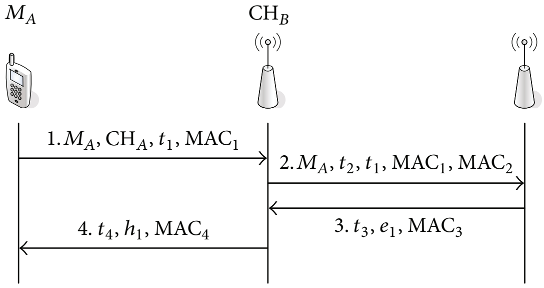


Figure 2: The re-authentication of mobile sensor node.



M_A
CH_B
1. M_A, CH_A, t_1, MAC_1
2. M_A, t_2, t_1, MAC_1, MAC_2
4. t_4, h_1, MAC_4
3. t_3, e_1, MAC_3

Figure 3: Re-authentication of mobile sensor nodes.

As Figure 2 shows, the whole process can be divided into two phrases. *Phrase 1* the mobile sensor node M_A first registers at the base station and then joins the network from the fixed cluster CH_A (home cluster). The main purpose of this phase is to make M_A initial trustily join the network. To ensure security, this phrase is realized by offline distribution. *Phrase 2* M_A moves to another new cluster CH_B (foreign cluster), and it should pass the legal identify authentication of CH_B before enjoying the resource of this cluster. To realize the mutual authentication of M_A and CH_B, we can use the trusted relationship among M_A, CH_A, and CH_B. We describe the further details of re-authentication as follows. The notation used throughout our scheme is shown in Table 1.

3.1. Predistribution for Mobile Sensor Node. Before roaming in the network, M_A should register in the base station and get the pre-distribution material by offline. Then, M_A joins the network from cluster head CH_A. After being authenticated by CH_A, M_A has the information including a hash number $H(I)$, a random number R_1, and the session key K_{CH_A, M_A}.

3.2. Mobile Sensor Node Reauthentication. After registering at the base station and trusted joining CH_A, M_A can roam in the network among clusters. As Figure 2 shows, after completing *Phase 1*, the mobile sensor node M_A moves to a foreign cluster CH_B, and M_A should pass the authentication of CH_B before communicating with other nodes in CH_B. The implementation mutual authentication of re-authentication protocol is shown in Figure 3.

```
Require: $M_A, t_2, t_1, \text{MAC}_1, \text{MAC}_2$.
Verify $(t* - t_2) \le \Delta t$.
if $t_2$ is valid then
    Compute $\text{MAC}_2* = (K_{\text{CH}_B,\text{CH}_A}, M_A||t_2||t_1||\text{MAC}_1)$.
    if $\text{MAC}_2* = \text{MAC}_2$ then
        Compute $\text{MAC}_1* = (K_{\text{CH}_A,M_A}, M_A||t_1||H(I))$.
        if $\text{MAC}_1* = \text{MAC}_1$ then
            Compute $e_1 = \{H(I), R_1\}_{K_{\text{CH}_A,\text{CH}_B}}$ and
             $\text{MAC}(K_{\text{CH}_A,\text{CH}_B}, t_3||e_1)$.
        end if
    end if
end if
```

Algorithm 1: The process executed by CH_A to *message 2*.

```
Require: $t_3, e_1, \text{MAC}_3$.
Verify $(t* - t_3) \le \Delta t$.
if $t_3$ is valid then
    Compute $\text{MAC}_3* = \text{MAC}(K_{\text{CH}_A,\text{CH}_B}, t_3||e_1)$.
    if $\text{MAC}_3* = \text{MAC}_3$ then
        Extract $H(I)$ and $R_1$.
        Generate random number $R_2$, compute $K_{\text{CH}_B,M_A}$
         as formula (1).
        Compute $h_1 = H(R_1) \oplus R_2$.
        Compute $\text{MAC}_4 = (K_{\text{CH}_B,M_A}, H(I)||t_4||h_1)$.
    end if
end if
```

Algorithm 2: The process executed by CH_B to *message 3*.

(1) When M_A moves to the new cluster CH_B, it first launches the authentication procedure to CH_B. M_A sends *the message 1*: M_A, CH_A, t_1, MAC_1 to CH_B, where t_1 is the timestamp, and $\text{MAC}_1 = (K_{\text{CH}_A,M_A}, M_A||t_1||H(I))$.

(2) When CH_B receives *message 1* at time $t*$, and CH_B first checks whether $(t * - t_1) \le \Delta t$. If the result is valid, since there is no shared information between CH_B and M_A, and CH_B would send *message 2*: M_A, t_2, t_1, MAC_1, MAC_2 to CH_A, where $\text{MAC}_2 = (K_{\text{CH}_B,\text{CH}_A}, M_A||t_2||t_1||\text{MAC}_1)$.

(3) Upon Receiving the *message 2*, the home cluster CH_A verifies message as Algorithm 1 and replies *message 3*: t_3, e_1, MAC_3 to CH_B.

(4) After receiving the *message 3*, CH_B verifies message as Algorithm 2. Then, CH_B sends the *message 4*: t_4, h_1, MAC_4 to M_A.

(5) Upon receiving the *message 4* from CH_B, the mobile node M_A executes Algorithm 3 to get the pair-wise key

$$K_{\text{CH}_B,M_A} = H\left(H\left(I\right)\|R_1\|\,R_2\right). \tag{1}$$

After generating the session key, M_A verifies the correctness of MAC_4. If the validation is right, the session key is right. During the communication with CH_B, $H(I)$ and R_1 should be updated as $H(I')$ and R_1' which are used as the authentication material for the further re-authentication. CH_B also sends these information to the base station.

For convenience, the role of home cluster is acted by the foreign cluster node through which the mobile sensor nodes have completed the re-authentication process. That means that after M_A completes re-authentication in the foreign cluster CH_B, CH_B is the new home cluster of mobile sensor node. When M_A moves to another foreign cluster CH_C, CH_B acts as the home cluster which responsibly completes the re-authentication between M_A and CH_C.

Taking the issue of tracking and protecting the privacy of M_A into account, we use the pseudonyms methods [20, 21] during the communication. The whole time of M_A in CH_B is divided in accordance with the time slice TS_j, and the length of each time slice is Δt, which means we can get C time slices. We denote $\text{PID}_{A,j}$ as the pseudonym of M_A in the time slice TS_j, where $\text{PID}_{A,j}$ is generated by two hash seeds $H(I)$ and R_2 as formula (2)

$$\begin{aligned} S_{1,j} &= H^j\left(R_2\right), \\ S_{2,j} &= H^j\left(H\left(I\right)\right), \\ \text{PID}_{A,j} &= H\left(S_{1,j} \oplus S_{2,j}\right). \end{aligned} \tag{2}$$

```
Require: $t_4, h_1, \text{MAC}_4$.
Verify $(t* - t_4) \le \Delta t$.
if $t_4$ is valid then
    Compute $H(R_1)$, and $R_2* = h_1 \oplus H(R_1)$.
    Compute $K_{\text{CH}_B,M_A}*$ as formula (1).
    Compute $\text{MAC}_4* = (K_{\text{CH}_B,M_A}*, H(I)||t_4||h_1)$.
    if $\text{MAC}_4* = \text{MAC}_4$ then
        $K_{\text{CH}_B,M_A}* = K_{\text{CH}_B,M_A}$.
    end if
end if
```

Algorithm 3: The process executed by M_A to *message 4*.

Notice that since CH_B knows R_2 and $H(I)$, so it can trace the messages sent by M_A in its communication range. While M_A moves to CH_C at TS_j (actually CH_C only knows a mobile sensor node named $\text{PID}_{A,j}$ joining its cluster), because CH_B does not have the materials to generate the pseudonyms, so it cannot trace the messages sent by M_A out of its communication range. By this way, we can protect the privacy of M_A.

4. Protocol Security Analysis

4.1. The Protocol Satisfies Forward Security. Suppose that the attacker gets the session key K_{CH_C,M_A} between the mobile sensor node M_A and cluster node CH_C. It is difficult for attackers to derive the session key used before such as K_{CH_B,M_A}. The session key between M_A and CH_B is determined by two random numbers R_1 and R_2. R_1 is produced in the last re-authentication cycle and is transmitted in the ciphertext. R_2 is transmitted by the XOR hash value h_1 in *message 4*. If the attacker wants to obtain the plaintext R_1, he must know the session key K_{CH_A,M_A} between M_A and CH_A. Thus, the problem is deduced into how to get the session key between M_A and the first cluster CH_A. K_{CH_A,M_A} is sent offline, which is assumed to be secure. R_2 is gotten by the hash and XOR of the hash value of R_1, and according to the irreversibility of hash, the problem of obtaining the plaintext of R_2 is derived to obtain the plaintext R_1. Even if attackers get the current session key of M_A, they cannot derive the previous session key of M_A through the previously analysis. The protocol satisfies forward security.

4.2. Mutual Identity Authentication. In our scheme, as there is no shared information between CH_A and CH_B, CH_B cannot verify the identity of M_A, so when CH_B receives *message 1*, it transfers the message to CH_A. CH_A helps CH_B authenticate the identity of M_A by computing MAC_1 through using the hidden $H(I)$. M_A authenticates the identity of the foreign cluster CH_B mainly through MAC_4 which also uses the hidden $H(I)$. If MAC_4 is right, we believe that CH_B has the right identity. By this way, we realize mutual identity authentication.

4.3. Prevent Man-in-the-Middle Attack. From the analysis of our scheme, an attacker can track or intercept *message 1* to act the mobile sensor node M_A and continue communicating with foreign cluster head. It makes the entire protocol go on running. Finally, feedback message (*message 4*) is gotten to extract the session key material. However, according to the analysis of forward security, R_1 and R_2 are not sent in plaintext. In order to attack the protocol, the previous session key should be known. And the whole problem is back to the security of K_{CH_A,M_A}. For man-in-the-middle attack, as mentioned in mutual identity authentication, mutual identity authentication ensures the correctness of the identity of the message sender. MAC used in every message ensures the message integrity. According to the general security assumption of MAC [17], attackers cannot construct a valid message to achieve communication. So the protocol can prevent man-in-the-middle attack.

4.4. Prevent Replay Attack. When the mobile node M_A applies to join registered foreign cluster, every message of our scheme has the current timestamp (t_1, t_2, t_3, t_4). The message received in Δt time, to some extent, can prevent replay attack. According to the session key generated in formula (1), the generation of session key selects new random number, which ensures the freshness of session key and prevents replay attack effectively.

4.5. Protect the Privacy of the Mobile Node. Since the communication of mobile sensor node uses the pseudonyms, attackers and other entities cannot distinguish them which protects the privacy of the mobile sensor node. But to the base station and cluster heads, they can track the mobile sensor node. After the mobile node joins the foreign cluster, the cluster head sends the $H(I)$ and R_1 to the base station, which helps the base station to track and manage the mobile sensor node. However, for the cluster head (such as CH_C), CH_C only knows that the pseudonyms of the mobile sensor node M_A is in its cluster. When the mobile sensor node M_A moves to a new cluster head (such as CH_D), CH_C does not know the pseudonyms of M_A, and it cannot track M_A. Therefore, the privacy preserving of mobile sensor node is conditional.

5. Protocol Performance Analysis

We give the performance analysis of our scheme in this section in terms of communication pass, message size, and

Table 2: The required number of communication passes.

	Han's scheme	Qiu's scheme	Our scheme
Node to Sink	$2n$	n	n
Sink to Sink	m	l	2
Sink to Node	1	1	1
BS to Sink	—	1	—

computation overhead. We also give the simulation of our scheme on the NS2 simulation platform and use the time delay to reflect the efficiency of our scheme.

5.1. Communication Pass. We compared the required number of communication passes with Han's [17] and Qiu's [18] schemes, since both of them propose the reauthentication protocols for mobile sensor nodes in WSN. Table 2 shows the comparison of communication passes for mobile node reauthentication, where n denotes the number of hops from M_A to the foreign cluster head (sink), m denotes the number of neighbor cluster heads (sinks) around the home cluster head, and l denotes the number of hops from foreign cluster (sink) to the base station which is used in Qiu's scheme.

Since Han's and our schemes use the relation among cluster heads to realize re-authentication for mobile sensor node, which do not need communication with the base station, in Qiu's scheme, when the node joins a new sink, it first sends the requirement message to the base station. Actually, the message is first sent to the foreign cluster head (n hops) and then to the base station via the foreign cluster (l hops) which incurs large communication overhead. The hole communication passes are $(n + l)$ hops.

Although the re-authentication of Han's scheme does not need communication with the base station, he prestores the authentication information in all surrounding neighbor cluster heads which are related with the number of neighbor cluster (m hops), while our scheme realizes the re-authentication by the tradition tripartite authentication, which results in less communication pass.

5.2. Message Size. The message size during the re-authentication process is quantified by the byte which is to show the communication overhead. We compare the message size with Han's. We use the base parameter setting of message as Han's [17] in Table 3.

From Table 4, we can see that our scheme has less message size of the whole re-authentication process. Notice that Han pre-stores authentication information in all surrounding neighbor clusters and we only consider that the number of authentication material of transmission size is 36 bytes, while the actual number may be more than 36.

During re-authentication for mobile sensor node, we reduce the message length transmission among the entities since the data transmission consumes much more energy than computation in WSN.

Table 3: The base parameter setting of message.

Notation	Length (byte)
MAC	4
Random number	8
Identity	1
Time stamp	8
Key size	16

Table 4: The required message size for re-authentication (byte).

	Han's scheme	Qiu's scheme	Our scheme
Node to Sink	$48n$	$15n$	$14n$
Sink to Sink	$\geq 36m$	$15l$	53
Sink to Node	86	22	20
BS to Sink	—	50	—
Total	$\geq (48n + 36m + 86)$	$15(n + l) + 72$	$14n + 73$

Table 5: The required message size for re-authentication.

	Han's	Qiu's	Ours
Encryption/decryption in total	4	2	4
Encryption/decryption in node	1	1	2
MAC generation/verification in total	8	4	8
MAC generation/verification by node	3	2	4

5.3. Computation Overhead. Computation overhead is quantified by the number of execution encryption algorithm. As Table 5 shows, the total number of encryption/decryption of our scheme is the same as Han's, both more than Qiu's scheme. Since the re-authentication of Qiu's scheme is based on the base station, our scheme and Han's scheme are based on the relation among clusters.

5.4. Protocol Simulation. We use NS2.29 to simulate our scheme and Han' scheme [17], since both of them realize mobile node re-authentication without requiring communication with the base station. We use the transmission delay to quantify the message size, which can reveal the availability and efficiency of our scheme. The simulation uses the mesh network topology, MAC layer uses the 802.15.4 protocol written by Zheng and Lee [22] for NS2, the routing layer uses the AODV routing protocol which has the shortest hops, the transportation layer uses the UDP protocol, and the application layer transmits the CBR packet. The message size is set as Table 4. The data transmission speed is 250 KB/S, which adopts the recommended beacon mode standard setting in reference [22].

Supposing the communication radius of the mobile sensor node and the common sensor nodes within the cluster to be 20 m, the communication radius of cluster head is 100 m. The computation delay of *message 1* and *message 4* in mobile sensor node is 6 ms and 3 ms [23], respectively, while the computation delay of *message 2* and *message 3* node is 1 ms for cluster head.

To reflect the comparison of Table 4, we design two groups of simulation for our scheme and Han's scheme [17]. The number of each group simulation is 100 times.

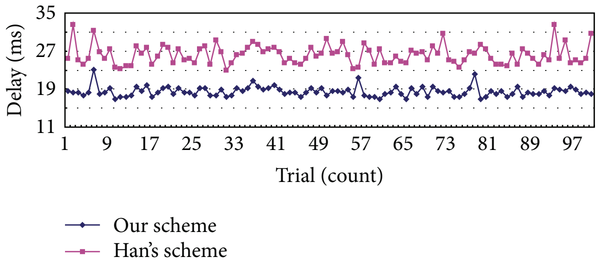


Figure 4: The time delay for $m = 1$, $n = 2$.

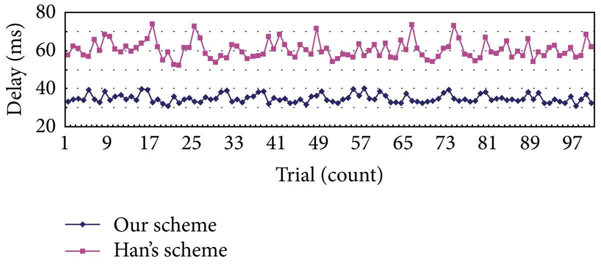


Figure 5: The time delay for $m = 1$, $n = 5$.

In Figure 4, $m = 1$, and $n = 2$. The simulation delay of Han's scheme is 26.217 ms, while our scheme is 18.432 ms. However, the time delay of our schemes simulation is not as good as the comparison in Table 4. Since there is an addition MAC layer head for each message, the time delay of simulation is not the same as the comparison of message size in Table 4. From Figure 4, we can know the whole delay of our scheme is less than Han's. On one hand, our scheme has less message size, on the other hand, we reduce the number of messages sending. The fluctuation of the simulation in Figure 4 is caused by $n = 2$. Since the message transmitted by the static nodes in cluster needs to consume transmission delay (when node transmits message, it will repeat calling the sending and receiving process, and seek the routing table, which leads to more delay time), and it results in the unstability of time.

In Figure 5, $m = 1$, and $n = 5$. The simulation delay of Han's scheme is 60.384 ms, while our scheme is 34.8608 ms. Compared with Figure 4, with the number of relay hops increasing, the advantage of our scheme is more obvious. This is due to less communication message size of mobile sensor node. Moreover, with the number of hops increasing, the instability of the simulation is more obvious.

6. Conclusion

The security problem brought by the mobile sensors in MWSN attracts more and more attention of researchers. In this paper, we propose a re-authentication protocol for the mobile node roaming among clusters. Our protocol can transfer the credibility among the clusters which can efficiently achieve the requirements of secure identity authentication and establish the forward secure pairwise key. Meanwhile, the base station can track the mobile trajectory and protect the privacy of the mobile sensor node. We also give performance analysis and simulation for our re-authentication protocol. The results and comparison show that our protocol achieves better security and has better performance on communication overhead, message size, and computation cost.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61202474 and the Natural Science Foundation of Jiangsu Province under Grant BK2011464.

References

[1] M. F. Younis, K. Ghumman, and M. Eltoweissy, "Location-aware combinatorial key management scheme for clustered sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 8, pp. 865–882, 2006.

[2] M. Eltoweissy, M. H. Heydari, L. Morales, and I. H. Sudborough, "Combinatorial optimization of group key management," *Journal of Network and Systems Management*, vol. 12, no. 1, pp. 33–50, 2004.

[3] S. Hussain, F. Kausar, and A. Masood, "An efficient key distribution scheme for heterogeneous sensor networks," in *Proceedings of the International Wireless Communications and Mobile Computing Conference (IWCMC '07)*, pp. 388–392, New York, NY, USA, August 2007.

[4] X. Du, Y. Xiao, M. Guizani, and H. H. Chen, "An effective key management scheme for heterogeneous sensor networks," *Ad Hoc Networks*, vol. 5, no. 1, pp. 24–34, 2007.

[5] S. A. Munir, B. Ren, W. Jiao, B. Wang, D. Xie, and J. Ma, "Mobile wireless sensor network: architecture and enabling technologies for ubiquitous computing," in *in Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops*, pp. 113–120, Niagara Falls, Canada, 2007.

[6] M. Rahimi, H. Shah, G. S. Sukhatme, J. Heideman, and D. Estrin, "Studying the feasibility of energy harvesting in a mobile sensor network," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '03)*, pp. 19–24, September 2003.

[7] M. Yarvis, N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu, and S. Singh, "Exploiting heterogeneity in sensor networks," in *Proceedings of the IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies (NFOCOM '05)*, pp. 878–890, Hillsboro, Ore, USA, March 2005.

[8] W. Wang, V. Srinivasan, and K. Chua, "Using mobile relays to prolong the lifetime of wireless sensor networks," in *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking (MobiCom '05)*, pp. 270–283, New York, NY, USA, 2005.

[9] B. Liu, P. Brass, O. Dousse, P. Nain, and D. Towsley, "Mobility improves coverage of sensor networks," in *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC '05)*, pp. 300–308, New York, NY, USA, May 2005.

[10] R. D. Pietro, C. Soriente, A. Spognardi, and G. Tsudik, "Collaborative authentication in unattended wsns," in *Proceedings of the*

2nd ACM Conference on Wireless Network Security (WiSec '09), pp. 237–244, New York, NY, USA, 2009.

[11] A. Perrig, R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler, "SPINS: security protocols for sensor networks," *Wireless Networks*, vol. 8, no. 5, pp. 521–534, 2002.

[12] W. Du, R. Wang, and P. Ning, "An efficient scheme for authenticating public keys in sensor networks," in *Proceedings of the 6th ACM International Symposium on Mobile ad hoc Networking and Computing (MobiHoc '05)*, pp. 58–67, New York, NY, USA, 2005.

[13] J. Ibriq and I. Mahgoub, "A hierarchical key establishment scheme for wireless sensor networks," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications (AINA '07)*, pp. 210–219, NiagaraFalls, Canada, May 2007.

[14] W. Zhang, H. Song, S. Zhu, and G. Cao, "Least privilege and privilege deprivation: Towards tolerating mobile sink compromises in wireless sensor networks," in *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '05)*, pp. 378–389, New York, NY, USA, May 2005.

[15] L. F. M. Vieira, U. Lee, and M. Gerla, "Phero-trail: a bio-inspired location service for mobile underwater sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 4, pp. 553–563, 2010.

[16] S. Agrawal, R. Roman, M. L. Das, A. Mathuria, and J. Lopes, "A novel key update protocol in mobile sensor networks," in *Proceedings of the 8th International Conference on Information Systems Security (ICISS '12)*, vol. 7671 of *Lecture Notes in Computer Science*, pp. 194–207, 2012.

[17] K. Han and K. Kim, "Untraceable mobile node authentication in wsn," *Sensors*, vol. 10, pp. 4410–4429, 2010.

[18] Y. Qiu, J. Zhou, J. Baek, and J. Lopez, "Authentication and key establishment in dynamic wireless sensor networks," *Sensor*, vol. 10, pp. 3718–3731, 2010.

[19] L. Eschenauer and V. Gligor, "A key management scheme for distributed sensor networks," in *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS '02)*, pp. 41–47, New York, NY, USA, 2002.

[20] S. Jiang, X. Zhu, and L. Wang, "A conditional privacy scheme based on anonymized batch authentication in vehicular ad hoc networks," in *Proceedings of the Wireless Communications and Networking Conference (WCNC '13)*, Shanghai, China, 2013.

[21] Y. Sun, R. Lu, X. Lin, X. Shen, and J. Su, "An efficient pseudonymous authentication scheme with strong privacy preservation for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 7, pp. 3589–3603, 2010.

[22] J. Zheng and M. Lee, "A comprehensive performance study of IEEE 802.15.4," in *Proceedings of the 4th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WIMOB '08)*, pp. 580–585, 2008.

[23] G. D. Meulenaer, F. Gosset, F. X. Standaert, and O. Pereira, "On the energy cost of communication and cryptography in wireless sensor networks," in *Sensor Network Operations*, chapter 4, pp. 218–237, Wiley-IEEE Press, New York, NY, USA, 2006.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 256852, 9 pages
http://dx.doi.org/10.1155/2013/256852

Research Article

An Efficient Data Aggregation Protocol Concentrated on Data Integrity in Wireless Sensor Networks

Liehuang Zhu, Zhen Yang, Meng Li, and Dan Liu

Beijing Engineering Research Center of Massive Language Information Processing and Cloud Computing Application, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Liehuang Zhu; liehuangz@bit.edu.cn

Received 5 March 2013; Accepted 30 April 2013

Academic Editor: Yulong Shen

Wireless sensor networks consist of a great number of sensor nodes with strictly limited computation capability, storage, communication resources, and battery power. Because they are deployed in remote and hostile environments and hence are vulnerable to physical attacks, sensor networks face many practical challenges. Data confidentiality, data integrity, source authentication, and availability are all major security concerns. In this paper, we focus on the very problem of preserving data integrity and propose an Efficient Integrity-Preserving Data Aggregation Protocol (EIPDAP) to guarantee the integrity of aggregation result through aggregation in sensor networks. In EIPDAP, base station can immediately verify the integrity of aggregation result after receiving the aggregation result and corresponding authentication information. However, to check integrity, most existing protocols need an additional phase which will consume a lot of energy and cause network delay. Compared with other related schemes, EIPDAP reduces the communication overhead per node to $O(\Delta)$, where Δ is the degree of the aggregation tree for the network. To the best of our knowledge, EIPDAP has the most optimal upper bound on solving the integrity-preserving data aggregation problem.

1. Introduction

Wireless sensor networks (WSNs) have many security-critical applications such as real-time traffic monitoring, wildfire tracking, or military surveillance. In a sensor network, thousands of low-cost sensor nodes collectively monitor an area within a certain range and report their own data to the base station which distributes a data query. However, this would incur high communication overhead which cannot be afforded by sensor nodes. Data aggregation [1, 2] mechanisms are proposed to reduce the power consumption. Data aggregation poses security threat; many secure data aggregation protocols [3, 4] have been emerging over these years, which prove to be secure and considerably improve the resource utilization.

Although data confidentiality could guarantee that legal parties obtain plain data without being leaked out to adversaries, it does not protect data from being altered [5–7]. In this paper, we focus on the problem of preserving data integrity through aggregation in sensor networks. Message authentication codes (MACs) are used in [8] to protect data integrity, while causing other problems, such as high communication overhead. In this paper, we present a provably secure sensor network integrity-preserving aggregation protocol based on the elliptic curve discrete logarithm problem for general networks with hierarchical aggregator topologies, assuming that adversaries are able to corrupt a (small) fraction of sensors. With the increasing of sensor node's computation capacity, public key cryptography, such as elliptic curve cryptosystems (ECC), is suitable for constrained environments such as WSN. In [9], authors propose secure data aggregation schemes using ECC to obtain data confidentiality and integrity in the data aggregation because of their smaller key size, faster computations, and reductions in processing power, storage space, and bandwidth. TinyECC is proposed by Liu and Ning [10] which provides ECC-based operations that can be flexibly configured and integrated into WSN applications.

An adversary can perform a variety of attacks. For example, a denial-of-service (DoS) attack can totally block the communication between sensor nodes and the base station. However, this attack is not concerned because it is detectable by the querier and solutions can be implemented to remedy this situation. In stealthy attack [4], the attacker's goal is to make the base station accept false aggregation results, which are significantly different from the true results determined by the measure values, while not being detected by the base station. Our goal is to prevent this kind of attack even when high-level aggregator is corrupted.

A number of protocols [11–13] have been proposed which focus on the problem that how can the base station obtain a good approximation of the aggregation result and how to obtain data integrity when a fraction of sensor nodes are compromised. One common sensor feature is the disproportionately high cost of transmitting information, as compared to performing local computation. For example, a Berkeley mote spends approximately the same amount of energy to compute 800 instructions as it does in sending a single bit of data. It thus becomes essential to reduce the number of bits forwarded by intermediate nodes, in order to extend the entire network's lifetime [14]. All the above schemes need to verify the integrity of aggregation result in an additional phase which consumes a lot of energy and causes network delay.

In this paper we propose EIPDAP, which can immediately verify the integrity of aggregation result after receiving the aggregation result and corresponding authentication information, hence significantly reducing energy consumption and communication delay which will be caused if the verification process is done through another query-and-forward phase.

The rest of the paper is organized as follows: in Section 2 we describe a survey of other approaches to integrity-preserving aggregation in sensor networks, in Section 3 more details about the problem we are trying to solve are discussed, in Section 4 we describe a new scheme that is, the centerpiece of our work, and in Section 5 the security properties and performance of our scheme are analyzed.

2. Related Work

There has been a number of works on preserving integrity in aggregation protocols for sensor networks. Many protocols have been proposed for the single-aggregator model [4, 13, 15]. But the aggregator in these schemes suffers from significantly high congestion and only reduces communications on the link between the aggregator and the base station. So this model is not scalable to large multihop sensor deployments.

Another significant work is introduced in [11]. The main idea of this approach is that each node sends its value, complement, and commitment up the aggregation tree and then a commitment would pass down the tree for a node to verify that if its value was added into the SUM aggregation and the complement of its data value was added into the COMPLEMENT aggregation. However, the scheme requires three phases. The delay aggregation strategy used in the second phase increases communication from $O(1)$ to $O(\log n)$, computation from $O(1)$ to $O(q \log n)$, where n is the number of nodes in the network and q is the number of forests in the commitment tree. The result-checking phase costs $O(\Delta \log^2 n)$ congestion. Frikken and Dougherty in [12] improves Chan's approach by reducing the maximum communication to $O(\Delta \log n)$.

A secure hop-by-hop data aggregation protocol SDAP for sensor networks is proposed in [13]. The authors believe that we should be more concerned about high-level nodes, since these nodes represent a large portion of the final result delivered to base station and there would be more catastrophic consequences if they are compromised. Hence, SDAP dynamically partitions the topology tree into multiple logical groups of similar sizes using a probabilistic approach, following the divide-and-conquer principle. In this way, fewer nodes are located under a high-level sensor node in a logical subtree resulting in reduced potential security threat by a compromised high-level node. SDAP introduces probability and attestation to the data result-checking; the communication required per node is $O(\log (n/n_g))$. Because SDAP just let part nodes be attested, attestation algorithm cannot find all compromised nodes. By adding attestation paths can increase the detection probability, but it will increase communication cost.

Aggregate message authentication codes introduced by Katz and Lindell (CT-RSA 2008) [8] provided a new perspective of preserving integrity. In their construction, aggregating MAC simply computes the XOR of all the MACs into one value, the size of which is the same as an ordinary MAC. After receiving all the data and the final aggregate MAC, the base station uses secret keys shared with each node to compute a new aggregate MAC from these data and compares it with the received aggregate MAC. Although it remarkably reduces communication overhead we have seen in former protocols [11–13] and is easy to perform, it suffers from the "mix-and-match" attacks [16] in which the adversary can easily forge several types of aggregate combinations.

In [9], the authors proposed a new algorithm using homomorphic encryption and additive digital signatures to achieve confidentiality, integrity, and availability for in-network aggregation in WSN. However, the protocol cannot resist stealthy attack. We discuss concrete attack on the protocol due to Albath and Madria [9] in the appendix.

Besides, there have been several protocols designed for preserving the confidentiality of the aggregation results [17–19]. This issue is orthogonal to our work and is not considered in this paper.

3. Problem Model

This section contains the definitions of basic problems and includes discussion on the nodes' setup, the security infrastructure, and the attack model.

3.1. Network Assumptions. We assume a query-based sensor network with a large number of sensors and a powerful base station with transmission ranges covering the whole wireless sensor network can broadcast messages to all nodes directly.

Before aggregation process, sensors will form a tree topology where base station locates at the root.

We further assume that the base station would broadcast an authenticated query before collecting data. If there is no aggregation tree, then an aggregation tree should be formed as the query has been sent to all nodes. Our protocol takes the structure of the aggregation tree as given. One method for constructing an aggregation tree is described in TaG [20].

Each node is sensing an integer value r that is in the range $(0, v]$ (we rule out "0" in defense of θ_i which we will explain later) for some application-based value v. The goal is to return the SUM result with two tags proving that the SUM result has not been forged (even in the presence of malicious nodes). Due to resource constrains, all readings need to be aggregated by aggregators while being transmitted over a multihop path.

3.2. Security Infrastructure. We assume that each node i has a unique identifier s_i, private keys $r_i, l_i \in Z_p$ and shares a private key sk_i and a private point $\theta_i \in$ cyclic elliptic group $E(Z_p)$ with base station. ECC domain parameters including the generator point $G \in E(Z_p)$ are preloaded in all nodes. In each node i we set two parameters α_i and β_i which will be used later:

$$\alpha_i = r_i G, \qquad \beta_i = r_i \alpha_i. \tag{1}$$

3.3. Attack Model and Security Goals. We consider a setting with a polynomially bounded adversary, which can physically access the sensors and read their interval values. The adversary is also restricted to corrupt a (small) fraction of nodes including the aggregators.

Once the adversary compromises a sensor node, it can obtain all the node's secret keys. An adversary can modify, forge, or discard messages or simply transmit false aggregate results, and its goal is to forge valid aggregate result to be accepted by the base station. The higher false aggregate result level is, more catastrophic consequence will be caused.

In this setting, we focus on stealthy attacks [4] where the attacker's goal is to make the base station accept false aggregate results while not being detected by base station. And our security goal is to prevent stealthy attacks. In particular, we want to guarantee that once the aggregate result has been accepted by the base station, it is indeed the real result aggregated by honest nodes.

Definition 1 (integrity-preserving aggregation algorithm). An aggregation algorithm is integrity-preserving if, by tampering with the aggregation process, an adversary is unable to induce the base station to accept any forged aggregate result.

Since if a sensor node is compromised, the adversary can obtain all its confidential information (e.g., cryptographic keys) and send false data without being detected. In this paper we will focus on the situation where an aggregator is compromised and see whether it can forge a valid aggregate result.

In this paper, however, we do not address the denial-of-service attack where the adversary prevents the querier from getting any aggregation result at all; because nodes' not responding queries clearly indicate that something is wrong and solutions can be implemented to remedy this situation.

4. Our Work

In this section, we present a new approach, especially aiming to preserve integrity of the aggregation result. We first give an overview of this approach and then present the details.

4.1. Overview. The design of our algorithm is based on the elliptic curve discrete logarithm problem. The overall algorithm consists of three main phases: query dissemination, aggregation-commit, and result-checking.

In query dissemination phase, the base station broadcasts the query to the network. An aggregation tree, or a directed spanning tree over the network topology with the base station at the root, is formed as the query sent to all the nodes, if one is not already present in the network. Then the path-keys and edge-key for each node encrypted with the secret key shared between base station and node are sent to the corresponding node. Path-key and edge-key are calculated by the base station according to the network topology. We show the detail of the calculation of the path-key in Section 4.2.

In aggregation-commit phase, each sensor node collects raw data and computes two corresponding tags before sending them to their own parent node in the aggregation tree. After receiving all the messages from all child nodes, aggregator performs modulo addition operations over the three items and forwards the result to high-level aggregators until the base station.

In the result-checking phase, the base station verifies the integrity of the SUM aggregation with two aggregation tags. Compared with Chan's and Keith's approach, ours does not require any dissemination from top root node down to the leaf nodes which causes congestion $O(\Delta \log^2 n)$ in Chan's approach and $O(\Delta \log n)$ in Keith's approach and energy cost in this phase.

4.2. The SUM Approach

4.2.1. Query Dissemination Phase. Before aggregation, the base station broadcasts an authenticated query to the network. The query request message contains a nonce N to prevent replay attack [1]. If there is no aggregation tree, an aggregation tree with the base station at the root will be formed as the query has been sent to all nodes. Then the tree information will be reported back to the base station. After the base station receives the tree information, it calculates the path-key for each node: for each aggregator or sensor node i, base station generates a key bs, and calculates edge key according to one-way hash function F, where

$$k_{i-j} = F_{\mathrm{bs}}\left(s_i, s_j, N\right), \tag{2}$$

and node i is the parent of node j. s_i and s_j are unique identifiers of node i and node j.

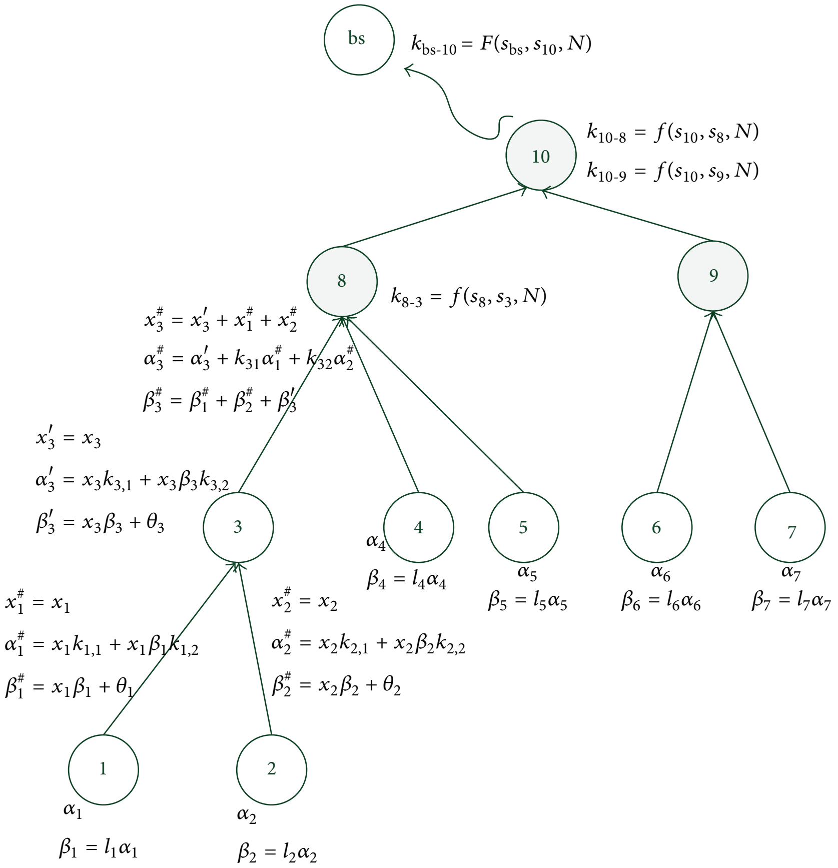

bs
$k_{\text{bs-10}} = F(s_{\text{bs}}, s_{10}, N)$
10
$k_{10\text{-}8} = f(s_{10}, s_8, N)$
$k_{10\text{-}9} = f(s_{10}, s_9, N)$
8
$k_{8\text{-}3} = f(s_8, s_3, N)$
9
$x_3^{\#} = x_3' + x_1^{\#} + x_2^{\#}$
$\alpha_3^{\#} = \alpha_3' + k_{31}\alpha_1^{\#} + k_{32}\alpha_2^{\#}$
$\beta_3^{\#} = \beta_1^{\#} + \beta_2^{\#} + \beta_3'$
$x_3' = x_3$
$\alpha_3' = x_3 k_{3,1} + x_3\beta_3 k_{3,2}$
$\beta_3' = x_3\beta_3 + \theta_3$
3
4
5
6
7
α_4
$\beta_4 = l_4\alpha_4$
α_5
$\beta_5 = l_5\alpha_5$
α_6
$\beta_6 = l_6\alpha_6$
α_7
$\beta_7 = l_7\alpha_7$
$x_1^{\#} = x_1$
$\alpha_1^{\#} = x_1 k_{1,1} + x_1\beta_1 k_{1,2}$
$\beta_1^{\#} = x_1\beta_1 + \theta_1$
$x_2^{\#} = x_2$
$\alpha_2^{\#} = x_2 k_{2,1} + x_2\beta_2 k_{2,2}$
$\beta_2^{\#} = x_2\beta_2 + \theta_2$
1
2
α_1
$\beta_1 = l_1\alpha_1$
α_2
$\beta_2 = l_2\alpha_2$

Figure 1: Aggregation phase. The nodes 1, 2, 3, 4, 5, 6, and 7 are sensor nodes, and the nodes 8, 9, and 10 are aggregators while node 3 works as both sensor node and aggregator. Without losing generality, we assume that every intermediate node is able to sense raw data and performs aggregation like node 3 does.

For each sensor node i, base station also calculates two path-keys $k_{i,1}$ and $k_{i,2}$ as follows:

$$k_{i,1} = \frac{\theta}{k_{\text{path}}}, \tag{3}$$

$$k_{i,2} = \frac{l}{k_{\text{path}}}, \tag{4}$$

where θ is a point in $E(Z_p)$, l is an integer and they are both chosen by base station to enable data aggregation and prevent stealthy/replay attacks.

If the path from base station to sensor node i is 1-2-3-i, then $k_{\text{path}} = k_{1\text{-}2}k_{2\text{-}3}k_{3\text{-}i}$.

Finally the base station with transmission ranges covering the whole wireless sensor network directly broadcasts to node i the path-keys and edge-keys encrypted with the secret key sk_i.

4.2.2. Data Aggregation Phase. In the query dissemination phase, each node has already identified their parents and the base station has the overall view of the aggregation tree.

In Figure 1, take paths BS-10-8-3-1 and BS-10-8-3-2, for instance, as sensor node, nodes 3, 1, and 2 each has a message, that is, to be passed to their parents. And the message has the following format:

$$\left\langle x_i^{\#}, \alpha_i^{\#}, \beta_i^{\#} \right\rangle, \tag{5}$$

where $x_i^{\#}$ is the SUM aggregation over all sensor nodes in the subtree; $\alpha_i^{\#}$ and $\beta_i^{\#}$ are the first and second tag, respectively.

For nodes 1 and 2:

$$\begin{gathered} x_1^{\#} = x_1, \qquad x_2^{\#} = x_2, \\ \alpha_1^{\#} = x_1 k_{1,1} + x_1\beta_1 k_{1,2}, \\ \alpha_2^{\#} = x_2 k_{2,1} + x_2\beta_2 k_{2,2}, \\ \beta_1^{\#} = x_1\beta_1 + \theta_1, \qquad \beta_3^{\#} = x_3\beta_3 + \theta_3. \end{gathered} \tag{6}$$

For aggregator/sensor node 3 with data x_3, it first computes α_3' and β_3' as a sensor node:

$$\begin{gathered} x_3' = x_3, \qquad \alpha_3' = x_3 k_{3,1} + x_3\beta_3 k_{3,2}, \\ \beta_3' = x_3\beta_3 + \theta_3. \end{gathered} \tag{7}$$

Then node 1 and 2 send their data and tags to node 3. After receiving all messages from its subtree, node 3 works as aggregator to perform aggregation:

$$
\begin{aligned}
x_3^{\#} &= x_1 + x_2 + x_3', \\
\alpha_3^{\#} &= k_{3\text{-}1}\alpha_1^{\#} + k_{3\text{-}2}\alpha_2^{\#} + \alpha_3', \\
\beta_3^{\#} &= \beta_1' + \beta_2' + \beta_3',
\end{aligned}
\tag{8}
$$

and sends $\langle x_3^{\#}, \alpha_3^{\#}, \beta_3^{\#} \rangle$ to node 8.

Aggregators 8 and 10 perform corresponding tasks:

$$
\begin{gathered}
x_8^{\#} = x_3 + x_4 + x_5, \qquad x_{10}^{\#} = x_8^{\#} + x_9^{\#}, \\
\alpha_8^{\#} = k_{8\text{-}3}\alpha_3^{\#} + k_{8\text{-}4}\alpha_4^{\#} + k_{8\text{-}5}\alpha_5^{\#}, \\
\alpha_{10}^{\#} = k_{10\text{-}8}\alpha_8^{\#} + k_{10\text{-}9}\alpha_9^{\#}, \\
\beta_8^{\#} = \beta_3^{\#} + \beta_4^{\#} + \beta_5, \qquad \beta_{10}^{\#} = \beta_8^{\#} + \beta_9^{\#},
\end{gathered}
\tag{9}
$$

where node 8 sends $\langle x_8^{\#}, \alpha_8^{\#}, \beta_8^{\#} \rangle$ to node 10, and node 10 sends $\langle x_{10}^{\#}, \alpha_{10}^{\#}, \beta_{10}^{\#} \rangle$ to base station.

4.2.3. Result-Checking Phase. The purpose of result-checking phase is to enable base station to verify that the integrity of SUM $x_{10}^{\#}$ has not been violated. The verification is performed as follows.

Base station checks if

$$
\beta_{10}^{\#} - l^{-1}k_{\text{bs-}10}\alpha_8^{\#} = \sum \theta_i \; - \; l^{-1}\theta x_{10}^{\#},
\tag{10}
$$

where $\langle x_{10}^{\#}, \alpha_{10}^{\#}, \beta_{10}^{\#} \rangle$ is sent by node 10 to base station; θ, l, $k_{\text{bs-}10}$, θ_i (i: from 1 to 7) are only known to base station and l^{-1} is the inverse of l modulo which is the order q of the elliptic curve group $E(Z_p)$.

Since

$$
\begin{aligned}
\beta_{10}^{\#} &= \beta_8^{\#} + \beta_9^{\#} \\
&= \beta_3' + \beta_4^{\#} + \beta_5^{\#} + \beta_6^{\#} + \beta_7^{\#} \\
&= \beta_1^{\#} + \beta_2^{\#} + \beta_3^{\#} + \beta_4^{\#} + \beta_5^{\#} + \beta_6^{\#} + \beta_7^{\#} \\
&= x_1\beta_1 + \theta_1 + x_2\beta_2 + \theta_2 + x_3\beta_3 + \theta_3 + x_4\beta_4 \\
&\quad + \theta_4 + x_5\beta_5 + \theta_5 + x_6\beta_6 + \theta_6 + x_7\beta_7 + \theta_7 \\
&= (x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5 + x_6\beta_6 + x_7\beta_7) \\
&\quad + (\theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6 + \theta_7), \\
l^{-1}&k_{\text{bs-}10}\alpha_{10}^{\#} \\
&= l^{-1}k_{\text{bs-}10}\left(k_{10\text{-}8}\alpha_8^{\#} + k_{10\text{-}9}\alpha_9^{\#}\right) \\
&= l^{-1}k_{\text{bs-}10} \\
&\quad \times \left(k_{10\text{-}8}\left(k_{8\text{-}3}\alpha_3^{\#} + k_{8\text{-}4}\alpha_4^{\#} + k_{8\text{-}5}\alpha_5^{\#}\right)\right. \\
&\quad \left. + k_{10\text{-}9}\left(k_{9\text{-}6}\alpha_6^{\#} + k_{9\text{-}7}\alpha_7^{\#}\right)\right) \\
&= l^{-1}k_{\text{bs-}10} \\
&\quad \times \left(k_{10\text{-}8}\left(k_{8\text{-}3}\left(\alpha_3' + k_{3\text{-}1}\alpha_1^{\#} + k_{3\text{-}2}\alpha_2^{\#}\right)\right.\right. \\
&\quad \left. + k_{8\text{-}4}\alpha_4^{\#} + k_{8\text{-}4}\alpha_5^{\#}\right) \\
&\quad \left. + k_{10\text{-}9}\left(k_{9\text{-}6}\alpha_6^{\#} + k_{9\text{-}7}\alpha_7^{\#}\right)\right) \\
&= l^{-1}k_{\text{bs-}10}k_{10\text{-}8}k_{8\text{-}3} \\
&\quad \times \left(k_{3\text{-}1}\left(\frac{l}{k_{\text{bs-}10}k_{10\text{-}8}k_{8\text{-}3}k_{3\text{-}1}}\right)x_1\beta_1\right. \\
&\quad + k_{3\text{-}2}\left(\frac{l}{k_{\text{bs-}10}k_{10\text{-}8}k_{8\text{-}3}k_{3\text{-}2}}\right)x_2\beta_2 \\
&\quad \left. + \left(\frac{l}{k_{\text{bs-}10}k_{10\text{-}8}k_{8\text{-}3}}\right)x_3\beta_3\right) \\
&\quad + l^{-1}k_{\text{bs-}10}k_{10\text{-}8} \\
&\quad \times \left(k_{8\text{-}4}\left(\frac{l}{k_{\text{bs-}10}k_{10\text{-}8}k_{8\text{-}4}}\right)x_4\beta_4\right. \\
&\quad \left. + k_{8\text{-}5}\left(\frac{l}{k_{\text{bs-}10}k_{10\text{-}8}k_{8\text{-}5}}\right)x_5\beta_5\right) \\
&\quad + l^{-1}k_{\text{bs-}10}k_{10\text{-}9} \\
&\quad \times \left(k_{9\text{-}6}\left(\frac{l}{k_{\text{bs-}10}k_{10\text{-}9}k_{9\text{-}6}}\right)x_6\beta_6\right. \\
&\quad \left. + k_{9\text{-}7}\left(\frac{l}{k_{\text{bs-}10}k_{10\text{-}9}k_{9\text{-}7}}\right)x_7\beta_7\right) \\
&= (x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5 + x_6\beta_6 + x_7\beta_7) \\
&\quad + l^{-1}\theta\left(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7\right).
\end{aligned}
\tag{11}
$$

We can say that base station accepts the SUM aggregation $x_{10}^{\#}$, or the two tags will only verify if all tags are generated by honest nodes and aggregated correctly along the path.

Again, note that l^{-1} is the inverse of l modulo q which is the order of the elliptic curve group $E(Z_p)$, and θ, l, $k_{\text{bs-}10}$ are only known to base station.

5. Analysis

This section discusses the security and congestion complexity of EIPDAP.

5.1. Overview. Once a node has been compromised, it is under the full control of the adversary which can record and inject messages as will. We also assume that the adversary can only corrupt a (small) fraction of nodes including the aggregators. Also, we do not concern denial-of-service attack. The following is the proof for security of EIPDAP.

Definition 2 (sensor node inconsistency). Let $\langle x_t^{\#}, \alpha_t^{\#}, \beta_t^{\#} \rangle$ be a message sent by sensor node t. There is an inconsistency at node t if either

(1) $\alpha_t^{\#} \neq x_t k_{t,1} + x_t \beta_t k_{t,2}$ or

(2) $\beta_t^{\#} \neq x_t \beta_t + \theta_t$.

Definition 3 (sensor node forgery). An adversary eavesdropping on sensor node i successfully forges a new message $\langle x_i^{*}, \alpha_i^{*}, \beta_i^{*} \rangle$ if

(1) $x_i^{*} \neq x_i^{\#}$,

(2) $\alpha_i^{*} = x_i^{*} k_{i,1} + x_i^{*} \beta_i k_{i,2}$,

(3) $\beta_i^{\#} = x_i^{*} \beta_i + \theta_i$.

Since once a sensor node is compromised, the adversary can obtain all its confidential information (e.g., cryptographic keys) and send false data without being detected; however, we do not address this kind of forgery here.

Lemma 4. *Let the final SUM aggregation received by the base station be $x_{final}^{\#}$, then $S_L + \mu \leq x_{final}^{\#} \leq S_L + \mu v$ where S_L is the sum of the data values of all the legitimate nodes, and μ is the total number of corrupted nodes.*

Proof. As the conclusion is obvious here, so we do not prove in detail. □

Lemma 5. *If elliptic curve discrete logarithm problem is hard, then it is not possible to forge a valid message as an honest sensor node for all eavesdropping probabilistic, polynomial adversaries.*

Proof. Let $\langle x_i^{\#}, \alpha_i^{\#}, \beta_i^{\#} \rangle$ be an internal message sent by sensor node i to its parent.

Say adversary is eavesdropping on node i. In order to forge a valid message $\langle x_i^{*}, \alpha_i^{*}, \beta_i^{*} \rangle$ for x_i^{*}, A can easily compute a valid $\alpha_i^{*} = x_i^{*} x_i^{\#-1} \alpha_i^{\#}$.

As $\beta_i^{\#} = x_i \beta_i + \theta_i = x_i y G$ for some integer y, a valid β_i^{*} should be computed as

$$\begin{aligned} \beta_i^{*} &= x_i^{*} \beta_i + \theta_i = x_i^{*} y^{*} G \\ &= x_i^{*} y^{*} x_i^{-1} y^{-1} \beta_i^{\#} G. \end{aligned} \tag{12}$$

Due to θ_i and β_i, adversary cannot forge β_i^{*} directly from $x_i^{*} \beta_i + \theta_i$ but to compute $x_i^{*} y^{*} x_i^{-1} y^{-1} \beta_i^{\#} G$. A has x_i^{*}, x_i, $\beta_i^{\#}$, and G; the factors it lack are y and y^{*}.

Calculating y from $\beta_i^{\#} = x_i y G$ and y^{*} from $\beta_i^{*} = x_i^{*} y^{*} G$ is ECDLP, which means calculating β_i^{*} from $x_i^{*} y^{*} x_i^{-1} y^{-1} \beta_i^{\#} G$ is hard.

Another concern rises when adversary keeps eavesdropping on sensor node i and records the messages sent by i. Assume that adversary has

$$\left\{ \left\langle x_{i\#j}^{\#}, \alpha_{i\#j}^{\#}, \beta_{i\#j}^{\#} \right\rangle \mid j \in [1, n] \right\}, \tag{13}$$

where $\langle x_{i\#j}^{\#}, \alpha_{i\#j}^{\#}, \beta_{i\#j}^{\#} \rangle$ represents the message $\langle x_i^{\#}, \alpha_i^{\#}, \beta_i^{\#} \rangle$, node i sends to parent in the jth query, and n is the number of queries. Note that in each query, every sensor node i chooses a new secret key l_i.

For all $j \in [1, n]$, adversary has

$$\beta_{i\#j}^{\#} = \beta_{i\#j} x_{i\#j}^{\#} + \theta_i. \tag{14}$$

Because the number of variable is the number of equations plus one, so adversary cannot solve equations in (14) to obtain $\beta_{i\#j}$ or θ_i.

In conclusion, the probability of an adversary successfully forging a new message $\langle x_i^{*}, \alpha_i^{*}, \beta_i^{*} \rangle$ when eavesdropping on sensor node i is negligible, completing the proof. □

Definition 6 (aggregator inconsistency). Let $\langle x_t^{\#}, \alpha_t^{\#}, \beta_t^{\#} \rangle$ be an internal message aggregated by node t with two children u and v. Let $\langle x_u^{\#}, \alpha_u^{\#}, \beta_u^{\#} \rangle$ and $\langle x_v^{\#}, \alpha_v^{\#}, \beta_v^{\#} \rangle$ be two messages from u and v. There is an inconsistency at node t if

(1) $x_t^{\#} \neq x_t' + x_u^{\#} + x_v^{\#}$ or

(2) $\alpha_t^{\#} \neq \alpha_t' + \alpha_u^{\#} + \alpha_t^{\#}$ or

(3) $\beta_t^{\#} \neq \beta_u' + \beta_u^{\#} + \beta_t^{\#}$.

Definition 7 (compromised aggregator forgery). An adversary which compromised a aggregator j successfully forges a new aggregate result $\langle x_j^{*}, \alpha_j^{*}, \beta_j^{*} \rangle$ if

(1) $x_j^{*} \neq x^{\#}{}_j$,

(2) $\alpha_j^{*} = \alpha_j' + k_{j-1} \alpha_{j1}^{\#} + k_{j-2} \alpha_{j2}^{\#} + \cdots + k_{j-l}\ \alpha_{jl}^{\#}$ ($\alpha_j' = 0$ if j does not sense data),

(3) $\beta_j^{\#} = \beta_j' + \beta_{j1}^{\#} + \beta_{j2}^{\#} + \cdots + \beta_{jl}^{\#}$ ($\beta_j' = 0$ if j does not sense data), assuming aggregator j has l children $j1, j2, \ldots, jl$.

Lemma 8. *If elliptic curve discrete logarithm problem is hard, then it is not possible to forge a valid aggregate result for all probabilistic, polynomial adversaries even when a high-level aggregator is compromised.*

Proof. We assume that aggregator 10 has been compromised where an adversary is in complete control of node 10, obtaining all secret keys of node 10. Now an adversary attempts to forge SUM aggregation and two corresponding tags after eavesdropping several aggregations and records

$$\left\{ \left\langle x_{10\#i}^{\#}, \alpha_{10\#i}^{\#}, \beta_{10\#i}^{\#} \right\rangle \mid i \in [1, n] \right\}, \tag{15}$$

where $\langle x_{10\#i}^{\#}, \alpha_{10\#i}^{\#}, \beta_{10\#i}^{\#} \rangle$ represents the message $\langle x_{10}^{\#}, \alpha_{10}^{\#}, \beta_{10}^{\#} \rangle$, node 10 sends to base station in the ith query, and n is the number of queries. Note that in each query, every sensor node j chooses a new secret key l_j.

For all $i \in [1, n]$, adversary has

$$\begin{aligned} \beta_{10\#i}^{\#} &= l^{-1} k_{\text{bs-10}} \alpha_{10\#i}^{\#} \\ &= \left(\theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6 + \theta_7 \right) \\ &\quad - l^{-1} \theta x_{10\#i}^{\#}. \end{aligned} \tag{16}$$

Now adversary tries to forge a new message $\langle x_{10}^{*}, \alpha_{10}^{*}, \beta_{10}^{*}\rangle$ which satisfies (14). Since node 10 is compromised, adversary has $x_8^{\#}, x_9^{\#}, x_{10}^{\#}, k_{10\text{-}8}, k_{10\text{-}9}, \alpha_8^{\#}, \alpha_9^{\#}, \beta_8^{\#}, \beta_9^{\#}$ in each query and the knowledge of the elliptic curve group $E(Z_p)$.

Case 1. Intuitively, adversary tries to obtain l, $k_{\text{bs-10}}$, θ, and $\sum \theta_i$ (i: from 1 to 7). However, this requires a powerful adversary which we do not concern here.

Case 2. Adversary tries to compute $l^{-1}\theta$ by multiplying $k_{i,1}$ (equals θ/k_{path}) and the inverse of $k_{i,1}$ (equals l/k_{path}). However, all path-keys and temporal key are encrypted before forwarding to nodes, so adversary cannot compute $l^{-1}\theta$ when node 10 only works as an aggregator.

Case 3. Aggregator 10 also senses data. So adversary can compute $l^{-1}\theta$ as in Case 2 and $l^{-1}k_{\text{bs-10}}$ which is the inverse of $l/k_{\text{bs-10}}$ modulo q. Now adversary has

$$\begin{aligned}&\beta_{10}^{\#} - \alpha_{10}^{\#\#}\\&\quad = \theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6 + \theta_7 - x_{10}^{\#\#}\\&\Longrightarrow \beta_{10}^{\#\#}\\&\quad = y_1\left(\theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6 + \theta_7\right) = y_2 G\end{aligned} \tag{17}$$

for some integer y_2. Similar to Lemma 4, since $\sum \theta_i$ (i: from 1 to 7) is kept secret from adversary and computing y_2 from $\beta_3^{\#\#} = y_2 G$ is ECDLP, then $\beta_{10}^{\#\#}$ cannot be forged either.

In all cases, adversary can only forge a new message $\langle x_{10}^{*}, \alpha_{10}^{*}, \beta_{10}^{*}\rangle$ with negligible probability, completing the proof. □

Theorem 9. *EIPDAP is integrity-preserving.*

Proof. From Lemmas 5 and 8, we know that EIPDAP is secure against sensor node forgery in the presence of an eavesdropper and aggregator forgery when an aggregator is compromised. Thus, EIPDAP is integrity-preserving, completing the proof. □

5.2. Congestion Complexity. The computational and memory costs are likely to be insignificant compared to communication [3, 14]. Higher computation surely causes more energy, but a Berkeley mote spends approximately the same amount of energy to compute 800 instructions as it does in sending a single bit of data [13, 20] in WSN.

Unlike general hard problems, there is no subexponential algorithm is known to solve the elliptic curve discrete logarithm problem (ECDLP), meaning that smaller parameters can be used in ECC than in other systems like RSA and DSA but with equivalent level of security. Because of their smaller key size, faster computations and reductions in processing power, storage space, and bandwidth, ECC is ideal for WSN. Although the use of elliptic curve cryptography incurs higher computational overhead than symmetric-key cryptography, our protocol is mainly designed to save energy.

In query dissemination phase, the base station collects aggregation tree information and broadcasts edge keys and path keys directly to the corresponding nodes. Collecting aggregation tree information costs each edge $O(1)$ congestion, and there is no congestion for sensor nodes and aggregators in broadcasting keys. In aggregation phase, each node forwards a message. The edge congestion in the aggregation tree is $O(1)$. In result-checking phase, all operations are done in the base station, so there is no congestion in the aggregation tree. Congestion complexity comparisons with Chan's scheme, Keith's scheme, and SDAP are shown in Tables 1, 2, and 3.

Table 1: Edge congestion in the aggregation tree comparison, n is the number of the nodes, and n_g is the group size.

	Query dissemination	Data aggregation	Result-checking
Chan's scheme	$O(1)$	$O(\log n)$	$O(\log^2 n)$
Keith's scheme	$O(1)$	$O(\log n)$	$O(\log n)$
SDAP	$O(1)$	$O(\Delta \log (n/n_g))$	$O(\Delta \log (n/n_g))$
EIPDAP	$O(1)$	$O(1)$	0

Table 2: Node congestion in the aggregation tree comparison, Δ is the degree of the aggregation tree.

	Query dissemination	Data aggregation	Result-checking
Chan's scheme	$O(\Delta)$	$O(\Delta \log n)$	$O(\Delta \log^2 n)$
Keith's scheme	$O(\Delta)$	$O(\Delta \log n)$	$O(\Delta \log n)$
SDAP	$O(\Delta)$	$O(\Delta \log (n/n_g))$	$O(\Delta \log (n/n_g))$
EIPDAP	$O(\Delta)$	$O(\Delta)$	0

Table 3: Aggregation tree congestion comparison.

	Query dissemination	Data aggregation	Result-checking
Chan's scheme	$O(n)$	$O(n \log n)$	$O(n \log^2 n)$
Keith's scheme	$O(n)$	$O(n \log n)$	$O(n \log n)$
SDAP	$O(n)$	$O(n)$	$O(n \log n)$
EIPDAP	$O(n)$	$O(n)$	0

By the comparison, we can conclude that EIPDAP has the minimum congestion and is much more energy efficient. Therefore it is much more suitable for power limited sensor networks.

6. Conclusion and Future Work

Protecting hierarchical data aggregation from losing integrity is a challenging problem in sensor networks. In this paper, we focus on the very problem of preserving data integrity and propose a novel approach to guarantee the integrity of aggregation result through aggregation in sensor networks. The main algorithm is based on performing modulo addition operation using ECC.

EIPDAP can immediately verify the integrity of aggregation result after receiving the aggregation result and corresponding authentication information, hence significantly reducing energy consumption and communication delay which will be caused if the verification phrase is done through another query-and-forward phase.

Compared with the other related schemes, our scheme reduces the communication required per node to $O(\Delta)$, where Δ is the degree of the aggregation tree for the network. To the best of our knowledge, our scheme has the most optimal upper bound on solving the integrity-preserving data aggregation problem. Based on the elliptic curve discrete logarithm problem, we prove that EIPDAP is integrity-preserving.

In the future, we will first further enrich EIPDAP in detail. Second, we will focus on the possibility of reducing the number of secret keys shared between sensor nodes and base station or the keys broadcast to all nodes. Third, based on the proposed algorithm, we may consider meeting other security requirements, like data confidentiality, source authentication, and availability.

We anticipate that our work provides new perspective on preserving integrity of hierarchical aggregation and encourages other researchers to consider this approach.

Appendix

Attack on the Julia-Sanjay Scheme

If the adversary has compromised a sensor node, then it can obtain the network wide integer k. With the k, it can modify any aggregated data received from its child nodes. For example, say the adversary has compromised a node with message $\langle s_i, \mathrm{enc}(m_i)\rangle$ received from its child i. In order to modify m_i to $m_i' = m_i + m_{\mathrm{forge}}$, the adversary can easily forge an encrypted message:

$$\mathrm{enc}\left(m_i'\right) = \mathrm{enc}\left(m_i + m_{\mathrm{forge}}\right) = \left\langle u_i, v_i + m_{\mathrm{forge}}\right\rangle, \quad \text{(A.1)}$$

given as $e\mathrm{nc}(m_i) = \langle u_i, v_i\rangle$. And it is also easy to forge a valid signature:

$$s_i' = k^{-1}\left(m_i + m_{\mathrm{forge}} + z_i * r\left(x\right)\right) \bmod p. \quad \text{(A.2)}$$

If the compromised node is in high level, this will cause more serious effects on the aggregation result since the aggregate result it handles represents large portions of the overall data in the WSN.

Acknowledgments

The authors would like to thank the anonymous reviewers and their coworkers for their valuable comments and useful suggestions.

References

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

[2] K. Akkaya, M. Demirbas, and R. S. Aygun, "The impact of data aggregation on the performance of wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 8, no. 2, pp. 171–193, 2008.

[3] L. Hu and D. Evans, "Secure aggregation for wireless networks," in *Proceedings of the Workshop on Security and Assurance in Ad Hoc Networks*, Orlando, Fla, USA, January 2003.

[4] B. Przydatek, D. Song, and A. Perrig, "SIA: Secure information aggregation in sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys '03)*, pp. 255–265, November 2003.

[5] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "Pda: privacy-preserving data aggregation in wireless sensor networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 2045–2053, 2007.

[6] T. Feng, C. Wang, W. Zhang, and L. Ruan, "Confidentiality protection for distributed sensor data aggregation," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '07)*, pp. 475–483, April 2007.

[7] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "PriSense: privacy-preserving data aggregation in people-centric urban sensing systems," in *Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM '10)*, pp. 758–766, IEEE, March 2010.

[8] J. Katz and A. Y. Lindell, "Aggregate message authentication codes. Topics in cryptology," in *Proceedings of the Cryptographers' Track at the RSA Conference (CT-RSA '08)*, Lecture Notes in Computer Science, pp. 155–169, Springer, 2008.

[9] J. Albath and S. Madria, "Secure hierarchical data aggregation in wireless sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '09)*, pp. 1–6, April 2009.

[10] A. Liu and P. Ning, "TinyECC: a configurable library for elliptic curve cryptography in wireless sensor networks," in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN '08)*, pp. 245–256, April 2008.

[11] H. Chan, A. Perrig, and D. Song, "Secure hierarchical in-network aggregation in sensor networks," in *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS '06)*, pp. 278–287, ACM Press, Alexandria, Va, USA, November 2006.

[12] K. B. Frikken and J. A. Dougherty, "An efficient integrity-preserving scheme for hierarchical sensor aggregation," in *Proceedings of the 1st ACM Conference on Wireless Network Security (WiSec '08)*, pp. 68–76, ACM Press, Alexandria, Va, USA, April 2008.

[13] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: a Secure hop-by-hop Data Aggregation Protocol for sensor networks," in *Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC '06)*, pp. 356–367, ACM Press, Florence, Italy, May 2006.

[14] C. Castelluccia, A. C. F. Chan, E. Mykletun, and G. Tsudik, "Efficient and provably secure aggregation of encrypted data in wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 5, no. 3, pp. 1–36, 2009.

[15] W. Du, J. Deng, Y. S. Han, and P. K. Varshney, "A witness-based approach for data fusion assurance in wireless sensor networks," in *Proceedings of the Global Telecommunications Conference (GLOBECOM '03)*, vol. 3, pp. 1435–1439, IEEE, December 2003.

[16] O. Eikemeier, M. Fischlin, J.-F. Götzmann et al., "History-free aggregate message authentication codes," in *Proceedings of the 7th International Conference on Security and Cryptography for Networks*, vol. 6280 of *Lecture Notes in Computer Science*, pp. 309–328, Amalfi, Italy, 2010.

[17] J. Girao, D. Westhoff, and M. Schneider, "CDA: concealed data aggregation in wireless sensor networks," in *Proceedings of the ACM Workshop on Wireless Security (WiSe '04)*, ACM Press, Philadelphia, Pa, USA, 2004.

[18] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous Systems -Networking and Services (MobiQuitous '05)*, pp. 109–117, IEEE Computer Society, San Diego, Calif, USA, July 2005.

[19] H. Çam, S. Özdemir, P. Nair, D. Muthuavinashiappan, and H. Ozgur Sanli, "Energy-efficient secure pattern based data aggregation for wireless sensor networks," *Computer Communications*, vol. 29, no. 4, pp. 446–455, 2006.

[20] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a tiny aggregation service for ad-hoc sensor networks," *SIGOPS Operating Systems Review*, vol. 36, pp. 131–146, 2002.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 416983, 13 pages
http://dx.doi.org/10.1155/2013/416983

Research Article

Incentive-Based Optimal Nodes Selection Mechanism for Threshold Key Management in MANETs with Selfish Nodes

Yuanbo Guo,[1,2] Jianfeng Ma,[3] Chao Wang,[3] and Kuiwu Yang[1]

[1] *State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450004, China*
[2] *Science and Technology on Information Assurance Laboratory, Beijing 100072, China*
[3] *Computer Science School of Xidian University, Xi'an 710071, China*

Correspondence should be addressed to Yuanbo Guo; yuanbo_g@hotmail.com

Received 28 February 2013; Revised 30 March 2013; Accepted 8 April 2013

Academic Editor: Liangmin Wang

Most of the previous work on threshold-cryptography-based distributed CA concentrates on the initial systems configurations and concrete protocols design, ignoring the efficiency and effectiveness of the key management service during its operation, and always assuming that there are honest nodes to carry out the service faithfully. This paper focuses on developing a selection mechanism in MANETs with selfish nodes, to dynamically select a coalition of nodes carrying out the threshold key management service optimally during system operation. First, we formulate the dynamic nodes selection problem as a combinatorial optimization problem, with the objectives of maximizing the success ratio of key management service and minimizing the nodes' cost of security and energy. Then, to ensure truth telling is the dominant strategy for any node in our scenario, we extend the payment structure of the classical Vickrey-Clarke-Groves (VCG) mechanism design framework and divide the payment into pieces to the nodes, with the consideration of their actual execution effectiveness. Simulations show that the proposed mechanism enjoys improvements of both the success ratio of key management service and lifetime of the network, as well as reductions of both the cost of participating nodes and compromising probability of MANETs, compared with the existing work.

1. Introduction

A mobile ad hoc network (MANET) is a network consisting of a collection of nodes capable of communicating without relying on a fixed infrastructure and is characterized by some of the features like lacking infrastructure, dynamic network topology, distributed operation, variable capacity links, use of low power devices, and so forth. This makes ad hoc networks financially viable and have tremendous potential for communications in battlefields, disaster recovery areas, and other environments such as collaborative computing and communications in smaller areas. For MANETs, public-key cryptography (PKC) is appealing in offering security support, due to its effectiveness in facilitating essential security services such as digital signatures and key management. However, the traditional public key infrastructure (PKI) supporting key management approaches require a global trusted certificate authority (CA) to manage public key certificates used to generate confidence in the legitimacy of public keys for the nodes of the network. This makes it difficult to deploy the PKI in MANETs, since this type of networks does not have any form of online or offline authority [1]. Even if the service node can be defined to act as an authority, maintaining such a centralized server and keeping its security and availability in such a dynamic network is a difficult task. Key management for MANETs therefore needs to mitigate the unreliability of basic CA services by taking on a distributed, self-organizing nature [2–9].

However, previous work on this subject mainly concentrates on the initial systems configuration and concrete protocols design of distributed CA itself and ignores the problem of how to select a threshold number of nodes from the set of all partial certificates during its operation with the consideration of attributes of all nodes in the network. Instead, a random selection scheme is often assumed or implicated. To the best of our knowledge, the only paper addressing the problem of optimal nodes selection for threshold key management in MANETs is [10], where the dynamic nodes selection process is formulated as a multiarm bandit problem. Then, an optimal selection scheme is proposed to select the best nodes to be used as private key generators (PKGs) from all available ones with the consideration of their security conditions and energy states. This scheme has nice features of decreasing network compromising probability and increasing network lifetime in MANETs.

There are still some problems suffering from the existing schemes, including the one proposed in [10]: (i) they do not consider the effectiveness of the key management. Given a crypto threshold k, more than k correct replies from nodes make a key management service successful. The success ratio must be kept at a high level under all circumstances to provide useful and effective key management services; (ii) they always assume that the nodes in MANETs cannot act rationally and strategically (i.e., each node follows the protocol specification by assumption).

In this paper, we present incentive compatible optimal nodes selection (ICONS), a mechanism which dynamically implements the optimal nodes selection for threshold key management based on the nodes' security and energy states truthfully in dominant strategies. Specifically, we formulate the dynamic nodes selection problem as combinatorial optimization problem [11], by combining two objectives of maximizing the success ratio of key management service and minimizing the nodes' cost of security and energy into a single weighted objective firstly. And then we extend the classical Vickrey-Clarke–Groves- (VCG-) [12] based mechanism design framework [13] to allow for implementing an objective function which is not quasi-linear and then divide the payment into pieces to the nodes according to their outcomes at current stage. The proposed mechanism not only enjoys the same nice features as the scheme in [10] (decreasing network compromising probability, lowering the energy cost, and prolonging network lifetime) but also achieves more performance benefits of increasing the success ratio of key management service and allowing the nodes in MANETs to remain truthful in the scenario where they act rationally and selfishly.

The rest of the paper is organized as follows. The next section reviews related work. Section 3 formulates the optimal nodes selection model for threshold key management in MANETs. Section 4 presents the incentive compatible optimal nodes selection mechanism and proves its correctness and truthfulness. The performance of our model is evaluated via detailed simulations in Section 5. Finally, this paper is concluded and discussed in Section 6.

2. Related Work

In this section, we review the related work in threshold-cryptography-based distributed CA (DCA) and mechanism design application in MANETs.

2.1. Threshold-Cryptography-Based DCA in MANETs. A DCA is realized through the distribution of the CA's private key to a number of shareholding nodes. The design of a DCA based on threshold cryptography is suggested in [14] firstly and then applied to solving the key management problem in MANETs in [2] by letting a set of nodes in the network share the system secret. From then, many DCA schemes in MANET have been proposed, which can be classified as partially or fully distributed certificate authorities [15].

In partially implemented DCA, services of the CA are distributed to a set of specialized server nodes using secret sharing. Each of these nodes can generate partial certificates and a user can create a valid certificate by combining enough number of these partial certificates. In [5], a cluster-based partially DCA architecture in MANETs is established. First, a cluster head assisted CA locating scheme is proposed to shift the responsibility of CA discovery from each user node to cluster heads, which greatly reduces service response time and system overhead. Then, a share update procedure is also proposed to resolve the multiple initializations problem and achieves fast systemwide update. The authors of [6] propose a partially distributed certificate management mechanism that can handle mobility of nodes for MANET. The mechanism segregates the roles of certification authority to keep with the dynamic mobility of nodes and handle rapid and random topological changes with minimal overhead. The mobile certificate authority (MOCA) key management framework is proposed in [7] based on threshold cryptography to provide authentication service for MANETs. MOCA utilizes a carefully selected set of mobile nodes to function as a collective certificate authority while the MOCA nodes are kept anonymous. Equipped with a novel routing protocol designed to support the unique communication pattern for certification traffic, MOCA achieves high availability key management and authentication service with intuitive metrics to measure the provided quality of service. Then, the authors of [16] extend the MOCA framework by proposing and evaluating a key management scheme that suits the dynamic nature of an ad hoc network. To enhance the robustness and security of the threshold key management scheme, the authors of [4] propose a secure and robust key management scheme (SRKM) based on threshold cryptography, making it more difficult for mobile adversaries to violate the secrecy of the private key of certification service, even if they compromise more than a threshold number of nodes.

In fully distributed CA, services of a CA are distributed to all nodes using secret sharing, and each of these nodes can generate partial certificates. Since almost all the neighbors of a requesting node hold shares of the DCAs private signature key, fully distributed CA reduces the communication delay and improves the availability. The authors of [8] distribute the functionality of conventional security servers, specifically the authentication services, so that each individual node can potentially provide certification services for other nodes in MANETs. Centralized management is minimized and the nodes in the network collaboratively self-secure themselves. Then, the authrs of [17] propose a modification to the scheme in [8] to make it suitable for a mobile ad hoc network in which forming a coalition of a large number of nodes is often difficult. The concept of redundancy in key shares is introduced to increase the probability of recreating the CA key for a node in a highly mobile network, by allocating more than one share to each node. In [3], a scheme called autonomous key management (AKM) is proposed to provide a self-organizing and fully distributed key management service, which uses hierarchical structure to ensure flexibility and adaptability and uses verifiable secret sharing (VSS) to resist active attacks. The authors of [9] propose a fully distributed trust model based on trust graph for mobile ad hoc networks, where nodes have a similar role and do not need to assign any special functions to a subset of nodes. This scheme allows users to fully control the security settings in the network and allows nodes to generate, store, and distribute their public key certificates without any central server or trusted party. The scheme is developed for open networks, in which nodes can join/leave the network without any centralized administration. The joining operation is performed by a coalition of member nodes to allow access to a new node.

2.2. Application of Mechanism Design. Mechanism design is the subfield of microeconomics and game theory that considers how to implement an optimal systemwide solution to problems that involve multiple self-interested players, each with private information about their preferences for outcomes [13]. It is a useful and powerful tool to design protocols in the environment where the players may deviate the given protocol specification if it is beneficial for them to do so, and has been used extensively in MANETs environments.

The work in [18] proposes ad hoc VCG, a reactive routing protocol for MANETs that is robust against individual selfishness of the communication nodes and achieves cost-efficiency and truthfulness. This scheme works well in the MANETs environment, where the communication nodes are assumed to be selfish and need to declare their cost of energy in order to compute a cost-efficient communication path. Following this approach, the authors of [19] present low overhead truthful routing protocol (LOTTO), a low-overhead truthful routing protocol for route discovery in MANETs with selfish nodes by designing incentives based on VCG mechanism [12], to prevent nodes from revealing fake information and ensure truth telling to be the dominant strategy among all nodes. In [20], a mechanism-design-based model is proposed to motivate nodes that do not belong to the confident community to participate in being selected as RA, by giving them incentives in the form of trust. An RA selection algorithm is also proposed in this paper to select nodes based on a predefined selection criteria function and nodes location. In [21], a novel surveillance mechanism is proposed to observe the packet-dropping behavior of suspicious insiders. It quantifies the threat level of the suspicious insiders and then realizes an incentive-compatible surveillance scheme to motivate the rational monitors to cooperate, by rewarding the cooperating monitors and punishing the violating monitors. The authors of [22] study the leader election in the presence of selfish nodes for intrusion detection in MANETs and propose an integrated solution for prolonging the lifetimes of mobile nodes and prevent the emergence of selfish nodes. Reputations are computed in [22] also by using the well-known VCG [12] mechanism design as a theoretical tool.

These existing studies clearly show that mechanism design becomes prevalent in many engineering applications in MANETs. It provides a rich set of mathematical tools and models to motivate the nodes to reveal truthfully their selection criteria function. However, there is no much work on applying mechanism design theory to threshold key management in MANETs, where the success of the key management task is highly dependent upon the distributed collaboration of a coalition of rational and selfish nodes.

3. Optimal Nodes Selection Model

3.1. System Models. MANET, in this study, is represented by an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N} = \{1, 2, \ldots, N\}$ is the set of wireless nodes including a leader, and $\mathcal{L}$ is the set of communication links between the nodes. To keep the security and energy information current, we divide the time axis into stages, which correspond to the time intervals between two continuous key management tasks of the MANET. The stages are indexed by the integers, so in each stage t only one threshold key management task can be completed by a coalition of nodes cooperatively.

At each stage, the leader takes the role of selecting the best nodes from $\mathcal{N}$ to act as server nodes based on the security and energy states of each node, and pays each selected node according to its completion of assigned task or not. Each server node has its own share of the private CA key and participates in the process of threshold key management in the current stage. Sometimes, we also call a server node a active node and call a nonserver node a passive node. The leader is not necessarily a special node in MANET. Instead, it can be elected dynamically from the nodes set $\mathcal{N}$ by a leader election algorithm. The aim of a leader election algorithm is to ensure that a suitable node in a network will be selected as the leader to perform a task whenever needed [23]. Since in this paper we mainly focus on developing an incentive compatible optimal nodes selection mechanism, to encourage each node in the system to be truth-telling, the details of the leader election are out of the scope here. There are several leader election researches that have been done for MANETs and wireless sensor networks [22, 24, 25].

3.1.1. Security Model. Denote by s_i^t the security state of a node $i\,(i \in \mathcal{N})$ at the stage t. Security state represents the security condition of a node and can be monitored by node itself with a local intrusion detection system (IDS). Assume each node i has a finite number of I_i states in security state space $\mathcal{S}$, and each security state s_i^t evolves according to an I_i-state Markov chain with one-step transition probability matrix as follows:

$$A_i^a = \left(\phi_{jk}\right)_{j,k\in I_i} = \Pr\left(s_i^{t+1} = k \mid s_i^t = j\right), \qquad (1)$$

where $a \in \{0, 1\}$ stands for an action. Action 1 means that the node is active, and action 0 means that the node is passive.

3.1.2. Energy Model. We represent each node i's energy state at stage t as e_i^t, which can also be detected locally. We assume that the continuous battery residual energy can be divided into discrete levels, denoted by $\mathcal{E} = (e_1, e_2, \ldots, e_h)$. To foresee the energy consumption at the current stage, we model the transition of energy levels of nodes in MANETs as a Markov chain with one-step transition probability matrix [26] as follows:

$$B_i^a = \left(\psi_{jk}\right)_{j,k\in\varepsilon} = \Pr\left(e_i^{t+1} = k \mid e_i^t = j\right). \qquad (2)$$

3.1.3. Node States Model. Note that both security state s_i^t and energy state e_i^t are independent of each other, so we can model the state of each node i in stage t as

$$\pi_i^t = \left(s_i^t, e_i^t\right). \tag{3}$$

The state set of π_i^t is represented as Π_i and we have $|\Pi_i| = |\mathcal{S}| \times |\mathcal{E}|$. The state π_i^t evolves with one-step transition probability matrix as follows:

$$P_i^a = \left[A_i^a \otimes B_i^a\right], \tag{4}$$

where A_i^a is security state transition probability matrix, B_i^a is energy states transition probability matrix, and $\otimes$ denotes the Kronecker product.

3.1.4. Cost Model. The costs associated with each node i at stage t are defined as security cost $\mu_l(s_i^t, a_i^t)$, from a potential compromise of the node, and the energy cost $\mu_e(e_i^t, a_i^t)$. The action adopted by node i at stage t is denoted as $a_i^t \in \{0,1\}$. Then, the instantaneous cost of node i at stage t is

$$z_i^t\left(s_i^t, e_i^t, a_i^t\right) = (1-\gamma)\cdot\mu_l\left(s_i^t, a_i^t\right) + \gamma\cdot\mu_e\left(e_i^t, a_i^t\right), \tag{5}$$

where $\gamma \in (0,1)$ is the weight factor for the two kinds of costs and could be adjusted according to circumstances. We then extend the cost model to more realistic settings by considering the network lifetime as follows:

$$z_i^t\left(s_i^t, e_i^t, a_i^t\right) = \frac{(1-\gamma)\cdot\mu_l\left(s_i^t, a_i^t\right) + \gamma\cdot\mu_e\left(e_i^t, a_i^t\right)}{e_i^t}, \tag{6}$$

where e_i^t is node i's energy level at stage t.

Since just one message is needed for a passive node to report its states to the leader, the cost of a passive node can be assumed as a constant in a given stage t. Then, the node i's cost at stage t can be denoted as

$$c_i^t\left(s_i^t, e_i^t\right) = \begin{cases} z_i^t\left(s_i^t, e_i^t, 1\right) & \text{if } i \text{ is a active node} \\ C^t & \text{if } i \text{ is a passive node,} \end{cases} \tag{7}$$

where constant C^t is the cost of a messages transmission in stage t.

If there are m active nodes at stage t (when (m, N) secret sharing is used in the threshold key management scheme), then the cost of all the nodes for key management is

$$Z(t) = \sum_{i=1}^{N} c_i^t\left(s_i^t, e_i^t, 1\right) = \sum_{j=1}^{m} z_{i_j}^t\left(s_{i_j}^t, e_{i_j}^t, 1\right) + (N-m)\cdot C^t, \tag{8}$$

where $\{i_1, i_2, \ldots, i_m\} \subset \mathcal{N}$ denotes the set of all active nodes at stage t.

3.1.5. System Value Model. Key management service consists of a set of tasks and procedures supporting the establishment and maintenance of keying relationships between authorized parties [27], such as new node authentication and admission, generation and distribution keying material, update/revocation/destruction of keying material, bootstrapping, and maintenance of trust in keying material. Without loss of generality, we assume that the key management task at each stage t has a certain value V^t to the system, which is determined by the expected gain that system could gain from successful completion of this task. For example, the task of joint authentication has a higher gain for the system to admit a new node to join the network than that of joint session key establishment for two existing peer nodes, because joint authentication might enlarge the network scale. So, the former can be assigned a value of 200 while the latter just be assigned 120.

In the presence of attacks, an active node may fail to complete its assigned task of key management to act as a server node. Let $p^t = \{p_1^t, p_2^t, \ldots, p_N^t\}$ be the vector of task completion at stage t, where p_i^t $(0 \le p_i^t \le 1)$ is node i's success ratio to fulfill its assigned task at stage t, and if there are m active nodes $\{i_1, i_2, \ldots, i_m\}$ which are selected by the leader to cooperatively complete the key management task at this stage, then the expected system value will be

$$\mathbb{E}\left(V^t\right) = \left(V^t \cdot \prod_{j=1}^{m} p_{i_j}^t\right). \tag{9}$$

We assume that there is a map $f : \mathcal{S} \to [0,1]$, defined from the security state s_i^t of each node i to the success ratio p_i^t; that is, $p_i^t = f(s_i^t)$.

3.2. Optimal Selection Model. We denote by $\mathcal{U}$ the class of all admissible nodes selection policies. The admissible policy $u \in \mathcal{U}$ is a $T \times N$ matrix, whose element of the tth row and the ith column is a_i^t, representing the action taken by node i in stage t. The optimal nodes selection policy u^* is the policy that achieves the system objective.

3.2.1. Cost. The total cost of system at stage t is defined in (8), and the optimization objective is to find the optimal policy u'^* to minimize this cost as

$$\begin{aligned} Z^* &= \min_{u'^* \in \mathcal{U}} Z\left(u'^*\right) \\ &= \min_{u'^* \in \mathcal{U}} \left[\sum_{a_i'^t = 1} z_i^t\left(s_i^t, e_i^t, 1\right) + (N-m)\cdot C^t\right] \\ &= \min_{u'^* \in \mathcal{U}} \left[\sum_{a_i'^t = 1} z_i^t\left(s_i^t, e_i^t, 1\right)\right] + (N-m)\cdot C^t. \end{aligned} \tag{10}$$

3.2.2. System Value. The expected value of system at stage t is defined in (9), and the optimization objective is to find the optimal policy u''^* to maximize this value as

$$V^* = \max_{u''^* \in \mathcal{U}} V\left(u''^*\right) = \max_{u''^* \in \mathcal{U}} \left[V^t \cdot \prod_{a_i''^t = 1} p_i^t\right]. \tag{11}$$

3.2.3. Optimal Nodes Selection Policy. Now, we have two important but conflicting objectives: minimizing the expected system cost (10) and maximizing the expected system value (11), and both have their own optimal policy u'^{*} and u''^{*}.

Hence, there is an intrinsic tradeoff between cost minimization and system value maximization. By introducing a new system parameter $\rho \in (0,1)$ to combine these two objective functions together into a single objective function and using the weighting method [28], we formulate this multiobjective programming problem as a combinational optimization problem as follows:

$$u^{*} = \arg\max_{u\in\mathcal{U}} \left[\rho \cdot V^{t} \cdot \prod_{a_i^t=1} p_i^t - (1-\rho) \cdot \sum_{a_i^t=1} z_i^t \left(s_i^t, e_i^t, 1 \right) \right], \tag{12}$$

where the value of weight factor $\rho \in (0,1)$ can be set according to the application. For example, ρ can be set to a value close to 1 in a battlefield MANET, in order to reflect the fact that the improvement of the success ratio of the key management task is more important than the reduction of the cost in the battlefield network. By contrast, in a civilian MANET, ρ can be set to 0 to reflect the fact that the reduction of the costs is more important than the improvement of the success ratio of the key management task in the civilian network.

Since $\rho \in (0,1)$, we have $(1-\rho) \in (0,1)$ and so we rewrite (12) as follows:

$$u^{*} = \arg\max_{u\in\mathcal{U}} \left[\left(\frac{\rho}{1-\rho} \right) \cdot V^{t} \cdot \prod_{i\in u} p_i^t - \sum_{i\in u} z_i^t \right]. \tag{13}$$

Then the coefficient $(\rho/(1-\rho))$ can be omitted for simplicity, and we have

$$u^{*} = \arg\max_{u\in\mathcal{U}} \left(W^{t} \cdot \prod_{i\in u} p_i^t - \sum_{i\in u} z_i^t \right). \tag{14}$$

By this we do not lose generality because when ρ is given, we can substitute "W^{t}" for "$(\rho/(1-\rho)) \cdot V^{t}$", so that the coefficient of the new variation is equal to 1.

4. Incentive Compatible Optimal Nodes Selection Mechanism Design

As stated before, mechanism design [13, 29] is concerned with the situation where a policy maker faces the problem of aggregating the individual preferences into a collective decision and the individuals' actual preferences are not publicly known and studies how to elicit this privately held information and how the information revelation problem constrains the way in which social decisions can respond to individual preferences. To implement optimal nodes selection objective defined in (14), we apply game-theoretic approach to mechanism design and formulate the nodes selection process at each stage t as a game where N mobile nodes in the MANET are the players. Based on this model, we can design incentives to encourage each node in revealing its true information and honestly participate in the threshold key management process.

We assume that all nodes in the MANET are owned by rational and strategically selfish individuals, whose objectives are to maximize their individual goals. For this reason, these nodes may not always participate honestly in threshold key management, since this might cause security compromising and consume the nodes' resources, including battery power, bandwidth, and CPU cycles. But as discussed in Section 3, the leader here can take the role of nodes selection and reputations payment loyally. In this study, we just deal with the battery power consumption and security compromising, but our model can be extended to include more general cases straightforward.

4.1. The Mechanism. In each stage t, the leader initiates the game by asking each node, including itself, to reveal its type. Then, each node i plays game by revealing its own private information based on its strategy ξ_i^t drawn from an available strategy set $\Xi = \{\mathit{Truth}, \mathit{Untruth}\}$, according to how much the node values its utility with a utility function. If the node's strategy is "*Truth*" then the node reveals the true type to the leader. If the node's strategy is "*Untruth*", then the node reveals a fake type to the leader. After receiving the revelations from each node, the leader takes these revelations as the input and makes its selection of m nodes out of $\mathcal{N}$ by using a given selection function. Each selected node is assigned a task to act as a server node and then cooperatively completes the current threshold key management task with other selected nodes. Finally, the leader pays the nodes by computing the payments vector $R^{t} = (R_1^t, \ldots, R_n^t)$ with a payment function. Payments are used to motivate nodes to behave in accordance with the mechanism goals.

In the rest of the paper, we use $\theta_i^t = (p_i^t, z_i^t)$ to denote the real type of node i at stage t, and use superscript "^" to denote the type which is revealed to the leader, so to differentiate it from what is privately known by node itself. And use "$-i$" to denote all the other nodes in nodes set $\mathcal{N}$ except i. Now we define ICONS mechanism which implements our optimal objective as follows.

4.1.1. Selection Function. Given the input of nodes' revealed type vector $\widehat{\theta}$ at stage t, the mechanism of choosing m nodes from nodes set $\mathcal{N}$ that maximizes the system's welfare can be formulated as follows:

$$K_{\mathcal{N}}^{*t}\left(\widehat{p}^{t}, \widehat{z}^{t}\right) = \arg\max_{K_{\mathcal{N}}^{t}\in\mathcal{N}^{m}} \left(W^{t} \cdot \prod_{i\in K_{\mathcal{N}}^{t}} \widehat{p}_i^t - \sum_{i\in K_{\mathcal{N}}^{t}} \widehat{z}_i^t \right). \tag{15}$$

A branch and bound method [30] can be applied to allow us to find the optimal set of m nodes $K_{\mathcal{N}}^{*t}$ from MANET's nodes set $\mathcal{N}$ as defined in (15), with reduced computational cost.

4.1.2. Payment Function. Let $q_i^t \in \{0,1\}$ denote the node i's completion of assigned key management task in stage t if this node is active, where 1 means success in completing this task and 0 means fail. The payment will be given to each node i

by the leader in the form of reputation [31], according to the following payment function:

$$R_i^t = \begin{cases} W^t \cdot \prod_{i \in K_{\mathcal{N}}^t} \widehat{p}_i^t - \overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t, \widehat{z}_{-i}^t\right) - \sum_{i \in K_{\mathcal{N}}^{*t}} \widehat{z}_i^t & \\ & \text{node } i \text{ is active and } q_i^t = 1 \\ -\overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t, \widehat{z}_{-i}^t\right) - \sum_{i \in K_{\mathcal{N}}^{*t}} \widehat{z}_i^t & \\ & \text{node } i \text{ is active and } q_i^t = 0 \\ C^t & \text{node } i \text{ is passive.} \end{cases} \tag{16}$$

To let the leader detect if a selected node completes its task or not, we follow the previous work on developing an integrated fault-intrusion tolerance framework [32, 33] and do not differentiate between malicious faults and normal server failures (e.g., node crash, network disconnection, power failure, etc.) in our scenario. The detection method proposed in [34] that detects the corrupted shares for the proactive secret sharing can be adopted here, by checking if the node participates in the process with the presence of an uncorrupted share of system's secret. A recent similar work is found in [35, 36], which uses Shamir's secret sharing scheme to detect malicious activities in the encrypted networks such as virtual private networks (VPNs) that encrypt and conceal network traffic.

Now, we get the utility function for node i at stage t as follows:

$$u_i^t\left(K_{\mathcal{N}}^{*t}, \widehat{\theta}_i^t, \widehat{\theta}_{-i}^t, z_i^t\right) = R_i^t\left(K_{\mathcal{N}}^{*t}, \widehat{\theta}_i^t, \widehat{\theta}_{-i}^t\right) - z_i^t, \tag{17}$$

where z_i^t is the instantaneous cost of node i at stage t, as defined in (7); $R_i^t(K_{\mathcal{N}}^{t*}, \widehat{\theta}_i^t, \widehat{\theta}_{-i}^t)$ is the payment defined in (16) given by the mechanism to the node i, when the coalition of m nodes selected from $\mathcal{N}$ to cooperatively complete the key management task at stage t is $K_{\mathcal{N}}^{t*}$; $\widehat{\theta}_i^t$ and $\widehat{\theta}_{-i}^t$ are revealed type(s) of node i and nodes set $\{-i\}$, respectively.

Then, node i's expected utility at stage t is

$$\overline{u}_i^t\left(K_{\mathcal{N}}^{t*}, \widehat{\theta}_i^t, \widehat{\theta}_{-i}^t, z_i^t\right) = \mathbb{E}\left[u_i^t\left(K_{\mathcal{N}}^{t*}, \widehat{\theta}_i^t, \widehat{\theta}_{-i}^t, z_i^t\right)\right]. \tag{18}$$

Note that, $\overline{u}_i^t(K_{\mathcal{N}}^{t*}, \widehat{\theta}_i^t, \widehat{\theta}_{-i}^t, z_i^t)$ is what the node i usually seeks to maximize. It reflects the amount of benefits gained by node i if it follows a specific strategy ξ_i^t at stage t. Nodes might deviate from revealing their truthful types if that could lead to a better payment. Therefore, our mechanism must be strategyproof where truth revealing is the dominant strategy, and thus the following standard properties are required to be satisfied [13, 29].

(1) Individual rationality holds when truthful nodes are guaranteed to have nonnegative expected utility. Formally, this condition holds, when for all i, θ_i^t, and $\widehat{\theta}_{-i}^t$,

$$\overline{u}_i^t\left(K_{\mathcal{N}}^{*t}, \theta_i^t, \widehat{\theta}_{-i}^t, z_i^t\right) \geq 0. \tag{19}$$

(2) Incentive compatibility holds when it is a dominant strategy for each node with truthful revelation. Formally, for all i, θ_i^t, $\theta_i'^t$, and $\widehat{\theta}_{-i}^t$,

$$\overline{u}_i^t\left(K_{\mathcal{N}}^{*t}, \theta_i^t, \widehat{\theta}_{-i}^t, z_i^t\right) \geq \overline{u}_i^t\left(K^t, \theta_i'^t, \widehat{\theta}_{-i}^t, z_i^t\right). \tag{20}$$

(3) No-free-riders holds if all nodes not selected to participate in the current key management have a revenue of 0.

The properties of *individual rationality*, *incentive compatibility*, and *no free riders* imply that, (1) the expected utility of a truthful node is always nonnegative; (2) each node will find no better option than to reveal their true type; (3) the nodes that are not selected to participate in the current key management have a revenue of 0. Therefore, all rational nodes that include the malicious ones will find that revealing their types untruthfully can never lead to a better payment than revealing their types truthfully, and sending no information to the leader can never lead to a better payment than reporting their types to the leader. Then all rational nodes will always report their types truthfully to the leader, since their objectives are to maximize their individual benefits.

Similarly in our mechanism, leader also need to maximize the system's welfare, and so we have the following definition.

Definition 1. A selection function is called socially efficient if the chosen selection $K^{*t} \in \mathcal{N}^m$ maximizes social welfare over $\mathcal{N}^m$; that is, for all $K_{\mathcal{N}}^t \in \mathcal{N}^m$,

$$\left(W^t \cdot \prod_{i \in K_{\mathcal{N}}^{*t}} \widehat{p}_i - \sum_{i \in K_{\mathcal{N}}^{*t}} \widehat{z}_i\right) \geq \left(W^t \cdot \prod_{i \in K_{\mathcal{N}}^t} \widehat{p}_i - \sum_{i \in K_{\mathcal{N}}^t} \widehat{z}_i\right). \tag{21}$$

Now, we have the following required property.

(4) A mechanism is called a socially efficient mechanism, if it has a socially efficient selection function.

To ensure truth elicitation from all the nodes, we need to prove that the presented mechanism is strategyproof.

4.2. Properties of Mechanism

4.2.1. Individual Rationality. Individual rationality means that the expected utility of a truthful node is always nonnegative. Truthful node i with its revealed type θ_i^t might either be selected to participate in completing the key management task at stage t or not, given other nodes $-i$'s revealed types vector $\widehat{\theta}_{-i}^t$. Now, we consider both cases as follows.

Case 1. Truthful node i is not selected to participate in the key management at stage t.

From (18) and (16), we know that node i's expected utility is 0 at this stage, because both its payment and its cost are C^t and hence we proved our claim.

Case 2. Truthful node i is selected to participate in the key management task at stage t.

From (18) and (16), we know that node i's expected utility in this case is

$$\begin{aligned} \overline{u}_i^t &= R_i^t - z_i^t \\ &= p_i^t \cdot \left(W^t \cdot \prod_{\substack{j \in K_{\mathcal{N}}^{*t}, \\ j \neq i}} \widehat{p}_j^t - \overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t, \widehat{z}_{-i}^t\right) - \sum_{\substack{j \in K_{\mathcal{N}}^{*t}, \\ j \neq i}} \widehat{z}_j^t \right) \end{aligned}$$

$$+\left(1-p_i^t\right)\cdot\left(-\overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right)-\sum_{\substack{j\in K_{\mathcal{N}}^{*t},\\ j\neq i}}\widehat{z}_j^t\right)-z_i^t$$
$$=p_i^t\cdot W^t\cdot\prod_{\substack{j\in K_{\mathcal{N}}^{*t},\\ j\neq i}}\widehat{p}_j^t-\overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right)-\sum_{\substack{j\in K_{\mathcal{N}}^{*t},\\ j\neq i}}\widehat{z}_j^t-z_i^t. \tag{22}$$

Since node i is truthful at this stage, we have $\widehat{\theta}_i^t = \theta_i^t$, that means $\widehat{p}_i^t = p_i^t$ and $\widehat{z}_i^t = z_i^t$. Therefore, (22) can be rewritten as

$$\overline{u}_i^t=\widehat{p}_i^t\cdot W^t\cdot\prod_{\substack{j\in K_{\mathcal{N}}^{*t},\\ j\neq i}}\widehat{p}_j^t-\overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right)-\sum_{\substack{j\in K_{\mathcal{N}}^{*t},\\ j\neq i}}\widehat{z}_j^t-\widehat{z}_i^t$$
$$=\left(W^t\cdot\prod_{j\in K_{\mathcal{N}}^{*t}}\widehat{p}_j^t-\sum_{j\in K_{\mathcal{N}}^{*t}}\widehat{z}_j^t\right)-\overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right) \tag{23}$$
$$=K_{\mathcal{N}}^{t*}\left(\widehat{p}_{i,-i}^t,\widehat{z}_{i,-i}^t\right)-\overline{K}_{-i}^{t*}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right)$$
$$=K_{\mathcal{N}}^{*t}\left(\widehat{p}^t,\widehat{z}^t\right)-\overline{K}_{-i}^{*t}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right).$$

According to the selection function defined in our mechanism, the first term in (23) quantifies the optimal welfare that can be obtained when node i's revealed type is its true one $\widehat{\theta}_i$, the vector of other nodes' revealed types is $\widehat{\theta}_{-i}$, and node i is involved in the selection. Similarly, the second term quantifies the optimal welfare that can be obtained when the vector of other nodes' revealed types is $\widehat{\theta}_{-i}$ but node i is not involved in the selection. Since node i's involvement can only improve the total welfare, we have $\overline{u}_i^t = K_{\mathcal{N}}^{t*}(\widehat{p}^t,\widehat{z}^t)-\overline{K}_{-i}^{t*}(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t)\geq 0$ and proved this property.

4.2.2. Incentive Compatibility. Incentive compatibility means that players will find no better option than to reveal their true type. We consider the node i and other nodes $-i$. Given revealed types of i and $-i$, $\widehat{\theta}_i^t = \{\widehat{p}_i^t,\widehat{z}_i^t\}$ and $\widehat{\theta}_{-i}^t = \{\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\}$, we need to show that node i cannot gain more from not revealing its true type than revealing its true type. Now given its true type θ_i^t at stage t, we consider two cases of this node by revealing its true types, namely, either selected or not.

Case 1. Node i is selected by the leader if it reveals its type θ truthfully at stage t. Then from the property of individual rationality, we know that by revealing its truth type at this case, node i can gain a utility $\overline{u}_i^t \geq 0$. Now we consider two subcases of node i's untruthful revelation.

(i) If it is not selected by the leader due to its untruthful revelation, then node i's utility at stage t will still be 0, and this make node i have no incentive to be untruthful at this subcase.

(ii) If it is selected by the leader to participate in completing the key management task at current stage, then from (22) and the payment function of our mechanism we know that, given other nodes' revealed types $\widehat{\theta}_{-i}^t$, node i's expected utility at current stage t just relates to node i's true type θ_i^t, no matter what type it had revealed. That is to say, untruthful revelation cannot make extra utility to node i, so node i has no incentive to be untruthful at this subcase also.

Case 1. Node i is not selected to participate in the key management task at stage t. From (16), we know that by revealing its truth type, node i could gain a utility of 0 at this case. Now we consider two subcases of node i with an untruthful revelation.

(i) Node i is still not selected to participate in the task with its untruthful revelation.

In this subcase, node i's utility is still 0, and this means that the utility of node i remains the same with an untruthful revelation of type $\widehat{\theta}_i^t$.

(ii) Node i is selected to participate in the task because of its untruthful revelation. From (22) and the payment function in our mechanism, we know that node i's expected utility $u_i'^t$ at this scenario will be

$$u_i'^t=\left[p_i^t\cdot W^t\cdot\prod_{\substack{j\in K_{\mathcal{N}}^{t*},\\ j\neq i}}\widehat{p}_j^t-\left(\sum_{\substack{j\in K_{\mathcal{N}}^{t*},\\ j\neq i}}\widehat{z}_j^t+z_i^t\right)\right]$$
$$-\overline{K}_{-i}^{t*}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right) \tag{24}$$
$$=K_{\mathcal{N}}'^t\left(p_i^t,\widehat{p}_{-i}^t,z_i^t,\widehat{z}_{-i}^t\right)-\overline{K}_{-i}^{t*}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right),$$

where (p_i^t,z_i^t) is node i's true type at stage t, and $K_{\mathcal{N}}'^t(p_i^t,\widehat{p}_{-i}^t,z_i^t,\widehat{z}_{-i}^t)$ is a new selection scheme over the vector of nodes' revealed types $(\theta_i^t,\widehat{\theta}_{-i}^t)$, with node i being selected to participate in completing the task at stage t, that is, $i\in K_{\mathcal{N}}'^t$.

Note that in this subcase, node i would not be selected by the leader if it revealed its true type, and so with its truthful revelation, node i whether involved in the selection or not will make no difference on system's welfare at this stage. Then we have

$$K_{\mathcal{N}}^{t*}\left(\widehat{\theta}_i,\widehat{\theta}_{-i}^t\right)=K_{\mathcal{N}}^{t*}\left(\widehat{p}^t,\widehat{z}^t\right)$$
$$=\overline{K}_{-i}^{t*}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right). \tag{25}$$

If we assume $u_i'^t > 0$ and substitute $\overline{K}_{-i}^{t*}(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t)$ with $K_{\mathcal{N}}^{t*}(p_i^t,\widehat{p}_{-i}^t,z_i^t,\widehat{z}_{-i}^t)$ in (24), then we can get

$$u_i'^t=K_{\mathcal{N}}'^t\left(p_i^t,\widehat{p}_{-i}^t,z_i^t,\widehat{z}_{-i}^t\right)-\overline{K}_{-i}^{t*}\left(\widehat{p}_{-i}^t,\widehat{z}_{-i}^t\right)$$
$$=K_{\mathcal{N}}'^t\left(p_i^t,\widehat{p}_{-i}^t,z_i^t,\widehat{z}_{-i}^t\right)-K_{\mathcal{N}}^{t*}\left(p_i^t,\widehat{p}_{-i}^t,z_i^t,\widehat{z}_{-i}^t\right) \tag{26}$$
$$>0,$$

where $K_{\mathcal{N}}^{t*}(p_i^t,\widehat{p}_{-i}^t,z_i^t,\widehat{z}_{-i}^t)$ is the optimal nodes selection scheme over the vector of nodes' revealed types $(\theta_i^t,\widehat{\theta}_{-i}^t)$ at stage t. This means that, given the revelation of node types $(\theta_i^t, \widehat{\theta}_{-i}^t)$, new selection scheme $K_{\mathcal{N}}^{\prime t}$ is more optimal than $K_{\mathcal{N}}^{t*}$. However, this contradicts the definition of $K_{\mathcal{N}}^{t*}$. Therefore, revealing an untruthful type $\widehat{\theta}_i$ cannot lead to a higher utility of node i in this subcase.

4.2.3. No Free Riders. This property can be derived from the payment function of our mechanism directly.

4.2.4. Socially Efficient. This property can be derived from incentive compatibility and the selection function of our mechanism directly.

5. Simulation Experiments and Results

In this section, we illustrate some of the performance benefits of our proposed model. To show efficient improvement of our model and to show the negative impact of selfish node with untruth telling, we evaluate the performance of our strategyproof optimal nodes selection model with respect to Yu's selection model [10] and random selection model [2–9].

To eliminate the effect of leader election phase's cost, a reasonable choice would be to run the leader election algorithm only once and follow a fixed number of stages after the initial election and thus amortizing the overhead through the many iterations of the key management tasks, similar to what is explained in [37]. In this way, the actual overhead of leader election would be neglected when considering the network lifetime.

For simplicity, we use three security states: *safe* (s), *vulnerable* (v), and *compromised* (c) and three energy states *high* ($b1$), *middle* ($b2$), and *low* ($b3$). There are a total of nine states for a node (s, $b1$), (s, $b2$), (s, $b3$), (v, $b1$), (v, $b2$), (v, $b3$), (c, $b1$), (c, $b2$), and (c, $b3$), as defined in (3). The state transition probability matrix stands for the probability of a node changes from one state to another. We define the basic security state transition probability matrix A1 and basic energy state transition probability matrix B1 for active nodes as follows: A1 = [0.92, 0.05, 0.03, 0.04, 0.92, 0.04, 0.01, 0.03, 0.96; 0.94, 0.04, 0.02, 0.02, 0.03, 0.95, 0.01, 0.03, 0.96; 0.93, 0.05, 0.02, 0.03, 0.94, 0.03, 0.03, 0.02, 0.95; 0.998, 0.001, 0.001, 0.001, 0.0998, 0.001, 0.001, 0.001, 0.999] and B1 = [0.98, 0.02, 0.00, 0.00, 0.98, 0.02, 0.00, 0.00, 1.00; 0.99, 0.01, 0.00, 0.00, 0.99, 0.01, 0.00, 0.00, 1.00]. In our simulation, each active node can choose its security state transition probability matrix and energy state transition probability matrix from A1 and B1 with an initialization function randominit. For example, when node i, with a security state of vulnerable and a energy state of middle in current stage, chooses [0.93, 0.05, 0.02, 0.03, 0.94, 0.03, 0.03, 0.02, 0.95] and [0.99, 0.01, 0.00, 0.00, 0.99, 0.01, 0.00, 0.00, 1.00] as its security state transition probability matrix and energy state transition probability matrix, respectively, then its security state in the next stage will be compromised with probability 0.30, will remain be vulnerable with probability 0.94, and will be snatched back to the safe state with probability 0.30; its energy state in the next stage will still be middle with probability 0.99 but low with probability 0.01. We also assume that when a node is passive, the transition probability is lower than that when the node is active, and so we define the basic security state transition probability matrix A2 and basic energy state transition probability matrix B2 for passive nodes as follows: A2 = [0.99, 0.01, 0, 0, 0.99, 0.01, 0.005, 0.005, 0.99; 0.999, 0.001, 0, 0, 0.999, 0.001, 0, 0.001, 0.999] and B2 = [0.99, 0.01, 0, 0, 0.99, 0.01, 0, 0, 1].

To conduct parameter-sensitivity analysis and check the impacts of various transition probabilities on the performance in the simulations, we change the transition probability from 0.98 to 0.82 with the matrix of [0.98, 0.02, 0.00, 0.00, 0.98, 0.02, 0.00, 0.00, 1.00; 0.96, 0.04, 0.00, 0.00, 0.96, 0.04, 0.00, 0.00, 1.00; 0.94, 0.06, 0.00, 0.00, 0.94, 0.06, 0.00, 0.00, 1.00; 0.92, 0.08, 0.00, 0.00, 0.92, 0.08, 0.00, 0.00, 1.00; 0.90, 0.10, 0.00, 0.00, 0.90, 0.10, 0.00, 0.00, 1.00; 0.88, 0.12, 0.00, 0.00, 0.88, 0.12, 0.00, 0.00, 1.00; 0.86, 0.14, 0.00, 0.00, 0.86, 0.14, 0.00, 0.00, 1.00; 0.84, 0.16, 0.00, 0.00, 0.84, 0.16, 0.00, 0.00, 1.00; 0.82, 0.18, 0.00, 0.00, 0.82, 0.18, 0.00, 0.00, 1.00]. Since the risk of damage to the network would be further increased if a compromised node is chosen to act as a server node, we set the cost of selecting a higher security state to lower values than that associated with a lower security state selection. Similarly, to balance the nodes' energy consumption and avoid the situation where there are some higher energy level nodes still in a dead network, the selection of a lower energy level node should have a higher cost than selecting a higher energy level node. Thus, we define the cost matrices for the simulation as follows: C1 = [5.5, 7.5, 11, 25, 31, 37, 43, 52, 63], C2 = [6, 8, 11, 23, 30, 35, 42, 50, 62], C3 = [6.5, 8.5, 12, 21, 28, 33, 40, 48, 60], C4 = [8, 10, 13, 18, 25, 30, 37, 45, 55], C5 = [8.5, 11, 15, 19, 27, 32, 38, 47, 58], C6 = [9, 10, 15, 20, 28, 32, 38, 48, 59], C7 = [9, 12, 16, 15, 20, 28, 33, 40, 50], C8 = [10, 13, 18, 16, 21, 28, 34, 42, 51], and C9 = [11, 14, 19, 17, 22, 29, 35, 43, 52], corresponding to the system state matrix [(s, $b1$), (s, $b2$), (s, $b3$), (v, $b1$), (v, $b2$), (v, $b3$), (c, $b1$), (c, $b2$), (c, $b3$)]. The system value of the threshold key management service in our simulation is defined as $V^t = 300$, $\rho = 0.25$, and then $W^t = 200$ for each stage.

5.1. Cost Reduction. First, we compare the costs in different models along simulation stages when a (3, 7) secret sharing scheme is used. Initializing each node with a state of (1, 1), we can see from Figure 1 that there is a distinct cost reduction of both our strategyproof optimal selection model and Yu's selection model over the random selection model. This is mainly due to the fact that the random selection model selects nodes without considering the cost and thus leading to a higher average cost. Figure 1 also demonstrates that for the two optimal selection models, there is a better performance of our selection model than that of Yu's selection model. This result indicates that, with the presence of selfish nodes, the normal nodes in the system must more often be active to carry out the duty of key management than the nodes in the system without selfish nodes and so will transfer into more cost states with a higher probability.

We also perform parameter-sensitivity analysis on the models by considering different crypto thresholds. The simulation is performed with 30 stages for 200 times and then the average costs of each stage are calculated. Figure 2 shows the costs comparison over our model, the random selection

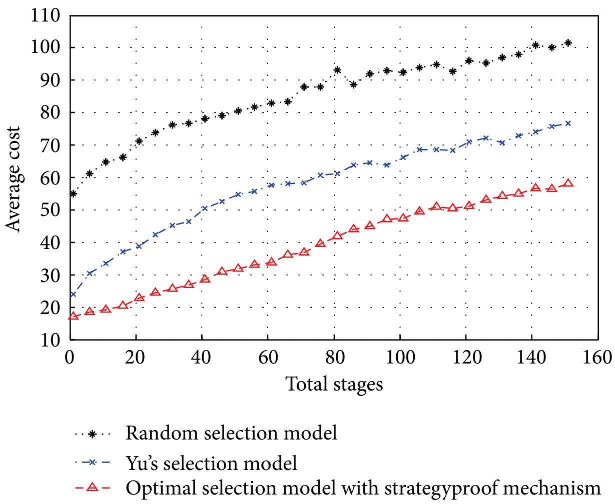


Figure 1: Average cost with different stages.

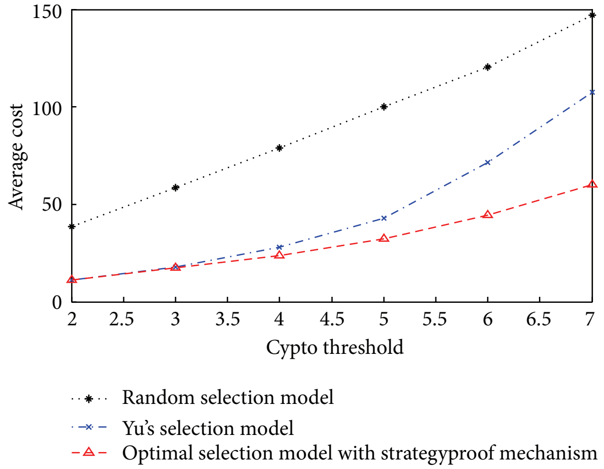


Figure 2: Average cost with different crypto thresholds.

model, and Yu's selection model, when there are 15 nodes participating in the key management service, all with initial node state of (1, 1), with the crypto threshold changing from 2 to 7. With the increase of the crypto threshold, the cost to perform the key management task increases due to more nodes that need to be active, but our model always has the lowest cost.

5.2. Network Lifetime Improvement. In these simulations, we investigate the network lifetime improvement of the proposed model. Let N be the number of nodes in the MANET, and we consider the crypto threshold as $(N-1)/2$ in the simulation. So when there are $N-(N-1)/2=(N+1)/2$ nodes run out of power, the network is regarded as dead. A node is considered running out of power, if it has run x stages on passive mode and y stages on active mode since entering into the *low* ($h3$) energy state, where $x+2y=30$ and $0 \leq x, y \leq 30$. We first check the performance when different energy transition probabilities are used. We set $N=7$ and let the energy transition probability of the active node be in the range from 0.88 to 0.98. The energy transition probabilities matrix for passive node remains unchanged in all circumstance as defined before, that is, B2 = [0.99, 0.01, 0, 0, 0.99, 0.01, 0, 0, 1]. As shown in Figure 3, our strategyproof optimal selection model and Yu's selection model always have longer network lifetime than the random selection model, because in random selection model the nodes are selected without considering the energy level and this leads the nodes with low energy to die fast.

Then we check the performance when different numbers of nodes are available in the network. From Figure 4 we can see that the key management service is distributed among more nodes and thus prolonging the lifetime of the network in all three models (random selection model, Yu's selection model, and our selection model), with the number of available nodes in MANET increasing from 5 to 35. Still the same as before, our optimal selection model shows consistent improvement over the other two models in this simulation.

5.3. Success Ratio Improvement. The success ratio is the probability that the leader successfully collects all the requested partial signature or partial authentication from a threshold number of nodes and then completes the assigned key management task. We assume that a node in safe state will always complete its assigned task with probability of 1, while a vulnerable node and a compromised node with probabilities of $1/\sqrt{2}$ and $1/\sqrt{3}$, respectively. First, we compare the average success ratios of threshold key management in different models along simulation stages when there are 7 nodes in the network, with the crypto threshold and selfish nodes number set to 3 and 2, respectively. Figure 5 shows that our strategyproof optimal selection model has a distinct success ratio improvement over the random selection model and also demonstrates a better success ratio than Yu's selection model. This is mainly due to the fact that neither the random selection model nor Yu's selection model selects nodes with the consideration of the success ratios of each nodes.

We then show the success ratio improvement when there are more nodes in the network. From Figure 6 we can see, with the number of available nodes in the network increasing from 7 to 19, that the success ratios of all three models become higher since there are more nodes with higher security state which can be selected from. The success ratio of our optimal selection model still is shown to be the highest in these circumstances.

5.4. Network Compromising Probability Reduction. Last but not least, we investigate the probability of the network being compromised by attacker(s) who is(are) attempting to assemble enough key information to deduce the master key of the system. In order to quantify and compare different models in our scenario, we will use as a metric the compromise probability that is defined as the probability that an attacker can recover the master key of the secret sharing scheme, after capturing enough nodes, and so is inversely proportional to the number of stages required by attacker(s) to capture the required number of nodes and then to compromise the network. Assume that the attacker knows all public parameters of the system, then when a (m, N) secret sharing

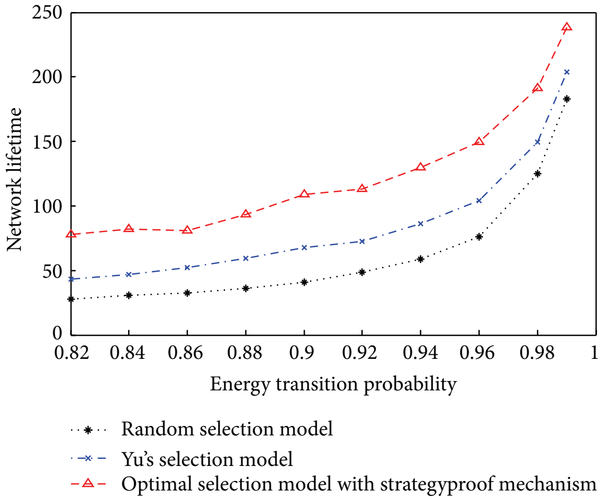


Figure 3: Network lifetime with different transition probabilities.

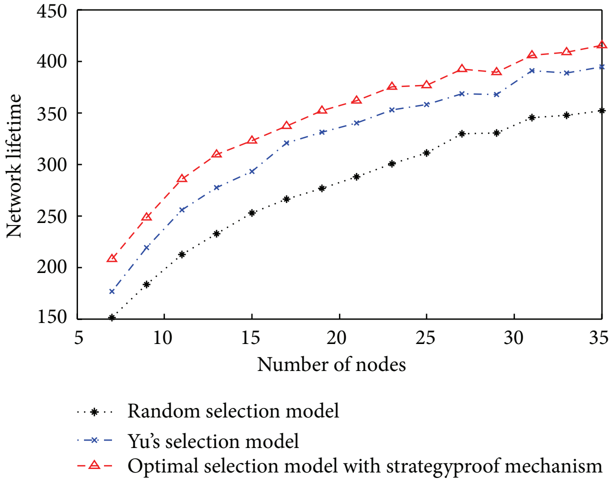


Figure 4: Network lifetime with different numbers of nodes.

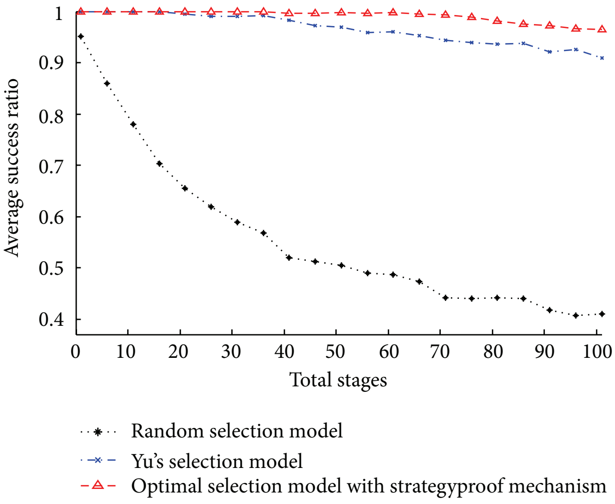


Figure 5: Average success ratio with different stages.

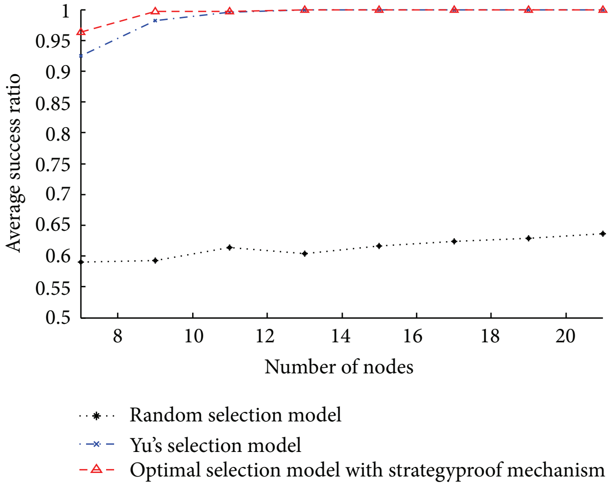


Figure 6: Average success ratio with different numbers of nodes.

scheme is used, the MANET is deemed compromised if $(N-m)$ nodes are captured by the attacker(s). In our simulation, a node is defined as captured, if it has run x stages on passive mode and y stages on active mode since it entered into the compromised security state, where $x + 2y = 30$ and $0 \leq x, y \leq 30$.

Firstly, we set $N = 7$, $m = 3$, and compare the network compromising probabilities when security transition probabilities are in the range from 0.76 to 0.98. The security transition probabilities matrix for passive node remain unchanged in all circumstance as defined before, that is, A2 = [0.99, 0.01, 0, 0, 0.99, 0.01, 0.005, 0.005, 0.99; 0.999, 0.001, 0, 0, 0.999, 0.001, 0, 0.001, 0.999]. Among the 7 nodes, we assume in the beginning, that 3 are in safe security state, 2 in vulnerable security state that, and 2 in compromised security state.

The results in Figure 7 indicate that the proposed selection model has lower network compromising probability than the random selection model and Yu's selection model, since our strategyproof optimal selection model tends to select nodes with higher security levels and thus keeping a balance of all node's security level with time. When the transition probabilities are closer to 1, the compromising probabilities of all models asymptotically approach 0. This is because the nodes in each model will keep their security state unchanged and so keep the networks safe in each stage. In Figure 8, we compare the network compromising probability when there are different numbers of nodes in the network. With an increase in the total number of nodes in the MANET, all the models show a downward trend in compromising probabilities because the key management service can be distributed among more nodes which will decrease the probability of each node transition into a lower security level. Similarly as before, our optimal selection model has lower compromising probability than the other two models.

From the simulation results in Figures 1–8, we can see that with the incentive compatible optimal nodes selection mechanism to encourage each node in the system to be truth-telling and to select a coalition of nodes with the purposes of maximizing the success ratio of key management service

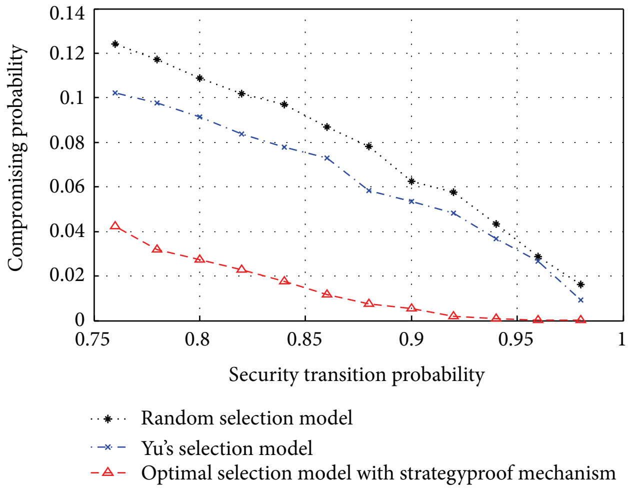


Figure 7: Compromising probability with different transition probabilities.

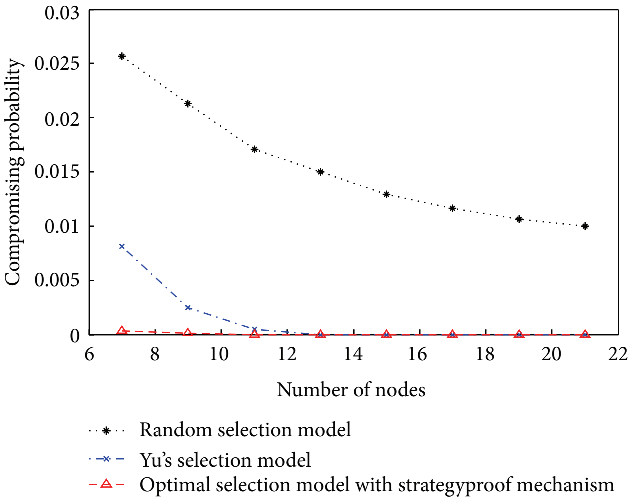


Figure 8: Compromising probability with different numbers of nodes.

and minimizing the nodes' cost, our optimal nodes selection model certainly has dramatic results in reducing the nodes's resource consumption and network's security compromising, while improving the lifetime of the network and the success ratio of the key management service dramatically, in the presence of selfish nodes.

6. Conclusion and Discussion

In this paper, focusing on the optimal nodes selection problem in presence of selfish nodes for threshold key management in MANETs during its operation, we formulated the dynamic nodes selection problem as a combinatorial optimization problem firstly, with the objectives of maximizing the success ratio of key management service and minimizing the nodes' cost of security and energy and then proposed the incentive compatible mechanism to implement the optimal nodes selection process in MANETs, to ensure that the truth telling is the dominant strategy and so prevent the emergence of selfish nodes. To the best of our knowledge, this is the first incentive-compatible mechanism for threshold key management.

In our scheme, although one of the nodes in the network needs to be specified as a leader, essentially there are differences between the leader node in this scheme and the PKG server in the centralized scheme [38, 39], and these make our proposed mechanism efficient and suitable for the MANETs. Specifically, (1) instead of being specified by the administrator in bootstrapping network phase and fixed during network lifetime, the leader can be elected/reelected periodically and/or when found to be failed, attacked, or run out of battery, it cannot reach any nodes in the system, and so forth and so eliminating the single point of failure. (2) The network's primary secret is not held by the leader itself but is split and shared by all the nodes by using secret sharing method. Therefore, no node in the network is required to be trusted and available to all other nodes. (3) The main task of key management is not performed by the leader but is performed by a threshold number of nodes selected at each stage elaborately, with the consideration of maximizing the success ratio of key management service and minimizing the nodes' cost of security and energy. In this way, our scheme improves both the success ratio of key management service and lifetime of the network and reduces both the cost of participating nodes and compromising probability. (4) The ICONS mechanism that we proposed cannot only be used in the scenario of threshold key management but also in other cooperation scenarios in MANET

For future work, we plan to extend this mechanism to a distributed setting [40]. Although we argued in Section 3 that it is reasonable to assume that there is always a node in the network to act as a leader, a practical and distributed selection model without a specific leader node will be more helpful to implement the nodes selection model in the real world.

Acknowledgments

This work was supported by the Excellent Youth Foundation of Henan Scientific Committee (104100510025), the National Natural Science Foundation of China (no. 60633020), and the Open Foundation from the Key Laboratory of Intelligent Computing and Signal Processing of the Ministry of Education (201107).

References

[1] S. Capkun, J. P. Hubaux, and L. Buttyan, "Mobility helps peer-to-peer security," *IEEE Transactions on Mobile Computing*, vol. 5, no. 1, pp. 43–51, 2006.

[2] L. Zhou and Z. Haas, "Securing ad hoc networks," *IEEE Network*, vol. 13, pp. 24–30, 1999.

[3] B. Zhu, F. Bao, R. H. Deng, M. S. Kankanhalli, and G. Wang, "Efficient and robust key management for large mobile ad hoc networks," *Computer Networks*, vol. 48, no. 4, pp. 657–682, 2005.

[4] K. Hamouid and K. Adi, "Secure and robust threshold key management (SRKM) scheme for ad hoc networks," *Security and Communication Networks*, vol. 3, Article ID 517534, 2010.

[5] Y. Dong, A. F. Sui, S. M. Yiu, V. O. K. Li, and L. C. K. Hui, “Providing distributed certificate authority service in cluster-based mobile ad hoc networks,” *Computer Communications*, vol. 30, pp. 2442–2452, 2007.

[6] D. Y. Lee and H. C. Jeong, “An efficient certificate management for mobile ad-hoc network,” in *Ad-Hoc, Mobile, and Wireless Networks*, T. Kunz and S. Ravi, Eds., vol. 4104 of *Lecture Notes in Computer Science*, pp. 355–364, Springer, Berlin, Germany, 2006.

[7] S. Yi and R. H. Kravets, “MOCA: mobile certificate authority for wireless ad hoc networks,” in *Proceedings of the 2nd Annual PKI Research Workshop (PKI ’03)*, 2003.

[8] H. Luo, J. Kong, P. Zerfos, S. Lu, and L. Zhang, “URSA: ubiquitous and robust access control for mobile ad hoc networks,” *IEEE/ACM Transactions on Networking*, vol. 12, no. 6, pp. 1049–1063, 2004.

[9] M. Omar, Y. Challal, and A. Bouabdallah, “Reliable and fully distributed trust model for mobile ad hoc networks,” *Computers and Security*, vol. 28, no. 3-4, pp. 199–214, 2009.

[10] F. R. Yu and H. Tang, “Distributed node selection for threshold key management with intrusion detection in mobile ad hoc networks,” *Wireless Networks*, vol. 16, no. 8, pp. 2169–2178, 2010.

[11] V. A. Leoni and G. L. Nasini, “On the computational complexity of combinatorial flexibility problems,” *International Journal of Computer Mathematics*, vol. 87, no. 15, pp. 3330–3343, 2010.

[12] L. Hurwicz and S. Reiter, *Designing Economic Mechanisms*, Cambridge University Press, Cambridge, UK, 1th edition, 2008.

[13] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*, Oxford University Press, Oxford, UK, 1995.

[14] T. Wu, M. Malkin, and D. Boneh, “Building intrusion tolerant applications,” in *Proceedings of the 8th Conference on USENIX Security Symposium*, vol. 8, p. 7, USENIX Association, 1999.

[15] M. Masdari, S. Jabbehdari, M. R. Ahmadi, S. M. Hashemi, J. Bagherzadeh, and A. Khadem-Zadeh, “A survey and taxonomy of distributed certificate authorities in mobile ad hoc networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, pp. 1–12, 2011.

[16] J. Sen, M. Girish Chandra, P. Balamuralidhar, S. G. Harihara, and H. Reddy, “A scheme of certificate authority for ad hoc networks,” in *Proceedings of the 18th International Workshop on Database and Expert Systems Applications (DEXA ’07)*, pp. 615–619, September 2007.

[17] D. Joshi, K. Namuduri, and R. Pendse, “Secure, redundant, and fully distributed key management scheme for mobile ad hoc networks: an analysis,” *Eurasip Journal on Wireless Communications and Networking*, vol. 2005, no. 4, pp. 579–589, 2005.

[18] L. Anderegg and S. Eidenbenz, “Ad hoc-VCG: a truthful and cost-efficient routing protocol for mobile Ad hoc networks with selfish agents,” in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom ’03)*, pp. 245–259, September 2003.

[19] Y. Wang and M. Singhal, “On improving the efficiency of truthful routing in MANETs with selfish nodes,” *Pervasive and Mobile Computing*, vol. 3, no. 5, pp. 537–559, 2007.

[20] A. Rachedi, A. Benslimane, H. Otrok, N. Mohammed, and M. Debbabi, “A secure mechanism design-based and game theoretical model for MANETs,” *Mobile Networks and Applications*, vol. 15, no. 2, pp. 191–204, 2010.

[21] D. Hao, Y. Ren, and K. Sakurai, “A game theorybased surveillance mechanism against suspicious insiders in MANETs,” in *Trusted Systems*, vol. 6802 of *Lecture Notes in Computer Science*, pp. 237–252, Springer, Berlin, Germany, 2011.

[22] N. Mohammed, H. Otrok, L. Wang, M. Debbabi, and P. Bhattacharya, “Mechanism design-based secure leader election model for intrusion detection in MANET,” *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 1, pp. 89–103, 2011.

[23] H. Garcia-Molina, “Elections in a distributed computing system,” *IEEE Transactions on Computers*, vol. 31, pp. 48–59, 1982.

[24] H. C. Chung, P. Robinson, and J. L. Welch, “Optimal regional consecutive leader election in mobile ad-hoc networks,” in *Proceedings of the 7th ACM SIGACT/SIGMOBILE International Workshop on Foundations of Mobile Computing (FOMC ’11)*, pp. 52–61, June 2011.

[25] Q. Dong and D. Liu, “Resilient cluster leader election for wireless sensor networks,” in *Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON ’09)*, pp. 108–116, IEEE Press, June 2009.

[26] P. Hu, Z. Zhou, Q. Liu, and F. Li, “The HMM-based modeling for the energy level prediction in wireless sensor networks,” in *Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications*, pp. 2253–2258, May 2007.

[27] P. Samarati, M. K. Reiter, and S. Jajodia, “An authorization model for a public key management service,” *ACM Transactions on Information and System Security*, vol. 4, pp. 453–482, 2001.

[28] K. Miettinen, *Non Linear Multi-Objective Optimization*, Kluwer Academic Publisers, Norwell, Mass, USA, 1999.

[29] N. Nisan, “Algorithms for selfish agents mechanism design for distributed computation,” in *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science (STACS ’99)*, pp. 1–15, Springer, Trier, Germany, 1999.

[30] M. J. Brusco and S. Stahl, *Branch-and-Bound Applications in Combinatorial Data Analysis*, Springer, New York, NY, USA, 2005.

[31] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara, “Reputation systems: facilitating trust in internet interactions,” *Communications of the ACM*, vol. 43, no. 12, pp. 45–48, 2000.

[32] A. Bessani, “From byzantine fault tolerance to intrusion tolerance,” in *Proceedings of the IEEE/IFIP 41st International Conference on Dependable Systems and Networks Workshops (DSNW ’11)*, pp. 15–18, 2011.

[33] Q. Nguyen and A. Sood, “A comparison of intrusiontolerant system architectures,” *IEEE Security Privacy*, vol. 9, pp. 24–31, 2011.

[34] A. Herzberg, S. Jarecki, H. Krawczyk, and M. Yung, “Proactive secret sharing or: how to cope with perpetual leakage,” in *Proceedings of the 15th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO ’95)*, pp. 339–352, Springer, Santa Barbara, Calif, USA, 1995.

[35] V. T. Goh, *Intrusion detection framework for encrypted networks [Ph.D. thesis]*, Queensland University of Technology, Brisbane, Australia, 2010.

[36] V. T. Goh, J. Zimmermann, and M. Looi, “Experimenting with an intrusion detection system for encrypted networks,” *International Journal of Business Intelligence and Data Mining*, vol. 5, pp. 172–191, 2010.

[37] M. Conti, R. Di Pietro, L. Mancini, and A. Mei, “Distributed detection of clone attacks in wireless sensor networks,” *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 5, pp. 685–698, 2011.

[38] R. Gennaro, S. Halevi, H. Krawczyk, T. Rabin, S. Reidt, and S. Wolthusen, “Strongly-resilient and non-interactive hierarchical key-agreement in MANETs,” in *Proceedings of the 13th European Symposium on Research in Computer Security*, pp. 49–65, Springer, 2008.

[39] H. Deng, A. Mukherjee, and D. P. Agrawal, "Threshold and identity-based key management and authentication for wireless ad hoc networks," in *Proceedings of the International Conference on Information Technology: Coding and Computing*, pp. 107–111, IEEE Computer Society, 2004.

[40] D. Monderer and M. Tennenholtz, "Distributed games: from mechanisms to protocols," in *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI '99)*, pp. 32–37, American Association for Artificial Intelligence, 1999.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 693639, 10 pages
http://dx.doi.org/10.1155/2013/693639

Research Article

Distributed Information Flow Verification Framework for the Composition of Service Chain in Wireless Sensor Network

Ning Xi, Jianfeng Ma, Cong Sun, Yulong Shen, and Tao Zhang

School of Computer Science and Technology, Shaanxi Key Laboratory of Network and System Security, Xidian University, China

Correspondence should be addressed to Ning Xi; xining@stu.xidian.edu.cn

Received 28 February 2013; Revised 13 April 2013; Accepted 15 April 2013

Academic Editor: Liangmin Wang

Dynamic service composition provides us with a promising approach to cooperate different sensor nodes in WSN to build complex applications based on their basic functions. Usually multiple nodes located in different regions provide data with different security levels, and it is critical to ensure the security of the information flow in the composite services. However, the energy-limited nature of sensor nodes in WSN poses a significant challenge for the centralized information flow verification with which the verification node needs to consume lots of computation and network resources. In this paper, we specify the security constraints for each service participant to secure the information flow in a service chain based in the lattice model and then present a distributed verification framework that cooperates different service participants to verify their information flow policies distributively. The evaluation results show a significant decrease on the verification cost of the single verification node, which provides a better load balance in each sensor node.

1. Introduction

WSN is the key enablers for the development of the Internet of Things (IoT), which is responsible for collecting surrounding context and environment information. In a service-oriented WSN [1, 2], multiple sensor nodes with different basic services, for example, data aggregation, data processing, and decoding, can cooperate with each other to develop new applications rapidly. However, because of the variety and regional characteristics of WSN, the data provided by the sensor nodes have different security levels. When services are composed together, data are transmitted among these nodes, respectively, where an operation in a node assigning high-level data to a low-level object would cause the information leakage with a serious impact on the public safety or personal privacy.

For example, a personal-health helper service can be provided for the healthy advice according to the body status and environments data. Most of the former work, mainly focus on the access control of the individual services [3, 4]. But in a service chain, data may be computed from its prior services which can result in the undesired information leakage. When the collection service is completed, the data collected by the wearable sensors and environmental sensors are delivered to the data processing node, such as mobile phone. Healthy information may leak to untrusted third party through the illegal operations during data processing. So the information flow security is one of the major concerns about the service composition in sensor network environments.

One issue in information flow security of the composite service is the dynamic dependence among various objects in different service participants. Accorsi and Wonnemann [5] use Petri nets to represent the workflow and detect information leaks in workflow descriptions based on static information flow analysis. But this work can only validate the information flow in fixed workflow with static input and output dependences. In service-oriented WSN, there are several candidate services with same functions where the dependences between input and output are different from each other. It is necessary for user to select appropriate service for the secure composition of the service chain. She et al. [6, 7] define transformation factors to measure how likely the output depends on the input data in different candidate services. But it is hard to define the LR, MR, and

HR transformation factors. Therefore, a suitable dependence model is required for the analysis of the information flow in different candidate services.

Another major issue for the information flow verification in WSN is the energy cost of the verification node. Zorgati and Abdellatif [8] and She et al. [9] propose the centralized verification approach against the information flow control policies to ensure an end-to-end security in wired network. However, in WSN, the sensor node is energy limited, and the centralized way consumes lots of energy of the verification node. Yildiz and Godart [10] propose an decentralized service composition approach considering the information flow policies in an inexpensive manner, but its policies are static. Based on the information flow type system, Hutter and Volkamer [11] specify the composition rules to control the security of dynamically computed data and their proliferation to other web services. But it costs extra energy of the sensor node to compile the service code again before the service execution.

In this paper, we present a distributed information flow verification approach applied on the composition of the service chain in wireless sensor network. Our contributions include the following. (1) For the dynamic dependences in service chain, we define the intra and inter dependences among different objects in composite service based on the PDG. (2) We specify the security constraints for each service participant based on the dependences and lattice model. (3) We propose a decentralized information flow verification approach to execute the verification process distributively to provide a better load balance of the sensor nodes in WSN.

The rest of the paper is structured as follows. Section 2 presents the basic definitions of the wireless sensor service system. Section 3 specifies the security constraints for each service participants based on the analysis of the information flow in the service chain. In Section 4, we propose the distributed information flow verification framework based on the secure information flow model. Section 5 evaluates the proposed verification approach. Section 6 concludes the paper.

2. Wireless Sensor Service System

A wireless sensor service system (WSS) is a large-scale distributed environment which consists of multiple wireless sensor nodes, public data resources and security authorities, which is shown in Figure 1. Sensor nodes in WSN can collect these resources, and provide different basic functions, such as data analyzing or processing, which are treated as various services in WSN. There is also a security authority for each data resources for the management of these data security levels which are used for the security verification. The service on each sensor node can be defined as follows.

Definition 1. Each service s_i is a tuple $s_i = \langle id_i, In_i, Out_i, F_i, Ce_i \rangle$, where id_i is the identifier of the service; In_i is the set of input of service; Out_i is the set of the output of service; F_i is composed of a sequence of actions $\langle a_1, a_2, \ldots \rangle$; Ce_i is the certificate of the service which specifies the security properties of service.

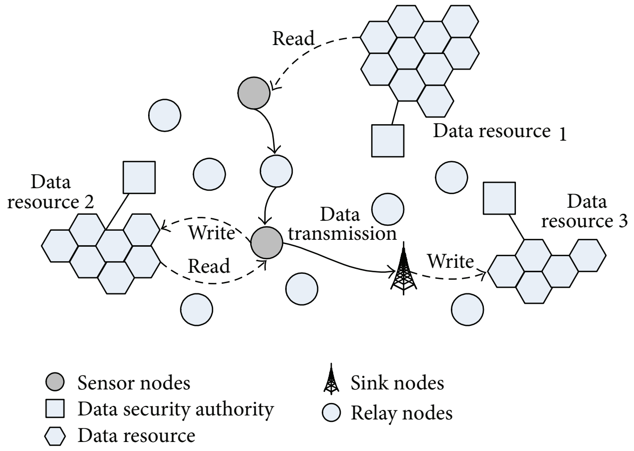


Figure 1: A wireless sensor service system.

In WSS system, various services are provided by different sensor nodes. These individual services can also be combined together to generate a more powerful service. During the execution of composite service, each service node collects data from its local storage or the public resources, processes the input data, and finally provides results to the sink nodes. On the other hand, these nodes may also update the local storage or store to the public data resources in WSS. A composite service can be denoted as a directed graph, where the vertex is the service component and the edge represents an composition relationship from one service to another. In this paper, we investigate a simplified composite service, the service chain, which is defined as follows.

Definition 2. A service chain s_c can be represented as a tuple $s_c = \langle CH, In, Out \rangle$ where CH is a sequence of services $\langle s_1, s_2, \ldots \rangle$; In is the set of input of s_c, s_c; Out is set of output of s_c.

In a service chain s_c, the predecessor of a service s_i can be denoted as s_{i-1}, and the successor of a service s_i is denoted as s_{i+1}. s_0 denotes the node who sends the initial request to s_1, and s_{n+1} denotes the sink node who receives the service result from s_n. Figure 2 shows a simple service chain model.

Due to the dynamic and heterogeneous sensor network environment, it is necessary to select appropriate service to satisfy the different requirements including QoS and security. In this paper, we focus on the verification of the information flow security in composite service chain and providing support for the security enforced selection of services in WSN.

3. Secure Information Flow Model

3.1. Security Label Model. For the information with different sensitivities, we use multilevel security labels to describe the security properties of objects o.

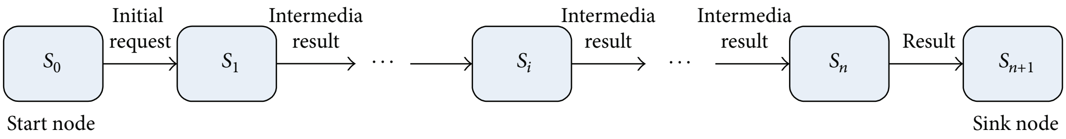


Figure 2: A service chain model.

Definition 3. Security label model is defined as a lattice $(SL, \leq)$, where SL is a finite set of security levels that is totally ordered by $\leq$.

The lattice model is widely used in government or military systems in which the security classes are determined solely from the four security levels: unclassified, confidential, secret, and top secret [12].

For a clear discussion, in this paper, we define that each object o has a provided and required security level, $\mathrm{Pr}(o)$ and $\mathrm{Re}(o)$, which specifies the read and write permissions possessed by o. The provided security labels of the objects can be given by the data owners, which are specified in certificates. And the required security labels of data objects will be computed according to the dependence of the input and output data.

3.2. Information Flow in Service Component. In a service chain, the information flow through s_i is shown in Figure 3. We consider a data flow model in which each service may read from a set of input data objects and write to a set of output data objects. The set of input objects of a service s_i includes all the objects that s_i receives from its predecessor s_{i-1} and all data objects obtained from the public data resources or stored in the local storage in sensor nodes. The set of output objects of s_i includes all the objects that s_i sends to its successor s_{i+1} and all the data objects that s_i updates to the public data resources and the local storage.

For the input information for s_i, there is $In_i = \{In_i^M, In_i^D, In_i^L\}$, where

(i) $In_i^M = \{In_{i,1}^M, In_{i,2}^M, \ldots, In_{i,n}^M\}$ is the set of all input objects that s_i receives from its predecessor $s_{\mathrm{i}-1}$;

(ii) $In_i^D = \{In_{i,1}^D, In_{i,2}^D, \ldots, In_{i,n}^D\}$ is the set of all input objects from the public data resources;

(iii) $In_i^L = \{In_{i,1}^L, In_{i,2}^L, \ldots, In_{i,n}^L\}$ is the set of all input objects located in local storage in sensor nodes.

For the output information for s_i, there is $Out_i = \{Out_i^M, Out_i^D, Out_i^L\}$, where

(i) $Out_i^M = \{Out_{i,1}^M, Out_{i,2}^M, \ldots, Out_{i,n}^M\}$ is the set of all output objects that s_i sends to its successor s_{i+1};

(ii) $Out_i^D = \{Out_{i,1}^D, Out_{i,2}^D, \ldots, Out_{i,n}^D\}$ is the set of all output objects that s_i updates to the public data resources;

(iii) $Out_i^L = \{Out_{i,1}^L, Out_{i,2}^L, \ldots, Out_{i,n}^L\}$ is the set of all output objects that s_i writes to the local storage in sensor nodes.

In order to validate the information flow in s_i, we need to analyze the relationships between the input and output objects. The output Out_i is computed from In_i during the execution of the service function F_i. The syntax of F_i is defined as follows:

$$\begin{aligned} & f ::= a; f, \\ & a ::= \mathrm{skip} \mid \mathrm{input}\,(in, e) \mid \mathrm{var} := e \mid a; a \\ & \qquad \mid \mathrm{if}\,(e)\ \ \mathrm{then}\ a\ \mathrm{else}\ a \\ & \qquad \mid \mathrm{while}\,(e)\ \ a \mid \mathrm{output}\,(out, e)\,, \\ & e ::= \mathrm{var} \mid e\,\mathrm{Re} \\ & R ::= + \mid - \mid = \mid < . \end{aligned} \tag{1}$$

A service function consists of a collection of activities, some of which are the control and computation operations, while some of which are responsible for receiving the inputs from different sources *in* and producing outputs data to the required objects *out*. We can establish the program dependence graph (PDG) [13] of F_i according to its syntax to analyze the relationships among different objects used in F_i. The PDG is defined as follows.

Definition 4. Program dependence graphs (PDG) is a directed graph $\langle V, \vec{E}\rangle$, where the expressions and the activities in F_i constitute the nodes of the graph and the edges express data and control dependences. A data dependence represented by an edge $a \rightarrow_d a'$ means that the activity a assigns variable x which is used in activity a'. A control dependence represented by an edge $e \rightarrow_c a$ means that the execution of a depends on the value of the expression e, which is typically a branch and loop condition.

Once a program dependence graph $\mathrm{PDG} = \langle V, \vec{E}\rangle$ has been constructed, program backward slice [14] is used to analyze the dependences among the different objects that are used in activities and expressions in PDG. Here we use $\mathrm{Dep}(o)$ to represent the obtained dependency set of an object o.

Based on the dependency set $\mathrm{Dep}(o)$, we can compute output object required security level according to the following equations: for $\forall u \in Out_i$,

$$\mathrm{Re}\,(u) = \bigsqcup_{\max} \mathrm{Re}\,(v)\,, \qquad v \in In_i \wedge v \in \mathrm{Dep}\,(u)\,. \tag{2}$$

Based on the previous equation, we can obtain the following.

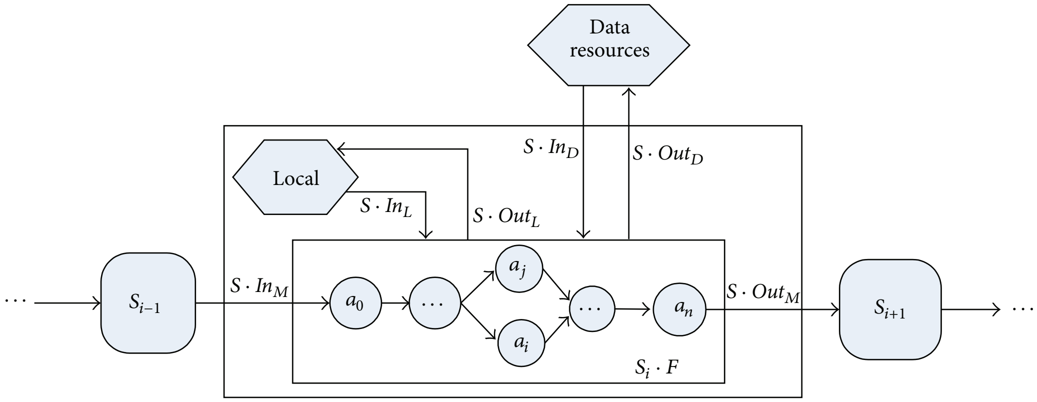


Figure 3: Information flow in service component.

Theorem 5. *For* $\forall u \in Out_i^M$, *there is*

$$\operatorname{Re}(u) \geq \operatorname{Re}(v), \quad \forall v \in In_i \wedge v \in \operatorname{Dep}(u). \tag{3}$$

Each service s_i has different levels of inputs and outputs. The value of the input objects with high-level security label may flow to the low-level output objects during the execution of the service and cause the information leak. Therefore, the definition of the secure information flow in service component is given as follows.

Definition 6. The information flow in service component s_i is considered secure if it satisfies that for $\forall u \in Out_{i,j}^D \cup Out_{i,j}^L$, there are

$$\operatorname{Pr}(u) \geq \operatorname{Re}(u) = \bigsqcup_{\max} \operatorname{Re}(v), \quad v \in In_i \wedge v \in \operatorname{Dep}(u). \tag{4}$$

The previous condition provides that there are no lower level objects in public resources and local storage storing the data with higher security level during the execution of each service.

3.3. Secure Information Flow in Service Chain. Consider the service chain $\langle s_1, s_2, \ldots, s_n \rangle$. The output data sent from s_i to s_{i+1}, Out_i^M, may be dynamically computed from some data stored by sensor node and public data resources, In_i^L and In_i^D, and some data received from s_{i-1}, Out_{i-1}^M. Out_i^M may be further processed by $s_{i+1}, \ldots, s_{j-1}$ and computed into Out_{j-1}^M which is delivered to service s_j, $j > i$. And the dependence between objects belonging to different service components is considered as the interservice dependence. The interservice dependence set of object u, $\operatorname{Dep}_{\text{inter}}(u)$, is defined as follows.

Definition 7. For objects $v \in s_i$ and $u \in s_j$ where $j > i$, v is in $\operatorname{Dep}_{\text{inter}}(u)$ which satisfies one of the following two conditions:

(1) $i = j - 1$:

$$\exists w_1 \in Out_i^M,\ w_2 \in In_j^M,\ w_1 = w_2$$

$$(v \in \operatorname{Dep}(w_1) \vee w_1 = v) \wedge (w_2 \in \operatorname{Dep}(u) \vee w_2 = u),$$

(2) $i \neq j - 1$:

$$\exists w \in s_k, i < k < j$$

$$v \in \operatorname{Dep}_{\text{inter}}(w) \wedge w \in \operatorname{Dep}_{\text{inter}}(u).$$

For two adjacent services s_i and s_j where $i = j - 1$, there are four cases that need to be considered. (1) For $v \in Out_i^M = u \in In_j^M$, there is an interservice dependence between u and v. (2) For $v \in Out_i^M$ and $u \notin In_j^M$, there is an interservice dependence between them if there are objects $w \in In_j^M$ and $v = w$ that u depends on. (3) For $v \notin Out_i^M$ and $u \in In_j^M$, u externally depends on v if there are objects $w \in Out_i^M$ and $u = w$ that depends on v. (4) For $v \notin Out_i^M = u \notin In_j^M$, if there are two objects $w_1 \in Out_i^M$, $w_2 \in In_j^M$ that $w_1 = w_2$, while data object u in s_j depends on w_2, and w_1 depends on v in s_i, we call $u$that externally depends on v.

For two services s_i and s_j where $j > i + 1$, if there is an object w in s_k, $i < k < j$ which w externally depends on v, while u externally depends on w, the dependence between u and v is the interservice dependence.

For a service chain s_c where $In_{sc} = \bigcup\{In_i^L \cup In_i^D\}$ and $Out_{sc} = \bigcup\{Out_i^L \cup Out_i^D\}$, $0 \leq i \leq n+1$, we use s_0 that denotes the start node which sends the initial request to s_1, and s_{n+1} denotes the sink node which receives the service results from s_n. And we assume that $In_0^M = \phi$, $Out_0^D \cup Out_0^L = \phi$, $In_{n+1}^L \cup In_{n+1}^D = \phi$, and $Out_{n+1}^M = \phi$.

Definition 8. The information flow in service chain s_c is considered secure if and only if it satisfies that for $\forall u \in Out_c$, there are

$$\begin{gathered} \operatorname{Pr}(u) \geq \bigsqcup_{\max} \operatorname{Re}(v) \\ v \in In_c \wedge (v \in \operatorname{Dep}(u) \vee v \in \operatorname{Dep}_{\text{inter}}(u)) \\ 1 \leq j < i. \end{gathered} \tag{5}$$

According to the definition of the secure information flow in s_i and s_c, we can obtain the following lemmas and theorems.

Lemma 9. *In a service chain $s_c\langle s_1, s_2, \ldots, s_m\rangle$, $\forall u \in Out_i^M$, $0 \le i \le m$ satisfies*

$$\begin{gathered}\operatorname{Re}(u) \ge \operatorname{Re}(v)\\ v \in In_j \wedge (v \in \operatorname{Dep}(u) \vee \operatorname{Dep}_{\text{inter}}(u)),\\ 0 \le j \le i.\end{gathered} \tag{6}$$

Proof. First, let $m = 1$, then there are two service components s_0 and s_1.

For $\forall u \in Out_0^M$, Theorem 5 provides that $\forall v \in In_0 \wedge v \in \operatorname{Dep}(u)$, and there is $\operatorname{Re}(u) \ge \operatorname{Re}(v)$.

And there is no interservice dependence in s_0, so the lemma is proved.

For $\forall u \in Out_1^M$, there are two cases to consider.

Case 1. $j = 1, \forall v \in In_1 \wedge v \in \operatorname{Dep}(u)$. Theorem 5 provides $\operatorname{Re}(u) \ge \operatorname{Re}(v)$.

Case 2. $j = 0, \forall v \in In_0 \wedge v \in \operatorname{Dep}_{\text{inter}}(u)$. In this case, the definition of the interservice dependence provides $\exists w_1$ and w_2 where

$$\begin{gathered}w_1 \in Out_0^M = w_2 \in In_1^M,\\ v \in \operatorname{Dep}(w_1) \wedge w_2 \in \operatorname{Dep}(u).\end{gathered} \tag{7}$$

Theorem 5 provides that

$$\operatorname{Re}(u) \ge \operatorname{Re}(w_2), \tag{8}$$

$$\operatorname{Re}(w_1) \ge \operatorname{Re}(v). \tag{9}$$

And there is

$$\operatorname{Re}(w_2) = \operatorname{Re}(w_1). \tag{10}$$

Based on (8), (9), and (10), we can get

$$\operatorname{Re}(u) \ge \operatorname{Re}(v). \tag{11}$$

In a conclusion, when $m = 1$, the lemma is proved.

Then we suppose that the lemma is true when $m = n - 1$; that is, for $\forall u \in Out_i^M$, $0 \le i \le n - 1$, there are

$$\begin{gathered}\operatorname{Re}(u) \ge \operatorname{Re}(v),\\ v \in In_j \wedge (v \in \operatorname{Dep}(u) \vee v \in \operatorname{Dep}_{\text{inter}}(u)),\\ 0 \le j \le i.\end{gathered} \tag{12}$$

And the case that $m = n$ is proved as follows: for $\forall u \in Out_n^M$, there are also two cases to consider.

Case 1. $j = i, \forall v \in In_n \wedge v \in \operatorname{Dep}(u)$. In this case, Theorem 5 provides $\operatorname{Re}(u) \ge \operatorname{Re}(v)$.

Case 2. $0 \le j < i, \forall v \in In_j \wedge v \in \operatorname{Dep}_{\text{inter}}(u)$. In this case, the definition of the interservice dependence provides $\exists w_1, w_2$ where

$$\begin{gathered}w_1 \in Out_{n-1}^M = w_2 \in In_n^M,\\ (v \in \operatorname{Dep}(w_1) \vee v \in \operatorname{Dep}_{\text{inter}}(w_1)) \wedge w_2 \in \operatorname{Dep}(u).\end{gathered} \tag{13}$$

Theorem 5 provides that

$$\operatorname{Re}(u) \ge \operatorname{Re}(w_2). \tag{14}$$

The previous assumption provides that for $0 \le i \le n-1$, there is

$$\operatorname{Re}(w_1) \ge \operatorname{Re}(v), \tag{15}$$

and there is

$$\operatorname{Re}(w_2) = \operatorname{Re}(w_1) \tag{16}$$

Based on (14), (15), and (16), we can get

$$\operatorname{Re}(u) \ge \operatorname{Re}(v). \tag{17}$$

In a conclusion, when $m = n$, the lemma is proved. □

Lemma 10. *If the information flow of each service in first m step of s_c is secure, $\forall u \in Out_i^D \cup Out_i^L$, $0 \le i \le m$ satisfies*

$$\begin{gathered}\operatorname{Pr}(u) \ge \operatorname{Re}(v),\\ v \in In_j \wedge (v \in \operatorname{Dep}(u) \vee \operatorname{Dep}_{\text{inter}}(u)),\\ 0 \le j \le i\end{gathered} \tag{18}$$

Proof. For $\forall u \in Out_i^D \cup Out_i^L$, there are also two cases to consider.

Case 1. $j = i, \forall v \in In_i \wedge v \in \operatorname{Dep}(u)$. In this case, the secure information flow Definition 6 provides $\operatorname{Pr}(u) \ge \operatorname{Re}(v)$.

Case 2. $0 \le j < i, \forall v \in In_j \wedge v \in \operatorname{Dep}_{\text{inter}}(u)$. In this case, the definition of the interservice dependence provides $\exists w_1, w_2$ where

$$\begin{gathered}w_1 \in Out_{n-1}^M = w_2 \in In_n^M,\\ (v \in \operatorname{Dep}(w_1) \vee v \in \operatorname{Dep}_{\text{inter}}(w_1)) \wedge w_2 \in \operatorname{Dep}(u).\end{gathered} \tag{19}$$

The secure information flow Definition 6 provides

$$\operatorname{Pr}(u) \ge \operatorname{Re}(u). \tag{20}$$

Theorem 5 provides that

$$\operatorname{Re}(u) \ge \operatorname{Re}(w_2). \tag{21}$$

And the Lemma 9 provides that

$$\operatorname{Re}(w_1) \ge \operatorname{Re}(v). \tag{22}$$

And there is

$$\operatorname{Re}(w_2) = \operatorname{Re}(w_1). \tag{23}$$

Based on (20), (21), (22), and (23), we can get

$$\operatorname{Pr}(u) \ge \operatorname{Re}(v). \tag{24}$$

In a conclusion, the lemma is proved. □

Theorem 11. *For a service chain s_c, if the information flow in each service component s_i is considered secure, the flow in the service chain is secure.*

Proof. Let $m = n + 1$, and the theorem is proved based on Lemma 10. □

4. Distributed Information Flow Verification Framework for Wireless Service Composition

4.1. Information Flow Verification Framework. For a service chain $s_c = \langle s_0, s_1, \ldots, s_n, s_{n+1}\rangle$, there are several candidate services but different implementations by developers for each service step s_i. In the distributed information flow verification framework, each sensor node is only responsible for validating its next-step candidate service node $s_{i,j}$, which can balance the energy cost on a single verification node. The distributed information flow verification framework is shown in Figure 4.

In our framework, Service Authorization Centre (SAC) is a trusted third party for service certificate generation before the deployment of the sensor node. There are two phases for the verification of the information flow: service certificate setup and service verification phase. The service certificate that specifies the security properties of the service, that is, the dependence between the service input and output, is first generated and signed by a SAC. During the service composition procedure, the service composer obtains the required service certificates, and verifies the information flow in candidate nodes. These two phases are detailed in the following sections.

4.2. Service Certificate Setup. Service certificate setup is the preparation phase of the verification process, which is shown in Figure 5. In this phase, service developer submits authorization request containing service function code in service node to SAC. And then the generated service certificate Ce is installed on the sensor node with the service. Considering the complexity and security of the service code transmission, the authorization process is executed by the offline mode between the service developer and SAC, which does not need to consume extra energy of the sensor node.

Definition 12. A service certificate Ce is a tuple $\langle CA, s, De\rangle$, where CA is the issuer, that is, SAC; s is the service identifier; De is the set of statements that describe the output data dependence.

The service certificate Ce specifies the attributes of the service including the service identifier, the dependence between input and output objects in the service function. Regarding the PDG construction of service function, SAC uses the algorithms presented in [13] to generate the PDG. Once a program dependence graph PDG $= \langle V, \vec{E}\rangle$ has been computed, a dependence set can be established for each node $x \in V$ by using intraprocedural backward slice [14], written $De(x)$ containing the set of all nodes in PDG from which x can be reached as follows: $De(x) = \{y \mid y \rightarrow_* x\}$. In this paper we mainly consider the dependences between the input and output objects in the PDG nodes; that is, $De_{in}(Out_{i,k}) = \{v \mid v \in De(Out_{i,k}) \wedge v \in In_i\}, 0 \le i \le n+1, 0 \le k \le |Out_i|$. For each $Out_{i,k} \in Out_i$, its input dependence is written into the certificate. Finally, the certificate is signed by SAC and sent to the service node. Then the service certificate setup phase is complete. The Algorithm 1 is shown as follows.

When there is a request for the service, the node needs to send its certificate to the composer for its information flow verification. The provided security levels of the public and local input data and output objects are also required to be sent to the verification node. If the realization of the service is changed, for example, a new version service is published, the service needs to be authorized by SAC again and reinstalled on the sensor node.

4.3. Service Verification. Service verification is a vital phase in which the verification node requires the service certificates and validates the candidate nodes against the information flow control policies. The verification procedure is shown in Figure 6. During the verification process, service composer s_{i-1}, required for the service certificate and the provided security levels of the public and local data and objects first. Then the composer computes the required security levels of the output objects and then validates whether they satisfy the security constrains.

4.3.1. Required Security Level Computation. According to the secure information flow definition in service chain, the required security levels of the data objects need to be computed first. The required security levels of the objects in each service $s_{i,j}$ are computed according to the following three computation rules (CR):

CR 1 For $\forall u \in In_i^D \cup In_i^L$, $\mathrm{Re}(u) = \mathrm{Pr}(u)$;

CR 2 For $\forall u \in In_i^M$, $\mathrm{Re}(u) = \mathrm{Pr}(v)$ where $v \in Out_{i-1}^M \wedge v = u$;

CR 3 For $\forall u \in Out_i$, $\mathrm{Re}(u) = \sqcup_{\max} \mathrm{Re}(v) \quad v \in In_i \wedge v \in \mathrm{Dep}(u)$.

CR 1 specifies that the required security levels of the input objects from public and local storage are equal to their provided security levels. CR 2 specifies that the required security levels of the input objects from predecessor are determined by that of the output objects in s_{i-1}. CR 3 specifies that the required security levels of the output objects are computed from that of the input objects that the output depends on.

4.3.2. Service Verification. During the service verification, the information flow control policy (IFCP) specifies how to validate a candidate service $s_{i,j}$. Based on the security label model and the definition of the secure information flow in each service, we define the information flow control policies in each service s_i as follows:

IFCP 1 For $\forall u \in Out_i^D$, $\mathrm{Pr}(u) \ge \mathrm{Re}(u) \Rightarrow$ true, $\mathrm{Pr}(u) < \mathrm{Re}(u) \Rightarrow$ false;

IFCP 2 For $\forall u \in Out_i^L$, $\mathrm{Pr(u)} \ge \mathrm{Re(u)} \Rightarrow$ true, $\mathrm{Pr(u)} < \mathrm{Re(u)} \Rightarrow$ false.

Based on the required security level computation rules and information flow control policies, verification node can validate the candidate sensor node $s_{i,j}$ in a service chain. The Algorithm 2 is shown as follows.

```
Input: Service $s_i$ $\langle id_i, dom_i, In_i, Out_i, F_i, Ce_i\rangle$.
Output: Service certificate of $Ce_i$ $\langle CA, s, De\rangle$.
\ \Var $(x)$ represents the variables objects in $x$ statement
$Ce_i \cdot s = id_i$
PDG $G\left\langle V, \vec{E}\right\rangle$
generatePDG($F_i, G$)
for each output node $x \in G \wedge \mathbf{Var}(x) \in Out_i$ do
    BS(x)=backwardSlice(x)
    for each $y \in \mathbf{BS}(x)$ do
        if $\mathbf{Var}(y) \in In_i$ then
            pushInto($Ce_i \cdot De(\mathbf{Var}(x), \mathbf{Var}(y))$)
        end if
    end for
end for
signature($Ce_i \cdot CA, SAC$)
return $Ce_i$
```

Algorithm 1: Service_Certificate_Set_Up().

```
Input: Re($Out^M_{i-1}$), Candidate Service Set $S_i$, Pr($In^D_{i-1}$), Pr($In^L_{i-1}$), Pr($Out^D_{i-1}$), Pr($Out^L_{i-1}$)
Output: Passed Service Set $S_{p,i}$.
\ \exOutput ($In^M_{i,j}$) represents $In^M_{i,j}$'s corresponding output in its predecessor
\ \filterService ($S_p, S_{i,j}$) represents filtering the unsatisfied candidate service $s_{i,j}$ from $S_p$
$S_{p,i} = S_i$
for each $s_{i,j} \in S_i$ do
    requestCert($s_{i,j}, Ce$)
    for each $u \in In^D_{i,j}, v \in In^L_{i,j}, w \in In^M_{i,j}$ do
        $\mathrm{Re}(u) = \mathrm{Pr}(u)$
        $\mathrm{Re}(v) = \mathrm{Pr}(v)$
        $\mathrm{Re}(w) = \mathrm{Re}(\mathbf{exOutput}(w))$
    end for
    for each $u \in Out^D_i, v \in Out^L_i$ do
        for each $w \in De(u)$ do
            $\mathrm{Re}(u) = \sqcup_{\max} \mathrm{Re}(w)$
        end for
        if $\mathrm{Pr}(u) < \mathrm{Re}(u)$ then
            filterService($S_{p,i}, S_{i,j}$)
            break;
        end if
        for each $w \in De(v)$ do
            $\mathrm{Re}(v) = \sqcup_{\max} \mathrm{Re}(w)$
        end for
        if $\mathrm{Pr}(v) < \mathrm{Re}(v)$ then
            filterService($S_{p,i}, S_{i,j}$)
            break;
        end if
    end for
end for
return $S_{p,i}$
```

Algorithm 2: Service_Verification().

4.4. Decentralized Information Flow Verification Algorithm for the Service Chain. For each step verification, verification node obtains the passed candidate service set $S_{p,i}$, then the verification node will notice these passed sensor nodes to verify the following candidate services. And there are three types of messages for the synchronization of the verification procedure, that is, *start_message*, *success_message*, and *fail_message*. *start_message* is used to allow the candidate service $s_{i,j}$ to execute the *Verify_ServiceChain*() procedure. When the nodes in service chain all pass the service verification process, *success_message* with the executable path is sent to inform its requestor s_0. During each step verification *fail_message* will be sent to the predecessor of the verification node when there are no candidate services passed

```
Input: Re($Out_{i-1}^{M}$), Candidate Service Set $S_{i+1}$.
Output: Secure Execution Path $P$
if $i \neq 1$ then
    wait start_message
end if
if $i = n$ then
    send success_message with secure execution path $P$ to the requestor
else
    push $S_{i,j}$ into the Execution Path $P$
    $S_{p,i+1} = Service\_Verification(Re(Out_{i-1}^{M}), S_{i+1})$
    if $S_{p,i+1} = \phi$ then
        send fail_message to its predecessor
    else
        for each $s_{i+1,k} \in S_{p,i+1}$ do
            send start_message to $s_{i+1,k}$
        end for
    end if
end if
```

Algorithm 3: Verify_ServiceChain().

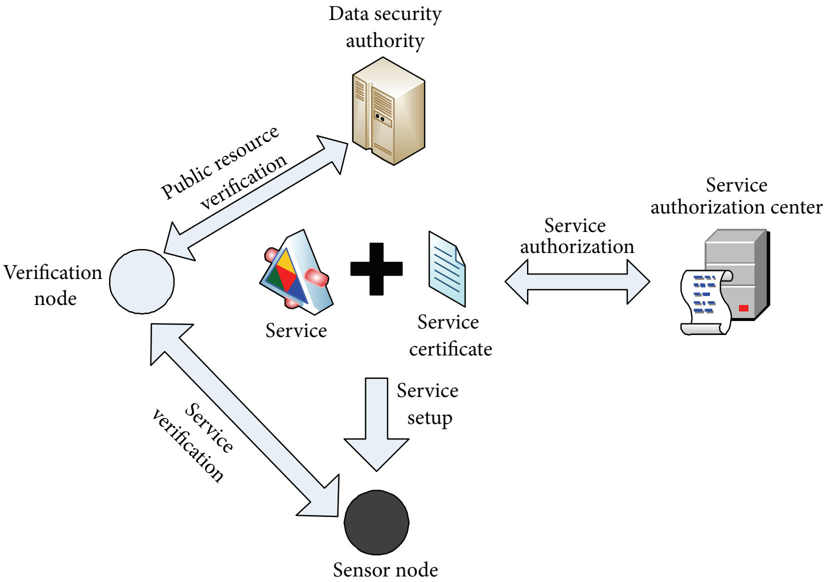


Figure 4: Decentralized information flow verification framework.

the verification in next step. The Algorithm 3 is presented as above.

5. Experiments and Evaluations

This paper studies distributed information verification framework for the service composition in WSN. Through the security analysis in Section 3, the information flow security can be ensured by the Theorem 11. In this section, we investigate the impact of distributed service verification on the sensor node's cost including verification time and communication effort. A centralized approach implements the service verification work by a single sensor node. We test both approaches with NS-3 [15] in multiple scenarios. Table 1 shows further details about the simulation configuration.

Figure 7 shows the computation time on the verification node. In the centralized way, time rises vastly with the increase of the candidate service number. That is because the execution paths that need to be verified are increased at an exponential rate. However, time increases slowly in the distributed way because there is no significant variations on the candidate nodes that each sensor node needs to verify.

Figure 8 shows the communication effort on the verification node. In Figure 8, the communication effort in the centralized way is evidently higher than that used the distributed way. That is because the single verification node needs to

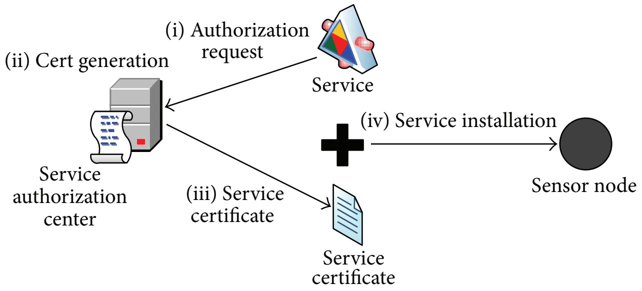

(i) Authorization request
(ii) Cert generation
Service
(iv) Service installation
Sensor node
Service authorization center
(iii) Service certificate
Service certificate

Figure 5: Service certificate setup phase.

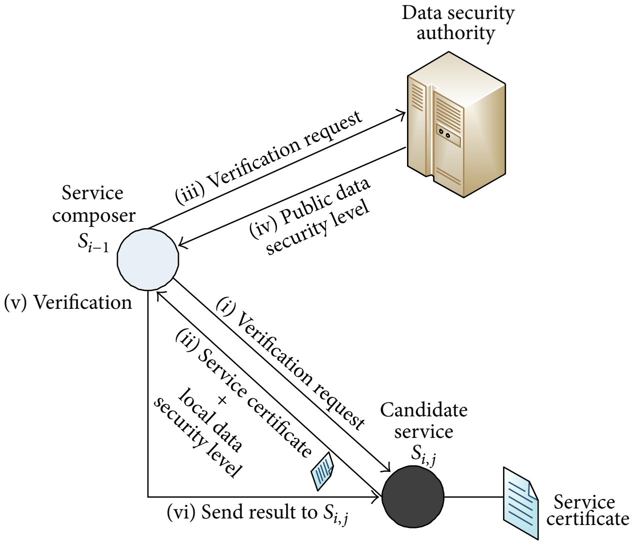

Data security authority
(iii) Verification request
(iv) Public data security level
Service composer S_{i-1}
(v) Verification
(i) Verification request
(ii) Service certificate + local data security level
Candidate service $S_{i,j}$
Service certificate
(vi) Send result to $S_{i,j}$

Figure 6: Service Verification Phase.

Table 1: Simulation Configuration.

General	
Simulator	NS-3
Field (m^2)	100 × 100
Radio type	Zigbee
Service step	4
Simulation duration (s)	1000
Random	
Node placement	
Node movement	
Security level	Unclassified, confidential, secret, and top secret
Controlled	
Candidate number	5–10
Verification mode	Centralized, decentralized

communicate with all other service nodes in centralized way, while it just needs to communicate with the next-step service nodes which can decrease the communication effort and save the energy of the sensor nodes.

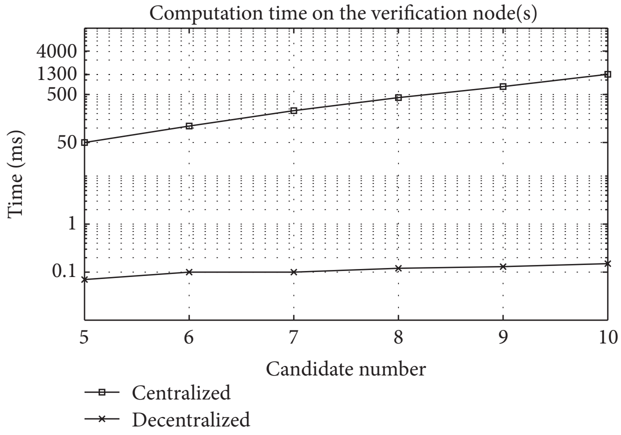


Figure 7: Computation time on the verification node(s).

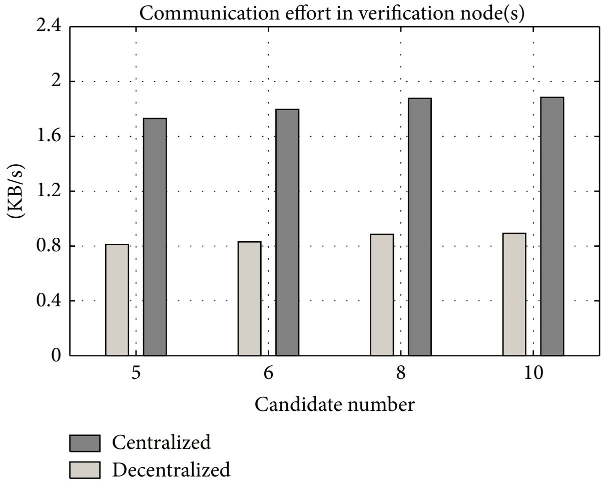


Figure 8: Communication effort on the verification node(s).

6. Conclusion

In this paper, we specify the security constraints for each service participant based on the partial order model and propose a decentralized information flow verification approach that cooperates each sensor node to verify the information flow security distributively and builds up secure service chains in wireless sensor environments. Through the simulation on NS-3, the result shows that this approach can decrease the cost of the sensor nodes effectively.

Acknowledgments

This work is supported by Program for the Key Program of NSFC-Guangdong Union Foundation (U1135002), Major National S&T Program (2011ZX03005-002), National Natural Science Foundation of China (60872041, 61072066), and the Fundamental Research Funds for the Central Universities (JY10000903001, JY10000901034, and K5051203010).

References

[1] A. Rezgui and M. Eltoweissy, "Service-oriented sensor-actuator networks," *IEEE Communications Magazine*, vol. 45, no. 12, pp. 92–100, 2007.

[2] D. Gračanin, M. Eltoweissy, A. Wadaa, and L. A. DaSilva, "A service-centric model for wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 6, pp. 1159–1165, 2005.

[3] E. Bertino, A. C. Squicciarini, and D. Mevi, "A fine-grained access control model for Web services," in *Proceedings of the IEEE International Conference on Services Computing (SCC '04)*, pp. 33–40, September 2004.

[4] R. Bhatti, E. Bertino, and A. Ghafoor, "A trust-based context-aware access control model for Web-services," in *Proceedings of the IEEE International Conference on Web Services (ICWS '04)*, pp. 184–191, July 2004.

[5] R. Accorsi and C. Wonnemann, "Static information flow analysis of workflow models," in *INFORMATIK, 2010-Business Process and Service Science-Proceedings of ISSS and BPSC*, vol. 177 of *Lecture Notes in Informatics*, pp. 194–205, 2010.

[6] W. She, I. L. Yen, B. Thuraisingham, and E. Bertino, "The SCIFC model for information flow control in web service composition," in *Proceedinds of the IEEE International Conference on Web Services (ICWS '09)*, pp. 1–8, July 2009.

[7] W. She, I. L. Yen, B. Thuraisingham, and E. Bertino, "Policy-driven service composition with information flow control," in *Proceedings of the IEEE 8th International Conference on Web Services (ICWS '10)*, pp. 50–57, July 2010.

[8] H. Zorgati and T. Abdellatif, "SEWSEC: a SEcure web service composer using Information flow control," in *Proceedings of the 6th International Conference on Risks and Security of Internet and Systems (CRiSIS '11)*, 2011.

[9] W. She, I. L. Yen, B. Thuraisingham, and S. Y. Huang, "Rule-based run-time information flow control in service cloud," in *Proceedings of the IEEE International Conference on Web Services (ICWS '11)*, pp. 524–531, 2011.

[10] U. Yildiz and C. Godart, "Information flow control with decentralized service compositions," in *Proceedongs of the IEEE International Conference on Web Services (ICWS '07)*, pp. 9–17, July 2007.

[11] D. Hutter and M. Volkamer, "Information flow control to secure dynamic web service composition," *Science Security in Pervasive Computing*, vol. 3934, pp. 196–210, 2006, Lecture Notes in Computer.

[12] D. E. Denning, "A lattice model of secure information flow," *Communications of the ACM*, vol. 19, no. 5, pp. 236–243, 1976.

[13] J. Ferrante, K. J. Ottenstein, and J. D. Warren, "The program dependence graph and its use in optimization," *ACM Transactions on Programming Languages and Systems*, vol. 9, no. 3, pp. 319–349, 1987.

[14] G. Snelting, T. Robschink, and J. Krinke, "Efficient path conditions in dependence graphs for software safety analysis," *ACM Transactions on Software Engineering and Methodology*, vol. 15, no. 4, pp. 410–457, 2006.

[15] Open Source Project, NS-3 Project, http://www.nsnam.org/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 693930, 13 pages
http://dx.doi.org/10.1155/2013/693930

Research Article

Expedite Privacy-Preserving Emergency Communication Scheme for VANETs

Long Chen,[1] Xuedan Jia,[2] Lixia Meng,[3] and Liangmin Wang[3]

[1] *Automotive Engineering Research Institute, Jiangsu University, Zhenjiang 212013, China*
[2] *School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China*
[3] *Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, Anhui University, Hefei 230039, China*

Correspondence should be addressed to Xuedan Jia; laura_j@163.com

Received 25 February 2013; Accepted 13 April 2013

Academic Editor: Yulong Shen

The existing communication schemes are often unusable in natural disasters and public emergencies. But requirements of information collection and data transmission in emergency scenario are very imperative. Thus, sensor networks ad hoc networks are required in the emergency communication systems. For example, rescue vehicles equipped with wireless communication devices, sensors, and cameras are regularly used to collect and transmit the real-time information for the rescue action. The paper focuses on security solutions for the vehicular ad hoc networks (VANETs) in the emergency communication cases, in which the communication infrastructures are not always available. An expedite privacy-preserving emergency communication (EPEC) scheme is presented for the vehicles to securely connect with the others in the neighbor area even when the trusted infrastructures are destroyed by the disaster. EPEC satisfies conditional privacy preservation requirements, in which both lightweight signature and batch verification are employed to provide efficiency. We also show the proof of the security, feasibility, and efficiency of our EPEC by the theoretical and experimental analyses.

1. Introduction

Information perception, transmission, and processing are big problems in emergency rescues. The wireless sensor networks and ad hoc networks are expected to be used in these scenarios, but the network deployment and resource sharing face new problems of node connection, network security, and human privacy. As we know, vehicles are often used in disaster rescues, such as in Sichuan Earthquake [1] and Tōhoku Earthquake [2]. Then, we can equip rescue vehicles with sensors, cameras, and wireless communication devices, which combine wireless sensor networks with mobile ad hoc networks and will be more applicable in real-time data collection, transmission, and processing. Vehicle communication in rescue action should maintain the primary requirements of security, privacy, and efficiency, which contain mutual authentication, conditional privacy preservation, internal attacks prevention, and expedite authentication in emergency communication. The communication units of these rescue vehicles based networks are similar to typical vehicular ad hoc networks (VANETs), but the network structures are different because there are no road-side units (RSUs). Thus, the new networks are the same as traditional VANETs without the supports of RSUs, and we study the communication scheme for the emergency scenario where the vehicles cannot connect with an RSU.

VANETs mainly consist of mobile vehicles equipped with wireless communication devices and the road-side units. As shown in Figure 1, vehicles can communicate with one another (V2V) and with the RSUs (V2R) by means of the Dedicated Short-Range Communication protocol [3]. Figure 1 also shows that VANET is a subnet under the architecture of the Internet of Things, for VANET is connected with a trust authority (TA) and many Application Servers through the Internet. VANETs can provide drivers with traffic information to defend against dangers and traffic congestions [4], as well as entertainment information to improve the driving experience [5]. Privacy preservation [6] and expedite authentication [7] are two critical issues in VANETs communication, where privacy is conditional

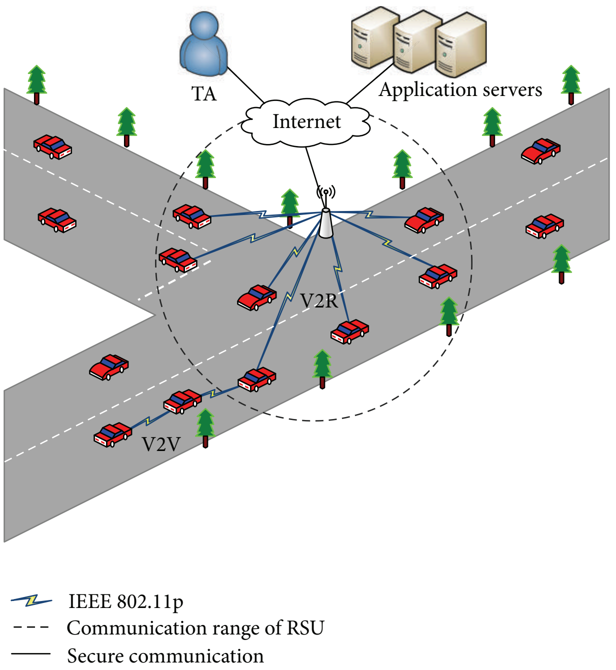


Figure 1: VANETs model.

as TA is allowed to reveal any entity's real identity and revoke it from the network. It is because on one hand, drivers would not be willing to join VANETs if their private information could not be well preserved. And on the other hand, malicious vehicles should be removed from the network timely. In addition, expedite authentication can reduce packet loss and ensure smooth operations of VANETs. Many communication schemes have been proposed for conditional privacy preservation and expedite authentication in VANETs. But these reported schemes are not suitable for emergency communication in the rescue scenario we described earlier, because they take RSU as an important part of VANETs.

We will present an expedite privacy-preserving emergency communication (EPEC) scheme for disaster rescues, where RSUs are unavailable. The remainder of this paper is organized as follows. Section 2 overviews the related works in VANETs. In Section 3, we describe the system model and give basic presuppositions used in the proposed scheme. The proposed EPEC is thoroughly described in Section 4. The security analysis of the proposed scheme is presented in Section 5. Performance evaluation is given in Section 6, followed by the conclusion in Section 7.

2. Related Works

Many related studies have been reported in VANETs based on different cryptographic systems, such as public key infrastructure (PKI) based signature, group signature, and identity-based signature. In terms of disaster rescue, the primary functional requirements of communication do not change, but these requirements should be realized without fixed infrastructures. Next, we analyze the related works in the aspects of EPEC's requirements aforementioned.

Mutual authentication is the primitive property required in VANETs communication, which can be achieved through digital signature. Raya and Hubaux proposed PKI-based schemes [16, 17] in 2005 and 2007, respectively to realize authenticity in VANETs. But PKI-based scheme is not suitable for VANETs because of the management overhead of certificates.

The commonly used techniques for conditional privacy preservation are group signature and Mix-zone pseudonym—changing pseudonym within specified region, that is, Mix-zone. Lin et al. [8] adopted short group signature [18], and TACK [9] and TARI [19] adopted group signature with verifier-local revocation [20], to realize conditional privacy preservation. For the shortcoming of the demand for group manager, ring signature [21] met conditional privacy preservation without group managers. However, what makes these schemes inapplicable to VANETs is that the verification cost for group/ring signature is very high. For authentication efficiency, Lu et al. [10] achieved conditional privacy preservation by pseudonym signature. Reference [22] aimed to establish Mix-zones at social points, allowing all vehicles in the Mix-zone to change pseudonyms at the same time. However, only when given a reasonable large number of vehicles in the Mix-zone, the privacy requirements can be well protected due to pseudonyms updating.

In aspect of internal attacks prevention, some existing schemes perform well while some fail to realize internal attacks prevention requirement. IBV [11] devised identity-based signature to realize unidirectional V2R authentication, which could not achieve internal attacks prevention or conditional privacy preservation requirements. Subsequently, TSVC [23] based on the TESLA [24] achieved fast authentication with internal attacks prevention using message authentication code (MAC). However, its drawback is that packet loss ratio increases with the speed of vehicles. ABAKA [12] was dedicated to entertainment services in VANETs, but internal attacks prevention requirement was not realized as users shared the same secret.

In addition to aforementioned requirements, verification efficiency, that is, expedite authentication, is another requisite in VANETs. One-by-one message verification is characterized by being simple to use. In RAISE [13], RSU verified messages one by one and broadcasted 128 bytes for each valid message, which caused severe inefficiency. Cooperative authentication is a method of raising verification efficiency. In COMET [13], verifiers verified message with probability of p, and if it was invalid, they notified the neighbors of the result. CMAP [14] chose verifiers based on location, and nonverifiers waited for the verifiers' results. Cooperative verification can raise verification efficiency, as vehicles do not need to verify every message received. However, due to the uncertainty of the vehicle speed and road conditions, the scalability and practicality of these schemes face questioning. Batch verification is another effective method to improve efficiency as it allows verifiers to authenticate a group of signatures at the same time. CPAS [15] and MLAS [25] verified messages in batch based on bilinear pairing operations. But bilinear pairing operation is of large computational overhead. In addition, for CPAS, there exists the problem of key escrow as private key generator (PKG) is essential to generate user private key. Malina et al. [26] adopted group signature supporting batch verification, and messages were classified in different priority level in order to improve verification efficiency. However, the group signature also suffers from high computational overhead. Even more important, all the previous schemes are not applicable to emergency communication.

Virtually, VANETs-based emergency communication system is a special kind of VANETs model without RSUs. But currently there is no related research. Overall, current works in VANETs cannot satisfy all the functional requirements in this scenario. And it is shown in Table 1, where “√” indicates “realized” and “×” indicates “unrealized”, respectively. It is obvious that no scheme is applicable to our scenario. To solve this problem, we present an expedite privacy-preserving emergency communication scheme. In terms of conditional privacy preservation, an exclusive secret key is established through secure protocol between the vehicle and TA in the reregistration phase, only allowing TA to track malicious vehicles. As for expedite authentication, lightweight signature and batch verification are combined to reduce computation and communication overhead, providing fast and efficient communication.

The proposed EPEC achieves the following five aspects of requirements. (1) *No Fixed RSU*. Vehicles can reregister with TA dynamically, and then proceed with mutual authentication in broadcast communication without fixed RSU. (2) *Conditional Privacy Preservation*. EPEC allows only trusted TA to trace vehicles' real identities. (3) *Internal Attacks Prevention*. The malicious vehicles can only cause limited damage to the whole networks. (4) *High Efficiency*. Lightweight signature and batch verification are combined to reduce computation overhead. And identity-based signature is adopted to save communication overhead. (5) *Vehicle Group Communication*. Vehicles can form a group to communicate timely with each other.

3. System Model and Preliminaries

In this section, we formalize the specific system model for VANETs-based emergency communication and the basic presuppositions used in the proposed scheme EPEC and identify the design objectives.

3.1. System Model. VANETs for disaster rescue are different from the regular vehicular network model in Figure 1. According to the accurate circumstance of disaster rescue, a two-layer network model is proposed, as shown in Figure 2. The upper layer comprises the trust authority TA and the Application Server, while the lower layer is composed of rescue vehicles including emergency communication cars and regular vehicles. TA is responsible for issuing real identification RID and public/private key pairs to all entities. Most importantly, TA is always trusted and can never be compromised. The Application Server is responsible for information analyses and feedbacks. The emergency communication cars contain ambulances, fire trucks, and so on, AMBs for short in EPEC. AMBs are allocated with powerful hardware facilities with longer communication range and stronger computation capability than regular vehicles. They are authorized by TA and can arrive at the target zone after the disaster, responsible for networking the regular vehicles.

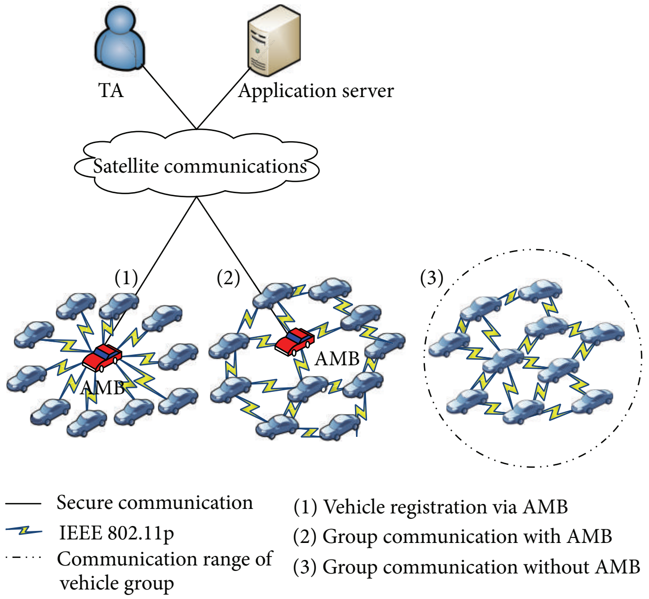

TA
Application server
Satellite communications
(1)
(2)
(3)
AMB
AMB
Secure communication
IEEE 802.11p
Communication range of vehicle group
(1) Vehicle registration via AMB
(2) Group communication with AMB
(3) Group communication without AMB

Figure 2: System model.

In the framework we proposed, there are two different kinds of communication: the reregistration communication and the broadcast communication, where the broadcast communication can also be divided into vehicle to AMB communication and vehicle group communication. The left-most group 1 in Figure 2 indicates vehicles reregistration with TA via AMB. Vehicles reregister themselves with the TA via AMB and get the exclusive secret key used to generate pseudonyms. Groups 2 and 3 show the communication with and without AMB available, respectively. EPEC does not require AMBs to cover the entire network. They may move to another place after networking the vehicles in some place. The dashed line indicates the group communication range, within which the vehicles communicate with each other by the wireless communication standard IEEE 802.11p. AMBs communicate securely with TA and the Application Server by satellite communications.

3.2. Basic Presuppositions. An elliptic curve is a cubic equation of the form $y^2 + axy + by = x^3 + cx^2 + dx + e$, where a, b, c, d, and e are all real numbers. In an elliptic curve cryptography (ECC) system, the elliptic curve equation is defined as the form of $E_q(a, b)$: $y^2 = x^3 + ax + b (\bmod q)$, over a prime finite field $\mathbb{F}_q$, where $a, b \in \mathbb{F}_q$, $q > 3$, and $4a^3 + 27b^2 \neq 0 (\bmod q)$. In general, the security of ECC depends on the difficulties of the following problems [27, 28]. So far, no polynomial algorithm is capable to solve these problems.

Table 1: Comparison of related works.

Schemes	Mutual authentication	Conditional privacy preservation	Internal attacks prevention	Authentication efficiency		Emergency communication		
				Primitive overhead	Batch verification	No RSU	Dynamic re-registration	Vehicle group communication
GSIS [8]	√	√	√	High	×	×	×	×
TACK [9]	√	√	√	High	×	×	√	×
ECPP [10]	√	√	√	High	×	×	√	×
IBV [11]	×	×	×	High	√	×	×	×
ABAKA [12]	√	×	×	Low	√	×	×	×
RAISE [13]	√	√	√	Low	×	×	×	×
COMET [13]	√	×	√	Low	×	√	×	×
CMAP [14]	√	√	√	High	×	×	×	√
CPAS [15]	√	√	√	High	√	×	×	×

Definition 1 (*Elliptic Curve Discrete Logarithm Problem (ECDLP)*). Given two points P and Q over $E_q(a,b)$, the elliptic curve discrete logarithm problem (ECDLP) finds an integer $x \in \mathbb{F}_q$ such that $x \cdot P = Q$.

Definition 2 (*Computational Diffie-Hellman Problem (CDHP)*). Given three points P, sP, and tP over $E_q(a,b)$ for $s, t \in \mathbb{F}_q$, the computational Diffie-Hellman problem (CDHP) finds the point $(st)P$ over $E_q(a,b)$.

3.3. Design Objectives. In terms of emergency communication, the VANETs system needs to satisfy all requirements in the condition of no fixed infrastructures. The precise functional requirements are presented as following.

Dynamic Reregistration without Fixed RSUs. As fixed RSUs may have been destroyed during the disaster, vehicles reregister themselves with TA via AMBs to get the system public parameters. However, AMBs do not cover the entire network and vehicle may need to update the secret key. So vehicles should be able to dynamically reregister with TA when AMB is available.

Mutual Authentication. AMB needs to authenticate itself to regular vehicles on arriving at the disaster area. And vehicles should also be authenticated when they reregister with TA through AMB. In addition, during broadcast phase vehicles need to authenticate each other to ensure that the messages are indeed sent by legitimate entities to guard against the impersonation attack. The mutual authentication is achieved by signature, which also enforces message integrity checking.

Conditional Privacy Preservation. The real identity of a vehicle should be hidden from any entity during the communication process in order to protect the sender's private information. But on the other hand, when vehicles are found to abuse the network or are in dispute for an accident, it is necessary to allow TA to trace back to the obligated vehicles' real identities and revoke them.

Internal Attacks Prevention. Different from reference [11, 12], legitimate vehicles holding their own secret key materials can get neither the private key nor the real identity of another licit vehicle. Even if some vehicles are captured by the attacker, the attacker cannot obtain other legitimate vehicles' secret key materials with the captured materials.

Efficiency. Because of the strict time restriction of message authentication in VANETs, communication schemes should be efficient in terms of small computational overhead and acceptable verification delay. In addition, the communication overhead of the security programs should be as small as possible considering the confined bandwidth.

4. EPEC: Expedite Privacy-Preserving Emergency Communication

In this section, we detail the expedite privacy-preserving emergency communication EPEC scheme for VANETs-based disaster rescue. EPEC presents expedite authentication for two communication patterns: EPEC1 for vehicle networking with AMB and EPEC2 for vehicle group communication without AMB. When AMB enters the disaster area, it regularly broadcasts beacon of its own identification. Vehicles receiving the beacon verify its authenticity and, if valid, reregister with TA via AMB. Specific to the requirement of conditional privacy preservation, a secure key agreement protocol is proposed to establish an exclusive secret key between the vehicle and TA during the reregistration phase. Then, vehicles self-generate pseudonyms for subsequent broadcast communication. In both vehicle to AMB communication and vehicle group communication, point multiplications instead of bilinear pairings are conducted for authentication, thus saving computational overhead. Besides, signatures can be verified in batch. The prime notations in EPEC are defined in Table 2.

Table 2: Notations.

Notations	Descriptions
AMB	Emergency communication car
V_i	Regular vehicle
$\mathrm{PK}_i/\mathrm{SK}_i$	Public/private key of V_i
h_1, h_2	One-way hash function $h_1, h_2: \{0,1\}^* \rightarrow \mathbb{Z}_q^*$
H	MapToPoint hash function $H: \{0,1\}^* \rightarrow G_1$
k_i	Secret key of V_i
ID_i	Pseudonym of V_i
CSK_i	Private key of V_i corresponding to ID_i
GSK	Group secret key

4.1. System Initialization. Before AMBs enter the disaster area, TA initializes the system to establish the public parameters. It is reasonable to assume that TA bootstraps the whole system, as a single authority VANETs model is under consideration in Figure 2. Specifically, in this phase, TA generates a cyclic additive group G_1 and a multiplicative group G_2 of the same prime order q, chooses two generators $P, Q \in G_1$, and then gets $g = e(P,Q)$. In addition, TA picks three secure cryptographic hash functions h_1, h_2, and H, where $h_1, h_2 : \{0,1\}^* \rightarrow \mathbb{Z}_q^*$ and $H : \{0,1\}^* \rightarrow G_1$. Then, TA randomly chooses $s \in \mathbb{Z}_q^*$ as its master key and sets $P_{\text{pub}} = sQ$. Finally, TA gets the system public parameters $\{G_1, G_2, g, P, Q, P_{\text{pub}}, h_1, h_2, H\}$. Vehicles and AMBs can download the system public parameters from TA.

4.2. EPEC1: Networking with AMB. AMBs enter the disaster area after the system initialization for dynamic vehicle reregistration, after which vehicles generate pseudonyms by themselves and conduct communication with AMB according to EPEC1.

4.2.1. Dynamic Reregistration. In the reregistration procedure, a secure protocol with secret key agreement is presented to establish a shared secret key between vehicle and TA via AMB. The protocol is secured with signature and encryption

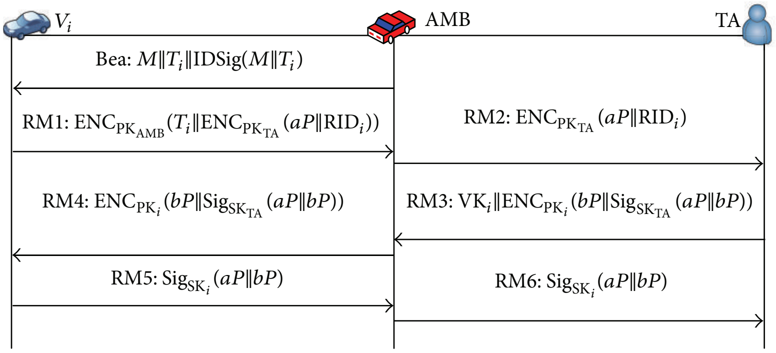

V_i
AMB
TA
Bea: $M\|T_i\|\text{IDSig}(M\|T_i)$
RM1: $\text{ENC}_{\text{PK}_{\text{AMB}}}(T_i\|\text{ENC}_{\text{PK}_{\text{TA}}}(aP\|\text{RID}_i))$
RM2: $\text{ENC}_{\text{PK}_{\text{TA}}}(aP\|\text{RID}_i)$
RM4: $\text{ENC}_{\text{PK}_i}(bP\|\text{Sig}_{\text{SK}_{\text{TA}}}(aP\|bP))$
RM3: $\text{VK}_i\|\text{ENC}_{\text{PK}_i}(bP\|\text{Sig}_{\text{SK}_{\text{TA}}}(aP\|bP))$
RM5: $\text{Sig}_{\text{SK}_i}(aP\|bP)$
RM6: $\text{Sig}_{\text{SK}_i}(aP\|bP)$

Figure 3: EPEC1: Vehicle reregistration protocol.

to prevent the intermediate intrusion attacks. The precise process is shown in Figure 3.

AMB periodically broadcasts its own identity information when entering some area with the following beacon message:

$$\text{Bea} = \langle M, T_i, \text{IDSig}(M \parallel T_i) \rangle,$$
$$M = \text{ID}_{\text{AMB}} \parallel \text{PK}_{\text{AMB}} \parallel \text{Cert}_{\text{TA}}(\text{PK}_{\text{AMB}} \parallel \text{ID}_{\text{AMB}}) \parallel \quad (1)$$
$$\text{other information},$$

where ID_{AMB} is the identity of AMB, $\text{Cert}_{\text{TA}}(\text{PK}_{\text{AMB}} \parallel \text{ID}_{\text{AMB}})$ is the certificate of public key PK_{AMB} signed by TA, T_i is the current timestamp, and $\text{IDSig}(M \parallel T_i)$ is the secure identity-based signature [29] with the private key $\text{CSK}_{\text{AMB}} = (h_1(\text{ID}_{\text{AMB}}) + s)^{-1}P$ to provide the origin authentication on the beacon. Concretely, $\text{IDSig}(M \parallel T_i)$ has the following form of $\text{IDSig}(M \parallel T_i) = (\alpha, \beta)$ as

$$\alpha = h_2(M \parallel T_i \parallel r),$$
$$\beta = (x + \alpha)\,\text{CSK}_{\text{AMB}}, \quad (2)$$

with a random number $x \in \mathbb{Z}_q^*$, and $r = g^x$.

As AMB's communication range is much larger than regular vehicles, V_i can receive the beacon Bea before AMB enters its communication range. It allows V_i to verify the beacon first and if valid, prepare reregistration message. The actual verification procedure is as follows.

(i) In regard to replay attack, V_i first checks the freshness of the beacon. Assuming that V_i receives the message Bea at T_i', V_i checks whether $\Delta T \geq T_i' - T_i$ is valid, where ΔT is the preset maximum transmission delay of the system. If the inequality does not hold, V_i discards the outdated message; otherwise, V_i continues the verification.

(ii) V_i verifies the validity of AMB's certificate Cert using TA's public key. If the following equation

$$\text{Verify}(\text{PK}_{\text{TA}}, \text{Cert}_{\text{TA}}(\text{PK}_{\text{AMB}})) = 1 \quad (3)$$

holds, V_i continues to verify AMB's signature on the message.

(iii) V_i checks the signature $\text{Sig}(M \parallel T_i)$ with

$$\alpha = h_2\left(M \parallel T_i \parallel e\left(\beta, h_1(\text{ID}_{\text{AMB}})Q + P_{\text{pub}}\right) g^{-\alpha}\right). \quad (4)$$

If (4) holds, the beacon is accepted and V_i continues its reregistration process. Otherwise, V_i waits for new AMB beacons. As aforementioned, all vehicles already have public/private key pairs, public key certificates, and real identities issued by TA. They reregister with TA using the key pairs and real identities. The process of reregistration and key agreement is secured with elliptic curve digital signature algorithm (ECDSA) and elliptic curve integrated encryption scheme (ECIES).

First, vehicle V_i randomly selects $a \in \mathbb{Z}_q^*$ and concatenates the random element aP and its real identity. Then, V_i encrypts the concatenation with the TA's public key. Finally, V_i encrypts the concatenation of the encryption and the time stamp with AMB's public key PK_{AMB} and sends it to the AMB in the following form:

$$\text{RM1} = \text{ENC}_{\text{PK}_{\text{AMB}}}\left(T_i \parallel \text{ENC}_{\text{PK}_{\text{TA}}}(aP \parallel \text{RID}_i)\right), \quad (5)$$

where ENC indicates the elliptic curve integrated encryption scheme ECIES, and the same as follows.

Receiving the first message RM1 from vehicle V_i at T_i', AMB decrypts the message and verifies its freshness. If the inequality $\Delta T \geq T_i' - T_i$ holds, it delivers the remaining part of the message, that is, RM2, to TA securely. Otherwise, it discards the message. Consider

$$\text{RM2} = \text{ENC}_{\text{PK}_{\text{TA}}}(aP \parallel \text{RID}_i). \quad (6)$$

TA decrypts the message and verifies the real identity RID_i. If RID_i is in the revocation list (RL), the message will be abandoned. Otherwise, TA randomly chooses $b \in \mathbb{Z}_q^*$, gets $k_i = abP$ and computes $VK_i = \text{RID}_i \oplus h_1(k_i)$. Next, TA signs the concatenation of aP and bP and then encrypts the signature and bP by V_i's public key. Then, the VK_i and the encryption are sent to AMB securely. Consider

$$\text{RM3} = VK_i \parallel \text{ENC}_{\text{PK}_i}\left(bP \parallel \text{Sig}_{\text{SK}_{\text{TA}}}(aP \parallel bP)\right). \quad (7)$$

In (7), Sig represents the ECDSA signature.

AMB stores the VK_i for vehicle authentication subsequently. Then, it passes the message RM4 to vehicle V_i. Consider

$$\text{RM4} = \text{ENC}_{\text{PK}_i}\left(bP \parallel \text{Sig}_{\text{SK}_{\text{TA}}}(aP \parallel bP)\right). \quad (8)$$

The vehicle verifies the TA's signature and, if valid, sends its own signature on aP and bP to the TA via AMB. Consider

$$\mathrm{RM5} = \mathrm{RM6} = \mathrm{Sig}_{\mathrm{SK}_i}(aP \parallel bP). \tag{9}$$

TA authenticates the vehicle V_i's signature. If it is valid, a shared secret key $k_i = abP$ between V_i and TA has been established, which is essential for real identity tracing of a malicious vehicle.

The secret key agreement protocol adopts ECDSA signature and ECIES encryption to prevent the intermediate intrusion attacks. Under the assumption of computational Diffie-Hellman problem, the adversary cannot calculate any information about the secret key k_i, ensuring that the real identity of the vehicle can only be tracked by TA.

4.2.2. Pseudonym Self-Generation. In the identity-based cryptography, the entity's public key can be generated based on the real identity. Different pseudonyms are used to sign messages to protect the vehicles from being tracked or associated.

The pseudonym ID_i comprises three parts: $\mathrm{ID}_{i,1}$, $\mathrm{ID}_{i,2}$, and ET_i, where $\mathrm{ID}_{i,1}$ and $\mathrm{ID}_{i,2}$ are the pseudonym material and ET_i is the life period of this pseudonym. Note that the pseudonym life period ET_i is predelimited in the system. To generate a pseudonym, the vehicle first selects a random number $r_i \in \mathbb{Z}_q^*$ to establish point $R_i \in G$, so that $R_i = (x_i, y_i) = r_iP$. The vertical and horizontal coordinates of each point R_i are integers within $\mathbb{F}_q$. Then, the vehicle generates its pseudonym as

$$\begin{gathered}\mathrm{ID}_{i,1} = R_i,\\ \mathrm{ID}_{i,2} = \mathrm{RID}_i \oplus h_1(k_i) \oplus \mathrm{H}(\mathrm{ID}_{i,1} \parallel \mathrm{ET}_i),\\ \mathrm{ID}_i = (\mathrm{ID}_{i,1}, \mathrm{ID}_{i,2}, \mathrm{ET}_i).\end{gathered} \tag{10}$$

And the corresponding private key CSK_i is

$$\begin{gathered}\mathrm{CSK}_i = h_1(\mathrm{ID}_i \parallel VK_i),\\ VK_i = \mathrm{RID}_i \oplus h_1(k_i).\end{gathered} \tag{11}$$

In the end, the vehicle V_i stores a list of the pseudonym ID_i with its corresponding private key CSK_i and the random point R_i. Notice that (1) it is essential to insert life period ET_i into every pseudonym to prevent attackers from abusing obsolete pseudonyms; (2) the pseudonym and the corresponding private key generation can be completed prior to the broadcast communication. Thus, the delay of signing a message does not include the time of generating the pseudonym and its private key.

4.2.3. Vehicle to AMB Communication

Signing. After the reregistration and the pseudonym generation, vehicles can send perceived information to AMB. And AMB verifies vehicles' messages in batch and delivers the results to the Application Server. Then, the feedbacks from the Application Server are forwarded to vehicles via AMB. Vehicle V_i's signature σ_i on message M_i has the following form

$$\sigma_i = r_i + h_2(M_i \parallel \mathrm{ID}_i)\,\mathrm{CSK}_i, \tag{12}$$

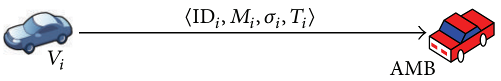


Figure 4: EPEC1: Vehicle to AMB communication.

where r_i is the random integer meeting $R_i = r_iP$. It is obvious that a vehicle needs to compute one hash function and one multiplication to sign a message. Compared to group signature and other identity-based signature, it can significantly save computational cost in the signature generation phase. Finally, V_i broadcasts the message in form of $\langle \mathrm{ID}_i, M_i, \sigma_i, T_i \rangle$ as shown in Figure 4.

Verification. It is computational costly to verify messages one by one. Batch verification allows verifiers to authenticate a number of messages at once, to save computational overhead and reduce verification delay. In EPEC1, vehicles' messages can be verified in batch. Given n distinct signatures $\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n$ received from $V_1, V_2, V_3, \ldots, V_n$, respectively, AMB first checks the pseudonym life period ET_i to prevent attackers from abusing the obsolete pseudonyms. Then, AMB checks the timestamps in the messages, verifies the freshness of every message, and deletes outdated ones. Next, AMB calculates the vehicles' private key CSK_i with the VK_i, which satisfies (10); $\mathrm{ID}_{i,2} = VK_i \oplus H(\mathrm{ID}_{i,1} \parallel \mathrm{ET}_i)$. Finally, AMB verifies all the signatures in batch. The specific batch verification process is introduced in detail as follows.

(i) AMB first checks the pseudonym life period ET_i to delete obsolete pseudonym signature.

(ii) For freshness, AMB checks the transmission delay. Assuming that AMB receives the message at T_i', AMB checks whether $\Delta T \geq T_i' - T_i$ is valid. If the inequality holds, AMB continues the verification; otherwise, it discards the outdated message. This step is done for every message.

(iii) AMB calculates the vehicle's corresponding private key according to (11), $\mathrm{CSK}_i = h_1(ID_i \parallel VK_i)$.

(iv) Verify all the signatures by

$$\left(\sum \sigma_i\right) P = \sum ID_{i,1} + \left(\sum h_2(M_i \parallel ID_i)\,\mathrm{CSK}_i\right) P. \tag{13}$$

If (13) holds, all the signatures are valid in the batch; otherwise, there is at least one invalid signature, which calls for invalid signature detection algorithm to find the invalid ones.

Note that AMB needs to find out the verification key of $V_1, V_2, V_3, \ldots, V_n$, by checking which of the stored VK_i satisfies (10); $\mathrm{ID}_{i,2} = VK_i \oplus H(\mathrm{ID}_{i,1} \parallel \mathrm{ET}_i)$. And the private key CSK_i needs to be calculated to achieve authentication during the verifying process. However, the security of our scheme is not destroyed. As even AMB knows V_i's private key, it is difficult for AMB to forge V_i's signature. This is because r_i is safe based on the ECDLP problem (in Definition 1), even $\mathrm{ID}_{i,1}$ is publicly known. In addition, vehicles change pseudonyms regularly and different pseudonyms are based

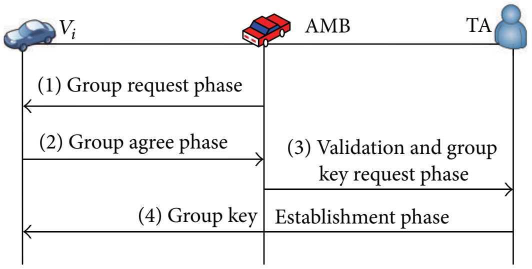


Figure 5: EPEC2: Group formation protocol.

on different random number r_i. Thus, AMB cannot get legitimate vehicle's secret key or forge its signature with the obsolete pseudonym and corresponding private key.

4.3. EPEC2: Vehicle Group Communication without AMB. As we do not assume that AMBs cover the whole network, they may forward to another place after networking the vehicles. So, vehicle group communication scheme is requisite for vehicle communication without AMBs.

4.3.1. Vehicle Group Formation. In this subsection, we present how to form a vehicle group with the help of AMB. The establishment of a group is divided into four stages as shown in Figure 5. Concretely, the specific procedure of group formation is detailed as the following.

(i) First of all, AMB generates the group request message $M = \{\mathrm{GR}, \mathrm{ID}_1, \mathrm{ID}_2, \mathrm{ID}_3, \ldots, \mathrm{ID}_n\}$ to start the group establishment, where GR indicates the group request and the group identity. AMB broadcasts the message in the form of

$$\mathrm{GR} = \langle \mathrm{ID}_{\mathrm{AMB}}, M, T_i, \mathrm{IDSig}(M \,\|\, T_i) \rangle, \tag{14}$$

where IDSig is the same identity-based signature as (2).

(ii) Vehicles receiving the message verify AMB's signature first and, if valid, check for their own pseudonym ID_j. If found, they generate agreed message $M_j = \{\mathrm{GA}, \mathrm{ID}_j\}$ and send it to AMB as follows:

$$\mathrm{GA} = \langle \mathrm{ID}_i, M_i, \sigma_i, T_i \rangle. \tag{15}$$

(iii) Receiving all the agreed messages, AMB verifies all the signatures in batch. If the batch verification fails, AMB suspends the protocol and broadcasts new group formation request. Otherwise, AMB sends group key request message GKR to TA applying for the group key. Consider

$$\mathrm{GKR} = \langle \mathrm{ID}_{\mathrm{AMB}}, M(\mathrm{GR}, \mathrm{ID}_1, \mathrm{ID}_2, \ldots, \mathrm{ID}_n), T_i, \mathrm{IDSig}(M, T_i) \rangle. \tag{16}$$

(iv) Upon receiving the group key request message GKR, TA verifies AMB's signature. If valid, they choose random number rand and get the group private key $\mathrm{GSK} = s \cdot \mathrm{rand}$. TA encrypts the shared group private key with each vehicle's secret key k_i, respectively, and sends the group key establishment message GKE to AMB. Consider

$$\begin{aligned} \mathrm{GKE} = \Big\langle & M\big(\mathrm{ID}_1, \mathrm{ID}_2, \ldots, \mathrm{ID}_n, E_{k_1}(\mathrm{GSK}), \\ & E_{k_2}(\mathrm{GSK}), \ldots, E_{k_n}(\mathrm{GSK})\big), \\ & T_i, \mathrm{Sig}_{\mathrm{SK}_{\mathrm{TA}}}(M, T_i) \Big\rangle. \end{aligned} \tag{17}$$

In the GKE message, E is a symmetric encryption algorithm. The message is delivered to the group member via AMB. The group member V_i first verifies TA's signature and then conducts decryption with its own k_i to get the group private key GSK. Finally, V_i calculates the group public key as $\mathrm{GPK} = \mathrm{GSK} \cdot P$.

4.3.2. Vehicle Group Communication without AMB. After group formation, vehicles can conduct mutual authentication within the group for real-time communication without AMB. With the pregenerated pseudonym ID_i and the newly generated group private key GSK, V_i generates the signature σ_i as

$$\sigma_i = r_i + h_2(M_i \,\|\, ID_i)\,\mathrm{GSK}. \tag{18}$$

The GSK is employed to generate group message signature for more efficient authentication. The group message is broadcasted in the format of $\langle \mathrm{GR}, \mathrm{ID}_i, M_i, \sigma_i, T_i \rangle$ as shown in Figure 6.

Receiving the message, the verifier V_j first checks the time validity as aforementioned. If the message is fresh, it comes to the signature verification phase as

$$\sigma_i P = \mathrm{ID}_{i,1} + h_2(M_i \,\|\, \mathrm{ID}_i)\,\mathrm{GPK}. \tag{19}$$

Note that the batch verification is also applicable to group communication, providing a much smaller computational overhead of two point multiplication operations. Thus EPEC2 provides much smaller authentication delay than other schemes, which will be detailed in Section 6.

4.4. Invalid Signature Detection. Invalid signature detection algorithm is essential for batch verification scheme. This is because invalid signatures could come from a variety of reasons, such as malicious vehicles, legitimate vehicle failure, or wireless channel interference. The batch verification will fail when there is one invalid signature in the batch. IBV [11] may suffer from severe inefficiency as it does not pay attention to invalid signatures. Once an invalid signature frustrates the batch verification, all the valid ones in the batch will be discarded. Therefore, invalid signature detection mechanism is necessary. ABAKA [12] adopted the binary search method for invalid signature detection. When the batch verification fails, messages in the batch are bisected, and verified, respectively until only one message left or valid. But it is inefficient to retest the messages for $d\lceil \log_2(n) \rceil$ times, where n and d are the total message number and

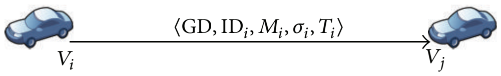

$\langle \mathrm{GD}, \mathrm{ID}_i, M_i, \sigma_i, T_i \rangle$
V_i
V_j

FIGURE 6: EPEC2: Vehicle group communication.

the invalid signature number in the batch, respectively. In order to reduce rebatch verification overhead, we adopt the Generalized Binary Splitting algorithm [30], which is the most efficient testing algorithm when d is not very great. In the worst case that all the d invalid signatures are divided into different subbatches during each section, the number of tests required is as follows:

$$\mathrm{NT}_{\mathrm{GBS}} = d - 1 + \lceil \log_2 (n * d) \rceil . \quad (20)$$

And the precise re-batch verification cost caused by invalid signature detection will be discussed in Section 6.

5. Security Analysis

We analyze the security performance of the proposed scheme EPEC in this section. According to the security objectives aforementioned, EPEC is evaluated in the following four aspects: dynamic reregistration without fixed RSUs, mutual authentication, conditional privacy preservation, and internal attacks prevention. And the efficiency of EPEC is analyzed in Section 6 in three different aspects. The security of EPEC is analyzed as follows.

Dynamic Reregistration without Fixed RSUs. AMBs authorized by TA enter the disaster area for networking the regular vehicles. In EPEC, AMBs don't cover the entire network. Vehicles dynamically reregister with TA when *AMB* is available. In addition, vehicles are able to update their secret keys via other AMBs.

Mutual Authentication. The proposed scheme EPEC securely achieves mutual authentication during the process of vehicles reregistration and broadcast communication. On arriving, AMB authenticates itself securely to vehicles by identity-based signature [29]. Next, vehicles authenticate themselves with ECDSA signature. The secure signatures guarantee mutual authentication among vehicle, AMB, and TA. In addition, the proposed lightweight pseudonym based signature enforces mutual authentication in broadcast communication. Pseudonym ID_i used during broadcast communication is self-generated with V_i's unique secret key k_i and real identity RID_i, which guarantees that no one else can forge V_i's pseudonym and signature. The signature also ensures that only the unmodified messages from legitimate senders are accepted. Because once the message content is distorted during the transportation, the signature verification will fail.

Conditional Privacy Preservation. The actual identity of a vehicle is concealed by the pseudonym in EPEC. As vehicle's secret key used to generate pseudonym is exclusive and can't be compromised because of the computational Diffie-Hellman problem. In addition, regular pseudonym changing prevents attackers from tracing a specific vehicle in the long term. But on the other hand, malicious vehicles should be revealed and revoked from the network in time. TA, and only TA is allowed to trace the real identity of a vicious vehicle. For example, once V_i is found misbehaving, the ID_i is reported to TA. Then, TA reveals the real identity RID_i through the following process:

$$\begin{aligned} &\mathrm{ID}_{i,2} \oplus h_1(k_i) \oplus H(\mathrm{ID}_{i,1} \parallel \mathrm{ET}_i) \\ &\quad = \mathrm{RID}_i \oplus h_1(k_i) \oplus H(\mathrm{ID}_{i,1} \parallel \mathrm{ET}_i) \oplus h_1(k_i) \\ &\qquad \oplus H(\mathrm{ID}_{i,1} \parallel \mathrm{ET}_i) \\ &\quad = \mathrm{RID}_i . \end{aligned} \quad (21)$$

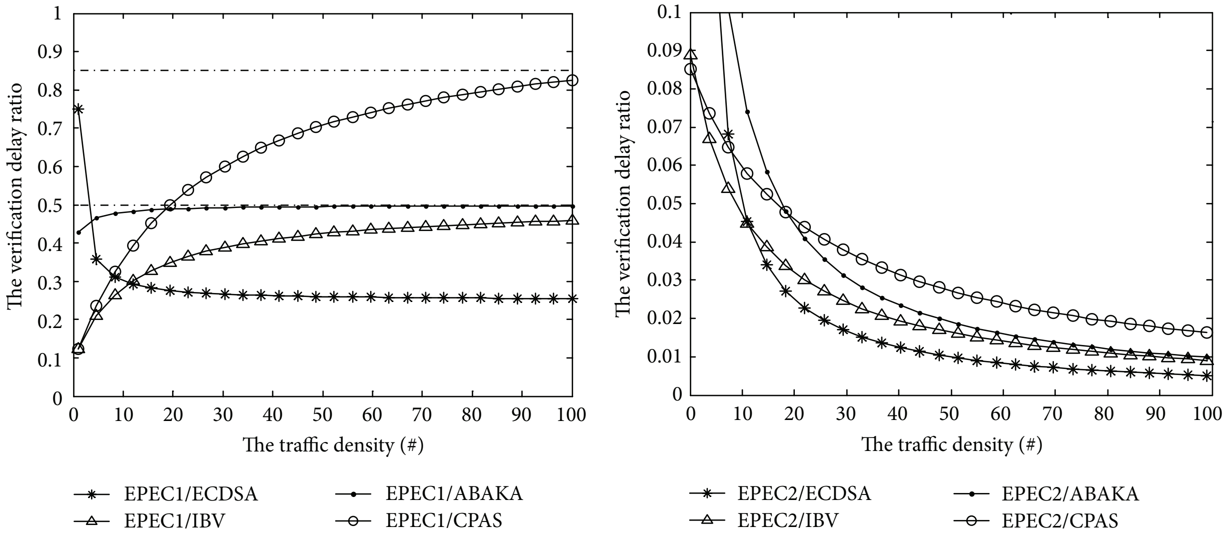
And it is TA that determines whether to revoke V_i from the system or not. The specific revocation mechanism is out of the scope of our paper.

Internal Attacks Prevention. Another important security property of EPEC is the ability to prevent internal attacks. Even if an attacker has captured some vehicles and got their private keys, the attacker still can't forge a valid signature of other legitimate vehicle, because it knows neither the private key nor the private secret used to reveal the real identity of the legitimate vehicle. In addition, the damage caused by the captured vehicles is also limited, because the tracking mechanism can quickly retrieve the real identities of these vehicles. TA can revoke the malicious vehicles from the network promptly.

6. Performance Evaluation

In this section, the effectiveness of EPEC is evaluated in terms of message verification delay, transmission overhead, and verification delay with invalid signatures. In the evaluation, EPEC is compared with four related typical schemes: IBV [11], ABAKA [12], ECDSA [13], and CPAS [15]. IBV is a typical batch verification scheme for unidirectional authentication based on bilinear pairing operations. ABAKA is a point multiplication based signature scheme with batch verification. Reference [13] presents a key establishment and authentication protocol RAISE based on ECDSA signature in the IEEE standard 1609.2. It is referred to as ECDSA in the evaluation. CPAS is a typical mutual authentication scheme with batch verification based on bilinear pairing operations.

6.1. Message Verification Delay. To evaluate and compare the schemes' verification delay, we first define the time complexity of the main cryptographic operations required in our EPEC and other schemes. Let T_{mtp}, T_{mul}, and T_{par} denote the time to perform a MapToPoint hash operation, a point multiplication over an elliptic curve, and a bilinear pairing operation, respectively. According to [31], T_{mul} is 0.6 ms, T_{mtp} is 0.6 ms, and T_{par} is 4.5 ms. It is apparent that the computational time of MapToPoint hash operation and point multiplication is much smaller than bilinear pairing operation. We don't consider the cost of one way hash function, which is only 2 microseconds. Table 3 shows



(a) EPEC1

(b) EPEC2

Figure 7: Verification delay ratio with different traffic density.

Table 3: Computational overhead.

Schemes	Authenticate a single message	Authenticate n messages
ECDSA	$4T_{\text{mul}}$	$4nT_{\text{mul}}$
IBV	$3T_{\text{par}} + T_{\text{mtp}} + T_{\text{mul}}$	$3T_{\text{par}} + nT_{\text{mtp}} + nT_{\text{mul}}$
ABAKA	$3T_{\text{mul}}$	$(2n+1)T_{\text{mul}}$
CPAS	$3T_{\text{par}} + 2T_{\text{mul}}$	$3T_{\text{par}} + (n+1)T_{\text{mul}}$
EPEC1	$T_{\text{mtp}} + 2T_{\text{mul}}$	$nT_{\text{mtp}} + 2T_{\text{mul}}$
EPEC2	$2T_{\text{mul}}$	$2T_{\text{mul}}$

Table 4: Transmission overhead.

Schemes	One message	n messages
ECDSA	181 B	$181n$ B
IBV	63 B	$63n$ B
ABAKA	80 B	$80n$ B
CPAS	105 B	$105n$ B
EPEC1	82 B	$82n$ B
EPEC2	84 B	$84n$ B

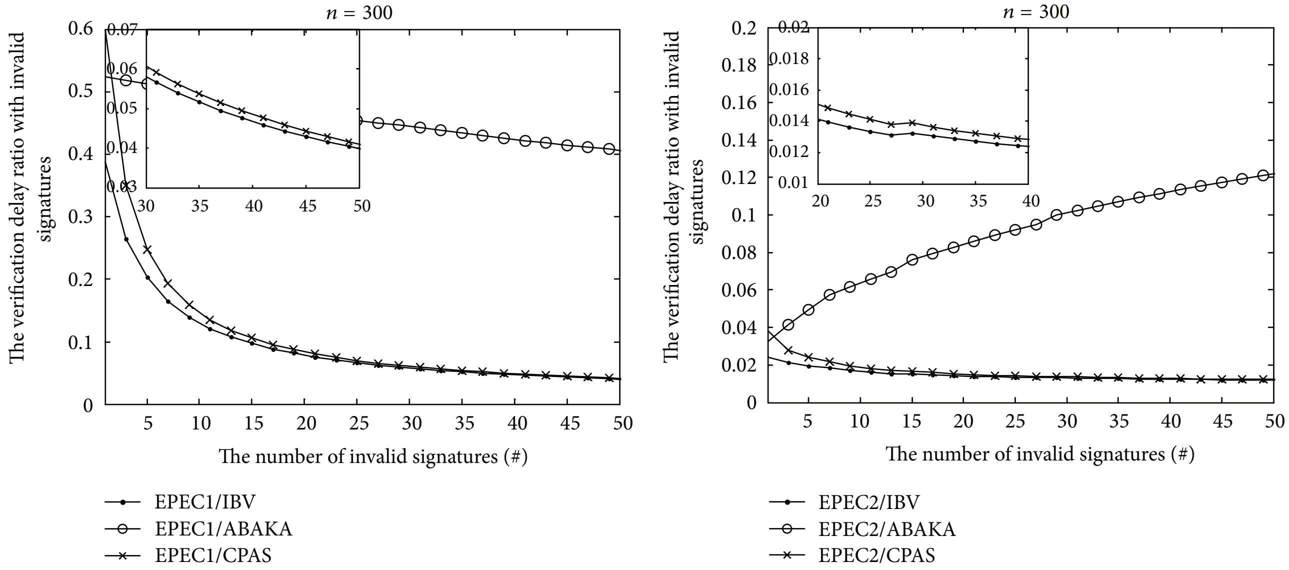
the computational overhead of all the schemes in terms of authenticating a single message and n messages.

Notice that IBV is an authentication scheme without mutual authentication and key agreement. So, for fairness, the schemes are compared in unidirectional authentication case without key agreement. The mutual authentication function of EPEC is analyzed in Section 5. And for EPEC, the broadcast authentication cost equals on the two sides. From the comparison in Table 3, we can see that the proposed scheme EPEC achieves the least computational overhead. ECDSA verifies n distinct signatures one by one; so, the n messages verification is the most inefficient. Although IBV adopts batch verification, the basic pairing operation is computational costly. ABAKA also verifies signatures with point multiplications, but the verification cost is much higher than our EPEC.

The computational overhead increases with the number of messages for all these schemes. The message verification delay ratio of these schemes is shown in Figure 7. It is apparent that EPEC is superior to all the other typical schemes, because EPEC adopts lightweight point multiplication to verify messages. Figure 7(a) indicates the relationship between EPEC1 and other schemes. It is obvious that the delay ratio between EPEC1 and ABAKA is always less than 0.50 regardless of the number of messages; the delay ratio between EPEC1 and ECDSA is approximately 0.25 when the number of messages is larger than 60. From Figure 7(b), we can see that the delay ratio between EPEC2 and other schemes approaches zero with the traffic density increasing, which is because the batch verification overhead has nothing to do with the number of messages for EPEC2.

6.2. Transmission Overhead. We consider the transmission overhead in terms of the signature and the certificate appended to the original message, while the message itself is not counted. The comparison of all the schemes is shown in Table 4. ECDSA signature is 42 bytes in length, but a certificate of 125 bytes must be transmitted along with each message, resulting in the total transmission overhead of 167 bytes. For IBV, the length of signature is 21 bytes, while pseudonym is 42 bytes. In ABAKA, the 20 bytes verification message and the 20 bytes material message all are authentication materials, resulting in a signature of 40 bytes. CPAS has 60 bytes signature, 41 bytes pseudonym, and 4 bytes message type information.

The transmission overhead of ECDSA is the largest among the five schemes, while IBV has the smallest. Since IBV adopts bilinear pairing cryptographic operations for



(a) EPEC1

(b) EPEC2

Figure 8: Verification delay ratio with invalid signatures.

signature, which is short in length but costly in verification. The overhead of ABAKA is a bit less than EPEC, which is because 2 bytes life period ET_i is added to the pseudonym in our scheme to prevent expired pseudonyms abuse. For group communication EPEC2, there is additional 2 bytes group ID information. Obviously, the transmission overhead increases linearly for all the schemes with the number of messages.

Table 5: Computational overhead with invalid signatures.

Schemes	First batch verification	Rebatch verification
IBV	$3T_{\mathrm{par}} + nT_{\mathrm{mtp}} + nT_{\mathrm{mul}}$	$3T_{\mathrm{par}}$
ABAKA	$(2n+1)T_{\mathrm{mul}}$	$1T_{\mathrm{mul}}$
CPAS	$3T_{\mathrm{par}} + (n+1)T_{\mathrm{mul}}$	$3T_{\mathrm{par}}$
EPEC1	$nT_{\mathrm{mtp}} + 2T_{\mathrm{mul}}$	$2T_{\mathrm{mul}}$
EPEC2	$2T_{\mathrm{mul}}$	$2T_{\mathrm{mul}}$

6.3. Verification Delay with Invalid Signatures. Batch verification can save computational overhead and provide low verification delay. However, the expense of the batch verification is that invalid signature may cause verification failure and re-batch verification is needed. As aforementioned, GBS is the most efficient invalid signature detection scheme when the number of invalid signatures is not very great. We adopt GBS, while the IBV and ABAKA use binary search algorithm to check the invalid signatures. Table 5 shows the computational overhead for first batch verification and one time re-batch verification. ABAKA and EPEC only need point multiplications for re-batch verification, while IBV and CPAS need three bilinear pairing operations.

The verification delay VD with invalid signatures in the worst case is computed with

$$\mathrm{VD} = 1 * \mathrm{VT}_{\mathrm{first}} + \mathrm{NT} * \mathrm{VT}_{\text{re-batch}}, \tag{22}$$

where $\mathrm{VT}_{\mathrm{first}}$ is the verification time needed for the first batch verification, NT is the number of tests needed to find all $d(d \geq 1)$ invalid signatures, and $\mathrm{VT}_{\text{re-batch}}$ is the verification time for one time re-batch verification. The verification delay for EPEC1 is

$$\begin{aligned}\mathrm{VD}_{\mathrm{EPEC1}} = {} & \left(n * T_{\mathrm{mtp}} + 2 * T_{\mathrm{mul}}\right) \\ & + \left(d - 1 + \lceil \log_2 (n * d) \rceil\right) * \left(2 * T_{\mathrm{mul}}\right),\end{aligned} \tag{23}$$

with $\mathrm{NT}_{\mathrm{GBS}}$ in (20). And for other schemes, binary search is adopted for invalid signature detection with the tests number of $\mathrm{NT}_{\mathrm{BS}} = d * \lceil \log_2 (n) \rceil$.

Figure 8 gives the verification delay ratio with the number of invalid signatures. In the experiment, the messages number n is set 300. It is apparently shown that our EPEC achieves the smallest verification delay with invalid signature detection compared to other three typical schemes. The efficiency of IBV and CPAS decreases markedly with the number of invalid signature increasing, because their re-batch verification requires three time-consuming bilinear pairing operations. EPEC2 achieves high efficiency even when the number of invalid signatures is large.

7. Conclusion

In this paper, we propose an expedite privacy-preserving emergency communication (EPEC) scheme for VANETs-based disaster rescue. EPEC fulfils two communication patterns: EPEC1 for vehicle networking with AMB and EPEC2 for vehicle group communication without AMB, respectively. By the theoretical and experimental analyses, we show that the proposed scheme has the following advantages. (1) *No Fixed RSU*. Vehicles are able to reregister themselves with

TA dynamically. (2) *Conditional Privacy Preservation*. EPEC allows only TA to trace malicious vehicle's real identity. (3) *Internal Attacks Prevention*. The captured vehicles can only cause limited damage to the whole networks. (4) *High Efficiency*. Lightweight signature and batch verification are combined to reduce authentication delay. And identity-based signature is adopted to save the delivery cost of public key certificates. To sum up, EPEC represents expedite authentication and satisfies the conditional privacy preservation requirement in emergency communication.

Acknowledgments

This work is supported by the Natural Science Foundation of China under Grant no. 61272074, the Natural Science Foundation of Jiangsu Province under Grant no. BK2011464, and the project of the Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education. And the author L. Wang is supported by the Disguised Researcher Program of Jiangsu Province of China (2012-wlw-020) as well as academic leader supported by Qinglan Project of Jiangsu Province of China.

References

[1] "2008 Sichuan Earthquake," http://en.wikipedia.org/wiki/2008_Sichuan_earthquake.

[2] "2011 Tōhoku Earthquake and Tsunami," http://en.wikipedia.org/wiki/2011_T%C5%8Dhoku_earthquake_and_tsunami.

[3] "Dedicated Short Range Communications (DSRC)," http://www.etsi.org/index.php/technologies-clusters/technologies/intelligent-transport/dsrc.

[4] USA Department of Transportation, "National Highway Traffic Safety Administration," Final Repot, Vehicle Safety Communications Project, 2006.

[5] S. Lee, G. Pan, J. Park, M. Gerla, and S. Lu, "Secure incentives for commercial ad dissemination in vehicular networks," in *Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 150–159, 2007.

[6] R. Lu, X. Lin, and X. Shen, "SPRING: a social-based privacy-preserving packet forwarding protocol for vehicular delay tolerant networks," in *Proceedings of IEEE Communications Society Conference on Computer Communications (INFOCOM '10)*, pp. 1–9, San Diego, Calif, USA, March 2010.

[7] A. Wasef and X. Shen, "EMAP: expedite message authentication protocol for vehicular ad hoc networks," *IEEE Transaction on Mobile Computing*, vol. 12, no. 1, pp. 78–89, 2013.

[8] X. Lin, X. Sun, P. H. Ho, and X. Shen, "GSIS: a secure and privacy-preserving protocol for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6 I, pp. 3442–3456, 2007.

[9] A. Studer, E. Shi, F. Bai, and A. Perrig, "TACKing together efficient authentication, revocation, and privacy in VANETs," in *Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '09)*, pp. 22–26, June 2009.

[10] R. Lu, X. Lin, H. Zhu, P. H. Ho, and X. Shen, "ECPP: efficient conditional privacy preservation protocol for secure vehicular communications," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 1903–1911, April 2008.

[11] C. Zhang, R. Lu, X. Lin, P. H. Ho, and X. Shen, "An efficient identity-based batch verification scheme for vehicular sensor networks," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 816–824, April 2008.

[12] J. L. Huang, L. Y. Yeh, and H. Y. Chien, "ABAKA: an anonymous batch authenticated and key agreement scheme for value-added services in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 248–262, 2011.

[13] C. Zhang, X. Lin, R. Lu, P. H. Ho, and X. Shen, "An efficient message authentication scheme for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3357–3368, 2008.

[14] Y. Hao, Y. Cheng, C. Zhou, and W. Song, "A distributed key management framework with cooperative message authentication in VANETs," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 3, pp. 616–629, 2011.

[15] K.-A. Shim, "CPAS: an efficient conditional privacy-preserving authentication scheme for vehicular sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1874–1883.

[16] M. Raya and J. P. Hubaux, "The security of vehicular ad hoc networks," in *Proceedings of the ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN '05)*, pp. 11–21, usa, November 2005.

[17] M. Raya and J. P. Hubaux, "Securing vehicular ad hoc networks," *Journal of Computer Security*, vol. 15, no. 1, pp. 39–68, 2007.

[18] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," in *Advances in Cryptology—CRYPTO 2004*, vol. 3152 of *Lecture Notes in Computer Science*, pp. 41–55, Springer, 2004.

[19] R. Chen, D. Ma, and A. Regan, "TARI: meeting delay requirements in VANETs with efficient authentication and revocation," in *Proceedings of International Conference on Wireless Access in Vehicular Environments (WAVE)*, 2009.

[20] D. Boneh and H. Shacham, "Group signatures with verifier-local revocation," in *Proceedings of the 11th ACM Conference on Computer and Communications Security (CCS '04)*, pp. 168–177, October 2004.

[21] B.-K. Chaurasua and S. Verma, "Conditional privacy through ring signature in vehicular ad-hoc networks," in *Transactions on Computational Science XIII*, vol. 6750, pp. 147–156, Springer, 2011.

[22] R. Lu, X. Lin, and T.-H. Luan, "Pseudonym changing at social spots: an effective strategy for location privacy in VANET," *IEEE Transaction on Vehicular Technology*, vol. 61, no. 1, pp. 86–96, 2012.

[23] X. Lin, X. Sun, X. Wang, C. Zhang, P. H. Ho, and X. Shen, "TSVC: timed efficient and secure vehicular communications with privacy preserving," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4987–4998, 2008.

[24] A. Perrig, R. Canetti, J. Tygar, and D. Song, "The TESLA broadcast authentication protocol," *RSA CryptoBytes*, vol. 5, no. 2, pp. 2–13, 2002.

[25] T. Chima, S. Yiua, L. Huia, and V. Lib, "MLAS: multiple level authentication scheme for VANETs," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1445–1456, 2012.

[26] L. Malina, J. Hajný, and V. Zeman, "Group signatures for secure and privacy preserving vehicular ad hoc networks," in *Proceedings of the 8th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pp. 71–74, New York, NY, USA, October 2012.

[27] F. Li, X. Xin, and Y. Hu, "Indentity-based broadcast signcryption," *Computer Standards & Interfaces*, vol. 30, no. 1-2, pp. 89–94, 2008.

[28] J. H. Yang and C. C. Chang, "An ID-based remote mutual authentication with key agreement scheme for mobile devices on elliptic curve cryptosystem," *Computers and Security*, vol. 28, no. 3-4, pp. 138–143, 2009.

[29] P. S. L. M. Barreto, B. Libert, N. McCullagh, and J. J. Quisquater, "Efficient and provably-secure identity-based signatures and signcryption from bilinear maps," in *Advances in Cryptology—ASIACRYPT 2005*, vol. 3788 of *Lecture Notes in Computer Science*, pp. 515–532, Springer, 2005.

[30] D. Du and F. Hwang, *Combinatorial Group Testing and Its Applications*, World Scientific, Singapore, 2nd edition.

[31] C. Zhang, P.-H. Ho, and J. Tapolcai, "On batch verification with group testing for vehicular communications," *Wireless Network*, vol. 17, no. 8, pp. 1851–1865, 2011.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 261264, 10 pages
http://dx.doi.org/10.1155/2013/261264

Research Article

A Trust Value-Based Spectrum Allocation Algorithm in CWSNs

Qing-Qi Pei, Zi Li, and Li-Chuan Ma

State Key Laboratory of Integrated Service Network, Xidian University, Xi'an 710071, China

Correspondence should be addressed to Qing-Qi Pei; xd_qqpei@foxmail.com

Received 1 March 2013; Revised 13 April 2013; Accepted 15 April 2013

Academic Editor: Liangmin Wang

With the increase of congestible frequency spectrums and the rapid development of wireless sensor networks (WSNs) applications, the new technology—combing cogitative radio technology with WSNs—called CWSNs will bring broad prospects to the field of radio and sensor networks. CWSN devotes itself to the solution of spectrum sharing in forms of networks, such as IEEE 802.11 and Bluetooth. A new algorithm is proposed for allocating the idle spectrum to secondary users (SUs). This algorithm uses "the last diminisher" algorithm, which appears in fair division models. It can help solve complicated problems in a simple fashion. This article combines the trust value and the method of spectrum allocation. The concept of trust value in reputation management is introduced. All the factors are applied to solve the realistic problems of spectrum allocation. With the growing number of sensor nodes, the stable throughput is of vital importance.

1. Introduction

With the global data congestion increasing by 50 percent per year, the wireless sensor networks (WSNs), one of the fastest growing wireless networks in recent years, are faced with network traffic as well. Tens of thousands of tiny sensor nodes are distributed in the wireless sensor networks to detect surrounding data. With the rapid development of the WSNs, this potential application scenario has gradually proceeded to reality, from industry to home to civilians as well as the military. New wireless communication techniques, such as ZigBee and 802.15.4, have already enabled the interoperability between products to ensure extendibility of the network and low consumption.

Most of the wireless sensor networks work in unauthorized frequency bands, such as ISM, which is a universal 2.4 G frequency band. This frequency band is shared by many popular technologies, such as Wi-Fi and Bluetooth. Therefore, with the increase of applications based on the WSNs, the frequency band has become increasingly congested. The sharing of unauthorized frequency bands has become the theme of extended research [1, 2]. Study has shown that the IEEE 802.11 [3] network will greatly impact the performance of the ZigBee network and 802.15.4 network if they work in the same frequency band [2].

Mitola first proposed the concept of cognitive radio in a paper [4] published in 1999. They represented a new language called radio knowledge representation language (RKRI), which can improve the flexibility of the personal wireless service. In the same year, Mitola and Maguire Jr. proposed the concept of spectrum pooling [5]. Soon afterward, he expounded his view that cognitive radio would be increasingly important with the shortage of spectrum in the doctoral defense held by the Royal Swedish Academy of Sciences in 2000 [6]. With probing spectrum holes and the adaptive modulation mode, cognitive radio offers efficient dynamic spectrum access and increases the utilization of spectrum in the frequency domain or time domain.

The growing demand for wireless communication has become a great challenge [7] for efficient use of spectrum. Therefore, cognitive radio was seen as a key technology which was focused on opportunistic spectrum access. The cognitive radio network is an intelligent wireless communication system which can sense the spectrum around and quickly adjust its inner parameter to achieve the reliable and efficient communication [8].

Dynamic spectrum access can be used in WSNs. Therefore, a new type of network emerges, which is the CWSN. Typical wireless sensor networks manage finite spectrum and transmit perceived data to adjacent nodes. A fundamental

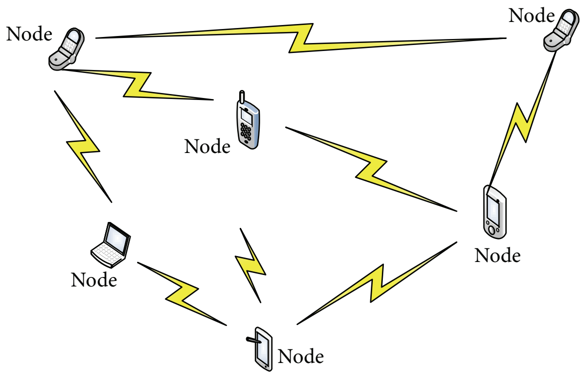


Figure 1: CWSN construction.

aim of the wireless sensor network is to transmit large perceived data in real time and synchronously on the premise that the fewest possible resources are used. For example, in a WSN of the medical environment, a lot of nodes are used to monitor various signals of a patient and transmit them in real time.

Identically, with the growing development of the cognitive radio technology, the study of the security of the technology in various forms of networks is deepened. The trust mechanism and trust management have an irreplaceable position in the study of the cognitive network's security. In a distributed wireless network, as there is no large base station, the security of the network depends on the cooperation between users.

As a hot area of wireless system research, trust management has an irreplaceable position in network defense and security. A crucial factor of trust management is the reputation value, and it is requisite to have a further analysis as well as discussion of the reputation value and combine it with necessary technology [9].

This way, CWSN is a new concept with the following advantages.

(1) Higher transmission range.

(2) Fewer sensor nodes required to cover a specific area.

(3) Better use of the spectrum.

(4) Lower energy consumption.

(5) Better communication quality.

(6) Lower delays.

(7) Better data reliability.

This paper is organized as follows. Section 2 introduces the concept of spectrum allocation and "the last diminisher" algorithm in the field of fair division. Section 3 introduces the new trust value-based spectrum allocation algorithm. Section 4 simulates the algorithm in different conditions and evaluates its performance. Section 5 summarizes the conclusion.

2. Related Work

2.1. CWSN Construction. The cognitive wireless sensor network (CWSN) is a new network form of cognitive radio technology based on the WSN. AS shown in Figure 1, sensor nodes have been around and achieve information interaction and data transmission by the wireless communication technology. Applying the cognitive radio technology on the basis of the traditional wireless communication technology both expands the networks beyond a single spectrum space and renders them a dynamic spectrum access feature. Without interfering with the communication of authorized users, the secondary users can choose free spectrum for data transmission in real time and self-adaptively. To solve the problem of low utilization rate of fixed distribution spectrum and frequency resource scarcity in the field of radio, it is needed to find more effective ways to fully perceive and use wireless spectrum resources.

Cognitive radio can perceive the wireless communication environment and change the system's working parameters in real time and self-adaptively by a learning decision algorithm. It has two objectives:

(1) to improve the utilization rate effectively,

(2) to improve the communication reliability effectively.

2.2. Related Spectrum Allocation. Cognitive radio spectrum allocation assigns spectrums to one or more designated subscribers according to the number of users accessing networks and service requirements. The main purpose of spectrum allocation is to choose and utilize idle spectrums effectively through an adaptive strategy. The dynamic spectrum allocation strategy can effectively increase the flexibility of wireless communication, avoid conflicts between the authorized user and unauthorized user, fairly share spectrum resources, and meet user's changing requirements for different services.

At present spectrum allocation has the following major models: graph coloring model, game theory model, auction model, and interference temperature model [9].

The graph coloring model in the graph theory is more mature when applied to planning and disposing frequency in the cell in the era of cellular communication. The available spectrum of the cognitive user is influenced by the location, work status, and the coverage of the main user due to its random use of the spectrum by the main user, which changes in the idle time.

The graph coloring model was proposed by Zheng et al. based on the graph theory. In the study of spectrum allocation, the network topology mapping of cognitive users is abstracted as the graph $G(V, E, L)$ [10], each vertex in the graph representing the wireless user, and each edge representing a conflict or interference between the pair of vertices. In particular, if two vertices are connected by an edge, it is assumed that the two nodes cannot simultaneously use the same spectrum. Further, each vertex is associated with a collection, which means the spectrum resources can be used by the area where the vertex is located. Because of the difference of each vertex's location, the resource collection associated with a specific vertex is distinct.

The spectrum allocation in the cognitive system can be modeled as the throughput of the game, the player being the cognitive user, game strategy being the choice of the channel, and the game effectiveness related to the quality of the channel. The mathematical description of distributing spectrums using gaming is

$$\Gamma = \left\{N, \{S_i\}_{i\in N}, \{U_i\}_{i\in N}\right\}, \tag{1}$$

where N represents the finite set of the players, S_i is the strategy set of the player i, and $i \in N$ is the strategy space. Define $S = \times S_i$; then $U_i : S \rightarrow R$ is the effective function set. Each player I in the game Γ has an effective function set U_i, which is the function of strategy S_i chosen by the player and adversary strategy S_{-i}.

Spectrum allocation has a lot of similarities to the auction in economics. For example, the objects which are bidden have a certain value, which is able to bring improvement in performance to the occupants. And they both have many competitors as well. So this theory can be applied to spectrum allocation in cognitive radio. Today the auction bidding theory has been proved to be an effective way to solve the problem of spectrum allocation.

The auction model is usually used in the centralized network structure. The central controller acts as an auctioneer and the secondary users as the bidder. In every auction, bidders bid a price for spectrum resources to meet their own demand, and the auctioneer will determine a winner as ruled. Compared with other allocation methods, the advantages of the auction bidding mechanism are that it is naturally dispersed and requires less signaling exchange and computational overhead.

Interference temperature equivalent to noise temperature is used mainly as a measure of the interference power and its bandwidth. It is introduced by FCC to quantify and manage interference. In this model, the cognitive facilities working in licensed bands can adjust the transmitter power and spectrum by measuring the interference environment to avoid interference to the authorized user exceeding the specified temperature threshold.

The interference temperature model uses the interference temperature as the decision threshold that is, used as the maximum cognitive user interference signal tolerated by the authorized users to make a judgment. In the working frequency, as long as the cognitive user's accumulated radio frequency power does not exceed the set value (the threshold value refers to the maximum interference value tolerated by some frequency in a certain band), the secondary user can use this band. Any secondary user using this band must ensure that the interference to the authorized user does not exceed the threshold value.

2.3. Fair Allocation. Fair allocation [11] is to allocate limited resources to several players fairly. Allocation could be easy when the resource is a certain material which is divisible. However, when resources are various and indivisible and players are of various interests, allocation could be difficult, for example, the allocation of 20 cows of different breeds for 3 players. Hu and Ibnkahla [12] has purposed a consensus-based protocol for spectrum sharing to address fairness in CWSN.

Fair allocation is a mathematical theory on the basis of idealization and practical issues. The realistic problem is the allocation of divisible resources or materials to players who are potential owners of goods. The core issue of fair allocation is that the allocation should be conducted by players, who know the value of goods through negotiation or with the aid of an intermediary rather than an intercessor.

The fair allocation theory provides a definite standard for the division of various kinds of fairness. It aims at providing a program to implement fair allocation or proving impossibility and studying the disagreements between the theory and real life [11, 13].

The assumptions about the valuation of the goods or resources are as follows:

(1) each player has this own opinion about the value of each part of the goods or resources,

(2) the value of a player of any allocation is the sum of his valuations of each part. Often just requiring the valuations to be weakly additive is enough,

(3) in the basic theory the goods can be divided into parts with arbitrarily small values.

There have been many allocation strategies for fair allocation, most of which conflict with one another. But they can still be combined. The fairness allocation mentioned above here means that every participant can gain goods with the smallest possible difference in amount or proportion.

Theoretically, spectrum resources can be divided into arbitrarily small channels and allocated to those who want to transmit data. Because of this, we borrow the "the last diminisher" algorithm and propose reputation based fair allocation for cognitive radio networks. This allocation strategy is capable of making on-demand and fair allocation for an arbitrary number of secondary users.

3. A Trust Value-Based Spectrum Allocation Algorithm in CWSNs

This chapter makes a detailed analysis of the spectrum allocation algorithm proposed in this paper. In the cognitive wireless sensor network, we assume that the user's reputation value has been obtained through the trust management mechanism. And in a process of our spectrum allocation, the reputation value is fixed and cannot be changed. Assuming that the needs of secondary users are not exactly the same and that the quality of channels is not exactly the same, either, it is rational for secondary users to choose the appropriate channel for data transmission. All the above ensures the effective utilization of idle channels and the reasonable and fair distribution of channels between secondary users.

The fairness this article proposes is not the easy division in geometry, but rather matching one's requirement with the piece of goods.

3.1. The Last Diminisher Algorithm. For an arbitrary number of participants, a balanced fair allocation algorithm named "the last diminisher" was proposed by Banach and Knaster in 1944 as follows.

(1) The first person cuts a slice they value as a fair share.

(2) The second person examines the piece.

If they think the piece is less than a fair share, they then pass on the piece unchanged.

If they think the piece is worth more than a fair share, they trim off the excess and lay claim to the piece. The trimmings are added back into the to-be-divided pile.

(3) Each remaining person, in turn, can either pass or trim the piece.

(4) After the last person has made his decision, the last person to trim the slice receives it. If no one has modified the slice, then the person who cuts it received it.

(5) Whoever receives the piece leaves with his piece, and the process repeats with the remaining people. Continue until only 2 persons remain; they can divide what is left by the divider-chooser method.

The fairness of the algorithm is that none of the players can diminish the amount of goods to a value smaller than the average value according to his or her will. It is said that the result of reduction might belong to oneself. Identically, in the case of denying the benefit of someone else, the piece of goods cannot be greater than average, either. Intellectually, everyone would divide goods into average pieces in their own opinion.

In the process of polling, the method of allocating and competing for "goods" on the basis of needs is coincidently parallel with the idea of pursuing optimization in the allocation of spectrums. When allocating spectrums, different users who need to transmit data have different requirements, such as the requirement for transmission rate and band-width. Meanwhile, it is well known that as a process of multiple users towards multiple users, the allocation of spectrums should be done in a certain order. For this reason, our original intention is to try our best to guarantee that all the users who need to transmit data are able to get appropriate channels by allocating channels in a certain order.

3.2. System Parameters. The node which needs data transmission is the center. All the nodes within the hop range form a group and then exchange reputation values at that time. In the period of time (denoted as Tt) when reputation values are exchanged, all nodes in this group adopt spectrum sensing, until the last node submits the sensing information and grades the perceived spectrum channel (this period of time is denoted as Ts). In this period of time ($\max\{Ts, Tt\}$), any node receiving information which needs to be forwarded will be counted as the same transmission.

It's not our responsibility to consider the production of these trust values. In order to ensure the stable operations of CRNs which support mobile computing and ubiquitous computing, the establishment of trust is an open, heatedly discussed, and challenging issue. A large amount of research has been done both at home and abroad. The authors in [14] proposed a Markov chain-based trust model for analyzing trust value in distributed multicasting mobile ad hoc networks. The method for developing a good trust management system for wireless sensor networks (WSNs) has been proposed in [15, 16]. In distributed CWSNs [17], the spectrum allocation algorithm proposed can be used together with the trust management to solve some complicated problems [18, 19].

The following three parameters are used to determine the user demand: the transmission rate R at the previous hop, the length L of the transmission data, and the vector property $Ns(R, L)$, which is defined to represent the main demands of each SU perceived.

The following three parameters are used to determine the channel level: the channel capacity C, the maximum transmission power P accepted by the channel, and the duration D of the channel. Set a vector $Nc(C, P, D)$ for each idle channel.

The time function is defined in (1), which is the ideal transmission time of the data length L in the capacity channel C

$$T_i = \frac{L}{C_i} \qquad \text{s.t.} \quad i \in (1, 2, 3, \ldots, N). \tag{2}$$

The channel capacity C is related to the transmission rate R at the previous hop. For $C > R$ in the transmission channel, in order to reduce the decrease of the transmission rate at this hop, choose those channels whose capacity is the largest.

The duration D of the channel is related to the length L of the transmission data. When in transmission, it is better to select those channels whose $D > T, T$ being the ideal transmission time of the data length$L$ in the capacity channel C. Therefore, a transmission time longer than T is essential.

As shown in Figure 2, because there is no central database to store the user's personal information in the distributed network, each SU is equipped with a fixed private agent. The agent comprises two parts: the utility management center and the personal database. The personal database is used to store the user trust value, demand vector, and degree of participation. The utility management center calls out the SU's information from the personal database, calculates the value of the utility function, and upgrades the information in the personal database.

3.3. A Trust Value-Based Spectrum Allocation Algorithm in CWSNs. Order the SUs according to the rules:

$$\begin{aligned} &V \triangleq \text{Max}: a \times t + b \times \|Ns\| \\ &\text{S.t.}: \quad (1)\ a + b = 1, \\ &\qquad\quad (2)\ a > b, \\ &\qquad\quad (3)\ a > 0.5, \end{aligned} \tag{3}$$

where V is the order value of each SUs. SUs decide the order to be allocated channels according to the order value from big to small.

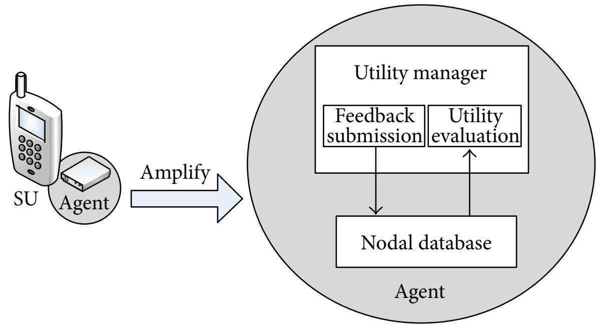


Figure 2: Agent architecture.

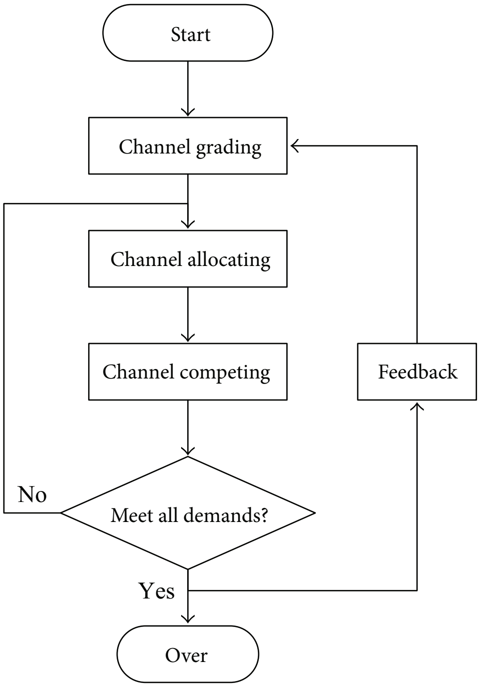


Figure 3: General flowchart of the algorithm.

The whole flowchart of the algorithm is shown in Figure 3. When one of the SUs has the need to transmit a series data, the algorithm starts. The agents of all channels that are unused calculate their own priority levels and decide which channel owns the better characteristics. Then the SUs begin to select the satisfying channel in the order of the order value. When one SU has selected a channel, other SUs whose order value is smaller than that one have the opportunities to give opinions about the match between the SU and the channel. The last one who holds the opposite attitude towards this match will have another chance of a priority selection. When all the needs of SUs are met, the algorithm is over. Those SUs who have been allocated a channel transmit the data, and the feedback of the behaviors of SUs will reflect SUs' order value.

3.3.1. Channel Grading. As is shown in Figure 4, channel grading is dividing and registering the channel in accordance with the assumption in Sections 3.1 and 3.2 of this section.

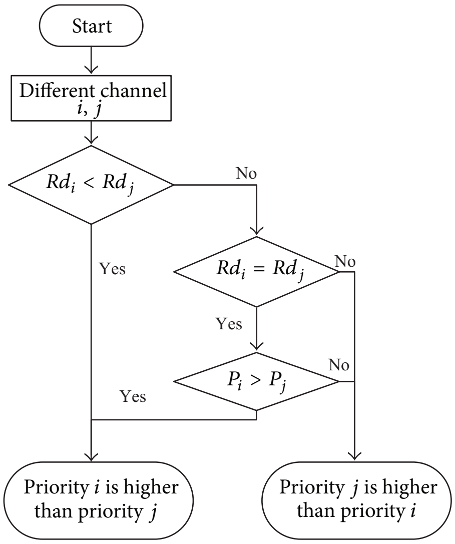

Start
Different channel i, j
$Rd_i < Rd_j$
No
Yes
$Rd_i = Rd_j$
No
Yes
$P_i > P_j$
No
Yes
Priority i is higher than priority j
Priority j is higher than priority i

Figure 4: Channel grading.

The channel with a higher level must have a better channel capacity and longer channel idle time. It will be requested by a large number of secondary users. And the low level user may have less channel capacity, shorter idle time, or less channel-allowed transmitter power. All in all, it has some parameters which cannot meet the demand of secondary users, or its channel is not welcome to the secondary users.

Suppose that the data length to be transmitted at one point is L and the transmission data of the previous hop is R. The number of channels perceived by the secondary user is N, C is the capacity of channel, its transmitted power is P, and the idle duration is D. $p(V_i)$ represents the channel's priority level.

The priority function is shown in (3):

$$Rd_i(L) = \frac{L}{C_i D_i} \quad \text{s.t.} \quad i \in (1, 2, 3, \ldots, N). \tag{4}$$

If $Rd_i < Rd_j$, $i \neq j$, $i, j \in (1, 2, 3, \ldots, N)$, then $p(V_i)$ is higher than $p(V_j)$. The smaller the ratio of transmission time to the channel idle time, the sooner the same amount of data will be transmitted completely, and thus, the probability of being affected accidentally (e.g., primary user comes back ahead) is reduced.

If $Rd_i = Rd_j$ and $P_i > P_j$, $i \neq j$, $i, j \in (1, 2, 3, \ldots, N)$, then $p(V_i)$ is higher than $p(V_j)$. The greater the channel transmission power, the higher the priority of the channel, of course.

3.3.2. Channel Allocation of SUs. This part is shown in detail in Figure 5. Assume there are M users in total, SU_m, $m \in (1, 2, 3, \ldots, M)$, the user credibility values have been known and stored from the highest to the lowest; that is, $t_1 > t_2 > \cdots > t_k > t_M$.

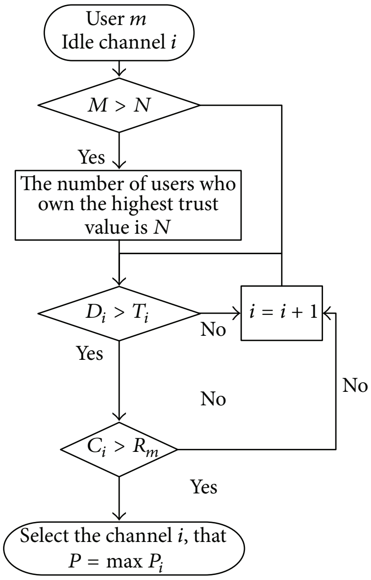

User m
Idle channel i
$M > N$
Yes
The number of users who own the highest trust value is N
$D_i > T_i$
No
$i = i + 1$
Yes
No
No
$C_i > R_m$
Yes
Select the channel i, that $P = \max P_i$

Figure 5: Channel allocation of SUs.

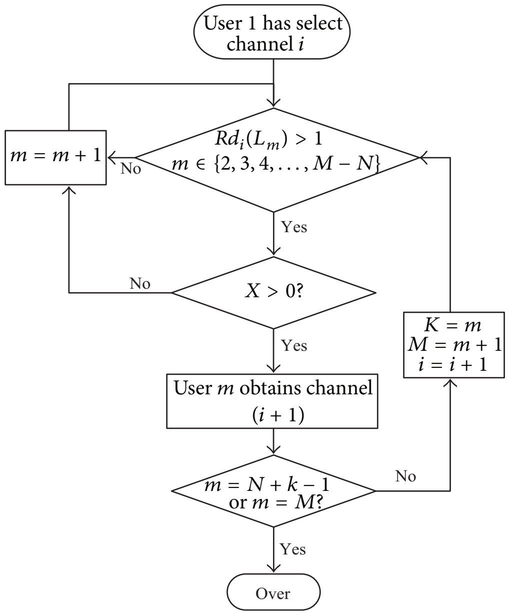

User 1 has select channel i
$Rd_i(L_m) > 1$
$m \in \{2, 3, 4, \ldots, M - N\}$
$m = m + 1$
No
Yes
No
$X > 0?$
Yes
$K = m$
$M = m + 1$
$i = i + 1$
User m obtains channel $(i + 1)$
$m = N + k - 1$ or $m = M?$
No
Yes
Over

Figure 6: Channel Competition in SUs.

Step 1. If $M > N$, which means the number of idle channels to be selected is small while the number of secondary users is large, to ensure that the secondary user with low reputation does not interfere with the channel of the user with high reputation, users who have high reputation participate in every selection of channels, and the rest of the users with low reputation will not be involved. If the low reputation user occupies the channel forcibly, its behavior will be recorded and his reputation is reduced or he is determined to be a malicious user according to the strategy.

Step 2. Choose $Rd_i(L) > 1$. When second users are transmitting data in a channel, we guarantee the time that second users spend in transmitting information, which means that the duration in which the channel is idle is longer than the time needed when transmitting data under perfect conditions.

Here, we assume that the time when the channel is idle has been gained in the active detecting method. For the reason that the throughput in the sensor networks is huge but the message included is about the external environment, the amount of data is small, and when taking security, efficiency, and confidentiality into consideration, different data types have different needs. Consequently, when choosing an appropriate channel, users have to choose those whose idle time is enough to do a complete data transmission.

Step 3. Under condition 1, we choose the channels with the characteristic $C_i > R_m$. In order to keep the speed of data transmission, we better choose the channels with $C > R$. In practical application, the actual data transmission speed is less than the channel capability. Sometimes we can promote the transmission speed comparing with the previous communication.

Step 4. Under conditions 2 and 3, we choose the transmission power $P = \max P_i$, $i \in (1, 2, 3, \ldots, N)$, which means that we select the one with the highest rate from available channels. Otherwise, we can choose the channel in any method.

Step 5. Set $i = i + 1$, turn to Step 1.

3.3.3. Channel Competition in SUs. The flowchart is shown in Figure 6. After the second user SU whose trust value is t selects channel i during the current channel choosing phase, the $M - N$ second users with the highest trust values compete for a channel. Assume the trust values are set from high to low as $t_1 > t_2 > \cdots > t_k > t_m$. And the second user with trust value t_k selects channel i.

Step 1. The second user with trust value t_m, ($m = 2, 3, 4, \ldots, M - N$) can decide whether to compete for channel i. If he decides to compete for this channel, turn to Step 2, or else turn to Step 4.

Step 2. If $Rd_i(L_m) > 1$, this means that the time needed for transmission is shorter than that in the perfect condition and turn to Step 3, or else the second user fails to use the channel and then turn to Step 1.

Step 3. Set the competition value in (4):

$$\begin{aligned} x &= \frac{\sqrt{L_m - L_k}}{\sqrt{R_m - R_k}}, \\ \text{s.t.} \quad & i \in (1, 2, 3, \ldots, N), \\ & m \in (2, 3, 4, \ldots, M - N). \end{aligned} \tag{5}$$

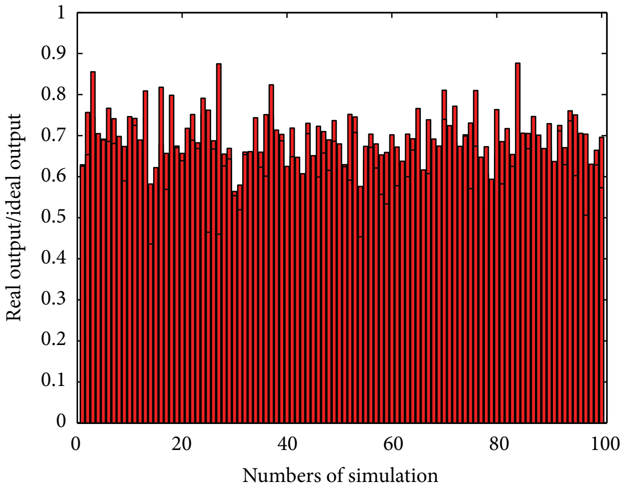


Figure 7: Throughout in 100 users, 10 channels, and 100 simulations.

If x exists and is greater than 0, the secondary user succeeds in this competition and accordingly can get the channel $i + 1$ whose priority is just less than i. Meanwhile, the channel i is allocated to the second user with trust value t_k and then turn to Step 4. Here it is required that the amount of data from competitors be huge and the transmission rate in the last hop be low. Through these adjustments, more secondary users will have the chance to obtain channels for transmission and do not need to wait for too long or be blocked.

Step 4. Set $k = m$, $m = m + 1$, and $i = i + 1$, and then turn to Step 1. If $m = N + k + 1$ or $m = M$, it means that all the available second users have been taken into this action and turn to Step 5.

Step 5. The competition is ended. All the second users that get channels can transmit messages in the channel, and other second users stay silent.

3.3.4. Feedback. Every agent of the nodes gathers the feedback information after a transmission. Whether to increase or decrease the trust value is decided by feedback information which has reached the threshold or not.

If the SUs who have been allocated an idle channel correspond and transmit data normally, increase the trust value by fifty percent. If the SUs have the following behaviors, decrease the trust value to a half:

(1) having nothing to do with the allocated channel,

(2) always in the competition,

(3) destroying the communication function of the network in the form of replay attack, Dos attack, Sybil attack and so forth,

(4) leaking users' privacy.

The trust value is related to order value, which decides the SUs' sequence in allocation. When an idle channel is allocated to an SU, the competing SUs cannot take it into account. That is to say, the trust value determines the ownership of the idle channels.

Identically, the trust value also plays an important role in the field of trust management, and some more in-depth research can be done in the near future.

4. Simulation and Evaluation

"The last diminisher" algorithm proposed by Banach and Knaster can solve the "Cake Cutting" problem impartially, but it cannot provide a satisfying solution for every user. Presume there is a piece of cake with multiple flavors, chocolate, cream, strawberry, and so forth, people who share the cake have different taste for cakes, and different parts of the cake have different values to different people. Geometrically equal distribution and simple equal distribution cannot solve the "Cake Cutting" problem. Therefore, this article uses "the last diminisher" algorithm for reference but does not distribute frequency spectrums totally depending on the algorithm. This article presents a method by which users can get satisfying frequency spectrums according to their particular needs.

Here, the frequency spectrum is the cake we mentioned before. We know different parts of the cake have different flavors. In other words, different idle channels have different parameters. Furthermore, users not only care about whether they can get a channel but also whether the channel they get can satisfy their needs. Therefore, "the last diminisher" algorithm cannot meet our requirements.

Banach and Knaster provided a useful thinking: polling for advice and then getting the "cut cake".

The frequency spectrum distribution method this article presents can ensure the following aspects:

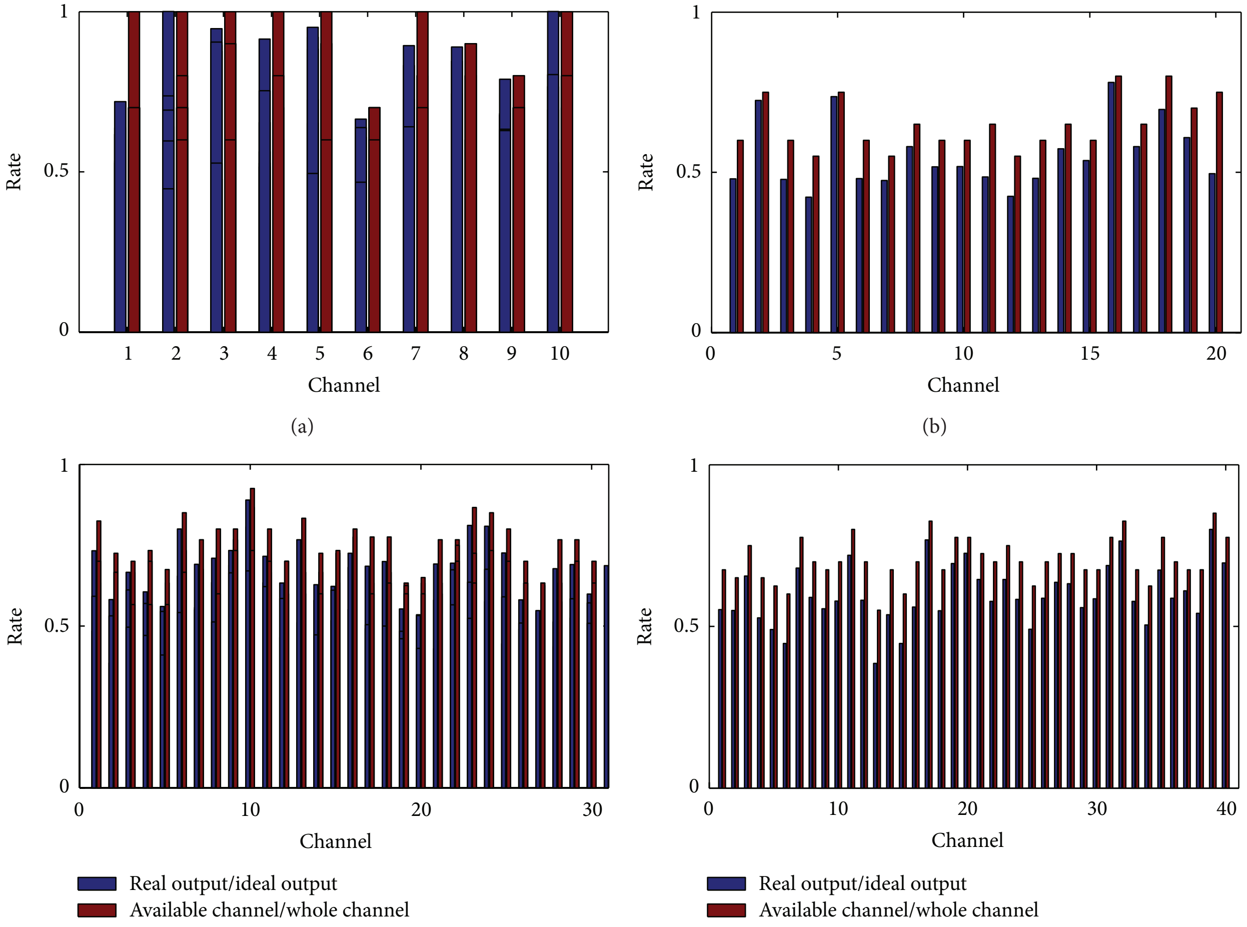
(1) user A with the highest reputation value is the first to select, and he is certain to get the channel i he selects. Even if user B disagrees with user A's selection of channel i and user A doesn't get the channel this round, user B will have to select his ideal channel from channels except i. In the next round, User A still possesses the highest reputation value, and thus he will definitely again select channel i,

(2) users who have high demands can select high-level channels; users who have low demands can select low-level channels,

(3) users with low reputation levels cannot get channels selected by users with high reputation values, and even if they finally get that channel, that channel is a degraded channel.

If users with a low reputation value or demands select a channel that users with high reputation value or demands need, then we regard these users as malicious users and reduce their reputation values according to the situation in the next round's reputation value calculation.

In order to further explain that our method has outstanding performance, we build a simulation environment. The number of secondary users and available channels are given, the sensor network has numerous and densely distributed



(a)

(b)

(c)

(d)

Figure 8: Throughput and utilization rate in 100 users 10, 20, 30, 40 and channels.

ground nodes, and thus the number of users are much larger than that of channels. In order to guarantee randomness and fairness of the simulation environment, every user has an initial random reputation value and random transmission data with the data length, data type, and transmission speed of the last jump. If no data is to be transmitted, then the data length is 0. For available channels, the channel capacity, idle duration, and maximum bearable transmission power of the channel are random in a certain range. Nearly all simulation parameters are random, which is most like the real sensor network environment.

We set the number of users as 100 and that of available channels as 20, and other parameters are random in a certain range and then perform 100 experiments. As is shown in Figure 7, the network's ratio of actual throughput to ideal throughput is generally a relatively stable value, 0.7. The throughput ratio should be 1 in the ideal situation, which means all idle channels can be used by secondary users, and the transmission rate equals the channel capacity.

When it comes to practical simulation, we need to allocate channels to users following our algorithm. Random parameters may lead to $D > T$ or $L = 0$ not being satisfied, and thus some idle channels cannot be selected and fully utilized by users because of their inappropriate parameters. This is also characteristic of the actual sensor network. We can see from Figure 7 that the simulation results and simulation count have no regularity, which means our algorithm will have stable and efficient throughput under any circumstances.

On the other hand, the utilization rate is a public focus. We combine the utilization rate and network throughput and conduct persuasive comparison. The utilization rate is the ratio of the number of channels allocated to users after one allocation to the total number of channels.

We set the number of users as 100 and that of channels as 10, 20, 30, and 40, respectively, and the results are shown in Figure 8.

(1) The pattern of the channel utilization rate is similar to that of throughput, which means the throughput value will be greater if the channel utilization rate is higher.

(2) There is no fixed relation between the channel utilization rate and the number of simulation. When the number of available channels is quite small but the total number of channels is big, all channels will be

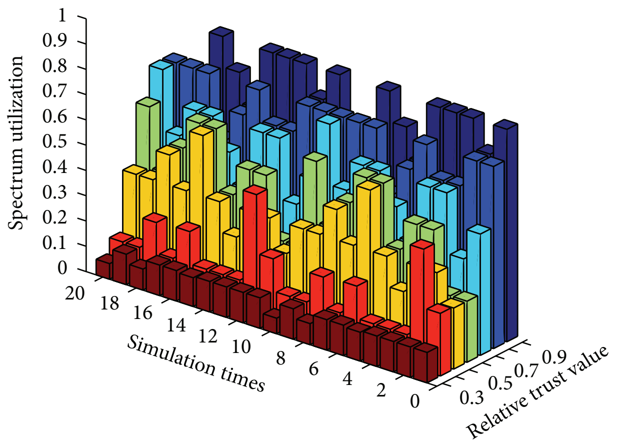


Figure 9: Spectrum utilization with trust values in 20 simulations.

distributed to secondary users, which is the spectrum utilization rate shown in Figure 8.

As is shown in Figure 9, in 20 simulations, the spectrum utilization, which is now equal to channel utilization, varies from the trust value. With the trust value increasing, the spectrum utilization has an obvious growth particularly. In our simulations, we set a large number of parameters randomly. In the 12th simulation, the spectrum utilization is highlighted when the trust value t is 0.4, which shows our simulation circumstance is close to reality.

This article analyzes and deals with the distribution of idle channels of different parameters to secondary users with different needs, with reference to the "Final Cutter" method. We basically guarantee every user can get the suitable spectrum, and the performance of our method is basically stable for the whole network. The variation of channel parameters and user needs caused by external disturbance will not influence the whole network's status.

5. Conclusion

This article proposes an algorithm for dynamic spectrum allocation in CWSNs, which integrates the idea of "the last diminisher" in the field of fairness allocation, the demand of SUs, and the channel characteristics.

At the beginning of the article, we give a brief introduction to the necessity of the CWSN spectrum allocation and put forward our main idea. In Section 2, we find works related to our article. The next two sections are our contributions to allocating spectrum recourses fairly, including the algorithm and simulation. The algorithm, aiming at complicated realistic sensor networks, has stable output and channel utilization.

The WSNs are facing increasingly congestible frequency spectrums and strait channels. In dealing with the inadequacy of spectrum resources, we still have a long way to go. Further researches need to be done on the cognitive WSNs from the aspects of trust management, spectrum sharing, power control, and so forth.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 61172068, the Program for New Century Excellent Talents in University (Grant no. NCET-11-0691), and the Fundamental Research Funds for the Central Universities (Grant no. K50511010003).

References

[1] I. Howitt and J. Gutierrez, "IEEE 802.15.4 low rate-wireless personal area network coexistence issues," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '03)*, vol. 3, pp. 1481–1486, 2003.

[2] D. Cavalcanti, R. Schmitt, and A. Soomro, "Achieving energy efficiency and QoS for low-rate applications with 802.11e," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, vol. 1, pp. 2145–2150, March 2007.

[3] IEEE 802.11 Standard, Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, (Reaff 2003) Edition, 1999.

[4] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Proceedings of the IEEE International Workshop on Mobile Multimedia Communications*, pp. 3–10, 1999.

[5] 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.

[6] J. Mitola, "Cognitive radio: an integrated agent architecture for software defined radio," *Doctor of Technology, Royal Institute of Technology*, pp. 271–350, 2000.

[7] X. Zhang and C. Li, "The security in cognitive radio networks: a survey," in *Proceedings of the ACM International Wireless Communications and Mobile Computing Conference (IWCMC '09)*, pp. 309–313, June 2009.

[8] K. Govindan and P. Mohapatra, "Trust computations and trust dynamics in mobile Adhoc networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 2, pp. 289–298, 2012.

[9] J. H. Cho, A. Swami, and I. R. Chen, "Modeling and analysis of trust management with trust chain optimization in mobile ad hoc networks," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 1001–1012, 2012.

[10] U. V. Catalyurek, J. T. Feo, A. H. Gebremedhin, M. Halappanavar, and A. Pothen, "Multithreaded algorithms for graph coloring," *Parallel Computing*, vol. 38, no. 10-11, pp. 576–594, 2012.

[11] E. N. Barron, *Game Theory: An Introduction*, Wiley-Interscience, 2011.

[12] P. Hu and M. Ibnkahla, "A Consensus-based protocol for spectrum sharing fairness in cognitive radio Ad Hoc and sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 370251, 12 pages, 2012.

[13] J. Lopez, R. Roman, I. Agudo, and C. Fernandez-Gago, "Trust management systems for wireless sensor networks: best practices," *Computer Communications*, vol. 33, no. 9, pp. 1086–1093, 2010.

[14] B. J. Chang and S. L. Kuo, "Markov chain-based trust model for analyzing trust value in distributed multicasting mobile Ad Hoc networks," in *Proceedings of the Asia-Pacific Services Computing Conference*, pp. 156–161, 2008.

[15] C. T. Clancy and N. Goergen, “Security in cognitive radio networks: threats and mitigation,” in *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom ’08)*, pp. 1–8, May 2008.

[16] S. Parvin, S. Han, L. Gao, F. Hussain, and E. Chang, “Towards trust establishment for spectrum selection in cognitive radio networks,” in *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications (AINA ’10)*, pp. 579–583, April 2010.

[17] Y. C. Liang, K. C. Chen, G. Y. Li, and P. Mahonen, “Cognitive radio networking and communications: an overview,” *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3386–3407, 2011.

[18] H. M. K. Alazemi and M. F. Uddin, “Fair resource allocation and DF relay selection for multiuser OFDMA-based cooperative networks,” *Wireless Networks*, pp. 1–12, 2013.

[19] Q. Zhang, S. Kota, V. Lau, W. Su, and A. Kwasinski, “Introduction to the issue on cooperative communication and signal processing in cognitive radio systems,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 1–4, 2011.