Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 839320, 8 pages
http://dx.doi.org/10.1155/2013/839320

Research Article

An Approach Based on Chain Key Predistribution against Sybil Attack in Wireless Sensor Networks

Chunling Cheng,[1,2,3] Yaqiu Qian,[1] and Dengyin Zhang[3]

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

Correspondence should be addressed to Chunling Cheng; chengcl@njupt.edu.cn

Received 8 July 2013; Accepted 12 September 2013

Academic Editor: Zhijie Han

In wireless sensor networks (WSN), Sybil attack can destroy the routing and data distributed storage mechanisms through fabricating identity information of legitimate nodes. This paper presents a chain key predistribution based approach against Sybil attack. To enhance the security of common keys between neighboring nodes, during the chain key predistribution phase, our approach uses a lightweight hash function to generate several chain keys by hashing the unique identity information of every node sequentially in the trusted base station. These chain keys construct a pool of chain keys. During the pairwise key authentication establishment phase, a node-to-node chain key based authentication and exchange (CK-AE) protocol is proposed, by which each node can share the unique pairwise key with its neighboring node. The CK-AE protocol is provably secure in the universally composable security model (UCSM). Finally, we analyze our approach from the resilience of network and the performance overhead, and the results show that our approach can not only enhance the ability of resilience to Sybil attack, but also reduce the communication overhead significantly at the cost of a certain amount of computational overhead.

1. Introduction

In recent years, with the wide application of wireless sensor networks (WSN), the increasing number of the terminal devices brings the explosive growth of data. The WSN is usually deployed in many areas, including traffic control, health care, environmental monitoring, and battlefield surveillance, where the sensor nodes obtain the useful information by gathering, storing, and analyzing the large number of data [1]. However, due to the openness of deployment, the limitation of resource and the Big Data environment, security emerges as a challenging issue in ubiquitous WSN. Sybil attack is a particularly harmful attack which can forge a large number of fake identities to disrupt the routing protocol, distributed storage, and malicious behavior detection [2]. Moreover, the entire communication of the network will even be destroyed. Therefore, it is particularly important to develop an effective approach to defend against Sybil attack.

So far, some defense approaches against Sybil attack in WSN are mainly based on a random key predistribution scheme [3]. In the random key predistribution scheme, each node is preloaded with a key ring (blocks of keys) randomly selected from a large pool of keys. After the deployment step, every node exchanges several keys which belong to its key ring with each of its neighbors. If two neighbor nodes share at least one key (common key), they establish a secure link and calculate their pairwise key (session key) which is one of the common keys. The process of establishing a pairwise key can authenticate the node identity; therefore, the fake identities of a Sybil attacker can be distinguished from the legitimate nodes. Furthermore, a random key predistribution scheme is implemented before deployment; thus the authentication of node identity can depend little on the trusted base station, which reduces the computation and communication overhead of dynamically generating pairwise keys.

Karlof and Wagner proposed a mechanism based on a unique common key for WSN against Sybil attack [4]. Every node shares a unique common key with a trusted base station. When two nodes need to communicate with each other, they first signal their desire to the base station. Then the base station verifies the identifications of the two nodes via the unique common key and distributes a pairwise key to them. After that, two nodes use the distributed key to verify each other's identity and establish a session link by a Needhan-Schroeder like protocol. During the establishment of session between two nodes, the base station must be requested to authenticate the node identities, which leads to a heavy load in the base station if the number of the nodes is large in a WSN. Besides, using the same distributed key may cause the situation that the attacker can easily capture the entire network by capturing only one node. Based on the idea of the pairwise key in [4], Newsome et al. proposed a random key predistribution scheme against Sybil attack in WSN [5]. In the scheme, a random set of keys or key-related information is assigned to each node. Then each node can discover the common keys it shares with its neighbors and the common keys will be used as a session key to ensure node-to-node secrecy. Nevertheless, the set of keys or key-related information is just determined by a pseudorandom function (PRF) in this scheme; therefore, the unsecure keys or key-related information may lead to unsecure session keys between the neighbor nodes. Once a Sybil attacker breaks the PRF, it will be able to quickly obtain the keys or key-related information in other nodes and disguise their identities by forging session keys. On the basis of random key predistribution scheme, Roberto et al. proposed a pairwise key establishment mechanism based on the node identity information in an attempt to protect the WSN from the Sybil attack [6]. However, due to the difficulty of constructing the specific node identity information, the mechanism is not practical and effective. Furthermore, Feng and Ma improved the Roberto's mechanism and presented a novel approach which is node identity witness information validation for random secret information predistribution to defend against Sybil attack [7]. But in this approach, the transmitter/receiver must send more messages during the process of establishing a secure link, which means more communication overhead.

Qian proposed an improved key predistribution scheme in which each node calculates the derived keys by using a hash function once [8]. This solution enhances the security of the original keys. However, the derived keys are calculated by every node after deployment. So, the computational overhead of the nodes is increased. In order to enhance the resilience of the network against attackers, Bechkit et al. proposed a novel hash-based key predistribution scheme [9]. Before deployment a hash function h is preloaded to the memory of each node. Then every node applies the hash function h to each key of its key ring. After that, the neighbor nodes calculate the pairwise keys by the hash function h and establish a secure link. However, in this solution the calculated pairwise keys are not unique. So the probability of successfully forging the pairwise key by a Sybil attacker is increased. If the forged pairwise key is the same as the legitimate one, the communication of the neighbor nodes will be disrupted by the Sybil attacker's fake identity.

As described above, the existing approaches play a certain part in defending against Sybil attack in WSNs, but some issues still exist, such as high communication overhead of the trusted base station, unsecure keys, high communication overhead of the process of node identity authentication, and nonunique pairwise key and so on. In this paper, we propose a chain key predistribution based approach to protect the WSN from Sybil attack. In the chain key predistribution phase, the chain keys are calculated with the entire node identity information through a lightweight hash chaining technique in the trusted base station, which conceals the original keys and makes the common chain keys more secure that preloaded in different nodes. Then a chain key pool which constitutes several hash chains of equal length is constructed. After deployment, we develop a node-to-node chain key based authentication and exchange (CK-AE) protocol in the pairwise key authentication establishment phase which is secure against the Sybil attackers. At last, we prove the security of our approach and make analysis for the connectivity, communication overhead, storage overhead, and computational overhead.

2. A Chain Key Predistribution Based Approach against Sybil Attack in WSNs

The proposed approach is divided into two phases: the chain key predistribution phase and the pairwise key authentication establishment phase. Two phases are described in detail below.

2.1. Chain Key Predistribution Phase. The issue of the unsecure keys usually makes the nodes vulnerable to the Sybil attack. If an attacker captures one node, it can get all the key information from its memory and impersonate other legitimate nodes. Such an attack can compromise not only adjacent links of compromised links but also external links that are independent of the compromised nodes. Therefore, in order to make the key information of nodes more secure, we generate several chain keys by hashing the unique identity information of every node sequentially based on a lightweight hash function H^{Nyb} [10] in the trusted base station. As known, the main characteristic of hash functions is that knowing a value of the chain it is computationally infeasible to determine the backward values. Therefore, our chain key predistribution scheme can conceal the original keys and enhance the resilience of common keys shared by the neighbor nodes. In our scheme, when an attacker captures one or more original keys it can only discover a forward value.

In the chain key predistribution phase, the generation of the hash chains is implemented in the trusted base station, and then the hash chains are constructed into a pool of chain keys. Take the generation procedure of the hash chain L_i for example. It is assumed that seed_i is an original key and the derived factor is a set of the node identities SIDSet (an ordered sequence of values). The set SIDSet is $\{\mathrm{SID}_1, \mathrm{SID}_2, \ldots, \mathrm{SID}_j, \ldots, \mathrm{SID}_N\}$ where SID_j $(1 \leq j \leq N)$

is the unique node identity information and N is the number of the nodes. The procedure of generating the jth chain key in the hash chain L_i is given as follows:

$$k_j^i = H^{\text{Nyb}}\left(k_{j-1}^i, \text{SID}_j\right), \quad \{i, j \in N^*,\ 1 \le j \le N\}. \tag{1}$$

N^* represents a set of positive integer. According to (1), $k_1^i = H^{\text{Nyb}}(\text{seed}_i, \text{SID}_1)$, $k_2^i = H^{\text{Nyb}}(k_1^i, \text{SID}_2)$, and so on. The last value of the hash chain k_N^i is $H^{\text{Nyb}}(k_{N-1}^i, \text{SID}_N)$. We define the value K_N^i as the hash chain tag Tag_i which is calculated from seed_i. Different hash chain tags are calculated from different original keys. The other hash values $(k_1^i, k_2^i, \ldots, k_{N-1}^i)$ are defined as the chain keys. To illustrate our idea about the generation procedure of the hash chain L_i, let us refer to Figure 1.

When a given hash value k_m^i needs to be verified in the above hash chain, it is feasible to calculate k_n^i $(0 < n < m)$ $m-n$ times through the hash function H^{Nyb}. Then we compare the output ${k_m^i}^*$ with the given hash value k_m^i. If ${k_m^i}^* = k_m^i$; it means that ${k_m^i}^*$ is a correct hash value. In order to construct a pool of chain keys, we select M different original keys $\{\text{seed}_1, \text{seed}_2, \ldots, \text{seed}_M\}$ from the trusted base station. By repeating the above generation procedure in Figure 1, M hash chains with the same length are generated. In our approach, the pool of chain keys is composed of M hash chains, which is shown in Figure 2. Each hash chain contains $N-1$ chain keys and the size of the pool K is $M * (N-1)$.

After the construction of the pool of chain keys, each node is preloaded with some chain key related information. The detailed performance step of the chain key distribution is given as follows. (1) Each node selects m different chain keys from the pool. It is noted that the method about selecting the chain keys from the pool is not limited. Moreover, the indexes of every selected chain key in the same hash chain are also preloaded in each node. Besides, if there is any chain key selected from one hash chain, the hash chain tag of this hash chain is also stored in the memory of the node. (2) Then the hash function H^{Nyb} is preloaded to the nodes.

Due to the limited resource of the sensor nodes, we apply the hash function H^{Nyb} which is based on WH function [11] to implement the computation of the chain keys. WH function not only has the traditional security of hash function clusters, but also gets a good efficiency about computation. Therefore, it is suitable to be applied to the actual hardware of the sensor nodes.

After the chain keys predistribution phase, all the sensor nodes are randomly deployed in a designated area. The trusted base station just participates in the first phase without communicating with the nodes in the second phase, therefore, the proposed approach has the advantage of reducing the high communication overhead of the base station. Figure 1 shows that every chain key is calculated from different original keys by hashing the unique identity information of nodes sequentially. So, the original keys and the node identities are concealed by such computation, and the resilience of shared pairwise keys between the neighbor nodes is also enhanced. When an attacker captures one or more original keys, it can not fake a derived chain key correctly with an unknown number of times.

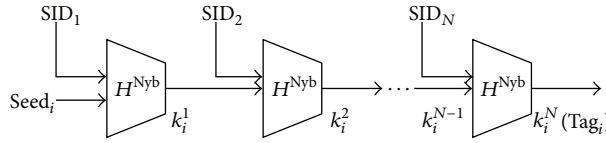

Figure 1: Generation procedure of the hash chain.

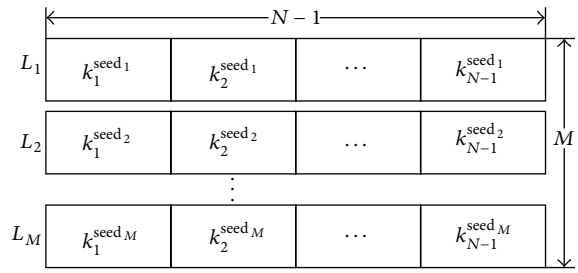

Figure 2: Composition of the pool of chain keys.

2.2. Chain Key Based Pairwise Key Authentication Establishment Phase. A session link established by a random key predistribution scheme usually leads to an important issue that the pairwise key is not unique [9, 12]. Thus, a Sybil attacker can easily disrupt the communication between the neighbor nodes. To solve this issue, we propose a node-to-node chain key based authentication and exchange (CK-AE) protocol in the pairwise key authentication establishment phase. By the CK-AE protocol, the neighbor nodes authenticate the identity information of the opposite node and then calculate a unique pairwise key to establish a secure session link for defending against the adversaries.

In the CK-AE protocol, one node stores multiple different key chains. For example, the ring of m chain keys of S_A is $\{k_{j,A}^i\}$, where $1 \le j \le m, 1 \le i \le M$. A protocol is usually executed distributedly, so we mark the different protocol instances as Sid. It is assumed that S_A and S_B are neighbors. The detailed description of CK-AE protocol is given as follows where "$\oplus$" represents XOR operation and "|" is connection operation. The parameters of hash function H^{Nyb} are exchangeable, which means $H^{\text{Nyb}}(x, y) = H^{\text{Nyb}}(y, x)$ [10].

(1) $S_A \rightarrow S_B$: after deployment, S_A broadcasts the packet $\{\text{Sid}_x, E(\text{Tag}_{i,A}, k_{j,A}^i \mid \text{Index}(j)_A \mid \text{SID}_A)\}$ to its neighbors. In this packet, Sid_x indicates the session identifier. $k_{j,A}^i$ is a chain keys which is selected from the ring of m chain keys of S_A. $\text{Index}(j)_A$ represents the corresponding index of chain key $k_{j,A}^i$. SID_A is the node identity information of S_A. $\text{Tag}_{i,A}$ is the hash chain tag which $k_{j,A}^i$ is selected from. In our approach, we use $\text{Tag}_{i,A}$ as the broadcast communication key. $E(k, M)$ means that M is encrypted by using the key k.

(2) $S_B \rightarrow S_A$: if S_B and S_A have the chain keys which are selected from the same hash chain, S_B can decrypt the received broadcast packet by using the preloaded hash chain tag $\text{Tag}_{i,B}(= \text{Tag}_{i,A})$. After decrypting the packet, S_B will use the hash function H^{Nyb} to authenticate the received chain key $k^i_{j,A}$. During authentication, S_B first finds the hash chain L_i that $\text{Tag}_{i,B}$ belongs to, then selects the chain key $k^i_{j,B}$ which is preloaded from the hash chain L_i and its corresponding index $\text{Index}(j)_B$ in L_i. Second, S_B calculates the difference between two indexes diff = $|\text{Index}(j)_B - \text{Index}(j)_A|$. (a) If $\text{Index}(j)_B \geq \text{Index}(j)_A$, S_B calculates $k^i_{j,B}{}^* = H^{\text{Nyb}}(k^i_{j,A}, \text{SID}_{j+1})^{\text{diff}}$ (using H^{Nyb} to calculate diff times). Then if $k^i_{j,B}{}^* = k^i_{j,B}$, the chain key $k^i_{j,A}$ is verified correctly by S_B. It is means that S_B knows S_A is a legitimate node; then it can establish a pairwise key k_{AB} with S_A. (b) If $\text{Index}(j)_B < \text{Index}(j)_A$, S_B calculates $k^i_{j,A}{}^* = H^{\text{Nyb}}(k^i_{j,B}, \text{SID}_{j+1})^{\text{diff}}$. Then if $k^i_{j,A}{}^* = k^i_{j,A}$, the chain key $k^i_{j,A}$ is verified correctly by S_B. It is also means that S_B knows S_A is a legitimate node; then it can establish a pairwise key k_{AB} with S_A.

After authenticating S_A, S_B selects a random number r_B and calculates the pairwise key $\text{key}_{AB} = k^i_{j,A} \oplus k^i_{j,B}$. Moreover, S_B calculates two temporary variables $X_B = k^i_{j,B} \oplus \text{SID}_B$ and $Y_B = \text{Index}(j)_B \oplus r_B$. At last, S_B delivers the packet $\{\text{Sid}_x, X_B, Y_B, \text{SID}_B, \text{diff}, r_B\}$ to S_A by unicast.

(3) S_A: after receiving the packet from S_B, according to the session identifier Sid_x S_A learns that the packet is a response message from S_B. Then S_A calculates $k^i_{j,B}{}^* = X_B \oplus \text{SID}_B$ and $\text{Index}(j)_B{}^* = Y_B \oplus r_B$. By judging the equation diff = $|\text{Index}(j)_B{}^* - \text{Index}(j)_A|$, S_A certifies S_B whether it is a legitimate node. If the equation is correct, S_B is a legitimate node, otherwise not. Finally, S_A calculates the pairwise key $\text{key}_{AB} = k^i_{j,A} \oplus k^i_{j,B}{}^*$ which can ensure a secure session link between the neighbor nodes.

In the above pairwise key authentication establishment phase, if S_B failed to decrypt the data packets with $\text{Tag}_{i,B}$, S_B should discard the packet. By this way, we can prevent the leakage of packet information if S_B is captured by an adversary. When the key chain cannot be authenticated, the nodes must discard the packet too, because the chain key may be forged by a Sybil attacker. In addition, after establishing the unique pairwise key, the nodes will delete the sent and received packets.

3. Security Analysis

The security of a protocol is usually defined and analyzed within the universally composable security model (UCSM) [13–15]. When a protocol P is proved to be secure in this model, it is able to ensure that when the protocol P runs in parallel with other protocols or runs separately, the protocol P is still secure. Generally speaking, the definition of the security of a protocol within the UCSM which means a UC-secure protocol is given as follows: the mutual information of protocol between the parties in any real model is computationally indistinguishable for the Environment (E) with the mutual information in the ideal model which a trusted party "ideal functionality F" exists in.

In this section, we first present the ideal functionality of the CK-AE protocol within the UCSM. Then we prove the security of the CK-AE protocol according to the definition within the UCSM. At last, the defense capability against Sybil attack of our approach is analyzed.

3.1. Ideal Functionality of the CK-AE Protocol. Based on the chain key predistribution scheme, we give our formulation of the ideal functionality $F_{\text{CK-AE}}$ for CK-AE protocol within the UCSM. The ideal functionality $F_{\text{CK-AE}}$ in the ideal model which describes the functions of the CK-AE protocol is shown as follows.

Ideal Functionality $F_{\text{CK-AE}}$. Functionality $F_{\text{CK-AE}}$ interacts with the parties S_A, S_B and an ideal adversary S via the following queries.

(1) Upon receiving a query {New Session, Sid, $k^i_{j,A}$, S_A, S_B, role} from party S_A: send {New Session, Sid, S_A, S_B, role} to S_B. In addition, if this is the first NewSession query, or if this is the second NewSession query and there is a record $(S_B, S_A, k^i_{j,B})$ then record $(S_A, S_B, k^i_{j,A})$ and mark it "fresh." Other conditions are ignored.

(2) Upon receiving a query {Test Session, Sid, S_A, $k^i_{j,A}{}^?$} from S: if there is a record of the form $(S_A, S_B, k^i_{j,A})$ which is "fresh," then do: if $k^i_{j,A}{}^? = k^i_{j,A}$, then mark the record "compromised" and reply to S {correct guess}. If $k^i_{j,A}{}^? \neq k^i_{j,A}$, then mark the record "interrupted" and reply to S with {wrong guess}.

The variable role in the message is included in order to let a party know if it is playing the initiator or responder role in the protocol. In (1) the ideal functionality describes the behavior of the initiator or responder role in the CK-AE protocol. And in (2) the ideal functionality describes a situation that the ideal adversary S tries to disguise the identity of party in the CK-AE protocol. Moreover, it explains that when one party is captured, the ideal functionality will know that the protocol is under attack and it will terminate the protocol.

3.2. Security of CK-AE Protocol. It is difficult to prove the security of a protocol within the UCSM without any security assumptions as premises. But many protocols can be proved secure within the UCSM based on a common reference string (CRS) model [16]. So, in this paper we prove the security of the CK-AE protocol based on a CRS model within the UCSM. The path to proof the security is as follows:

firstly construct a scene of simulation, then prove that the mutual information of the CK-AE protocol between the real model and ideal model is computationally indistinguishable for E.

Theorem 1. *Based on the CRS model, the CK-AE protocol is secure within the UCSM.*

Proof. Assume that A is a real-model adversary which participates in the interaction with the real parties which are running the CK-AE protocol. Then we construct an ideal-model attacker S (simulator S) for such ideal functionality $F_{\text{CK-AE}}$. Based on the CRS model, S gets some common information about the parties, such as Tag_i which the parties hold. Because A is equal to its duplicate copy $\widetilde{A}$ and $\widetilde{A}$ can call S to work, S is able to get all the information that A has submitted. Now Attacker S simulates the following two situations.

(1) S simulates that A disguises the legitimate node S_A. At first, S gets the fake information $\text{diff}^?$ and $k_{j,A}^{i\,?}$ that A submits. According to the CRS model, S guesses the chain key $k_{j,B}^{i\,?}$ which S_B holds then submits it to the ideal functionality $F_{\text{CK-AE}}$. If S guesses correctly, $\text{diff}^?$ and $k_{j,B}^{i\,?}$ (or $k_{j,A}^{i\,?}$) could be used repeatedly by it; then S submits $k_{j,B}^{i\,?}$ (or $k_{j,A}^{i\,?}$). S must guarantee that the output $k_{j,B}^{i\,*?}$ (or $k_{j,A}^{i\,*?}$) which is calculated by $k_{j,B}^{i\,?}$ (or $k_{j,A}^{i\,?}$) and $k_{j,B}^{i\,*}$ (or $k_{j,A}^{i\,*}$) is indistinguishable.

(2) S simulates that A disguises the legitimate node S_B. S disguises S_A to submit $\text{diff}^?$; then according the fake variable information X_B, Y_B and a random number r_B which are submitted by A, S guesses the chain key $k_{j,A}^{i\,*}$ of S_A and submits it to the ideal functionality $F_{\text{CK-AE}}$ based on the CRS model. If it guesses correctly, X_B and Y_B can be used again by S; then S submits $\text{diff}^?$ and $\text{Index}(j)_B^{*?}$. Similarly, the simulator S must guarantee that the output $\text{diff}^?$ (or $\text{Index}(j)_B^{*?}$) which is calculated by $X_B(Y_B)$ and diff (or $\text{Index}(j)_B^{*}$) are indistinguishable.

In this paper, we prove the indistinguishability of mutual information in ideal model and in real model by contradiction. Assumed that Environment (E) can distinguish the behavior of the simulator S; thus the following two cases exist. (1) When E gets the correct chain keys ($k_{j,A}^i$ and $k_{j,B}^i$) and is able to distinguish the chain keys submitted by the CRS model from the correct ones, in accordance with Theorem 1 in [12], the probability that this case happens is negligible. (2) E has got the difference information diff. If the number of the ring of chain keys m conforms to (2), and according to the Lemma 2 when some nodes are captured, the probability that the adversary obtains the difference information of the indexes of two chain keys in any other uncaptured nodes is negligible. In conclusion, the mutual information of CK-AE protocol between the parties in the ideal model is computationally indistinguishable for the E with the information in the real model. In other words, the CK-AE protocol is a UC-secure protocol. Proof ends. □

3.3. Analysis of Defense Capability against Sybil Attack. The resilience of the chain keys (the impact to secure session links between nodes brought by the exposed chain key in other captured nodes) is a significant security requirement in our approach. The stronger resilience of the chain keys is the more effective capability against Sybil attackers our approach has. The chain keys are predistributed before deployment; therefore, based on the Lemma 2 if m (the size of the ring of chain keys) conforms to (1), then the probability that attacker obtains other nodes' chain key information by capturing some nodes is negligible.

Lemma 2 (see [17]). *Assuming that the total number of nodes in WSN is N and the size of the ring of secret keys in one node is m. If the size of the pool K meets the inequality $K \geq N \log N$, and three variables K, N, and m meet* (2)*, then the wireless sensor network which is based on the secret keys predistribution is connected well. Furthermore, when some nodes are captured by adversaries, the impact to the secret keys in other uncaptured nodes is negligible:*

$$\frac{m^2}{K} = c\frac{\log N}{N}. \tag{2}$$

In our approach, we definite the resilience of chain keys of nodes as follows: the probability that each chain key is selected to each node from the pool K (equal to $(N-1) * M$) of chain keys is m/K. If a Sybil attacker captures w nodes, the probability that one of the chain keys in the uncaptured nodes is just equal to one of the chain keys in w nodes is $1-(1-m/K)^{\omega}$. Moreover, the probability that an attacker tries to forge the chain keys of legitimate nodes is $2^{-\lambda}$ [10], so the resilience of the chain keys is $2^{-\lambda} * (1-(1-m/K)^{\omega})$.

Figure 3 shows the relationship between the number of captured nodes and success attack rate in Newsome's approach [5], Feng's approach [7], HC approach [9], and our approach when the total number of nodes in WSN is 800 and the size of chain key pool is 10000.

It can be seen from Figure 3 that with the increasing number of the captured nodes, the difference of four approaches becomes more obvious. The success Sybil attack rate of first three different approaches all increase obviously, but the curve which represents our approach still stays stable. For example, when the number of captured nodes is 300, the success Sybil attack rate of the first three approaches is 67.91%, 2.12%, and 0.88%, respectively. But in our approach the probability is only 0.065%. The reason is that the hash function H^{Nyb} conceals the node identity information and the original keys by means of calculating some times depending on the difference of two indexes. Therefore, the defense capability against Sybil attack of our approach is stronger than that of three other approaches.

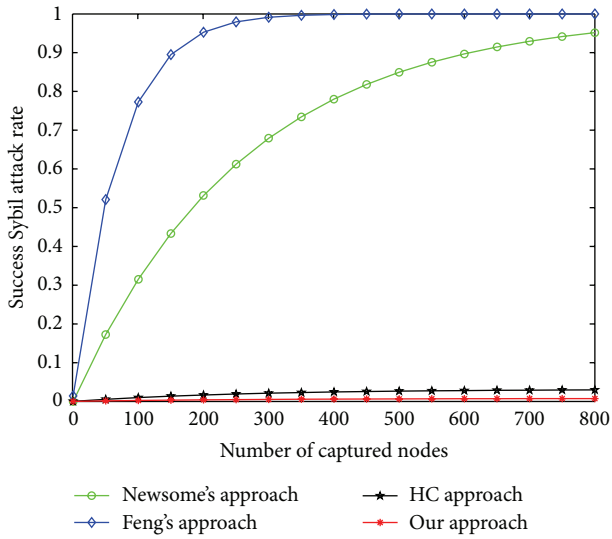

Figure 3: Defense capability against Sybil attack.

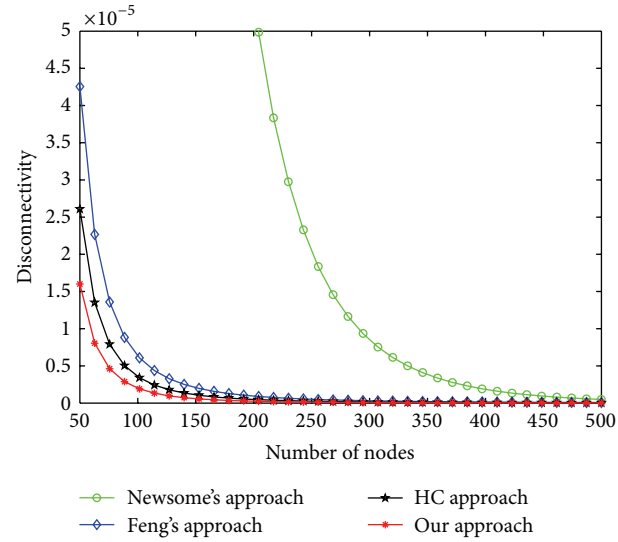

Figure 4: Disconnectivity of the network.

4. Performance Analysis

The performance of defense approaches against adversaries in WSNs can be evaluated on the basis of several criteria, such as connection, communication overhead, storage overhead, and computational overhead. In this section, we describe the performance evaluation of our approach and compare it with Newsome's approach [5], Feng's approach [7], and HC approach [9] in the above four aspects.

4.1. Connectivity. The secure connectivity (the probability that two neighbor nodes are able to establish a secure link) is considered as fraction of secured links among all possible links in the network. Thus, all defense approaches against Sybil attack based on a key predistribution scheme must ensure the connectivity of the network [18]. In our approach, if two neighbor nodes obtain the chain keys which are from the same hash chain (they share this hash chain), there is a secure link between them. Let us assume that the total number of network is N and each node is predistributed with a ring of m chain keys which are selected from a pool K at random. Besides, the pool consists of M hash chains and there are $N-1$ chain keys in each hash chain. If S_A randomly selects m chain keys from any t hash chains, S_B only can randomly select m chain keys from the other $M-t$ hash chains which S_A does not select from; therefore the disconnectivity of the network is described as follows:

$$P\left[\text{disconnectivity}\right] = 1 - P\left[\text{connectivity}\right] = \sum_{t=1}^{M-1} \frac{\binom{t\times(N-1)}{m}\binom{(M-t)\times(N-1)}{m}}{\binom{M\times(N-1)}{m}^2}. \quad (3)$$

Figure 4 shows the comparison of the connectivity of network between the above three approaches and our approach. In (3), the parameters K, N, and m should conform to Lemma 2 and $M = K/(N-1)$.

As can be seen from Figure 4, the proposed chain key predistribution scheme in our approach makes the network get a better connectivity than the others. For example, when the parameter N is 100 and accordingly K and m should be 200 and 14, respectively, the disconnectivity of the other three approaches is 6.32%, $6.32 * 10^{-4}$%, and $3.51 * 10^{-4}$%, respectively. However, it is only $2.13 * 10^{-4}$% in our approach. Besides, each node just need preload 14 chain keys that can make the network get a good connectivity; however, the other approaches need more than 20 chain keys preloaded in the memory of one node. Therefore, in the same case the proposed chain key predistribution scheme is more effective in reducing the storage requirement of the nodes. What causes a better connectivity is that hashing the unique identity information of every node sequentially to construct a pool of chain keys enhances the correlation of the common keys hold by different nodes and reduces the storage requirement of the ring of chain keys.

4.2. Communication Overhead. Communication represents the massage exchange between the nodes in WSNs, which consumes the limited energy resource of network. It is one of the important factors to evaluate the performance of defense approaches. We compare the communication overhead of our approach with that of the other three approaches. Based on the energy-consuming model in [19], we carry out simulations under the ONE software. Experimental parameters are set as follows: node deployment area = 1400 m * 1400 m, original number of nodes = 100, step = 20, communication range = 100 m, transmit speed of the node interface = 250 kBps, and the radio dissipates E_{elec} = 50 nJ/bit and ε_{amp} = 100 pJ/bit/m^2 for the transmit amplifier. The experimental results are shown in Figure 5.

In Feng's approach, two more messages need to be exchanged by the transmitter and receiver in the process of establishing the pairwise key; therefore, the complexity of

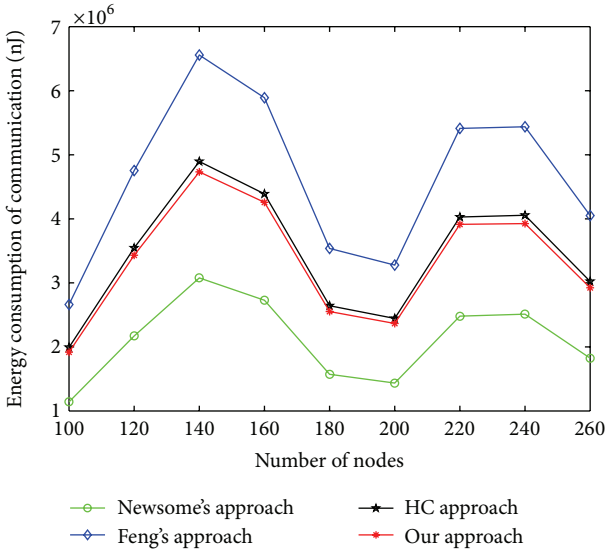

Figure 5: Energy consumption of communication.

communication is $O(m + 2)$. In Newsome's approach, the complexity of communication is $O(m)$. Because HC approach maintains the same pairwise key establishment phase with the Newsome's approach which is based on the key identifier exchange, it introduces the same complexity of communication as Newsome's approach. Only one message exchange needed to establish the pairwise key in our approach, so the performance of the CK-AE protocol is better than Feng's approach obviously. The complexity of communication in our approach is $O(m)$.

As can be seen from Figure 5, with the increasing number of nodes, the energy consumption of communication in our approach is always lower than that of HC approach, which is because the bits of the transmitted/received message are shorter than that of HC approach under the same complexity of communication. But compared with Newsome's approach, our approach needs to transmit/receive a little more bits of message (such as Tag_i) in order to conceal the original keys and enhance the resilience of network against the Sybil attack. In addition, Figure 5 shows that the energy consumption of communication increases or decreases irregularly, which is caused by the random deployment of sensor nodes in the simulation. When the number of nodes in WSN is 140, the number of neighbor nodes is the largest. Thus the energy consumption of communication is larger than other cases in Figure 5.

4.3. Storage Overhead. In all key predistribution scheme based defense approaches, each node needs to store a certain amount of keys. At first, we analyze the storage requirement of each node in Newsome's approach. $M_{\text{size}} = m \times (k_{\beta_i\,\text{size}} + \beta_{i\,\text{size}})$. $k_{\beta_i\,\text{size}}$ is the size of a given key, $\beta_{i\,\text{size}}$ represents the size of the key identifier, and m is the number of keys. Similarly, HC approach requires the same storage memory as the Newsome's approach. In Feng's approach, the required storage memory is $K \times Z(k_i)_{\text{size}} + m \times w_j(k_i)_{\text{size}}$. Due to K (the size of the pool) accumulated values are needed to be stored; therefore when K is a large number, a lot of storage memory will be occupied. In our approach the required memory for each node is $m \times (k^i_{j\,\text{size}} + \text{Index}(j)_{\text{size}} + \text{Tag}_{i\,\text{size}})$, where $k^i_{j\,\text{size}}$ is the size of chain keys, $\text{Index}(j)_{\text{size}}$ is the size of the corresponding index, and $\text{Tag}_{i\,\text{size}}$ is the size of the tags of the selected hash chain. Due to the additional storage of the tags, the occupied storage memory is more than that of the Newsome's and HC approaches but less than that of Feng's approach. It is noticed that when $\text{Tag}_{i\,\text{size}} = 128$ bits and m is less than 17 according to (1), the additional storage memory is 2176 bits (272 B) at most. However, the size of existing sensor nodes' memory is much bigger than 272 B. For example, Micaz has a memory of 512 KB which is equipped with CC2420. Therefore, our approach is suitable for the exiting nodes.

4.4. Computational Overhead. In the pairwise key establishment phase, the complexity of the computation for transmitter/receiver in the Newsome's, Feng's and HC approaches is $O(m)$, $O(m + 2H \times m)$ and $O(H \times m)$, respectively, where m is the number of messages that is transmitted or received and H is a pseudorandom function. In the CK-AE protocol of our approach, in order to determine a unique pairwise key between the neighbor nodes, the transmitter encrypts the message once then the receiver decrypts the message and applies the hash function H^{Nyb} a number of times (diff times) to compute the unique pairwise key. The average times of the above hash computation is $(2/N(N+1)) \sum_{i=0}^{N-1}((N-i)(N-i-1)/2)$, so the complexity of the computation of transmitter is $O(m + m \times E)$ and that of receiver is $O(m \times H + N)$. By contrast, our approach consumes a little more energy for computation than Newsome's and HC approaches, but less than that of the Feng's approach. But current studies have shown that the energy consumption of communication is much larger than that of computation [20]. Therefore, our approach is more suitable for the resource-limited WSN because of that the approach can get a better resilience of network against Sybil attack.

As shown above, the proposed approach not only enhances the security of the keys, which brings a better connectivity and stronger resilience of network but also reduces the significant communication overhead. Our approach solves the existing issues well, but it may incur unbalanced computational overhead.

5. Conclusion

The existing defense approaches are likely to be unsuitable to solve the problems for Sybil attack in WSN, such as high communication overhead of the trusted base station, unsecure keys, high communication overhead of the process of node identity authentication, and nonunique pairwise key. Focusing on these problems, we propose an approach based on chain key predistribution to defend against the Sybil attack in this paper. First, several chain keys are generated by hashing the unique identity information of every node

sequentially in the trusted base station, and they are constructed to a pool of chain keys. Then through the CK-AE protocol, neighbor nodes authenticate with each other and establish a unique pairwise key to protect the WSN from fake pairwise keys by Sybil attackers. This approach is provably secure against Sybil attack. The theoretical analysis shows that compared with other approaches, our approach can enhance the resilience of the WSN and reduce the communication overhead.

Acknowledgments

The paper is sponsored by the Research and Innovation Projects for Graduates of Jiangsu Province (nos. CXZZ12_0483 and CXLX12_0481), the Science and Technology Support Program of Jiangsu Province (no. BE2012849), the National Natural Science Foundation of China (no. 61071093), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (yx002001).

References

[1] F. Xiao, J. K. Liu, and J. Guo, "Novel side information generation algorithm of multiview distributed video coding for multimedia sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 582403, 7 pages, 2012.

[2] P. Anitha, G. Pavithra, and P. Periasamy, "An improved security mechanism for high-throughput multicast routing in wireless mesh networks against Sybil attack," in *Proceedings of the IEEE International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME '12)*, pp. 125–130, Tamilnadu, India, 2012.

[3] M. A. Simplício, P. S. L. M. Barreto, C. B. Margi, and T. C. M. B. Carvalho, "A survey on key management mechanisms for distributed wireless sensor networks," *Computer Networks*, vol. 54, no. 15, pp. 2591–2612, 2010.

[4] 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.

[5] J. Newsome, E. Shi, D. Song, and A. Perrig, "The Sybil attack in sensor networks: analysis & defenses," in *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks (IPSN '04)*, pp. 259–268, April 2004.

[6] R. Di Pietro, L. V. Mancini, and A. Mei, "Random key-assignment for secure wireless sensor networks," in *Proceedings of the 1st ACM Workshop on Security of Ad Hoc and Sensor Networks*, pp. 62–71, ACM, October 2003.

[7] T. Feng and J. Ma, "New approach against Sybil attack in wireless sensor networks," *Journal on Communications*, vol. 29, no. 6, pp. 13–19, 2008.

[8] S. Qian, "A novel key pre-distribution for wireless sensor networks," *Physics Procedia*, vol. 25, pp. 2183–2189, 2012.

[9] W. Bechkit, Y. Challal, and A. Bouabdallah, "A new class of Hash-Chain based key pre-distribution schemes for WSN," *Computer Communications*, vol. 36, no. 3, pp. 243–255, 2012.

[10] K. Nyberg, "Fast accumulated hashing," in *Proceedings of the 3rd Fast Software Encryption Workshop*, pp. 83–87, Springer, Berlin, Germany, 1996.

[11] K. Yukdrl, J. Kaps, and B. Sunar, "Universal hash functions for emerging ultra-low-power networks," in *Proceeding of the Communications Networks and Distributed Systems Modeling and Simulation Conference*, pp. 89–95, San Diego, Calif, USA, 2004.

[12] 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.

[13] T. Wu, Y. Tseng, and T. Tsai, "A revocable ID-based authenticated group key exchange protocol with resistant to malicious participants," *Computer Networks*, vol. 56, no. 12, pp. 2994–3006, 2012.

[14] R. Canetti and M. Vald, "Universally composable security with local adversaries," in *Security and Cryptography for Networks*, pp. 281–301, Springer, Berlin, Germany, 2012.

[15] R. Dowsley, J. Müller-Quade, A. Otsuka, G. Hanaoka, H. Imai, and A. C. A. Nascimento, "Universally composable and statistically secure verifiable secret sharing scheme based on pre-distributed data," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 94, no. 2, pp. 725–734, 2011.

[16] Y. Tian, J. Ma, and C. Peng, "Universally composable mechanism for group communication," *Chinese Journal of Computers*, vol. 35, no. 4, pp. 645–655, 2012.

[17] D. Roberto, V. Manceng, and M. Alessandro, "How to design connected sensor networks that are provably secure," in *Proceedings of the 2nd IEEE International Conference on Security and Privacy for Emerging Areas in Communication Networks*, pp. 259–266, Baltimore, Md, USA, 2006.

[18] Y. Wang, Y. Liu, and H. Jin, "The study on key pre-distribution methods for wireless sensor networks," *Physics Procedia*, vol. 25, pp. 560–567, 2012.

[19] 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 Sciences (HICSS-33)*, pp. 1–10, January 2000.

[20] Z. G. Wan, Y. K. Tan, and C. Yuen, "Review on energy harvesting and energy management for sustainable wireless sensor networks," in *Proceedings of the 13th IEEE International Conference on Communication Technology (ICCT '11)*, pp. 362–367, Jinan, China, September 2011.