Hindawi Publishing Corporation
Journal of Sensors
Volume 2016, Article ID 2098680, 11 pages
http://dx.doi.org/10.1155/2016/2098680

Research Article

RESH: A Secure Authentication Algorithm Based on Regeneration Encoding Self-Healing Technology in WSN

Wei Liang,[1] Zhiqiang Ruan,[2,3] Yuntao Wang,[4] and Xiaoyan Chen[1]

[1]*Department of Software Engineering, Xiamen University of Technology, Xiamen, Fujian 361024, China*
[2]*Department of Computer Science, Minjiang University, Fuzhou 350108, China*
[3]*Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Fuzhou 350116, China*
[4]*Institute of Information Engineering Chinese Academy of Sciences, Beijing 100093, China*

Correspondence should be addressed to Zhiqiang Ruan; rzq_911@163.com

Received 18 March 2016; Accepted 11 May 2016

Academic Editor: Fei Yu

In the real application environment of wireless sensor networks (WSNs), the uncertain factor of data storage makes the authentication information be easily forged and destroyed by illegal attackers. As a result, it is hard for secure managers to conduct forensics on transmitted information in WSN. This work considers the regeneration encoding self-healing and secret sharing techniques and proposes an effective scheme to authenticate data in WSN. The data is encoded by regeneration codes and then distributed to other redundant nodes in the form of fragments. When the network is attacked, the scheme has the ability against tampering attack or collusion attack. Furthermore, the damaged fragments can be restored as well. Parts of fragments, encoded by regeneration code, are required for secure authentication of the original distributed data. Experimental results show that the proposed scheme reduces hardware communication overhead by five percent in comparison. Additionally, the performance of local recovery achieves ninety percent.

1. Introduction

In recent years, wireless sensor network (WSN) is widely used to human life in various areas. The protection for individual privacy becomes increasingly prominent. In the area of medical care, various sensors are attached to human body in order to collect information of patients. The identity and signs data of patients are regarded as privacy and need protection [1]. WSN as a new way for information collection and processing is an interdisciplinary field of sensor technology, network communication, biological medicine, computer technology, and so forth. Nowadays, WSN becomes a hotspot in academia and industry [2]. Due to its features of small size, high flexibility, and low power, WSN is rapidly used in pervasive computing and system on chip, as shown in Figure 1. In WSNs it is used to cluster member nodes that take part in long distance data transmission to a base station (BS). However, the secure transmission and distribution of sensitive data in WSN require deep investigation in confidentiality and integrity of data transmission.

In previous transmission technologies, the fault-tolerant ability and resistance against node capturing are much lower. In communication, if the transmitted data is attacked, the security will be hardly ensured. Existing network recovery aims at single node: that is, the data in only one fault node can be restored each time. Multiple fault nodes are common in real application. Obviously, healing of single node will cause high communication bandwidth. Because encoded information of nodes is correlated and the correlation of fault nodes is not used in recovery of single node. Recently, many researchers have conducted work on healing technology for multiple fault nodes. The problems including key management, message authentication, secure time synchronization, and intrusion detection are considered in their research. Consequently, secure communication of data in WSN has been widely concerned [3].

In secure transmission of WSN, Benenson et al. proposed a secure authentication scheme in WSN based on asymmetric encryption [4]. Inner encryption of wireless network is

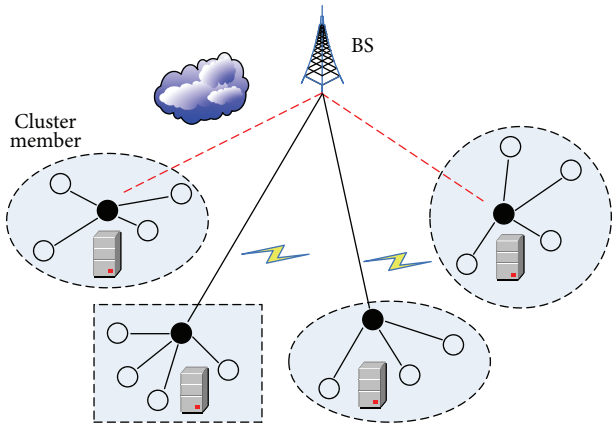


Figure 1: Clustering-based routing topology.

utilized for secure protection. After that, the scheme uses certificate authority for access control of the client. n neighbor nodes are selected as verifier. In this case, it is possible to verify the users by using (n, t) secret sharing method. Wang et al. [5] deployed a private wireless sensor network to monitor the whole vehicle network. The vehicle-mounted communication mode and the position of communication event are available. Besides, they have also conducted plenty of meaningful work in secure wireless vehicle network. Goyal et al. [6] proposed an access control strategy by allowing secret key to express any monotonous control tree. A user applies to a credibly authorized party for a secret key. The authorized party decides which characteristic combination in cryptograph can be decrypted by user. This strategy has added the expressive ability of KP-ABE, but the secret key of a user should be assigned in advance. Sahai and Waters [7] firstly presented a characteristic based encryption method and used it for access control. The encrypting party connects data with a series of characteristics. The secret key, assigned to user by the credible third party, is related to access structure of the characteristic set. The secret key reflects the privilege of user. The message is encrypted by using the characteristic. The key which satisfies the characteristic can only be used in decryption. However, this scheme cannot be popularized due to its lower expression of semantic. Bethencourt et al. [8] proposed another characteristic based encryption method. In this method, secret sharing is used in encryption stage to realize strict access control. The secret key is connected with related characteristic set. There is an access structure in the cryptograph. If the characteristic of secret key satisfies the access structure, it can be used in decryption. Otherwise, the decryption is rejected. The drawback of scheme in [9] is the requirement of polynomial interpolation to reconfigure the key. So, many complex operations of matching and exponentiation will be performed in decryption.

The authors in [9] have realized multiauthority attribute based encryption, which greatly reduced the computation overhead at stages of encryption and key generation. The security of encryption depends on hash function. Actually, no real random numbers are generated. In this case, the security of the proposed scheme is lower than that of SW scheme. Cheung and Newport utilized random elements instead of secret sharing to realize strict access control [10]. In this scheme, the sizes of cryptograph and key increase linearly with the growth of the number of characteristics. So, this scheme has lower efficiency. Carbunar et al. [11] investigated privacy content in WSN by query and proposed a SPYC protocol. This protocol considers that previous query mechanisms in WSN are lack of protection for user privacy, which may cause privacy leaking in transmission. Sheng and Li [12] presented a distributed data storage and query scheme to protect data and query range from being known by base station. But it cannot cope with collusion attack of sensor nodes and storage nodes. Subramanian et al. [13] introduced anonymous medium nodes to hide the incredible data origin. It can protect privacy of data type and query when a few normal nodes, storage nodes, and anonymous nodes are captured at the same time. However, selection of medium nodes is random and unpredictable. If the medium node is far away from the original node and destination node, it will cause unnecessary communication overhead. Additionally, data type is limit and there is one-to-one mapping between data type and conversion type. Attackers could find the mapping relationship by capturing a number of nodes. Finally, invalidation of medium node will make the path of data transmission lose efficacy.

Recently, researchers focus on secure encoding scheme with self-recovery. In these schemes, sensor nodes can receive important privacy data even when the data is attacked. Pawar et al. [14] proposed a secure scheme by restoring nodes dynamically. In this work, the authors list a few security threats of distributed storage system based on network encoding. On this basis, an eavesdropper model against illegal attacks is proposed. The scheme has good ability to resist collusion attacks. The authors in [15] proposed a fault-tolerant encoding scheme, as shown in Figure 2. The scheme integrates (n, k)-RS encoding with simple XOR operation. It mainly aims at high efficient restoration of single node. Actually, the scheme has improved immunity of data transmission from interference by decreasing the data transmission rate. n $(n > k)$ code words are generated by encoding k original data and distributed to n path for transmission. If there are multiple invalid paths, the destination node can restore k original data with the received m $(m > k)$ code words.

The error correction coding has strong fault-tolerant ability and low data redundancy, which is suitable for secure data transmission. Kim et al. used linear block code to construct secure wireless data transmission [16]. Dulman et al. [17] developed an error correction coding (ECC) based data transmission scheme by making a balance between data reliability and communication overhead. Djukic and Valaee [18] proposed a secure protocol (DCDD) at transport layer in data collection oriented WSN. ECC is utilized in oriented diffused routing protocol, which improves reliability by ten percent and greatly reduces the delay. In [19], RDP coding mixed with redundancy is utilized to accelerate data restoration. Furthermore, diagonal redundancy based cross-recovery is used. A half is restored by using redundancy of counterdiagonal and the other is regarded as shared data. This scheme reduces overhead of restoring bandwidth in

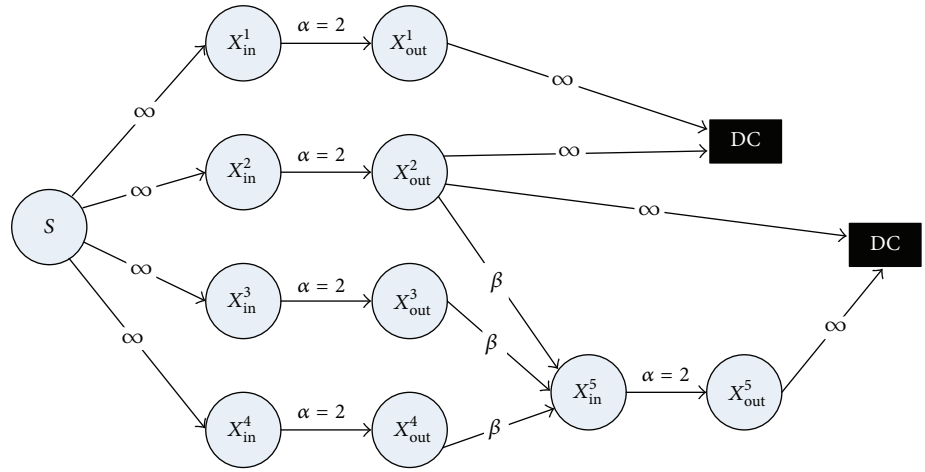


Figure 2: Multipath transmission based on error correction coding.

wireless nodes. The work of [20] proposed a method of combining scrambling technology with ECC to realize both confidentiality and reliability in wireless communication. The scheme overcomes burst error and has good security. But the communication overhead is large.

On the basis of the above studies, there are two issues on secure data transmission and data healing in WSN. On one hand, data storage in WSN is under security threat and can easily be attacked by dynamical tampering. Attackers could modify part of data content after capturing nodes. On the other hand, the usage rate of sensor nodes is limited. It may cause large performance overhead in transmission.

In this work, a secure fault-tolerant model in WSN is introduced. According to this model, the authors have designed an authentication scheme based on regeneration encoding self-healing technology. This scheme realizes secret dividing and content restoration of data on nodes in WSN. It can authenticate the integrality of data transmission without participation of original data. When wireless nodes suffered capture attack or tampering attack, the data can be restored with enough data fragments. After that, the secure authentication can be realized. The experimental results show that the proposed scheme has features of low complexity, high ability against capture attacks, and low overhead.

2. Preliminaries

In WSN, some nodes may lose efficacy if they are attacked. Thus, the invalid nodes will affect the reliability of data. In this work, we propose to solve data healing and security authentication by using regeneration code and secret sharing.

2.1. Regeneration Code. Regeneration code is a local encoding technology by combing (n, k)-RS code and simple XOR operation. It can restore arbitrary two missing data and has MDS feature to arbitrary n and k. By comparing with copying technology, regeneration code provides better efficiency of network storage and reliability of transmission. In traditional WSN, the operations of encoding and decoding are complex because computation is on the basis of finite field. So, large bandwidth overhead is required for node restoration.

We assume the restoration degree of invalid data in invalid node to be d. As known from RS encoding, if a redundant data in WSN is invalid, other k data blocks are required in order to restore the invalid one. Meanwhile, it causes communication overhead of k times than that of the invalid block.

The increase of invalid data of nodes in WSN will enlarge overhead of data transmission and cause lots of instable security factors. To address the issues of communication overhead and security threat, Dimakis et al. proposed a scheme by using regeneration code [21]. The transmitted file is set to be S. It is separated into two parts: $S = [S(1), S(2)]$, $S(i) \in F^{1\times k}$, $i \in \{1, 2\}$. F is finite field. $S(1)$ and $S(2)$ can, respectively, be encoded as a vector with the length of n by using (n, k)-RS code, $X = S(1) \cdot G$, and $Y = S(2) \cdot G$. Here, $G \in F^{k\times n}$ is a MDS matrix generated by (n, k)-RS code. With any k blocks encoded by regeneration code, a vector $C = X+Y$ is calculated through XOR operation. The value of C can be used to construct original data $S(1)$ and $S(2)$.

Firstly, the nodes in WSN should satisfy the (n, k) feature in encoding technology. In other words, encoded information is stored in n nodes and can tolerate n-k faults. Generally, X regeneration code is required for distributed network storage. Regeneration code is an array to tolerate two faults. The simple structure is shown in Table 1. When one node (two nodes) is (are) invalid, the restoration can be realized through simple XOR operations. The decoding and update can achieve the optimal. So, it is called the optimal regeneration code because this regeneration code could correct a few faults in data transmission. The regeneration code combines multi-faults-tolerant RS code and X encoding technology. It realizes the features of (n, k) and simple restoration. RS code offers a restoration for n-k faults. For single fault or double faults, the use of X encoding in RS code can achieve better performance in data healing.

X regeneration code is founded on polynomial, which has small local reparability [22]. Firstly, the generation rule based

Node 1	Node 2	Node 3	Node 4
x_1	x_2	x_3	x_4
y_2	y_3	y_4	y_1
z_3	z_4	z_1	z_2
$x_4 + y_4 + z_4$	$x_1 + y_1 + z_1$	$x_2 + y_2 + z_2$	$x_3 + y_3 + z_3$
$x_2 + y_4 + z_2$	$x_3 + y_1 + z_3$	$x_4 + y_2 + z_4$	$x_1 + y_3 + z_1$

Node 5	Node 6	Node 7	Node 8
x_5	x_6	x_7	x_8
y_6	y_7	y_8	y_5
z_7	z_8	z_5	z_6
$x_8 + y_8 + z_8$	$x_5 + y_5 + z_5$	$x_6 + y_6 + z_6$	$x_7 + y_7 + z_7$
$x_6 + y_8 + z_6$	$x_7 + y_5 + z_7$	$x_8 + y_6 + z_8$	$x_5 + y_7 + z_5$

Figure 3: The structure of X regeneration code when $n = 8$.

Table 1: The general structure of simple regeneration code.

Node 1	Node 2	$\cdots$	Node $n-2$	Node $n-1$	Node n
x_1	x_2	$\cdots$	x_{n-2}	x_{n-1}	x_n
y_2	y_3	$\cdots$	y_{n-1}	y_n	y_1
s_3	s_4	$\cdots$	s_n	s_1	s_2

on leading diagonal is utilized to divide redundant blocks at the first row. After that, the redundant blocks at the second row are divided by using counterdiagonal. We use 8 sensor nodes for illustration, as shown in Figure 3. For node 1, the redundant block is generated by performing xor operation on x_4, y_4, z_4. The same operation is performed on x_3, y_5, z_2 to get the second block. The data at the diagonal are collected as a redundant group: for example, $\{x_1, y_l, z_1, x_1 + y_1 + z_1\}$. If a fault occurs in one node, the redundant group will be utilized to reconfigure the damaged data. For example, when a fault occurs in node 1, it is possible to restore x_1 by downloading y_1 from node 5, z_1 from node 4, and $x_1 + y_1 + z_1$ from node 3. It is the same case to damaged data in other nodes. For nodes from 1 to 8, the recovery only needs connecting several surviving nodes. In distributed storage system of wireless sensor nodes, the superiority of using X regeneration code will be more obvious if there are a large number of nodes.

2.2. Thought of Secret Sharing. Shamir [23] proposed a secret sharing method based on Lagrange interpolation formula. It utilizes expressiveness of coplanar points to construct a reconfigurable polynomial function. The subkey and secret data are correlated into a class with the same attribute. The content of any item can be restored with other items. The scheme has strong security. But several conditions are satisfied in use of this scheme.

(1) A large enough prime number q and positive integer s are selected, $q > s$.

(2) $(m_i, m_j) = 1$ $(\forall i, j,\ i \neq j)$, $\forall i, (q, m_i) = 1$: that is, m_i cannot be the integral multiple of q.

(3) $N = \prod_{i=1}^{k} m_i > q\prod_{i=1}^{k} m_{n-i+1}$; N is the product of the top k numbers of m_i.

(4) The condition in (1) shows the secret data s less than q. But in (3), if N/q is greater than the product of selected $k-1$ numbers of m_i, the random number is A, $0 \leq A \leq [N/q] - 1$. The number of q and A can be made public.

(5) Let calculation function for data distribution be $y = s + Aq$. Due to $0 \leq A \leq [N/q] - 1$, we have $Aq \leq N - q$. Besides, $q > s$, $s - q < 0$, we derive $y = s + Aq \leq s + N - q < N$. So, there must be a $y < N$ and y is a secret. The key can be calculated by solving the equation set $y_i \equiv y \bmod m_i$, $i = 1, \ldots, n$.

(6) When (m_i, y_i) is ith subkey, the set of $\{(m_i, y_i)\}_{i=1}^{n}$ could construct a (k, n) secret sharing scheme in

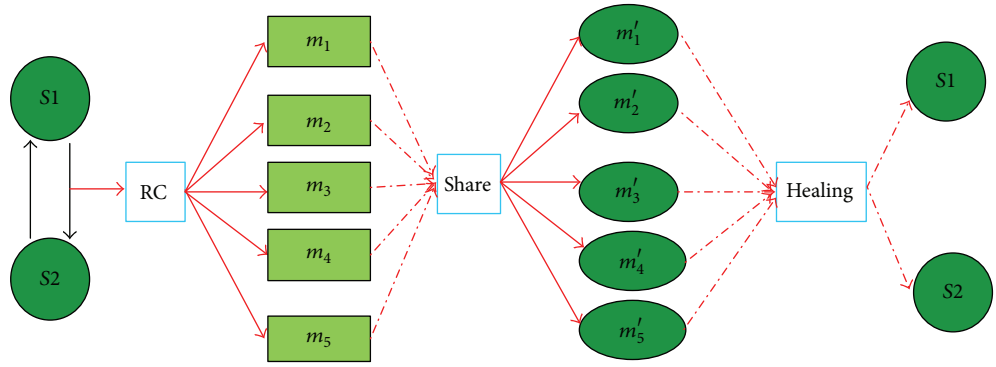


Figure 4: Secret sharing mechanism.

WSN. According to (5), $y \equiv y' \bmod N'$ is calculated. Here, we have $N' = \prod_{j=1}^{k} m_{i_j} \geq N$. y is the unique solution of congruence equation modulus N', $N' \geq N$, $y < N$. So, $y = y'$ is unique. The secret data s is calculated through $s = y - Aq$.

Figure 4 illustrates the thought of secret sharing. The transmitted data $S(1)$ and $S(2)$ are encoded by regeneration code with the production of m_1, m_2, m_3, m_4, and m_5. The information is further shared into n fragments. In the receiving end, the reliability can be authenticated through several fragments. It mainly depends on regeneration code for data restoration. If several data blocks are damaged, it will be possible to restore the original $S(1)$ and $S(2)$.

3. Regeneration Code Based Authentication Scheme

In this section, we introduce the healing mechanism after encoding the transmitted data in WSN. This scheme is resilient to illegal attacks. In real WSN, nodes may face security threats in terms of storage and computation. The following attacks are assumed to be suffered. (1) Illegal attackers intercept information from the communication flow in WSN. (2) Illegal attackers can randomly capture a few nodes. After that, they will get the key in these nodes and crack information in other nodes. (3) After capturing some nodes, the attackers could remove, modify, or forge the collected data in real WSN. The data is damaged. In this case, it is unable to trace the attacks.

To illustrate the security and reliability of the authentication scheme, we define some parameters in WSN in Notations. Here, the security involves confidentiality and completeness of data. The reliability is that some invalid or captured nodes will not affect normal running of the system. Furthermore, the ability of fault-tolerance represents that the damaged data could be restored through X regeneration encoding when random faults occur or some data blocks are modified.

3.1. Structure of Regeneration Code. A WSN is deployed in area of πR^2. The data is encoded by regeneration code. After that, the data in nodes is randomized. The regeneration encoding technology is fully utilized to get the encoded data. The procedure is described as follows.

(1) The information S in wireless sensor node is encoded into binary string and divided into groups. On the basis of regeneration encoding model, each group Si is transformed into decimal number and fatherly encoded by (5, 3) RS code. In Galois Field GF(24), the bit number of each information symbol is set as $m = 4$. (5, 3) RS code represents that five information symbols relate to two error correcting bits. In this case, three information symbols are a unit for RS encoding. Si is transformed into binary string T with length of 12 (padding zero on left when the bits are insufficient).

(2) For $S(x) = (S1, S2, \ldots, Si)$, each four bits are transformed into an element in Galois Field GF(24). After that, a sequence G in GF(24) is produced. Each row represents a sequence of elements for Si in Galois Field.

(3) The elements of each row in G are encoded and thus a matrix γ is produced. C_{ij} $(i = 1, 2, \ldots, n;\ j = 1, 2, 3, 4)$ denotes the data block before RS encoding and D_{ij} $(i = 1, 2, \ldots, n;\ j = 1, 2)$ is the error correction code of ith group of data after encoding. Each row in γ is a RS block BK_i $(i = 1, 2, \ldots, n)$. These blocks are randomly distributed and stored in nodes of wireless network.

(4) The data in WSN is denoted by W, which is transformed into encoded sequence B with the length of λ after decoding. Let the length of a pseudorandom sequence P be λ. We perform XOR operation on both the sequences. Finally, the distributed data fragments $W = \{W_i \mid 0 \leq i < \lambda\}$ based on regeneration code are produced.

3.2. Implementation of Secret Sharing. The encoded data fragments M_i based on regeneration code are shared and then distributed. The threshold secret sharing [24] and regeneration encoding technology [25] are utilized in data encryption on nodes in WSN. The concrete implementation is described as follows.

We assume N nodes to form an undirected graph $G(V, E)$ in WSN. The collections of nodes and edges are, respectively, denoted by $V = \{v_1, v_2, \ldots, v_N\}$ and $E = \{e_1, e_2, \ldots, e_N\}$. Each node is denoted by v_i $(1 \le i \le N)$, which has d_i neighbors. These nodes are organized as a collection NB_i. We produce a random session key k_s for v_i and compute hash value $H(D, K)$. After that, S_i and $H(D, K)$ are encoded by using k_s. Furthermore, the session key k_s is encrypted by public key Pk_u. Finally, two parts of data are produced, respectively, M_i and M_j. After that, v utilizes (m, n) secret sharing technology and regeneration code and divides M into n fragments, denoted by M_{ij} $(1 \le i \le n,\ 1 \le j \le n)$. Meanwhile, E is divided into m parts in order to construct

$$M(x) = E_0 + E_1 x + \cdots + E_{m-1} x^{m-1}. \tag{1}$$

When v acquires n fragments S_j $(1 \le j \le n)$, the equation $S_j = M(a^j)$ is satisfied. Here, $n \le 2^p - 1$. Finally, v selects n neighbor nodes from NB_i and realizes secret sharing to each neighbor node. The shared secret keys are denoted by M_i' and M_j'.

3.3. Distribution of Regeneration Code. The data in original node S is encoded into M, which is fatherly shared into multiple blocks. We randomly select $n - 1$ $(n \le N)$ nodes as initial distribution nodes. For the $(j+1)$th distribution node, the shared key $k_{i,j+1}$ could be calculated by asymmetrical secret key pair ID_s/K_s and ID_{j+1} [26]. The reserved data fragment of original node is $S_{i,n-1}$. The encrypted data fragment S_{ij} is sent to the $(j+1)$th node, $1 \le i \le \lceil \lceil T/m \rceil / k \rceil$, $0 \le j \le n-2$. The routing between original node and storage node cannot be determined in advance. When the $(j+1)$th node receives the response, the related key $k_{i,j+1}$ is calculated by ID_s/K_s and ID_{j+1}. Thus, $M'_{i,j+1}$ is produced. The above steps are repeated until all the data fragments are sent to the nodes.

3.4. Self-Healing Technology Based on Regeneration Code. When a part of data in wireless sensor node is attacked, the authentication data D is always damaged. In this section, we introduce a scheme to restore the damaged authentication data. Thus, the completeness of data in WSN can be ensured.

Let G be the generation matrix of regeneration code. If a receiving end receives k $(0 \le k \le n)$ blocks without errors, it is able to restore the original data. We assume the blocks from the jth node. If there is no error, k blocks in ith group could be restored by solving the following equation:

$$\left[u_{i_0}, u_{i_1}, \ldots, u_{i_{(k-1)}}\right] \widehat{G} = \left[c_{ij_0}, c_{ij_1}, \ldots, c_{ij_{k-1}}\right]. \tag{2}$$

Here,

$$\widehat{G} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a^{j_0} & a^{j_1} & \cdots & a^{j_{k-1}} \\ \left(a^{j_0}\right)^2 & \left(a^{j_1}\right)^2 & & \left(a^{j_{k-1}}\right)^2 \\ & & \vdots & \\ \left(a^{j_0}\right)^{k-1} & \left(a^{j_1}\right)^{k-1} & & \left(a^{j_{k-1}}\right)^{k-1} \end{bmatrix} \tag{3}$$

$\widehat{G}$ is established by element a in primitive field and index of c_{ij_ε}, $0 \le \varepsilon \le k-1$.

The authentication of data in wireless sensor nodes is performed in stages. We assume there are l errors at lth stage. If decoding fails or the decoded data cannot pass verification of CRC, too many errors may occur. The regeneration code cannot correct all errors. Two redundant symbols are required to correct an error. So, $k + 2l$ symbols are required at lth stage. The procedure of recovery is described as follows.

(1) Let $i = k$, randomly select k wireless sensor nodes, and detect the encoded data $c_i = (c_{j_0}, c_{j_1}, \ldots, c_{j_{(k-1)}})$. Set $r_i = c_i$.

(2) Calculate $u = r_i \widehat{G}^{-1}$ to produce the data blocks of ith group. If u passes verification of CRC, the CRC codes are removed to get original data u_0. Otherwise, go to step (3).

(3) Set $i = i + 2$. Select two symbols c_{i_1} and c_{i_2} from the nodes s_1 and s_2 that have not been accessed. They are added behind the received symbols to get a new code, $c_i \leftarrow c_{i-2} \cup \{c_{i_1}, c_{i_2}\}$. k symbols are produced by decoding the new code. It repeats until a failure occurs or $i \le n - 1$.

(4) $i \ge n - 1$ demonstrates too many errors and a failed decoding. In this case, it shows a message of failed decoding. Otherwise, it enters the next stage and performs step (2).

Recovery of data contents needs k subkeys at least. So, exposure of r $(r \le k - 1)$ subkeys will not leak the whole content. If the data of the nodes is lost or damaged, it can be successfully restored if there are k valid fragments. According to the sharing mechanism in Section 2.2, $\forall u \ge k$ number of $m_{i1}, \ldots, m_{iu}$ in $m_1, m_2, \ldots, m_n$, we have $H(s \mid m_{i1}, m_{i2}, \ldots, m_{iu}) = 0$. If $m_{i1}, \ldots, m_{iu}$ are known, the uncertainty of S is zero; that is, the content of S can be completely determined.

(5) On the basis of the above steps, an authorized user could directly restore the authentication data $D(C)$ from S_j. After that, S is restored with S_j by using Lagrange interpolation. In other words, if m fragments of authentication data $D(C)$ are collected from n nodes, we will effectively restore original data S in transmission.

3.5. Authentication Based on Regeneration Code. The participants of RC based authentication involve data distributer, data owner, and verifier. The distributer is responsible for encoding and sharing the data into n independent redundant blocks. These blocks are distributed to n data owners. The verifier takes charge of verifying completeness of data. Nodes in WSN can participate in authentication with both the identities of data owner and verifier.

The encoded data fragments are denoted by $D_1, D_2, \ldots, D_n$, $n \le 2^p - 1$. Each fragment has k symbols, denoted by $\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ik}$. The length of a symbol is p. All operations are performed in finite field $\mathrm{GF}(2^p)$. Completeness verification for node S has the following steps.

(1) $s_i \parallel h$ is encrypted by symmetric key K_{ij}, which is the shared key of d_i and d_j $(i \ne j)$. After that, the encrypted $s_i \parallel h$ is sent to d_j.

(2) After receiving all of $s_i \parallel h$ $(i \ne j)$, d_j decodes them with K_{ij} $(i \ne j, i \in [1, n])$. The equation $h_i = g^{s_i}$

is verified. If not satisfied, there are some errors in data fragments from d_i. Otherwise, the recovery continues to use Langrage interpolation $f(0) = D_i = \sum_{i=1}^{k} s_i \prod_{j=1, j\neq i}^{k} (j/(j-i)) \bmod p$ to restore d_i.

The regeneration code based authentication includes the following steps.

(1) Generation of Validation Information. Original node randomly selects u ($u \leq n$) elements $\beta_1, \beta_2, \ldots, \beta_u$ in finite field GF(2^p). n different odd-even check numbers $P_1, P_2, \ldots, P_u$ are produced, $P_j = \sum_{i=1}^{n} \beta_j^{i-1} S_i$, ($i \in [1, n]$, $j \in [1, u]$).

(2) Distribution. Original node v distributes u odd-even check numbers and n data fragments to n randomly selected neighbors in NB_i. For instance, original node selects P_j ($j \in [1, u]$) in u odd-even check numbers and S_i ($i \in [1, n]$) in n data fragments. After that, they are sent to a randomly selected neighboring node. Due to $u \leq n$, some nodes may only receive the data fragments other than odd-even check numbers. But some other nodes may receive both of them, which can be the verifiers. There are u verifiers among n nodes. The data is encrypted to avoid interception.

(3) Inquiry. Assume that the data owner of fragment $\{P_j, \beta_j\}$ wants to verify the data completeness. Firstly, an inquiry message $\{w_i, R_i, a, r\}$ is forecasted to all of data owners. r represents the number of required messages, $r \leq 2^p - 1$. a is a randomly selected element in GF(2^p).

(4) Response. After receiving the inquiry message, the nodes with $\{R_i, S_i\}$ ($i \in [1, n]$) will calculate r messages $\Phi_{(a,r)}(S_i) = (\Phi_{a^1}(S_i), \ldots, \Phi_{a^r}(S_i))$ and replies to w_i by broadcasting. The response message is encrypted to avoid interception.

(5) First Verification. With the returned $\Phi_{(a,r)}(S_i)$ ($i \in [1, n]$), node w_i verifies the equation $\Phi_{a^r}(P_j) = \Theta_j(\Phi_{a^t}(S_1), \ldots, \Phi_{a^t}(S_n)) = \sum_{i=1}^{n} \beta_j^{i-1}(S_i)$, $t \in [1, r]$. There are r equations. Any unsatisfied equation shows that modification or errors occur.

(6) Second Verification. The node w needs to store the detected data during a period of time. Each node stores data packages from different origins or data fragments at various stages. Any node could perform the first verification to all data fragments from the same node at any time. At different periods, w produces data fragments S_i and S_i'. The verifier sends an inquiry message to verify both of the fragments. When the inquiry message at ith round is received, the message $\Phi_a(S_i \parallel S_i')$ is returned to the verifier. The odd-even check numbers for the verifier are, respectively, P_i and P_i', $i \in [1, n]$. If all responses are received, the verifier calculates $\Phi_a(P_i \parallel P_i') = \Phi_a(P_i) + a^k \Phi_a(P_i')$ and performs verification. The proposed scheme could perform the second verification to the stored data during a period of time. No matter what the number of fragments is, the produced messages have the same size. It has effectively saved storage and communication overhead in procedure of authentication.

4. Performance Analysis

4.1. Overhead. In authentication of data in WSN, original node firstly calculates a hash value $h = g^{D_i} \bmod p$, n-order polynomial with degree of $k - 1$, and n hash values h_i. n data fragments are encrypted. Finally, the n hash values and data fragments are randomly sent to other nodes in network. The whole computation overhead in authentication is caused by decoding and hashing. In storage, each node stores n hash values. The communication cost is $n^*|s_i|$ because each authentication returns n data fragments. So, data distribution and verification at each round require n times of calculations. The overhead in communication is large.

Before generation of data fragments, original node needs to calculate a hash value H and perform two symmetrical encryptions, respectively, $H(D, K)$ and $\{k_s\}_{PK_u}$. The generation of data fragments requires calculating two polynomials. (m, n) RS code is utilized to encode $E = \{D_i, h(D_i, k_s)\}_{k_s}$ into n symbols. m is the number of data fragments and n is the number of selected neighbor nodes. Each data fragment is supposed to include c symbols. So, there are nc operations on polynomial. The original node utilizes (m, n) threshold secret sharing to get n symbols from $\{k_s\}_{PK_u}$. Finally, the check codes of all data fragments are produced. Let l be the length of data before encoding. The computation overhead of the original node includes hash of data with length of l, two symmetrical encryptions, $n(c+1)$ operations on m-order polynomial, and t odd-even checks. The computation overhead for the owner of each fragment is one decryption.

The verification of data fragments is in the finite field GF(2^p), $p = 8$. After generation of data fragments, the original node removes original data and sends n fragments of $\{\mathrm{ID}_v \parallel R_i \parallel S_i \parallel \beta_i \parallel P_i\}_{K_{v,w_i}}$ to neighbor node i. Here, S_i contains k symbols. So, the communication overhead in distribution is almost $n(2k + 3)q$. Obviously, the storage overhead for each data owner is $(2k + 3)q$. After all of the fragments are received, the owner of each fragment performs verification of completeness. This procedure needs to calculate odd-even check and message with length of k. Suppose that data owner w conducts completeness verification. Firstly, an inquiry message $\{w_i, R_i, a, r\}$ is broadcasted to all of the owners. The communication overhead is $4p$. Each data owner calculates a digest and returns it to the verifier after receiving the inquiry message. The length of this digest is same to that of a symbol. A symbol with length of k is calculated. The response is $\Omega_a(S_i)$. So, the communication overhead is p.

4.2. Security. Illegal users make security attacks by deploying sensor nodes or capturing nodes in WSN. The deployed nodes pretend to be the real nodes in WSN, steal confidential information, or launch false data injection. Besides, if multiple nodes are captured, the attackers could send plenty of false data. In this case, the network resource, such as energy, bandwidth, computation ability, and storage space, will be exhausted rapidly.

When the data is attacked, we need to restore and authenticate the damaged data through the proposed scheme. In this section, the security of the proposed scheme against attacks of forging, tampering, or collusion is analyzed.

(1) Ability against Forging Attacks. The sender always sends data with his pseudonym to the receiver. Other nodes cannot counterfeit the pseudonym of u. Otherwise, it is unable to pass verification of authenticator. The difficulty to counterfeit identity of u equals that of attacking SHA-1 hash function.

(2) Ability against Tampering. On one hand, the key in encryption is generated by using regeneration code. The privilege of decoding is under control. So, it is unable for anyone, including the sender, to tamper the information. On the other hand, attackers know nothing about the key. So, they cannot counterfeit or tamper the contents. The key is generated in authentication for multiple nodes. A single node authentication cannot realize tampering.

(3) Ability against Collusion Attacks. WSN is self-organized. So, it is possible to realize collusion attack. The sensitive data is divided into k parts and authenticated in k nodes with high credibility. It reduces the dependency on the third party. To get the secret, attackers need to restore $k-1$ order polynomial $F(x)$. On the basis of Lagrange interpolation, a successful attack means enough interpolating points are required. In other words, k nodes conspired at least. But the conclusion of so many nodes is much difficult. On the other hand, it makes the proposed scheme have ability against noncooperation of n-k nodes in procedure of recovery. So, the proposed scheme has high robustness in WSN.

4.3. Anonymity. Usually, it is unable to connect each authentication to real identity of the node. The data sent to authenticator is pseudonym. If no anonymity is stolen, attackers know nothing about real identity of authentication nodes. Of course, the real information in the nodes cannot be traced. Each authentication request utilizes multicast. So, attackers cannot destroy communication anonymity of medium nodes. Furthermore, authentication does not expose real identity of nodes. So, the attacked nodes cannot get other information. It offers good protection to sensitive information of wireless nodes.

4.4. Traceability. The data distribution system of nodes could trace behavior of attacks by using communication record. In other words, the nodes need to verify the key through cooperation of arbitration nodes before authentication. In this round of authentication, the arbitration nodes have sent the identity information to each node. The true identity is related to false identity. The relevance is kept all the time in authentication. Even if the attackers find resources in nonneighboring nodes through anonymous attack, the credentials could find the trace of attacks. In this case, the third institution can track out attackers on the basis of the information from the attacked nodes.

5. Experimental Result

5.1. Stimulation. The experiment is realized by C++ language and developed at visual C++ platform. In the secure network environment of this paper, the secret sharing mechanism is utilized to establish a secure recovery model based on regeneration code. Here, the initial threshold value is set as 0.5. The number of nodes is set at 500 in Network Simulator-2 (NS-2). These nodes are deployed within an area of 500 m × 500 m. Each node has initial energy of 2 J. In experiment, the dead nodes will exit network immediately [26]. The nodes are set at the highest level of protection in simulation. Illegal attackers are unable to perform successful attack on the signed nodes. Only the common nodes in network may be attacked. In this section, we have considered several common attacks and compared the performance against these attacks.

5.2. Analysis of Experimental Results

5.2.1. Security. Assume that N wireless sensor nodes are deployed within the area of πR^2. R denotes the transmission range. The positions of all nodes are supposed to obey two-dimensional Poisson distribution. So, when the communication radius of node s is denoted by r, the probability to include $n-1$ nodes within the area of h hops could be nearly calculated as follows:

$$\Pr\left[\chi(h) = n-1\right] = \frac{e^{-\theta}\theta^h}{h!}. \tag{4}$$

Here, $\theta = Nr^2/R^2$. h is regarded as a probability less than the expected value; that is, $\Pr[\chi(h) \geq n-1]$. The number of storage nodes is random, but the average value could be denoted by $\overline{n} = Nh^2r^2/R^2$. The probability of data hiding is calculated as the following formula:

$$P = 1 - \left(\frac{N'}{N}\right)^k \binom{\overline{n}}{k}. \tag{5}$$

Possibly, some common nodes in wireless network are selected as interception nodes. Illegal attackers attempt to intercept data package through decoding. In this experiment, the number of data packages is set as 500 at each time. Five illegal interception experiments are conducted for evaluation. Figure 5 shows the results. We observe the number of successful interception attacks in various schemes. The proposed RESH scheme has good ability against illegal interception by comparing to schemes in [27, 28].

5.2.2. Overhead. Figure 6 shows comparison of three schemes in overhead on storage and communication. The proposed scheme selects a few of key nodes for storage, which saves overhead on calculation and communication. The CADS scheme in [27] is based on discrete logarithm with complex calculation. Random walk in [28] has the highest communication cost because it utilizes broadcasting within the whole network. In the proposed scheme, the data in communication is compressed and has the ability of recovery due to the use of regeneration code. The results have verified effectiveness and availability in recovery. By analyzing, CADS scheme has higher detection rate, but the overhead on communication and calculation is much higher. For the proposed RESH scheme, it achieves higher detection rate, lower communication cost, and lower storage overhead.

We conduct experiments to evaluate time overhead. The downloading nodes and names of packages are randomly

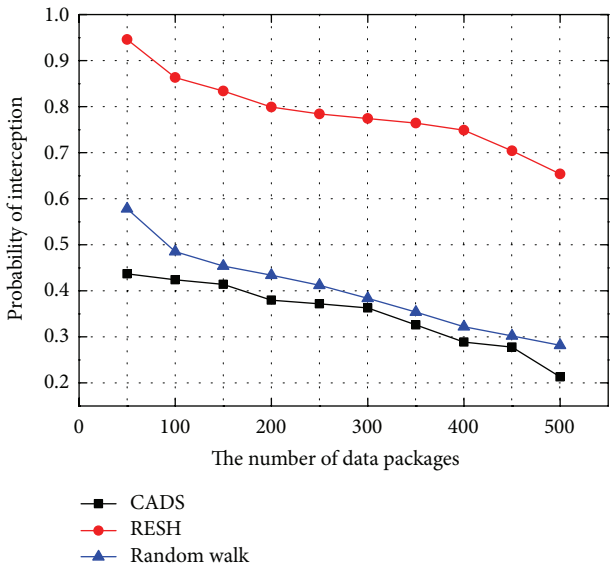
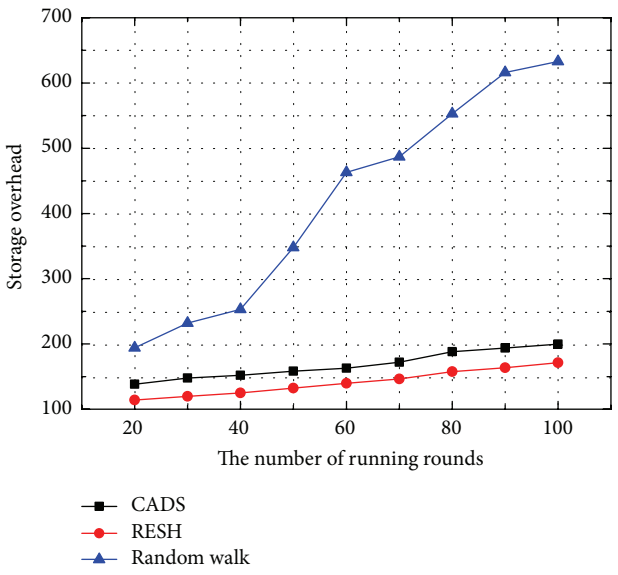
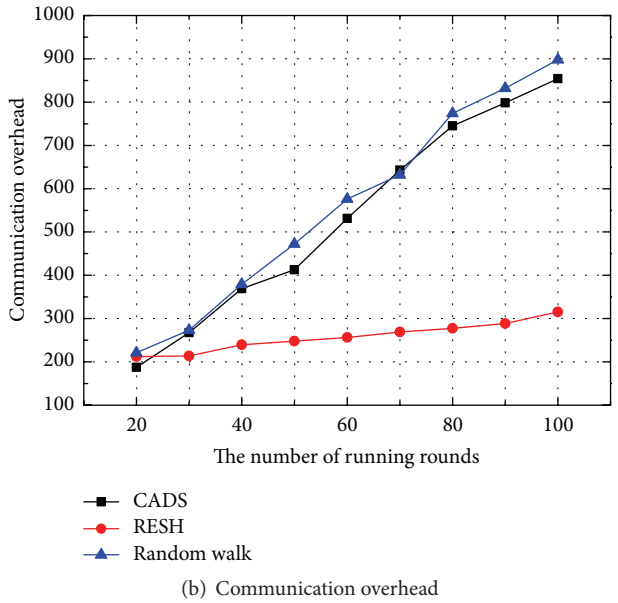


Figure 5: Comparison in probability of interception.



(a) Storage overhead



(b) Communication overhead

Figure 6: Overhead comparison for three schemes.

determined by system. The experiment considers the cases with different numbers of nodes. In Figure 7, we compare time cost of downloading data packages from nodes. The proposed scheme considers data communication based on regeneration code. It costs slightly more time by comparing to other schemes. But practically, it is valuable to exchange for privacy protection with less time cost.

5.2.3. Recovery Ability. The performance of recovery in wireless communication is evaluated by the minimum Hamming distance $d_{\min}$ of any two code words based on regeneration code. Odd-even check is utilized to restore the fault data. It enhances security in data transmission and storage. Due to the expandability of distributed cloud storage system, linear locally repairable codes (LRC) could restore encoded data through extended code and shortened code [29]. This method reduces the locality of restoring nodes by local and global redundancy. RS code is based on polynomial calculation and has lower locality of restoration. In Figure 8, we have compared the recovery ability of the proposed scheme to that of schemes based on RS encoding [30] and LRC encoding. The proposed scheme has good ability to restore fault nodes. The security and reliability of data transmission in wireless network are encouraging. The scheme based on regeneration code realizes real-time local healing when faults occur in nodes. If two nodes are fault at the same time, LRC based scheme [29, 31] needs to be transformed into RS encoding in restoration. It greatly decreases recovery ability of encoding algorithm. In our scheme, healing encoded information only requires connecting several local nodes. Once two nodes occurring faults, the recovery ability can also achieve 90 percent.

6. Conclusion and Prospect

This work considers secure transmission of data in WSN and proposes a model of completeness verification for WSN. Before data authentication, the data is divided into data fragments and stored in various nodes. On this basis, a secure authentication scheme based on recovery technology and regeneration code is designed in WSN. The scheme can restore damaged data. Besides, it saves communication overhead and processing time. The main contributions are as follows. (1) Regeneration encoding and healing based on threshold scheme are combined to achieve good performance in self-healing. (2) The proposed scheme has good performance in local recovery. In future, we continue to study secure transmission of privacy data. The concentration is to find and protect contents that users are interested in. Fast and secure forensics of privacy data in WSN will be also investigated. Besides, we will focus on effective distribution

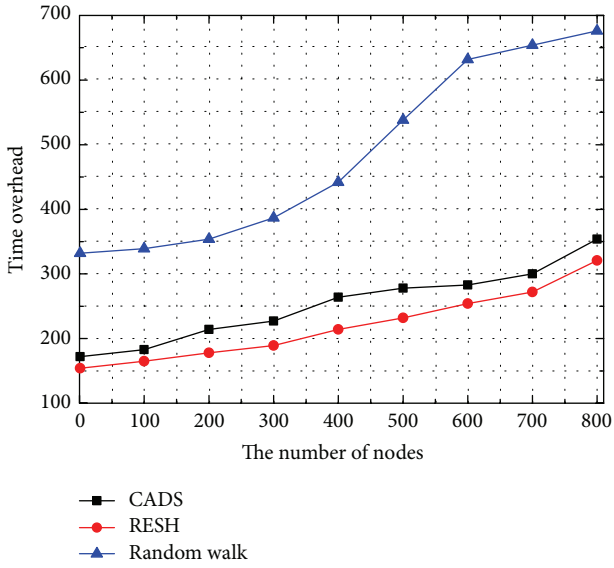


Figure 7: Comparison of time overhead.

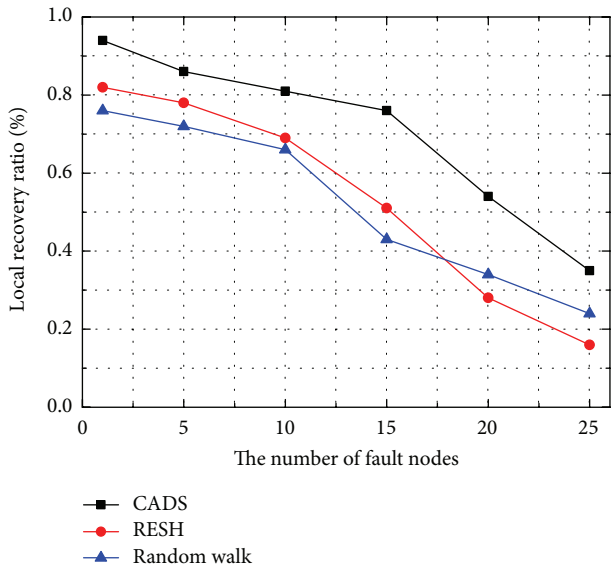


Figure 8: Comparison in local recovery ratio.

and secure communication of secret key in wireless network when the encoded content in network could be restored.

Notations

πR^2:	Deployment area of wireless network
k_s:	Session key
Bi:	Encoded information
S:	Content of original node
D:	Authentication information
$E = \{e_1, e_2, \ldots, e_y\}$:	Edge collection
P:	Secure hiding probability
C:	Power consumption in communication.

Competing Interests

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

Acknowledgments

This work is supported by the National Nature Science Foundation of China (Grant 61572188), the Natural Science Foundation of Fujian Province (Grant no. 2014J05079), the research project of Minjiang University (Grant no. MYZ14007), the research project of Fuzhou Science and Technology Office (Grant no. 2015-G-52), and the Research Project supported by Xiamen University of Technology (YKJ15019R).

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–105, 2002.

[2] Z. Chen, X. Li, B. Yang, and Q. Zhang, "A self-adaptive wireless sensor network coverage method for intrusion tolerance based on trust value," *Journal of Sensors*, vol. 2015, Article ID 430456, 10 pages, 2015.

[3] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on Database Systems*, vol. 30, no. 1, pp. 122–173, 2005.

[4] Z. Benenson, N. Gedicke, and O. Raivio, "Realizing robust user authentication in sensor networks," in *Proceedings of International Workshop on Realworld Wireless Sensor Networks (REALWSN '05)*, pp. 54–58, Stockholm, Sweden, 2005.

[5] S. Wang, Q. Sun, H. Zou, and F. Yang, "Detecting SYN flooding attacks based on traffic prediction," *Security and Communication Networks*, vol. 5, no. 10, pp. 1131–1140, 2012.

[6] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS '06)*, pp. 89–98, Alexandria, Va, USA, November 2006.

[7] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Advances in Cryptology—EUROCRYPT 2005: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22–26, 2005. Proceedings*, vol. 3494 of *Lecture Notes in Computer Science*, pp. 457–473, Springer, Berlin, Germany, 2005.

[8] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *Proceedings of the IEEE Symposium on Security and Privacy (SP '07)*, pp. 321–334, Berkeley, Calif, USA, May 2007.

[9] M. Chase, "Multi-authority attribute based encryption," in *Proceedings of the 4th Theory of Cryptography Conference (TCC '07)*, pp. 515–534, Amsterdam, The Netherlands, 2007.

[10] L. Cheung and C. Newport, "Provably secure ciphertext policy ABE," in *Proceedings of 14th International Conference on Computer and Communications Security (CCS '07)*, pp. 456–465, Alexandria, Va, USA, November 2007.

[11] B. Carbunar, Y. Yu, L. Shi, M. Pearce, and V. Vasudevan, "Query privacy in wireless sensor networks," in *Proceedings of the 4th Annual IEEE Communications Society Conference on Sensor,*

Mesh and Ad Hoc Communications and Networks, pp. 203–212, IEEE, San Diego, Calif, USA, June 2007.

[12] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 457–465, Phoenix, Ariz, USA, April 2008.

[13] N. Subramanian, K. Yang, W. Zhang, and D. Qiao, "ElliPS: a privacy preserving scheme for sensor data storage and query," in *Proceedings of the IEEE INFOCOM*, pp. 936–944, IEEE, Rio de Janeiro, Brazil, April 2009.

[14] S. Pawar, S. El Rouayheb, and K. Ramchandran, "On secure distributed data storage under repair dynamics," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '10)*, pp. 2543–2547, IEEE, Austin, Tex, USA, June 2010.

[15] J. Luo, M. Shrestha, L. Xu, and J. S. Plank, "Efficient encoding schedules for XOR-based erasure codes," *IEEE Transactions on Computers*, vol. 63, no. 9, pp. 2259–2272, 2014.

[16] S. Kim, R. Fonseca, and D. Culler, "Reliable transfer on wireless sensor networks," in *Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON '04)*, pp. 449–459, Santa Clara, Calif, USA, October 2004.

[17] S. Dulman, T. Nieberg, J. Wu, and P. Havinga, "Trade-off between traffic overhead and reliability in multipath routing for wireless sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '03)*, pp. 1918–1922, New Orleans, La, USA, March 2003.

[18] P. Djukic and S. Valaee, "Reliable and energy efficient transport layer for sensor networks," in *Proceedings of 49th IEEE Global Telecommunication Conference (GLOBECOM '06)*, San Francisco, Calif, USA, December 2006.

[19] G. Wang, X. Liu, S. Lin, G. Xie, and J. Liu, "Generalizing RDP codes using the combinatorial method," in *Proceedings of the 7th IEEE International Symposium on Networking Computing and Applications (NCA '08)*, pp. 93–100, July 2008.

[20] M. Zhang, Z. Wang, and M. Guo, "A method of combining scrambling technology with error control coding to realize both confidentiality and reliability in wireless M2M communication," *KSII Transactions on Internet and Information Systems*, vol. 6, no. 1, pp. 162–177, 2012.

[21] A. G. Dimakis, P. B. Godfrey, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 2000–2008, IEEE, Anchorage, Alaska, USA, May 2007.

[22] V. R. Cadambe and A. Mazumdar, "Bounds on the size of locally recoverable codes," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 5787–5794, 2015.

[23] A. Shamir, "How to share a secret," *Communications of the Association for Computing Machinery*, vol. 22, no. 11, pp. 612–613, 1979.

[24] A. K. Das, P. Sharma, S. Chatterjee, and J. K. Sing, "A dynamic password-based user authentication scheme for hierarchical wireless sensor networks," *Journal of Network and Computer Applications*, vol. 35, no. 5, pp. 1646–1656, 2012.

[25] A. Mazumdar, V. Chandar, and G. W. Wornell, "Update-efficiency and local repairability limits for capacity approaching codes," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 976–998, 2014.

[26] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[27] Z. Ruan, X. Sun, W. Liang, D. Sun, and Z. Xia, "CADS: co-operative anti-fraud data storage scheme for unattended wireless sensor networks," *Information Technology Journal*, vol. 9, no. 7, pp. 1361–1368, 2010.

[28] Y. Lin, B. Liang, and B. Li, "Data persistence in large-scale sensor networks with decentralized fountain codes," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 1658–1666, IEEE, Anchorage, Alaska, USA, May 2007.

[29] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with local regeneration," in *Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT '13)*, pp. 1606–1610, IEEE, Istanbul, Turkey, July 2013.

[30] T. Ernvall, T. Westerback, R. Freij-Hollanti, and C. Hollanti, "Constructions and properties of linear locally repairable codes," *IEEE Transaction on Information Theory*, vol. 62, no. 3, pp. 1129–1143, 2016.

[31] Z. Ruan, H. Luo, and Z. Chen, "Improving reliability of erasure codes-based storage paradigm under correlated failures for wireless sensor networks," *International Journal of Communication Systems*, vol. 29, no. 5, pp. 992–1011, 2016.