VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2016 ISSN(e): 2309-3978;ISSN(p): 2411-6246
Volume 4, Number 1, January-December, 2016

pp. 07-20

A SURVEY OF REPROGRAMMING SECURITY IN WIRELESS SENSOR NETWORK

DEPENG CHEN[1], D HE[1*], F AHMAD[2]

[1]School of Computer Science and Software Engineering, East China Normal University, Shanghai, China
[2] University of Central Punjab, Lahore Pakistan
Email: dhe@sei.ecnu.edu.cn

Revised March 2016

ABSTRACT. *Wireless sensor network is widely used around us, which makes a great contribution to our daily life. In order to use such peculiar network well, security aspect should be taken into consideration, especially in some vital areas such as military field. Code image dissemination is one of the most important functions in wireless sensor network, so ensuring the security of this process is profound. Thus, in this paper, we first introduce the basic concepts of wireless sensor network and its security issue in code image dissemination. Threats and vulnerabilities of code image dissemination are pointed out to illustrate why attentions should be paid to this process. Then we describe the major code image dissemination protocols. These protocols are often used to disseminate code image in wireless sensor network, and they contain some common techniques. Thus, according to these features, we give the six most important techniques used in this area and illustrate how to use these methods to prevent the network from attacks. Moreover, we also give some open research issues in this domain, which could act as a guide for the latter scholars.*
Keywords*:* Code image dissemination; attacks; security; wireless sensor network.

1. Introduction. Wireless sensor network (WSN) is a self-organizing network system, composed of a large number of low cost, small sizes, low power consumption, with awareness, computing, wireless communication capabilities of sensor nodes. Sensor network can collaborate to real-time monitor, perception and gathering all kinds of environmental or monitoring object information and processing, finally to transfer information to users, so that we can combine the physical world with the virtual information world.

Once a WSN is deployed in a specific environment, it is often necessary to update sensor nodes' configuration or install new software. For example, software in sensor nodes often carries bugs, so we need to fix these bugs. And more often, when we want to sense some other features of a particular application, we should add new functionality. Yet, it is very hard to do this one by one due to the quantity of sensor nodes and its own characteristics. Thus, reprogramming protocols are needed to be used. As WSN is often applied in harsh areas, it is easy to be attacked. The disclosure or tamper of the information may lead to the interests of the state, enterprises and individuals nursing big losses. So, some useful measures should be taken to protect the security of this process.

Because sensor networks have limited resources, distributed, self-organizing, multiple hops and data-centered characteristics, which not only make sensor networks face a greater risk of security that the traditional network environment security scheme cannot be directly applied to sensor networks, but also bring a lot of obstacles to the design and implementation of a sensor network security solution. According to our investigation, we are the first ones to present the state of art review on security of code image dissemination in WSN and point out the most important techniques used in this process.

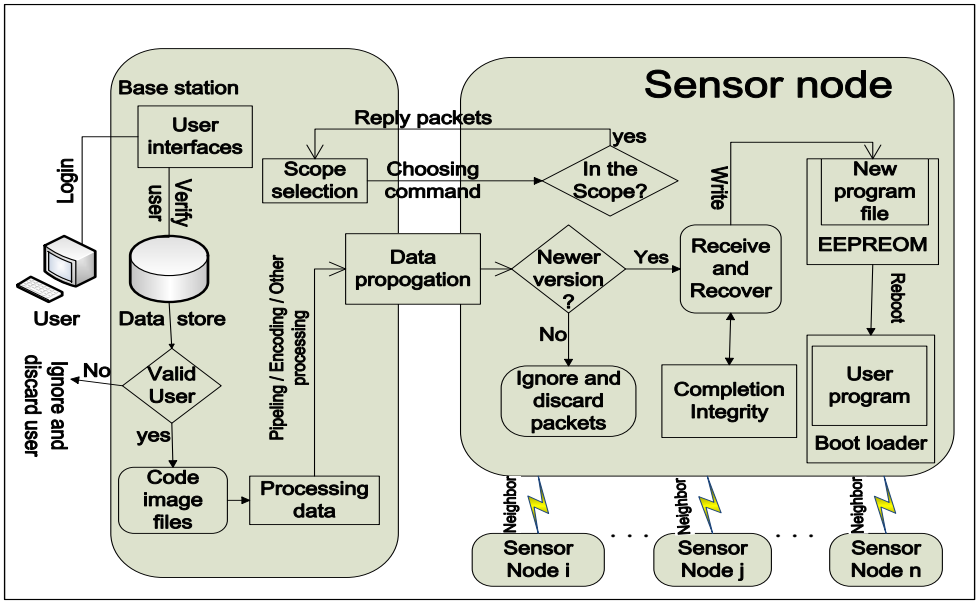

Fig-1 Code image dissemination in WSN

Here, we point out several typical challenges in designing secure reprogramming protocols and mechanisms.

(1) The Constraint of Sensor Node: Resource Constrain is one of most important features in WSN, which makes common data dissemination security scheme can hardly be used directly. As described in section I, sensor nodes has limited energy and memory, thus, most of public-key cryptography algorithm will become useless because of the poor.

(2) The Open and Mobile Characteristics of Sensor Network: Wireless network has many excellent features, such as easy to deployment and access. However, the open and mobile character is a double-edge sword. Compared with traditional network, adversary can easily access wireless network and launch insider or external attacks.

(3) Centralized Dissemination: Most data dissemination protocols assume that only base station can disseminate code image. If the base station is broken, the whole network will destroy. So, it is dangerous when there is only a single entity can disseminate code image. Thus, in order to relieve this constraint, some other architecture needs to be utilized, such as distributed schema, to realize secure code image distribution.

(4) Deployed in Harsh Environment: since a WSN is scalable, a sensor node may join and withdraw from network more frequently than other networks. So, it is very hard to describe and define the security perimeter. Moreover, sensor nodes are often deployed in harsh environment, and it is difficult to check sensor nodes one by one. What's more, sensor nodes in such environment are easy to be compromised, which makes it more difficult to secure the whole network.

Typically, the following requirements will be satisfied if one wants to protect the security of the process of code image dissemination. We summarize security goals and requirements of reprogramming in sensor networks from [4] [5] [15] [16] as follows:

(1) Integrity of code images: Before the source program images are received by sensor nodes, the sensor node should verify its integrity to test whether it has been tampered or not. If the code images have been tampered, the sensor node should discard it. Otherwise, receive the program image.

(2) Freshness: To make sure that the sensor nodes install the newest version software, so it is vital to control the sensor nodes to discard old code image and guarantee the code image consistency.

(3) Auto-recovery: In a WSN, a small part of sensor nodes break down is a very common thing. Thus, when some of the nodes cannot work well, to decrease the loss, we should make sure that this will not affect the whole network's function. The network could only discard these nodes and recovery by itself.

(4) High link quality: The link quality is particular important for code image dissemination. In a WSN, every node must receive the entire code image. Thus, to ensure the success of code image receiving, we should first

design effective scheme to enhance the link quality.
(5) DoS (Denial of Service) attacks resilient: Since WSN is easy to access, it is common to use DoS attacks to consume sensor nodes' energy. For example, the adversary may exploit compromised nodes to launch DoS attacks. So, in order to have a long service time, DoS attacks resilient should be taken into account first.
(6) Traceable: In a security schema, every sensor node should have a unique identity. When a sensor node transmits code images, other sensor node can verify its identity to determine whether to receive the data or not. On the other way, when a hostile attack has been detected, it is easy to trace the attack source.
(7) Small memory requirements: Sensor nodes often have limited memory and computing power. So the secure algorithms should be light-weighted.
(8) Energy efficiency: Sensor nodes often deployed in hash environment, it is difficult to supply electric power. Thus, a security scheme should not take too much energy.
(9) Distributed: As the network becomes larger and larger, it is inefficiency to disseminate code images only depending on the base station. Multiple authorized networks should be established to support distributed dissemination code image, which can transmit code images alone without involving base station.
(10) Access control: To prevent sensor nodes from being totally controlled by any of the network users, any access to the network should be controlled under some rules.
(11) Avoiding conflict: In a large sensor network, there may be a lot of users can reprogram a sensor node. Thus, some mechanisms should be introduced to prevent the dissemination conflict.
(12) Scalability: WSN contains a large number of sensor nodes. Each sensor node may join or withdraw the network, so it is important to ensure that the change of topology of WSN will not have bad effects on the whole network.

2. Reprogramming Protocols in WSN.
There are a variety of protocols and mechanisms in code image dissemination. According to their major characteristics, we classify these protocols and mechanisms into two large parts. One is non-security type and the other is security type. And then, we will choose some of the most representative protocols in both areas to illustrate how to disseminate code image and how to ensure the security of code image dissemination in WSN.

Protocols with Non-security. Trickle is a very famous flooding mechanism for propagating and maintaining code updates in WSN [6], and it is a basic mechanism of many other reprogramming protocols. Many code image distributions mechanisms are based on it [7]. Trickle uses a local self-regulation method called "polite gossip". Here "polite gossip" means that if a sensor node monitors an advertisement with the same code image as it has for several times(this number can be set according to the actual application) in an interval, it will keep quite. Whereas, when a sensor node monitors an older version advertisement, it will broadcasts code image to its neighborhood. Trickle gets both rapid propagation and low energy cost by tuning the interval of sending advertisement. When there is a new code image, it will decrease the interval.

However, to break the time to intervals can often lead to packets lost, so at this time Trickle has to transmit more messages. In order to solve the packets lost problem, many protocols use negotiation approach [8]. SPIN is the reprogramming protocol based on negotiation [9]. The negotiation protocol uses three-way handshakes between the sender nodes and receiver nodes and it includes the following three types of information: ADV, REQ and DATA. When a source node has received a new code image, it will send an ADV to other neighbor sensor nodes to inform them that there is new version software. And when a sensor node receives the ADV from the source node, it will send back REQ message to ask for the message needed. The source node sends the code image (i.e. DATA) to the sensor nodes after receiving the REQ from the neighbor sensor nodes.

MNP and Deluge are the two of the most popular protocols using negotiation way to deal with packets lost. Deluge [10] is based on Trickle and combined the negotiation approach. In Deluge, data is divided into pages and then a page divided into several packets. Using segmentation technique makes the large data objects easier to be transmitted as opposed with Trickle. Negotiation approach can reduce the redundancy message and solve packets lost problem. At the same time, Deluge exploits negotiation to select sender node by using the most recent advertisement. MNP [4] also uses negotiation approach to sender selection. Each source node holds a variable ReqCtr to count the number of distinct requests. Furthermore, when a node receives a download request which contains ReqCtr, the receiver can utilize this value to decide

which nodes should transmit the code image first. And the difference between MNP and Deluge is that MNP set a sleep state to reduce the active radio time [11]. In MNP, when a source node accomplishes the code image transmission, it will transfer to sleeping state for a while. This strategy benefits the energy consumption in WSN.

Table 1. The characters of reprogramming protocols

Protocol	Hierarchy	Multi-hop	Scope	Size of code	Pipeling	Networkcoding	Distributed
XNP [8]	No	No	All	Entire	No	No	No
Trickle [6]	No	Yes	All	Entire	No	No	No
Deluge [10]	No	Yes	All	Difference	Yes	No	No
MNP [4]	No	Yes	All	Entire	Yes	No	No
SPINS [46]	Yes	Yes	All	Entire	Yes	No	No
Seluge [12]	No	Yes	All	Difference	Yes	No	No
SDRP [15]	No	Yes	All	Entire	No	No	Yes
DiCode [16]	No	Yes	All	Entire	No	No	Yes
SenSeOP [17]	No	Yes	Selected	Entire	No	No	No
AdapCode [56]	No	Yes	All	Entire	No	Yes	No
Rateless Deluge [57]	No	Yes	All	Entire	Yes	Yes	No
LR-Seluge [1]	No	Yes	All	Entire	Yes	No	No

Protocols with Security. The above protocols are the basic protocols in network reprogramming. However, all of these reprogramming protocols focus on improving the efficiency and reliability of the process, but none of them take security problem into consideration. In order to prevent adversaries from attacking the code image distribution operation, many researchers proposed secure scheme which are based on the previous non-security protocols.

Seluge [12] is a model of secure protocol which extends to Deluge. It provides the integrity authentication for code image and can also resist to DoS attacks. Seluge exploits the efficiency propagation mechanism of Deluge and at the same time provides immediate authentication for each receipt packet. So, this rapid authentication and efficiency propagation of the code image can be a very efficient way to defect the DoS attacks by exploiting authentication delay. Seluge contains three main parts, and the details are illustrated as follows:

(1)Immediate authentication of code image packets

(2)Authentication of advertisement and SNACK packets

(3)Mitigating DoS attacks against signature packets

Seluge is a de facto standard of secure code image distribution and many other secure schemes are based on it. However, there is still some other secure problems need to be solved. In a lossy and hostile environment, packets loss occurs frequently. Thus, the ARQ (Automatic Repeat Request) protocol such as Deluge is not suit for this environment. Therefore, Zhang et al. proposed a secure schema to satisfy the need of both loss-resilient and attack-resilient called LR-Seluge [1]. LR-Seluge mainly exploits a limited number of predetermined redundant packets and chain relationships between encoded packets and original packets to decrease the gap between the loss-resilient and attack resilient. Unlike other loss-resilient reprogramming scheme, LR-Seluge uses a rateless way to encode the original packets. Different from Seluge [12] to combine the packets one-to-one correspondence between the two adjacent pages, LR-Seluge chains the original packets of one page and the encode packets of the next page.

Most of the code image protocols have to receive the packets in order, which may take much overload and energy when packets arriving out of order. In [13], the authors proposed an out-of-order-tolerant secure code image distribution. In their work, they combined a rateless erasure codes called Fountain codes [14] with immediate authentication to achieve out of order tolerant property as well as DoS protection. Rateless erasure codes are a very efficient way to transmit message in poor quality of WSN, since it does not need to response to such ARQ when a packet has lost. Because when exploits Fountain codes to encode and restore code image, a sensor node only need to receive sufficient encoded packets to recover source code

image. Moreover, before sending the encoded packets, the authors first produce a signature of all original packets. This will benefit the authentication of the integrity of the code image after receiving enough encoded packets.

To discuss the probability of distributed schema using in WSN to manage the propagation of code image, He et al. proposed a protocol called SDRP [15], which is the first proposed protocol for distributed code image dissemination. Almost all the existing code image distribution protocols are based on the centralized approach in which only the base station has the authority to disseminate the code image. SDRP realized the goal to disseminate code image by multiple authorized network users. To make the users in the network accountable and traceable, SDRP maps the identity and code image dissemination privilege into a public/private key pair. Thus, when a user wants to send new code image to the sensor nodes, the receiver nodes will verify the identity of the user and then decides whether to accept or drop the packets. And later, He et al. proposed another protocol named DiCode [16] to prevent DoS attack, which is an extension of SDRP. DiCode disseminate the code image page by page as Deluge, so it authenticates the packets rapidly upon it receives the right current page. Moreover, DiCode exploits the message specific puzzles, which can defect fake signature message efficiently. And just likes Seluge, DiCode also exploits cluster keys to resist DoS attacks making use of advertisement or SNACK packet. Thus, DiCode can provide a DoS-Resistant functionality..

3. **Current solutions.** Currently, there are lots of research papers to satisfy these security needs. From these mechenisms, we have identified six most common techniques used in code image dissemination of WSN. Many reprogramming protocols are the combination of these techniques.

Based on the Hash Chain. Hash chain is a secure and light weighted approach to protect the security of WSN. Thus, it is an appropriate way to use hash chain to verify the user's identity in WSN. Hash chain has many different types, but there are two basic types: one-way hash chain and multi-way hash chain. As the name implies, multi-way hash chain compose of the one-way hash chain and more complex. So, we will mainly introduce one-way hash chain

$$K_0 \xleftarrow{h(x)} K_1 \xleftarrow{h(x)} \cdots \xleftarrow{h(x)} K_{b-1} \xleftarrow{h(x)} K_b$$

Fig. 2 One way hash chain

One-way hash chain is based on a hash function H (.) with the features that it is very easy to compute the hash value, whereas pretty difficult to calculate its inverse function H−1 (.). In order to analyze the most essential facet of one-way hash chain, we give a hash chain with the length b just shown as below (as shown in Fig. 2). The first value K0 is a randomly seed, which is generated by sensor node. And then use function H (.) b times to calculate K1, K2...Kb. In a WSN every node has a unique node ID. When a sensor node B receive a message from other node A (or base station), the node B will verify A's legitimacy. This process has two steps. The firstly step is to search whether the entity A exists. If the receiver node B finds the entity, then go to the second step. In the second step, the sensor node B calculates whether H (Kb-1) is equal to Kb. If the two values equal to each other, then the node B will receive the message and update the hash chain about node A. And the next time when the node A sends message to B again, B will check whether H (Kb-2) is equal to Kb-1 and so on. So, every time the sensor node has different key value to verify the validity of the sensor nodes. Thus, it can protect the outer attacker even if the adversary gets the current hash value.
The above one-way hash chain is very simple. More often, many other messages will be added to this method. For example, in [18], the authors use multiple one-way hash chain which just extends the dimension of the one way hash chain. At the same time, some other literature [19] [20] add "salt" in one-way hash chain to make it security stronger. The "salt" here means to use other random or special message (such as user identity) to associate with the hash chain. Thus, the hash value is more complex and difficult to solve.

Secure code image distribution mechanism based on the hash chain can provide fast authentication of the integrity of the data, which in theory can effectively deal with competitors using DoS and DDoS (Distributed DoS) attacks. However, this approach must depend on good communication channel. Once the channel conditions get worse, data page can't be carried out in accordance with the established order receiving, it is unlikely to validate data integrity. Meanwhile, when the hash function is not so complex, it will be cracked by a method which uses a table called Rainbow Table to encrypt the key. Rainbow table first put

forward by Hellman in 1980 [21]. Now researchers find they can use rainbow table to crack password. For example, in [22] [23] rainbow table used to crack the password costs shorter time as compared to brute force method. Though it will take a lot of storage to store the whole rainbow table, it still deserves our attention. However, this problem can be solved by adding salt and organize the hash function, such as in [24]. Therefore, in practical application we can make use of the advantages of this technology in the sensor network identity authentication. Meanwhile, we also need to improve the method to make it in the case of receiving page do not match the order, also can go smoothly.

Based on Merkle Hash Tree. Merkle hash tree is also a very efficient technique to ensure the security of code image dissemination. Though we need to use more memory to hold the extra message generating by the constructing of the tree, the structure of Merkle hash tree makes it easy to authenticate the data packets. Since many scholars believe that it can resist to quanta computer algorithm, it now becomes an alternative to traditional digital signatures Such as RSA or MD families [25]. Due to these advantages, researchers apply it in the security area of WSN.

Shen et al. [26] firstly use Merkle signature scheme in WSN to realize access control technique. Wang et al. [27] use Merkle hash tree as a lightweight pre-authentication in two-factor user authentication schemes. According to the authors' experiment [27], it shows that using Merkle hash tree to authentication can adapt dynamically to DoS attacking scenarios. Meanwhile Li et al. employs the Merkle hash tree technique to secure smart gird communication, which can resilience to the replay attack, the message injection attack, the message analysis attack, and the message modification attack [28]. And Zhang et al. use one-way key chains and Merkle hash trees to defeat Sybil attacks [29], in which they uses a two-level Merkle hash tree to create certificates. In article [30], the authors pay much attention on the overheads of the protocol, so they use hierarchical hash tree to authenticate. By selecting proper parameters, such as the height of the tree, can optimize Hierarchical Hash Tree and the authentication overhead can be reduced to O (N). Therefore, it is very efficient in code image dissemination. To authenticate large amounts of data, such as a real-time data flow, Berbecaru et al. also combine Merkle hash tree with PKC (public key cryptosystem) to make it more effective and secure [31]. Meanwhile, in [32], Veronika Kondratieva and Seung-Woo Seo studied the problem of optimizing the authentication tree struc-ture for sensor network environments, which can help WSN save a lot of energy in authentication phase. For distributed transport protocol security, Dvir et al. use Merkle Trees and Hash Chains, this shows that their protocol can resist attacks against the reliability and energy efficiency [33]. And in [34], Przydatek et al. brings in Merkle hash tree to commit to a set of values. Their method just takes sub linear communication between the aggregator and the user.

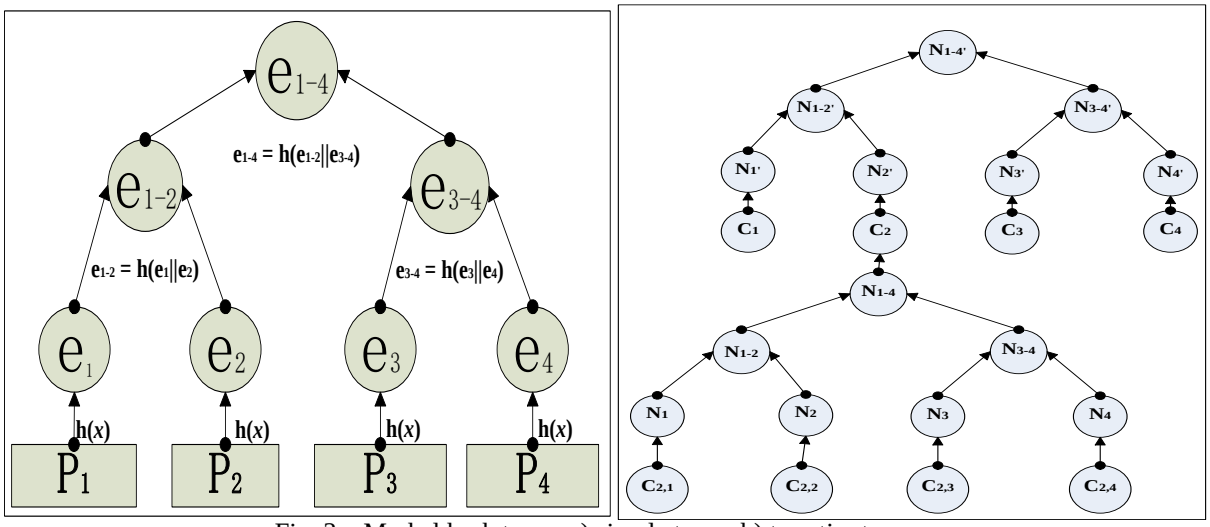

Fig. 3 Merkel hash tree a) simple tree b) two-tier tree

From the above analysis, we can see that Merkle hash tree is a very useful and efficient technique to apply in authentication of WSN. In different applications, there are different forms of Merkle hash tree. In order to analyze the basic principle of Merkle hash tree, we here introduce a very simple form (as shown in Fig. 3 part a) Merkle hash tree to verify the data integrity.

The main idea of Merkle hash tree is to build a tree structure based on a one-way hash function h (.) [28], the leaf nodes are the data items which can be verified through its authentication path information. In order to illustrate the work of Merkle hash tree, we will use a simple example. Here, we have four packets P1, P2, P3 and P4 to disseminate in the sensor network. Before disseminate these packets, we first build a Merkle hash tree based on the hash values of these 4 packets, i.e., hi = h (Pi), i=1, 2, 3, 4. Next we use hi as leaf node to construct hash tree. The values of internal nodes are obtained from the hash values of two child nodes. For example, the value of h (e1-2) is h (h (e1) ||h (e2)) and h1-4 = h (h (e1-2) ||h (e3-4)). This process will iterate until the root node is created. Each node can be verified with h (e1-4) and its authentication path information. Before disseminate the data items, the base station will first broadcast the signature h (h (e1-4)) to announce other sensor nodes that a new round dissemination starts. Then, when a sensor node received the packet Pi, it can immediately verify the authentication of this packet. For example, the packet P1 can be authenticated by a sensor node only through the data item P1 and its authentication path information (h (e2), h (e3-4)). Then, the sensor node can immediately check the validity of P_1 only by computing whether h (h (h (P_1) ||h (e_2)) ||h ($e_{3\text{-}4}$)) equals to h ($e_{1\text{-}4}$) [35]. Since the authentication path information is stored in sensor nodes and the concatenation operation is easy to calculate, the Merkle hash tree technique is light-weighted.

Another form of Merkle hash tree is shown in Fig. 3 part b. In a WSN, there may be hundreds or thousands of sensor nodes deployed in an extreme environment. Thus, a lot of users need to be authenticated. In order to organize these nodes in an efficient and energy less way and to increase scalability of Merkle hash tree scheme, we often use multi-layer Merkle hash tree. The following figure depicts a two-tier Merkle hash tree. From the diagram we can see that each leaf in the upper layer consists of a Merkle hash tree, and this feature will increase the scalability of Merkle hash tree based scheme. From the above introduction we can see that based on Merkle hash tree is able to take advantage of the advantages of small storage space to protect the data stored in harsh environment. Therefore, this technology applied in the field of code image distribution is quite effective. But we also noticed that, although this technique to a large extent reduces the quantities of stored data and at the same time ensure the security of the data dissemination, we find that the data in the actual use of the recovery phase need to consume a lot of memory, which is a great burden for sensor nodes. Wireless sensor nodes due to limited memory, often do not have good data computing capability. Therefore, we need to improve the method, such as exploit cache levels to improve hash tree [36] [37], to increase the amount of calculation of node per unit time.

Based on Fountain Codes. In WSN, the instability of communication channel be-tween sensor nodes and the base station and accessibility to a wireless channel characteristics, make the data in the process of transmission is not only faced with data intercepted and tampered threats, but also face the risk of transmission error. However, the traditional data distribution with forward error correction coding, in the event of error, can automatically correct, but for a short burst error, the error correction ability is limited, can't timely correction [38]. This is likely to make the attacker using this channel interference, hinder the node to receive data, so as to make the node receive wrong data message to achieve the goal of destruction the code image dissemination process.

Code image dissemination in WSN is often confronted with data error, so in order to make this process more secure, we need to provide a reliable code image transmission. Here, we can use Fountain codes to solve this problem. Fountain codes [39] first proposed in 1998 is a very efficient technique for transmitting data out of order. Foundation codes is a very convenient method in data dissemination, so many researchers introduce this technique in WSN data dissemination to ensure its security. To ensure reliable delivery in WSN during data dissemination, we need make the wireless channel to be tolerant to delivery error. A lot of works focus on using Fountain codes to solve this problem in WSN. In [40], the authors employed fountain codes to improve the fault tolerance and persistence of data in WSN. As the number of sensor nodes scales up, their experiment shows that the decoding performance is excellent and the decoding complexity is low when compared with other random liner codes. The sensor nodes in WSN often have constraint energy. Thus, a data dissemination scheme with energy-efficient is extremely essential in a sensor network. This work was done by [41]. SYNAPSE is a protocol for reprogramming WSN exploiting rateless Fountain Codes [42]. Unlike other de facto standard reprogramming protocol such as Deluge [10] using Automatic Repeat reQuest (ARQ) techniques to implement error recovery and SIMAGE [43] adjusting optimal payload size of packet according according to link quality to reduce the number of retransmission, SYNAPSE utilizes Fountain Codes to make it more efficient to recover data error. This work is very useful in large scale receivers. Later, they extended SYN-APSE and Proposed SYNPASE++ [44]. In SYNPASE++, they use a genetic optimization technique

with Fountain Codes encoding distribution to speed up the decoding phase and decrease its calculation time. Moreover, in order to make most use of the advantage of fountain codes, they combine novel channel access and pipeline technique. And it is very useful for releasing the hidden terminal problem and reduces the collision in channel. From the papers above, we can see that the Fountain Codes is very useful in error recovery and when combined with other technique, such as pipeline technique, it will help decrease the delay of data dissemination. And this is very important for data dissemination security. Because many adversaries may exploit propagation delay to launch attacks such as DoS attack.

The general fountain codes can be described as following. LT codes is the first fountain codes and it is pretty simple [14], so we will analyze its principle. Fig. 4 shows the simple encoding and decoding process of LT codes. S1, … , S5 are the resource data and P1 , ... , P7 are the encoded packets. The encoded packets are a little bit more than resource packets. The fountain codes technique contains two steps: encoding and decoding. In encoding phase, firstly divide the whole data into k parts; in the figure we get 5 parts. Then randomly choose the degree d (here, degree means the number of the combined source packets, for example the degree of P1 and P2 are 2 and 1. Because P1 is encoded by 2 source packets and P2 only encoded by a single packet) of the encoding symbol from a degree distribution [14]. At last, choose d (between 1 to k) source data packets to combine encoded packets by using XOR operation.

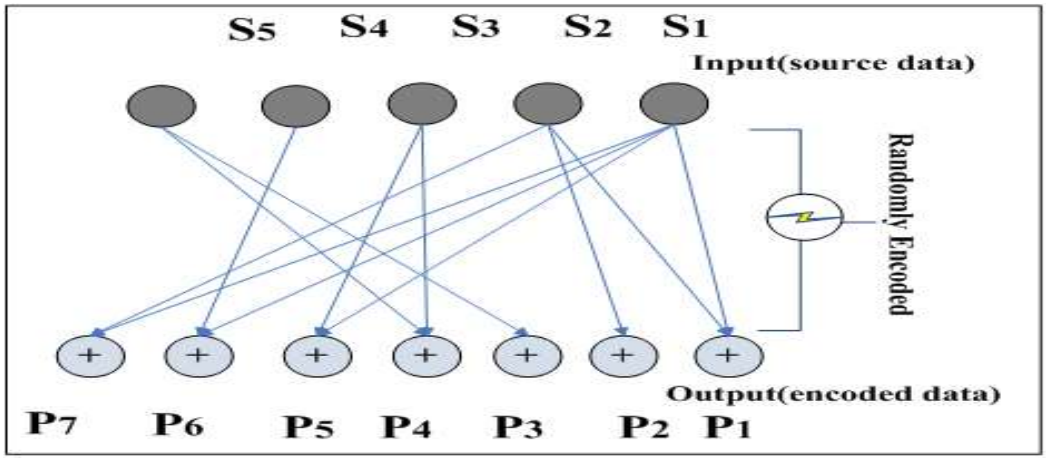

Fig. 4 An example of basic Fountain code

The decoding process can be depicted as following:
1) Receive a certain number of encoded groups and then choose the packet whose degree is one. Here, we choose P_2 and P_3. Thus, S_2 and S_5 can be decoded out.
2) Use the decoded packets to do XOR operation with the connective encoded data packets and remove the connection between them. Then, let the results represent the encoded packets.
3) Repeat steps 1) and 2), until the decoder stops.
If all the original data packets have been restored, decoding succeeds; otherwise, the decoding fails. Thus, it has to receive more coding group to continue.

However in the process of the actual data transmission, digital fountain codes with larger redundant coded packet, there are some decoding failure probability, and the existing need further improve decoding algorithm. At the same time, the degree distribution is extremely important. When the degree distribution is not as good as possible, there may be no encoded packet with 1 degree. And sometimes the decoding phase may become very slow because of the redundancy message. So, in [14], the authors introduce Ideal Soliton distribution and amend it to make this distribution more robust. Thus, in order to use Fountain codes in WSN data dissemination, we need to combine other method and design more efficient code image dissemination scheme.

Based on cryptography technology. Strong cryptography can be employed to ensure confidentiality, integrity, authentication and non-repudiation, which can be accomplished through a combined of symmetric key algorithms and cryptographic hash functions [45]. However, traditional secure communication cryptography methods may not work well in WSN as the limitation of sensor nodes.

Code image dissemination often requires high confidentiality and data integrity, so cryptography technology is an excellent choice. Since WSN has its own features, such as the limitation in memory, storage and energy, selecting the proper cryptographic is very important. In order to provide the security code image distribution and enhance the data confidentiality and integrity of the certification in WSN, many researchers proposed ECC and SHA algorithm. Although there are other methods applied in code image dissemination to encrypt and authenticate metadata, here we focus on the most famous methods ECC and SHA.

ECC and RSA are two of the most epidemic public-key cryptography. Many works in WSN exploit them to guarantee the security of data dissemination. Wander et al. compared the energy cost of authentication and key exchange between RSA and ECC on an 8-bit micro-controller platform [45]. Their experiments indicated that public-key cryptography is viable in WSN and ECC is superior to RSA .The use of ECC over RSA can lead to significant savings in public-key communication costs and also the amount of data transmitted and stored.

Based on Message Specific Puzzle. A large number of approaches [10], [12], [16], [46], [47] exploited digital signatures or μTESLA methods for broadcast authentication. However, digital signatures and μTESLA-based techniques are vulnerable by DoS attacks [47], since the receiver node cannot authenticate the broadcast packets immediately and attackers may use this weakness to launch DoS attacks.

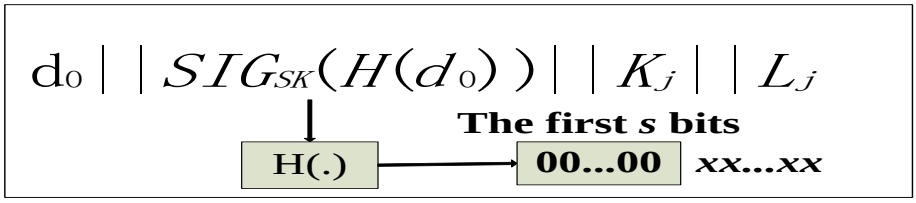

Fig. 5 Message specific puzzle

In order to act against the above denial of service attacks, researchers use a weak authentication method called message specific puzzle [47] to solve this problem. The specific resistance process is as follows: A message specific puzzle is derived from a one-way key chain. Here, a puzzle is generated by the puzzle key. Fig. 5 shows an example of message specific puzzle, where d0 denotes the code image packet, SIG_{SK} (H (d_0)) represents the signature created by the base station, K_j is j-th key chain and L_j is the puzzle solution. Here, a valid puzzle solution is a value that after using the hash function H (.) to the message specific puzzles d_0|| SIG_{SK} (H (d_0)) || K_j||L_j, the first s bits are all becoming 0. Here, the parameter s represents the strength of the puzzle. When sensor nodes meet a specific attack against signature packets, each node first use H (.) and j-th key chain K_j to test whether the puzzle key is valid and haven't been used before. If the answer is negative, the receiver will only drop the signature packet. Otherwise, verify the puzzle solution. Only if the solution is right, all of the nodes of the same area can verify the signature information. Otherwise, it will not verify the signature information so that it can withstand the DoS attacks against signature packets.

Generally, to solve a puzzle in the solution domain needs brute search. However, to test the method validation is very quickly. Moreover, the message specific puzzle method is set based on sensor node resources timeouts. To this end, the enemy must have enough resources in a certain period of time to calculate enough solutions of the puzzle. Therefore, specific message puzzle method can be applied to the sensor to the field of code image dissemination security.

Based on Secure Network coding. Network coding is a very important technique for data dissemination, which break the routine of communication style. In general communication styles, the relay nodes only transmit the message and do not process it. However, using network technique, the relay nodes in the network not only deliver the message, but also encode the message to compress the metadata. Thus, through network coding, the network transmission capacity will be improved, which will reach the ideal transmission condition defined by max-flow min-cut theorem. Fig. 6 part a shows the difference between traditional routing method and using network coding routing method

In WSN, network coding will become the vital technology of the next wireless network. Meanwhile, many researchers exploit the nice features of network coding to ensure the information security. *Cai et al.* first

incorporated network coding and data security [48]. In their work, they proposed a construction of secure linear network codes, in which they employed a certain graph-theoretic to prove that their model could protect the net-work from the wiretappers. Though the model is not that difficult, this work made a good start. Since then many corresponding work had already started. In [49], the authors evaluate the cost of secure network coding. Their experiments showed that the network coding is more efficient than traditional routing and their schema can make a low probability for the wiretapper to recover the captured data. Further, Cai et al. proposed another paper about the security condition for multi-source linear network coding [50]. Compared with their last work in [48], this paper put forward an algebraic method to give a necessary and sufficient condition about security of the linear network coding. And their algebra theories show that a linear network code does not rely on the source distribution, which will give us a guide to design secure linear network codes. In a tutorial paper [51], more details about the basic theory of the linear secure network coding were given. Furthermore, they analyzed some other secure properties and secure model for secure network coding. The main purpose of secure network coding is to force the wiretapper hear no source message from the radio.

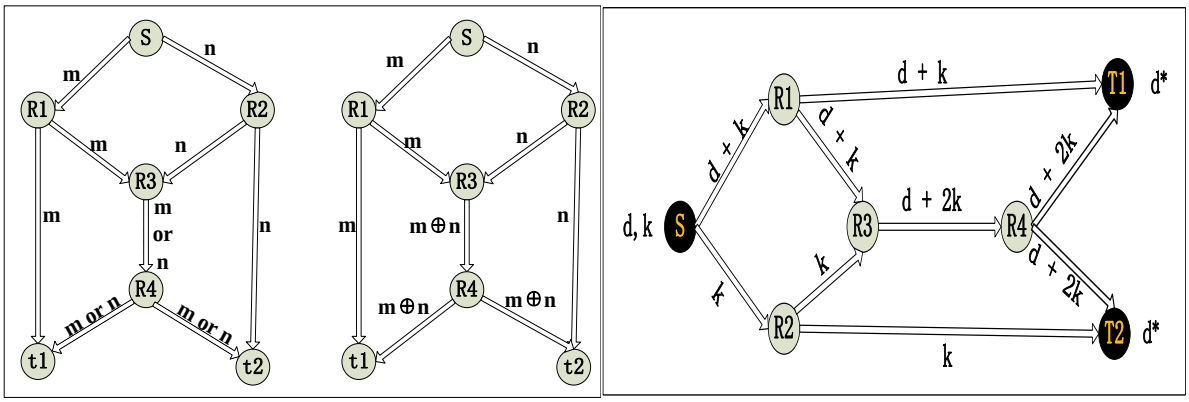

Fig. 6 a) An example of network code b) A simple illustration of secure network code

Secure network coding is a very potential technique for wireless network, since the wireless network is easier to be attacked. Although secure network coding can be alternative of routing transmission to protect the data confidentiality, but it also has some threats and challenge to be overcome. For example, the basic construction of network coding is vulnerable to pollution attacks, where an adversary blocks the transmission channel by inserting fake message. Network coding exploits the operation to mix the packet flow. A single bit error package may lead to the full packages to be polluted. Therefore, only using XOR operation to encode the packages cannot ensure the security of data message as an adversary may inject corrupted packets to the nodes in the network. Confronted with this problem, there are two choices: one is to detect the existence of the adversary and another one is combined other techniques to correct the errors or resilient to the error packets [52].

4. **Open research issues.** The security of code image dissemination process is often a key concern in WSN as lots of nodes needed to be managed. Although many research efforts have been made in this area, there are still some vital problems needed to be overcome.

Resist Interference. The signal interference is a vital problem of the security of WSN. Since WSN is an open network, it is very easy to be affected by extra signal interference. The signal interference will force the whole network collapse. The two sensor nodes in the wireless network cannot communicate with each other, so disseminate the code images to sensor nodes is impossible. Thus, the software will be delayed to update and the sensor nodes could not collect timely information. So, it is very important to design secure scheme to resist interference.

Network Coding. Network can help us improve communication capacity, but it will also bring pollution attacks problem. For example, an attacker may inject malicious code image packets into WSN, when these

packets are received by sensor nodes, it will become more difficult to decode the original packets. Thus, it will waste sensor nodes a lot of resource and energy to process these fake packets. So, there is much work to do in order to handle this problem. With network coding techniques, the sensor network will deliver more packets at one time. That is to say, network coding methods can virtually increase network throughput. However, coding methods' efficiency also affects the security of code image dissemination. If the decoding process takes too much time or resource, it may become a new security hole.

Formal Methods and Tools for Code Image Dissemination Protocols. In the literature, almost all the researchers stated that their approaches are secure enough, but there are no formal standards to prove that whether an algorithm or a protocol is secure or not. That's pretty hard to do this work. However, there are some formal methods we may apply to solve these formal security analysis and verification. For example, coq [53] and Isabelle [54], we can use this interactive theorem prover to do some formal verification. As far as we are concerned that researchers have proved the famous four color theorem with Coq [55]. So, we may exploit these tools to test whether the protocol or the scheme satisfy the secure requirements. The secure scheme can be extracted some theorem and then use some tactics to prove them. This is a very potential research direction.

Low Power Consumption Algorithms. Secure algorithms should consider the energy constraint of wireless sensor nodes, since it is difficult to charge the batteries. Thus, it is very essential that using low power consumption algorithms. For example, we can exploit "sleep "state in an algorithm if the sensor node does not need to do other activity.

Simple Scheme is the Best. When designing a secure schema, it is not reasonable to use complex methods. As far as we are concerned, if the method is used in WSN area, it almost useless as complex methods may lead the network work inefficiency.

Conclusion. In this paper, we first analyze the features of WSN, and then talk about the security aspect of it. We have analyzed several normal attacks in WSN and described the current work against these attacks. Later, we have showed the major non-security and security reprogramming protocol in WSN. These protocols could help us understand how to disseminate code image and how to protect this process against attackers. In order to comprehend the salient features of WSN, we have discussed the challenge when using these security techniques. After that, we have summarized six most useful techniques and illustrated how to exploit them in code image dissemination. However, WSN is still in the stage of development and many proposed protocols still need to be improved. Hence, it is urgent to research some efficient methods to reduce communication traffic, computation time, and storage overhead at the same time ensures the security of code image dissemination to the sensor nodes.

Acknowledgment. This research is supported by a strategic research grant from City University of Hong Kong [Project No. 7004225], the Pearl River Nova Program of Guangzhou (No. 2014J2200051), the National Science Foundation of China (Grants:51477056 and 61321064), the Shanghai Knowledge Service Platform for Trustworthy Internet of Things "(No. ZF1213), the Shanghai Rising-Star Program (No. 15QA1401700), a visiting scholar project of the State Key Laboratory of Power Transmission Equipment & System Security and New Technology (No. 2007DA10512713406), and the Specialized Research Fund for the Doctoral Program of Higher Education. D. He is the corresponding author of this article.

REFERENCES

[1] Yanchao Zhang and Rui Zhang.(2011). LR-Seluge: Loss-Resilient and Secure Code Dissemination in Wireless Sensor Networks. International Conference on Distributed Computing Systems. IEEE, pp. 497-506.

[2] J. Jeong, S. Kim, and A. Broad..(2003)."Network Reprogramming,"[Online]. Available: http://www.tinyos.net/tinyos-1.x/doc/Xnp.pdf.

[3] Patrick E. Lanigan, Rajeev Gandhi, and Priya Narasimhan.(2006) .Sluice: Secure dissemination of code updates in sensor networks. In 26th IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 53-53.
[4] Sandeep S.Kulkarni and Limin Wang..(2005). Mnp: Multihop network reprogramming service for sensor networks. In Proceedings of 25th IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 7-16.
[5] Jaleel Shaheen, Diethelm Ostry, Vijay Sivaraman, and Sanjay Jha..(2007). Confidential and secure broadcast in wireless sensor networks. In the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) ,pp. 1-5.
[6] PA Levis, N. Patel, D. Culler, and S. Shenker..(2003). Trickle: A Self-Regulating Algorithm for Code Propagation and Maintenance in Wireless Sensor Networks. Computer Science Division, University of California, pp. 15-28.
[7] Prabal Dutta, et al.,.(2006). Trio: enabling sustainable and scalable outdoor wireless sensor network deployment," In Proceedings of the 5th international conference on Information processing in sensor networks, pp. 407-415.
[8] Q. Wang, Y. Zhu, and L. Cheng..(2006). Reprogramming Wireless Sensor Networks: Challenges and Approaches. IEEE Network, Vol. 20, No. 3, pp.48-55.
[9] J. Kulik, W. Heinzelman, and H. Balakrishnan..(2002). Negotiation-based Protocols for Disseminating Information in Wireless Sensor Networks. Wireless Networks, Vol. 8, pp. 169-185.
[10] J. W. Hui and D. Culler..(2004). The dynamic behavior of a data dissemination protocol for network programming at scale. In Proceedings of the Second International Conferences on Embedded Network Sensor Systems, pp. 81-94.
[11] Kulkarni, Sandeep, and Limin Wang.(2009). Energy-efficient multihop reprogramming for sensor networks. ACM Transactions on Sensor Networks (TOSN) Vol. 5, No. 2, pp. 1-40.
[12] Sangwon Hyun, Peng Ning, An Liu, and Wenling Du..(2008). Seluge: Secure and DoS-Resistant Code Dissemination in Wireless Sensor Networks. In *Proceedings of the 7th international conference on Information processing in sensor networks*, pp. 445-456.
[13] Zeng Yong, Wang Xin, Dong Lihua, Ma Jianfeng, and Liu Zhihong..(2012). Out-of-Order-Delivery-Tolerant Secure Code Dissemination with Fountain Codes in Wireless Sensor Networks. In the 8th International Conference on Computational Intelligence and Security (CIS), pp. 683-686.
[14] M. Luby..(2002). LT codes. In Proceeding of the 43rd Annual IEEE Symposium on Foundations of Computer Science, pp. 271-280.
[15] Daojing He, Chun Chen, Sammy Chan, and Jiajun Bu..(2012). SDRP: A Secure and Distributed Reprogramming Protocol for Wireless Sensor Networks. IEEE Transactions on Industrial Electronics, Vol. 59, No. 11, pp. 4155-4163.
[16] Daojing He, Chun Chen, Sammy Chan, and Jiajun Bu. (2012). DiCode: DoS-Resistant and Distributed Code Dissemination in Wireless Sensor Networks. IEEE Transactions on Wireless Communications, Vol. 11, No. 5, pp. 1946 – 1956.
[17] Nils Aschenbruck, Jan Bauer, Jakob Bieling, Alexander Bothe, and Matthias Schwamborn..(2012). Selective and Secure Over-The-Air Programming for Wireless Sensor Networks," 21st International Conference on Computer Communications and Networks(ICCCN), pp. 1-6.
[18] D. He, S. Chan, Y. Zhang, and H. Yang..(2014). Lightweight and Confidential Data Discovery and Dissemination for Wireless Body Area Networks. IEEE Journal of Biomedical and Health Informatics, Vol. 18, No. 2, pp.440-448.
[19] C. Teat and S. Peltsverger..(2011). The security of cryptographic hashes. In Proceedings of the 49th Annual Southeast Regional Confference, pp. 103-108.
[20] D. He, S. Chan, S. Tang, and M. Guizani..(2013). Secure data discovery and dissemination based on hash tree for wireless sensor networks. IEEE Transactions on Wireless Communications, Vol. 12, No. 9, pp. 4638-4646.
[21] M. E. Hellma. (1980). A cryptanalytic time-memory trade-off. IEEE Transactions on Information Theory, Vol. 26, No. 4, pp. 401-406.
[22] H Kumar, S Kumar, R Joseph, D Kumar, et al. (2013). "Rainbow table to crack password using MD5 hashing algorithm," 2013 IEEE Conference on Information & Communication Technologies (ICT), pp. 433-439.

[23] Orhun KARA and Adem ATALAY. .(2009). Preimages of hash functions through rainbow tables. 24th International Symposium on Computer and Information Sciences (ISCIS), pp. 304-309.
[24] Ruhma Tahir, Huosheng Hu, Dongbing Gu, Klaus McDonald-Maier, and Gareth Howells. (2013). Resilience against brute force and rainbow table attacks using strong ICMetrics session key pairs. 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA), pp. 1-6.
[25] J. Buchmann, L. C. C. Garcia, et al.. (2006). CMSS-an improved Merkle signature scheme. In Progress in Cryptology (INDOCRYPT), Springer Berlin Heidelberg, pp. 349-363.
[26] SHEN Yu-long, MA Jian-feng, and PEI Qing-qi. .(2007). An Access Control Scheme in Wireless Sensor Network. In Proc. 4th IFTP International Conference on Network and Parallel Computing Workshops (NPC), pp. 362-367.
[27] F. Wang, Y. Zhang, Y. Xu, L. Wu, and B. Diao. (2014). A DoS-resilient enhanced two-factor user authentication scheme in wireless sensor networks. International Conference on Computing, Networking and Communications (ICNC), pp. 1096-1102.
[28] H. Li, R. Lu, B. Yang, and X. Shen. .(2014). An Efficient Merkle-Tree-Based Authentication Scheme for Smart Grid. IEEE Systems Journal, Vol. 8, No. 2, pp. 655-663.
[29] Q. Zhang, P. Wang, Douglas S. Reeves, and P. Ning. (2005). Defending against Sybil Attacks in Sensor Networks. 25th IEEE International Conference on Distributed Computing Systems Workshops, pp. 185-191.
[30] L. Yang, S. Li, Y. Zhang, and G. Liu. (2014). Low-overhead authentication method for reprogramming protocol based on rateless codes in wireless sensor networks. International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 881-886.
[31] Berbecaru, Diana, and Luca Albertalli. (2008). On the performance and use of a space-efficient merkle tree traversal algorithm in real-time applications for wireless and sensor networks. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WIMOB), pp. 234-240.
[32] V. Kondratieva and Seung-Woo Seo. (2007). Optimized Hash Tree for Authentication in Sensor Networks. IEEE Communications Letters, Vol. 11, No. 2, pp. 149-151.
[33] A. Dvir, L. Buttyan, and T. V. Thong. (2013). SDTP+: Securing a distributed transport protocol for WSNs using Merkle trees and Hash chains. IEEE International Conference on Communications (ICC), pp. 2073-2078.
[34] Przydatek, Bartosz, Dawn Song, and Adrian Perrig.(2003). SIA: Secure Information Aggregation in Sensor Networks. In Proc. of the 1st ACM international conference on Embedded networked sensor systems, pp.255-265.
[35] D. He, S. Chan, and M. Guizani. (2012). Small data dissemination for wireless sensor networks: The security aspect. IEEE Wireless Communications, Vol. 21, No. 3, pp. 110-116.
[36] K. Shashi Prabh and Tarek F. Abdelzaher. Energy-conserving data cache placement in sensor networks. ACM Transactions on Sensor Networks (TOSN), Vol. 1, No. 2, pp. 178-203.
[37] B. Gassend, GE Suh, D. Clarke, M. Van Dijk, and S. Devadas. (2003). Caches and Hash Trees for Efficient Memory Integrity Verification. IEEE 9th International Symposium on High Performance Computer Architecture (HPCA), pp. 295-306.
[38] MacKay and David JC .(2005). Fountain codes. IEE Proceedings-Communications, Vol. 152, No. 6, pp. 1062-1068.
[39] John W. Byers, Michael Luby, and Michael Mitzenmacher. (1998). A digital fountain approach to reliable distribution of bulk data," ACM SIGCOMM Computer Communication Review, Vol. 28, No. 4, pp. 56-67.
[40] Y. Lin, B. Liang, and B. Li. (2007). Data Persistence in Large-Scale Sensor Networks with Decentralized Fountain Codes. 6th IEEE International Conference on Computer Communications, pp. 1658-1666.
[41] M. Busse, T. Haenselmann, and E. Wolfgang. (2007). Energy-Efficient Data Dissemination for Wireless Sensor Networks. Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops. PerCom Workshops, pp.301-306.
[42] Rossi, Michele, et al. (2008). SYNAPSE: A Network Reprogramming Protocol for Wireless Sensor Networks Using Fountain Codes. 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, pp. 188-196.

[43] Ramalingam K. C., Venkatachalam Subramanian, A. Selcuk Uluagac and Raheem Beyah.(2012). SIMAGE: Secure and Link-Quality Cognizant Image Distribution for Wireless Sensor Networks. IEEE Global Communications Conference, Anaheim, CA, pp. 616-621.
[44] M. Rossi, G. Zanca, et al. (2010). Synapse++: Code Dissemination in Wireless Sensor Networks Using Fountain Codes. IEEE Transactions on Mobile Computing, Vol. 9, No. 12, pp. 1749-1765.
[45] A. S. Wander, N. Gura, H. Eberle, V. Gupta, and S. C. Shantz, "Energy analysis of public-key cryptography for wireless sensor networks," In Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications. Percom 2005, March 2005, pp.324-328.
[46] Perrig A, Szewczyk R, Tygar J D, et al.. SPINS: Security protocols for sensor networks. Wireless networks, Vol. 8, No. 5, pp. 521-534.
[47] A. Liu, P. Ning, and C. Wang. (2009). Lightweight remote image management for secure code dissemination in wireless sensor networks,"In IEEE INFOCOM, pp. 1242-1250.
[48] N. Cai and R. W. Yeung. (2002). Secure network coding," In Proceedings of IEEE International Symposium on Information Theory.
[49] J. Tan and M. Medard. (2006). Secure Network Coding with a Cost Criterion. IEEE 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, pp. 1-6.
[50] S-YR Li, Raymond W. Yeung, and Ning Cai. (2003). Linear network coding. IEEE Transactions on Information Theory, Vol. 49, No. 2, pp. 371-381.
[51] N. Cai and T. Chan. (2011). Theory of Secure Network Coding. Proceedings of the IEEE, Vol. 99, No. 3, pp. 421-437.
[52] S. Jaggi, M. Langberg, S. Katti, Ho Tracey, D. Katabi, M. Medard, and M. Effros. (2007). Resilient network coding in the presence of Byzantine adversaries. 26th IEEE International Conference on Computer Communications, pp. 616-624.
[53] Huet, Gérard, Gilles Kahn, and Christine Paulin-Mohring. (1997). The Coq Proof Assistant A Tutorial. Rapport Technique 178.
[54] Nipkow, Tobias, Lawrence C. Paulson, and Markus Wenzel, eds. (2002). Isabelle/HOL: a proof assistant for higher-order logic. Springer Science & Business Media Vol. 2283.
[55] G. Gonthier. A computer-checked proof of the four colour theorem". [Online]. Available: http://research.microsoft.com/enus/um/people/gonthier/4colproof.pdf.
[56] Hou I H, Tsai Y E, Abdelzaher T F, et al. (2008). Adapcode: Adaptive network coding for code updates in wireless sensor networks. The 27th Conference on Computer Communications. INFOCOM,.
[57] Hagedorn A, Starobinski D, Trachtenberg A. (2008). Rateless deluge: Over-the-air programming of wireless sensor networks using random linear codes. In Proceedings of the 7th international conference on Information processing in sensor networks. IEEE Computer Society, pp. 457-466.