
# FALSE DATA FILTERING IN WIRELESS SENSOR NETWORKS

Ze LUO[1], Lingzhi ZHU [2, 3*], Yunjie CHANG [2*], Qingyun LUO[2], Guixiang LI[2], Weisheng LIAO[2]

[1] Department of Electrpnical and information Engineering , Hunan Institute of Technology, Hengyang 421002, Hunan,
[2] Department of Computer and Information Science, Hunan Institute of Technology, Hengyang 421002, Hunan,
[3]School of Information Science and Engineering, Central South University, Changsha 410083, Hunan, China
Emails:2245275@qq.com, lingzhi0825@163.com*, changyunjie@aliyun.com*

***Submitted: June 13, 2016*** ***Accepted: Oct.6, 2016*** ***Published: Dec.1, 2016***

***Abstract- Wireless Sensor Networks (WSN) is often deployed in hostile environments, and the attacker can easily capture nodes to inject false data to the sensor network, which can cause serious results. This paper has studied various false data filtering techniques recently in wireless sensor network. Based on encryption technology, we have analyzed and compared the difference of various existing filtering strategies, then have pointed out the merits and demerits of them in detailed. At last, we give the developing trend of false data filtering, which provides a strong foundation for the further research.***

**Index terms: wireless sensor networks; false reports; out-dated packets; MAC.**

# I. INTRODUCTION

Wireless sensor networks (WSN) which consist of a large number of sensor nodes are widely used in many important fields, from environment monitoring to scientific data collection, to medical use and military applications [1]. When sensor nodes with limited resources are deployed in unattended and hostile environments, the adversary may comprise and reprogram some sensor nodes. Once the adversary successes to compromise a large number of senor nodes, they would disclose all the secret information and abuse launch false report injection attacks [2], or replay data injection attacks [3]. The filtering schemes may become ineffective or even useless. Defending against such attacks in WSN is of essential importance, because these illegitimate data not only cause false alarms but also may drain out the constrained resources of the sensors.

# II. FALSE DATA

a. the definition of false data

In general network, the security goals include data confidentiality, integrity and authentication. But in consequence of the particularity of WSN nodes and applications, its security goals can be divided into confidentiality, integrity, authentication, availability and non-repudiation. These characteristics of WSN includes: limited storage space and computing power, lack of late node layout prior knowledge, limited communications bandwidth and limited energy and so on. These characteristics make the attacker capture sensor nodes, fabricating false event, malicious tampering with data packets being sent or send duplicate data and so on, these are these false data problem in WSN.

In recent years, some researches focused on preventing such false data, and have proposed many ways to the problem of false data. The false data filtering mechanism in WSN can be divided into two types: false report injection attacks and replay data injection attacks. Report false data injection attacks means that after sensor nodes are compromised by the malicious attackers, the malicious attackers will inject false sensing reports, maliciously modify reports or replay routing information, in addition to Sinkhole Attacks, Selective Forwarding Attack,

Sybil Attacks, Hello Flood Attacks and Hello Ask Attacks. Replay attack refers to that the middle nodes intercept legitimate messages and repeatedly send to the target node. To prevent replay attack, the time stamp can be implemented in report, and the present study is focused on the prevention of false data injection attacks.

b. False data filtering method

The basic framework of forwarding filtration includes four sections: key distribution management, data report generating, forwarding filtration and Sink checking. Key distribution management is to establish key-sharing relation between nodes. The establishment of the key sharing relationship directly affects the filtration efficiency and energy consumption, and then the key management is the core of filtering mechanism. Data report generating refers to multiple detection nodes to perceive an event at the same time, and central nodes make use of key to encrypt event, generated a data report. Forwarding filtration is first to examine whether a data package attaches *t* MAC from different detecting nodes when nodes receive the data package，and then regenerate a MAC with the stored key corresponding to the data package and compare whether it is the same as the MAC to be checked in the data package. If the detection at any step is not passed, the data package is abandoned right away. If nodes do not store the corresponding secret key, the data package is directly forwarded. SINK has advantages of global key information, abundant energy, powerful computing capacity and enough storage capacity. As the final barrier, SINK can eliminate all missed false data of forwarding filtration.

Fortunately, the research has made some development on the identification and filtration of false data in the WSN. Domestic and foreign scholars have proposed a number of false data filtering methods, its main purpose is to filter false data at the same time, as far as possible to reduce overhead caused by this method, reduce the energy consumption early. To solve false data filtering problem in the WSN, Scholars focus on the following aspects:

b. ⅰ filtering schemes based on the symmetric key

In the filtering schemes based on the symmetric key, both sending and receiving data must use the same key for encryption and decryption of the data. The scheme has some characteristics,

a small computational complexity, easy to implement and low energy consumption and so on. In view of energy efficient and practical point, it is interested in these schemes for limited resources of sensor networks. It is the classic filtering schemes, MAC authentication [4], PVFS mechanism [5-6], SIEF mechanism [7] and so on, because their resistance to false data ability strong.

b. ii filtering mechanisms based on the public key

Filtering mechanisms based on the public key use different keys for encryption and decrypt the data, or forwarding authentication. The mechanism has high security, the use of more flexible and easy to implement authentication information, but their safety need to do a lot of computation, which is a heavy burden to less storage space, lower computing power and energy limited WSN. Although some scholars have been properly optimized, the cost is still too large, Therefore，These schemes need further research and optimization. Based on public key technology, it is the classic filtering schemes, location-based key [8-9], elliptic curve key [10] and LEDs mechanism [11], and so on. Because they can prevent collusion attacks, security better.

b. iii filtering c based on the group key

Filtering schemes based on the group key is to distribute the key to the group as a unit, which is different from the distribution management of a single key. And then schemes regularly updates set of keys to prevent compromise node continued to attack, but the update algorithm of the group key is proposed too costly to maintain. In the filtering schemes based on the group key, DRA mechanism [13] is relatively good.

b. iv filtering mechanisms based on the digital watermarking

Filtering mechanisms based on the digital watermarking are the authentication information, such as the identity of the sensor node embedded in the data packet. Some scholars have tried to use a digital watermark to guarantee the reliability and confidentiality of the wireless sensor network to collect data. However, the cost of digital watermarking technology is low, but the embedded watermark information requires multiple nodes cooperate to complete.

Therefore the digital watermarking technology is still in its early stage of exploration, a lot of technology is not yet mature.

b. v filtering mechanisms based on the time synchronization with time-varying parameters

An attacker exploits compromised nodes to inject false data into the network repeatedly, and it also leads to a sensor network of energy wasted. Time synchronization technology is based on highly synchronized in time, use the timestamp as the main means of identification data certification. Filtering schemes based on interference technique can resist the compromised nodes collusion attack, low overhead. SPINS, Zigbee and SRAR mechanisms can detect and filter out-dated packets by the sink nodes, but they uses so complex encryption and decryption algorithm that the forwarding nodes can not filter out-dated packets in the forwarding process, and it is not conducive to energy saving node. TSPC requires that all sensor nodes maintain time synchronization, but it is obviously difficult to practical application of energy and limited wireless sensor networks.

b. vi filtering mechanisms based on Trust management

Trust management mainly assesses the trust value of each node in WSN, and quantifies all relevant information about the affected node trust value, node behavior, interaction records node with other nodes and the views of other nodes and so on. This schemes use appropriate calculation model to go the trust value of the node, when the trust value of a node is below a certain threshold, this scheme thinks the node has been compromised. Thereby isolating the active node compromise, to prevent its network continues to launch false data injection attacks. Tana et al put forward a trust routing for location-aware sensor networks called TRANS. TRANS is a trust routing based on geographic information in WSN, locates and identifies a suspicious position, the suspect nodes removed from the routing list, or node bypasses the routing path suspicious nodes, enabling secure routing based on trust. Crosby et al propose a framework for trust based cluster head election in wireless sensor networks [24], the scheme introduces trust management mechanism into the election of cluster head node and takes the methods of redundancy and challenge-response, as far as possible to ensure that the elected cluster head node as a trusted node. Ganeriwal et al have suggested reputation

based framework for sensor networks, named RFSN [14],

The mechanism evaluates the credibility between nodes, and then establishing a trusted network environment is made up of trusted nodes.

b. vii filtering mechanisms based on the interference polynomial

The interference polynomial technique refers to the use of a quadratic polynomial instead of key data encryption and decryption, in polynomial parameter introduction interference number is in order to prevent the collusion attack.

## III. Typical false data filtering mechanisms analysis

a. SEF

To the problem of false data filtering, Fan Ye *et al.* firstly presented a mechanism called SEF [3] in wireless sensor networks. In the SEF mechanism, a global key pool is divided into multi-partitions of secret key and every partition includes *m* secret keys. Before deployment randomly, each node can choose a partition, and then selects arbitrary *k* keys from the partition to store. If any event happens, multi-nodes for detection generate a report including *t* different MAC collaboratively. In the process of transmitting the data package, the intermediate node which owes the same key partition with the detection node can check up an MAC in the data package using a probability of *k* / *m*. Finally, all the false packages can be filtered. However, there are two severe problems in the SEF mechanism. First, if a compromised node caches some legal data and injects a large number of copies into the network, then these copies would all be transmitted to sink, leading to a waste of energy. Second, if the key has not been bound with the surveyed area, once the attacker captures *t* different key partitions, he can fabricate false data packages which can't be distinguished by transfer nodes.

As illustrated in Figure. 1, we assume that six nodes $(S_1 \ldots S_6)$ are compromised by the adversary. When $t$ is equal to five, each node which has a different key partitions, locates in different geographic areas. But the attacker can still use them synergistically fake a false

report $R$ , and then send it to the neighbor nodes of $S_1$ by a sensor node $S_1$. Yet the forwarding nodes and sink nodes are not able to detect a false report *R*, which cause the waste of energy.

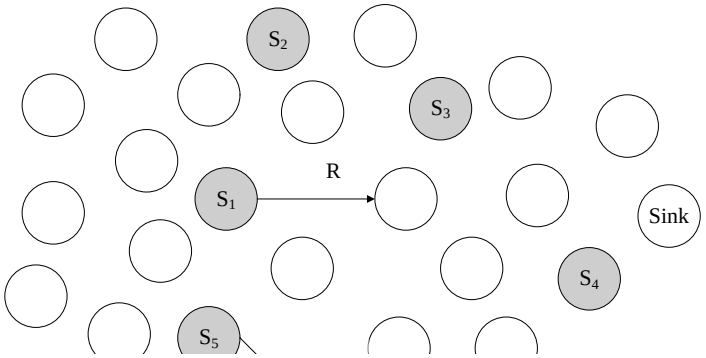


Figure.1 multiple nodes collaborative faking false data in WSN

b. Improved SEF

The filtering mechanism based on SEF mainly includes IHA[4], GRSEF[5], MDSEF[6] and AERF[7] and so on. Zhu et al firstly have suggested a step-by-step authentication mechanism (called IHA). In IHA, the sensor nodes self form clusters, and assume that each cluster includes $T$ +1 sensor nodes. After forming clusters, a path was established from every cluster head to sink node. In the path a cooperative relationship was set up between the nodes at a distance of $T$ +1 hop counts. When an event took place, every sensor node utilizes the private key sharing with sink node and the pair-wise key sharing with the downstream cooperative nodes, to produce 2 MACs. Cluster head nodes collect MACs of $T$ +1 sensor nodes to generate data report. In the process of forwarding, each node checked and corrected MAC brought by the upstream cooperative nodes. After successful verification, a new MAC formed and replaced the verified MAC by means of the key sharing with the downstream cooperative nodes. Compared with SEF, IHA achieves significant improvement in the anti-attack capability and the probability of false data filtering. However, once the route changes, the uncertainty of the data authentication also will fail to re-establish the route and distribute keys due to large maintenance costs. Therefore it is unbearable to energy constrained sensor

networks. Finally, the key is directly transmitted between nodes after deployment of sensor networks, which is likely to cause the key leakage.

b.i GRSEF

Yu et al. brought up a false data filtering scheme based on groups (called GRSEF). Before deployment, GRSEF use localization algorithm estimating network deployment size and shape of the region, and network parameters such as network topology and a key shared with sink nodes for each node. After deployment, we would divide all the sensor nodes into t groups, which make sure that each position is covered by different partitions of key. Then, the same key are distributed to the same group of nodes by the coordinate axis way. In the process of forwarding, forwarding nodes makes use of the pre-shared key to check MAC in the data report. Finally, Sink nodes filters all the missed false packages. GRSEF need equip every node the expensive positioning facility (like GPS) to estimate the deployment size and shape of sensor network, but the energy cost of this expensive positioning facility is too much.

b.ii MDSEF

Yang *et al*. [5] put forward a filtering mechanism, MDSEF, based on and en-route forward. A global key pool is divided into multiple sub-sets in MDSEF, and each set is also divided into multiple key groups. Each node takes part in each group of each set and selects some key from these groups for storage. This multi-dimensional key distribution mechanism can make nodes cover by multiple sub-sets simultaneously, and overcome the problem that individual nodes can not successfully generate data packets in the existing algorithm. The mechanism also proposes a coordinate axis strategy to correlates with geographic regions and a set of multiple keys, and then the key acquisition method based on geographic location is given from this strategy. According to a distributed stepwise refinement group joining algorithm, each node selects the key groups. Although MDSEF effectively improve the performance of key coverage and filtration efficiency, both the group joining algorithm and the axis of rotation based algorithm brought greater computational and communication energy consumption for energy limited wireless sensor networks.

b. iii AEFS

Naresh et al. [6] proposed an active filtering scheme for false data called AEFS to cope with f alse data injection attacks and DoS attacks simultaneously in WSN. Each node was firstly initi alized with a hash chain. It then distributes its verifying key to some intermediate nodes. After sending a data report, the sensing nodes all immediately disclose their keys, enabling the for warding nodes to check the corresponding reports. The scheme uses a so-called Hill Climbing algorithm to distributing keys with which the nodes lies closer to the source possess stronger filtering capacity than others. Furthermore, AEFS utilizes the actual property of broadcasting i n wireless communications to deal with

DoS attacks. The scheme is able to identify and filter out false data earlier with a low require ment in memory and computing. However, the Hill Climbing method also incurs a huge com munication cost.

b. iv TFPF

Figure2 assume that the nodes *S*1 and *S*6 are compromised. When *t* is equal to 5, if the attacker forges a fake bag $R : (e, M_1, M_2, M_3, M_4, M_5)$ and sends to its neighbor nodes, R will be filtered. But when the attacker injects false data package *R* to sensor networks by compromised nodes *S*2, forwarding nodes in the path from *S*2 to sink nodes can not filter false data package *R*, so that false data package *R* eventually is transferred to sink nodes, resulting in waste of networks energy.

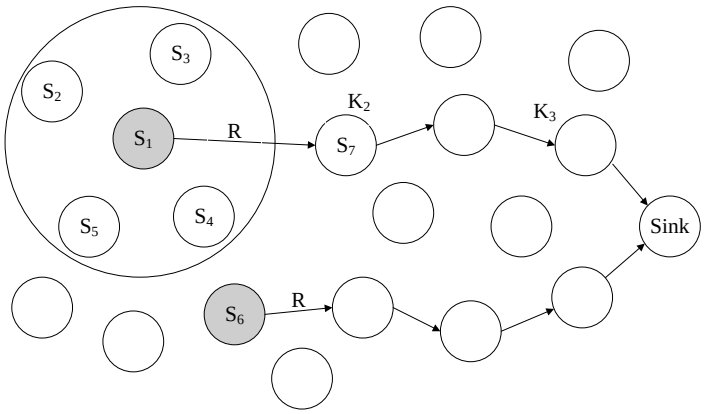


Figure2 false data injection from non-forwarding areas

Considering the problem of false data from non-forwarding areas, Zhu *et al*. puts forward a false data filtering mechanism based threshold in wireless sensor networks. In this scheme, each node establishes a path to sink node after deployment. When a report was generated for a sensed event, it must carry *t* MAC from different detecting nodes and two security threshold parameters. When a forwarding node S has received the data report R , the data report R is validated as the following steps:

$$R: \{e, S_1, S_2, ..., S_t, M_1, M_2, ..., M_i, Bin_v, T_v, T_c, flag\} \quad (1)$$

Firstly, a forwarding node checks whether the status flag of the data report is true. If the flag is true, there is no need to validate the data report, and then the forwarding node just forwards data report. If the flag is false, a forwarding node checks whether $t\{S_v, M_v\}(1 \le v \le t)$ tuples exist in the data report R . If so, the data report R can be dropped right away. When the number of MAC meets the demand, the stored key index table is checked. If the same key as the one in the data report R is not stored, it stores no keys. $T_c$ should be added 1. Next, is $T_c = T_c - max$ true? If it is true, it means $R$ has had continuous transmission of $T_c$ - $max$ hops but not validated them. We have concluded that the data package is the false one injected by the attacker from the non-forward zone of the compromised node. So R can be immediately discarded, and the verification process is over. The probability of filtering the false packages within one hop is

$$P_{tf\_1} = \frac{1}{\alpha + 1} \times P_v + \frac{\alpha}{\alpha + 1} \times \frac{1}{t} \quad (2)$$

And the probability of filtering the false packages within *h* hops is

$$P_{tf\_h} = 1 - (1 - P_{tf\_1})^h \quad (3)$$

The length of data package is

$$I_{r_{tf}} = I_y + (I_m + I_n) \times t + 3I_f + I_b \quad (4)$$

When transmission distance is *H* hop, the energy consumptions can be showed as followed:

$$E_{tf} = \left[ 1 + \frac{I_m + I_n}{I_r} \cdot t + 3I_f + I_b \right] \times \left[ H + \beta (H - \sum_{i=1}^{H-1} P_{tf}) \right] \quad (5)$$

But there are some problems about distributing keys after deployment:

1) The cost of communication link is too much.

2) It is not secure. Once the cluster head nodes are captured on the course of key distribution, the leakage of key will cause the security mechanism invalid.

3) Key distribution will take a long time to coverage, and network fails to carry through in-situ monitoring and data sense.

b. v PFPF

To the insecurity of post-deployment keys distribution, Zhu *et al*. presents a forwarding path-independent filtering scheme (called PFDF). In PFDF, a global key pool is constructed according to the expected keys-sharing degree, and each sensor node is given a unique *ID* before deployment. Suppose the node number in the network is *N* and the key sharing degree expected to realize is $n/N$ in the practical application, establish a global key pool as big as $m$, $m = n/N$.

$$G = \{K_i : 0 \le i \le m - 1\} \quad (6)$$

Delay $K_i$ to gi to load each node, as shown in figure 3.

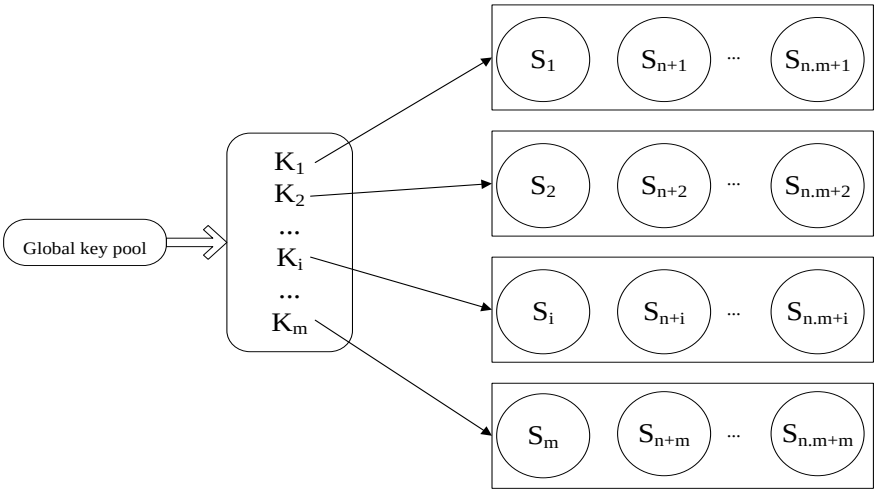


Figure 3 the distribution process of key

After key distribution is finished, every sensor node encrypts *e* with keys to create $M_i : K_i(e)$, and the node number and MAC are sent to the center node, which selects out *t* MACs caused by the nodes from different groups and data package *R* forms.

$$R : \{e, S_1, S_2, \cdots, S_t, M_1, M_2, \cdots, M_t\} \tag{7}$$

When an attacker injects from any zone in the network, the probability that a transmitted forgery package is filtered within one hop is,

$$P_{pf\_1} = \frac{t - N_c}{m} \tag{8}$$

and the probability that transmitted forgery packages is filtered within $h$ hops is

$$P_{pf\_h} = 1 - (1 - P_{pf\_1})^h \tag{9}$$

the length of data package

$$I_{r_{pf}} = I_y + (I_m + I_n) \times t \tag{10}$$

$$E_{pf} = \left[1 + \frac{I_m + I_n}{I_r} \cdot t\right] \times \left[H + \beta(H - \sum_{i=1}^{H-1} P_{pf\_i})\right] \tag{11}$$

c. Two-tiered wireless sensor networks

Because two-tiered wireless sensor networks have good extensible ability, they are regarded as the direction for the future of wireless sensor network (WSN).

c. i PVFS

Li et al came up with a filtering scheme based on cluster organization and voting mechanism, named PVFS [6]. In PVFS, each cluster is made up of this cluster covering nodes. The scheme establishes a shortest path from each cluster head nodes to sink nodes. Each cluster heads nodes were forwarding nodes. The key of a node in originating cluster was stored at the probability $d_i/d_0$. $d_0$ and $d_i$ were hop counts from originating cluster or forwarding cluster to sink node. Once an event happened, the sensor node brought out a Vote (The function of

Vote is similar to MAC). Data report was just produced by the Votes which were generated when cluster head collected *t* nodes in cluster. Upon forwarding, the forward cluster head verified data at a certain probability. Yet, much bigger semi-diameter for communication was needed between cluster heads than common nodes in order to forward data, which caused cluster heads to run out of their energy very quickly.

### c. ii RSFS

Ma et al proposed a sink verifying and filtering mechanism which is not limited by threshold (called RSFS). RSFS assumes that there are two types of nodes in the network, the first category is the general performance limited sensor nodes and the second category is the computing and communication ability of the cluster head node, and will not be compromised. The networks regions is divided into the same grid, after deployment, all nodes in the same grid form a cluster and sink nodes are most centrally located area, forming a star topology between cluster head nodes and sink nodes. Each node must share a common secret key with sink nodes, and after deployment establishes key shared relationship with neighboring nodes and all ordinary nodes of the same cluster by the pair-wise key mechanisms. Each cluster head nodes also share with eight neighbor cluster heads nodes by the pair-wise key mechanisms. When the unexpected event occurs, the node uses shared pair-wise key to encrypt data to generate MAC, and sends the MAC to the cluster head nodes. Compared to SEF and IHA, in the RSFS, cluster head nodes converges reported data and generated MAC of all nodes in the cluster, and then generates a data report. In the forwarding process, the forwarding nodes don’t verify MAC in the data report. Finally, sink nodes have received the aggregation results, verify and check the results and additional information of MAC for filtering false data. The filtering mechanism is not limited threshold, but the false data packet must be transferred to sink nodes to be detected, but can not be detected and filtered by forwarding nodes, it is not conducive to saving energy to sensor network. In addition, this mechanism assumes that the cluster head can communicate directly with sink nodes. Therefore it is not suitable to large-scale wireless sensor networks for assumptions too high.

### c. iii STEF

Krauss et al put forward a forwarding filter mechanism based on secure tickets (called STEF). The scheme, based on the common query-response model in the data collection mechanism, uses light weight one-way function to design the ticket. Sink node periodically sends a query to the perception area, which includes a ticket, nodes receive query automatically and become the temporary cluster head nodes, and the establishment of a dynamic cluster within its communication range, for data collection. Each sensing node uses a pre-shared secret key with sink node to generate a MAC and sends to the cluster head nodes. And then, the cluster head nodes collect MAC of sensing node and add the ticket to data to generate data reports，and sends data reports to sink nodes.

In the forwarding process, each intermediate node can use the pre-stored factor to verify the correctness of ticket. If the authentication fails, the data report is discarded, otherwise forwarding data report. Finally, when a data report is received, sink nodes use the shared secret key to verify correctness of each MAC and the ticket.

STEF utilizes a tickets model to identify and filter out false data, improving the probability of false data filtering. However, the scheme ignores that the forward node is also able to forge a fact, so only to the fake data of the source node, the false data filtering rate is high but the scheme is less secure in STEF. In addition, to the network node deployment sparse, the scheme cannot make sure that all nodes in the dynamic cluster can perceive simultaneously to the same incident, which can not construct a "t" legitimate data report.

c. iv Efficient false data filtering mechanism

Zhao, Zhu et al proposes an efficient and key security false data filtering mechanism. Before nodes deployment, the key server distributes each sensor node in a different key from global key pool $G = \{K_i : 0 \le i \le N - 1\}$. The key is called to the primary key of the node. The scheme make use of the pair-wise key management mechanism [7, 8] to achieve the confidentiality of the information transmission between cluster heads and ordinary nodes, and encrypts data report $R$ and Sends to the next hop cluster head.

$$R : \{e, N_1, N_2, ..., N_T, MAC_1, MAC_2, ..., MAC_T\} \qquad (12)$$

After the deployment of network node constituting clusters, we suppose that the cluster head

set is $CH_{\mathrm{v}} = \{C_1, C_2, ..., C_{\mathrm{i}}\}$, because the number of nodes in each cluster are L, and the initial energy of a node is the same, wherefore the initial energy of all the nodes in each cluster is the same. By using the algorithm in [20], the cluster head node set $CH_{\mathrm{v}}$ replaces node set V as input, which constructed the cluster head tree to maximize the network life cycle, as shown in Figure 4.

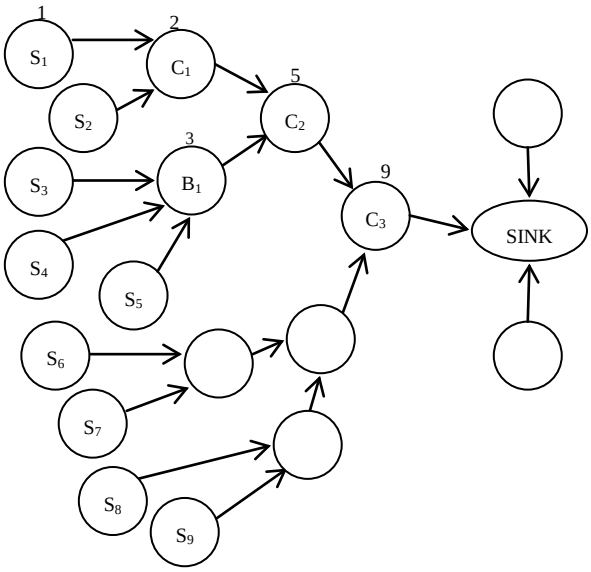


Figure.4 Cluster head spanning tree

each cluster nodes $(C_{1,}C_{2,},...,C_m)$ of burden factor is calculated as

$$\begin{cases} When\ C_i\ is\ a\ leaf\ \ in\ the\ tree, & Then\ \ B_i = 1; \\ Otherwise,\ note\ the\ sons\ of\ \ C\ \ as\ \ C_m (1 \le m \le n), & B_i = \sum_{m=1}^{n} B_m . \end{cases} \tag{13}$$

The probability that the forgery package of source cluster $S$ is filtered within $H$ hops is,

$$M_h = \sum_{i=1}^{h} R_S^{C_i} \tag{14}$$

$$\begin{cases} \mathrm{If}\ M_h = L, & p_h = 1; \\ If\ \ M_h < L, & p_h = \frac{T - N_C - 1 + M_h}{L}. \end{cases} \tag{15}$$

Making $L_{\mathrm{r}}$ signifies the length of data report without using any safe scheme, $L_{\mathrm{n}}$ signifies the length of sensor node ID , and $L_{\mathrm{M}}$ signifies the length of the MAC. Then the length of

data package in the scheme can be signified by

$$L_{\mathrm{r}} = L_{\mathrm{r}} + (L_{\mathrm{n}} + L_{\mathrm{m}}) \times T \tag{16}$$

and the expense of transmitting"1 false data + $\beta$ outdated data " is $E$ .

$$\begin{aligned} & E_{SEF} = (1 + \frac{L_M + L_n}{L_r} \cdot T)(H + \beta \frac{1-(1-p_0)^H}{p_0}), \\ & here\ p_0 = \frac{k_0 (T - N_c)}{N} \end{aligned} \tag{17}$$

$$\begin{cases} If\ M_{H_0 - 1} \le L - (T - N_c - 1) \le M_{H_0},\ and\ H_0 \le H, \\ Then\ E = [1 + \frac{(L_M + L_n)}{L_r} \cdot T] \times [H + \beta (H_0 - \sum_{i=1}^{H_0 - 1} p_i)]; \\ If\ M_H < L - (T - N_c - 1), \\ Then\ E = [1 + \frac{(L_M + L_n)}{L_r} \cdot T] \times [H + \beta (H - \sum_{i=1}^{H-1} p_i)]. \end{cases} \tag{18}$$

d. A one-way function based filtering

Zhou et al. put forward a false data filtering scheme based on one-way function FFRF [16]. In FFRF, sensor nodes can be divided into probe nodes and check nodes, and each node is preset one-way function, based on the node ID to produce a one-way Hash chain $C_1, C_2, ..., C_t$. And then, each node will publish the initial value to the public, and intermediate nodes randomly choosing some to store the hash value of the source node (called authentication hash value). In the process of forwarding, intermediate nodes using the pre-store Hash value to verify the correctness, and also make use of shared symmetric key to verify this packet of MAC, to filter the false data. When Sink receives the packet, on the one hand, it can revalidate MAC and hash values to filter the false data, on the other hand it also can verify the exclusive OR of each MAC in the data packets to determine whether the source node is reliable, which can judge the compromise approximate location of the node. There are some problems in FFRF. Firstly, it is very easy to capture sensor nodes to get $r$ different key partitions and Hash value of $f$ nodes, so as to fail to filter false data packet and break through the security scheme.

To these problems, liu *et al*. put forward false data filtering scheme based on the MAC and one-way function. After deployment, the initial key and hash value is pre-allocated to sensor nodes, each data report included with the MAC and fresh hash values from $t$ detecting nodes, and forwarding nodes verify the logicality of the relative positional relationship of the detecting nodes, the correctness of MAC and hash values, and the freshness of these hash values.

As is shown in Figure.5, here we assume that the forwarding path from the source node $A_1$ to Sink is described as: $Path(A_1)=\{A_1, A_6, A_7, A_8, sink\}$ where neighbor nodes of $A_1$ are $A_2,\ldots,A_5$ nodes. Forwarding node $A_6$ pre-stores the key $K_1$ and hash value $h_1^1$ from node $A_1$, and node $A_7$ stores the key $K_2$ and hash value $h_2^1$ from node $A_2$.

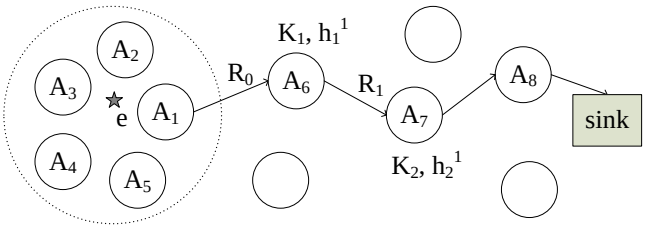


Figure.5 the procedure of verifying outdated and false reports

The probability of checking up one of these $t-N_d$ nodes is,

$$p_{b\_i}=\frac{t-N_d}{num(S)+1}\cdot\frac{c_0-i}{c_0} \tag{19}$$

Within *H* hops, the probability of filtering the false packages is

$$p_b(H)=1-\prod_{j=1}^{H}(1-\frac{t-N_d}{num(S)+1}\cdot\frac{c_0-j}{c_0}) \tag{20}$$

And the probability of the outdated package being filtered is,

$$p_{a\_i}=\frac{t}{num(S)+1}\cdot\frac{c_0-i}{c_0} \tag{21}$$

$$p_a(H)=1-\prod_{j=1}^{H}(1-p_{a\_j})=1-\prod_{j=1}^{H}(1-\frac{t}{num(S)+1}\cdot\frac{c_0-j}{c_0}) \tag{22}$$

The length of data package in this scheme could be signified by

$$L_{r0} = L_r + 2L_s + (L_n + L_k) \times t \tag{23}$$

Within $H$ hops, the expense of transmitting"1 false data + $\beta$ outdated data " is

$$E_0 = L_{r_0} \cdot [(1 + \sum_{i=2}^{H} (i \cdot p_{b\_i} \cdot \prod_{j=1}^{i-1} (1 - p_{b\_j})) + \beta \cdot \sum_{i=2}^{H} (i \cdot p_{a\_i} \cdot \prod_{j=1}^{i-1} (1 - p_{a\_j}))] \tag{24}$$

e. Filtering scheme based on the location

Ayday *et al*. put forward the key scheme and neighbor authentication scheme based on the location, and then further propose the threshold digital signature method of filtering false data mechanisms based on the location (called LCNS)[17].

After collecting the data, the aggregation node uses the secret sharing algorithm to gather data into share, and then pass each share to sink by the "cell by cell" way. Because the data packet contains a hash tree, the forwarding node can verify share. Therefore disadvantage of this mechanism is that it needs GPS and robotics, and the expenses of establishment and maintenance of pair-wise key are too high.

Wang *et al*. use elliptic curve key technology to improve the existing false data filtering algorithm based pair-wise key and propose PDF program [18]. PDF divides system secret into multiple share by the linear network coding way, and each share contains only part information of system secret. Before deployment, each node randomly selects a share for loading. After detection of emergencies, multiple sensor nodes unite and generate a key by each share saved in nodes, and then use this key to encrypt the perceived information to generate a digital signature.

Therefore, the sensor nodes respectively generate a digital signature attached to the back of the data packet to send. In the forwarding process, the intermediate node can use a pre-stored share verifying the correctness of the data packet, which will filter out false data. Although this scheme improves security, it does not apply to the smaller density and mobility of larger networks on account of its relatively large energy consumption.

REN *et al*. [12] proposed false data filtering mechanism based on geographic location and

public key mechanism to achieve end-end data safe transmission in WSN. This scheme divides the whole network into non-overlapping cell, apply the linear network coding, transmit data by the cell by cell methods and ensure the security of data by the threshold sharing technology and public key technology. After deployment, each node makes use of GPS positioning algorithm to obtain the location of node, and then independently produce three types of secret key based on the geography information: the first is the shared secret key between the nodes and sink which ensures communication secure between the nodes and sink, and such key has two key. The second is the cell key between the nodes and other nodes in cell, and the third is the different key which is shared by between nodes and the nodes in the forwarding path to ensure end to end data security. After detection of unexpected events, each node uses three key generate corresponding MAC and the MAC is embedded in the data packet for transmission. In the forwarding process, each node respectively verify corresponding MAC packet, which will filter out false data.

To the threshold value t limit issue, liu proposes the false data filtering scheme based on geographical location. After the sensor network deployment, node distributes geographic information to part of neighboring nodes, and each report must contain data t MAC and geographic information from different partition keys detection node. Forwarding nodes verify not only the correctness of MAC and geography information in the data packet and the legality of geographic location, which can effectively filter the collaborative forged false data of compromise nodes in different geographic areas, but the scheme needs to equip nodes with expensive GPS positioning tool.

f. Defending outdated data

Outdated data is also a kind of false data, and it will also lead to waste of energy networks. To defending this problem, Perrig *et al*. put forward the SNEP and $\mu$TESLA algorithms to check up outdated data in the SPINS protocol. The basic idea of the scheme is that the sender and target respectively maintain a counter and the counter are updated each time. After detection of unexpected events, the sender uses a hash function to encrypt the counter attached to the data packets and sends along. When the destination nodes accept the data packet, it uses a decryption function to get this counter, and determine the packet freshness by verifying the

correctness of the counter. Similarly, Zigbee protocol also uses the counter as time-varying parameters to detect outdated data in wireless sensor networks.

Some researchers propose the use random numbers to wireless sensor networks to detect outdated data solutions SRAR [17]. In SRAR, the sender use a pre-set hash function to encrypt random number, and then send attached secret information to the data packet. When the target node receives the data packet, it decrypts random number and verifies the data packet.

Chen *et al*. proposed an outdated data detecting scheme based on the time synchronization technique (named TSPC) [14]. In TSPC, all sensor nodes to maintain time synchronization, the sender embed timestamps in the packet, and the target node authenticates the time stamp.

g. Filtering scheme based on digital watermarking

Digital watermarking technology is kinds of data encryption technology developed recently, has been widely used for content authentication, secret communication and copyright protection, and have achieved good results. Compared to the traditional digital encryption technology, digital watermarking technology is that the secret information is embedded into the data, and it keeps the original characteristics of the data, the watermark information can change with the change of data. So we can extract the embedded watermark, and verifies its correctness and completeness to judge whether data has been forged or tampered. Therefore, some scholars try to use digital watermarking to ensure the reliability and security of collecting data in wireless sensor network. Feng first uses digital watermarking to protect the intellectual property of the sensing data mechanism [15]. Peng discusses the application of digital watermarking in the data security protection of wireless sensor network, and makes use of SPPW technology to achieve the embedding of the watermark information [13]. Zhang combines spread spectrum technology with digital watermarking technology to solve the security of wireless sensor network aggregation problem [16]. Kleider *et al*. applies digital watermarking technology to the data authentication of wireless sensor network, and digital watermarks are embedded in the process of data collection and data processing, which ensures the reliability of the data by the verification of sink node to the watermark information [17].Yi *et al*. applied cooperation watermark technology and multiple semi-fragile watermarking

models to filter false data in the wireless sensor network [14].

Although the consumption of digital watermarking technology is small, the embedding of the watermark information needs multiple cooperative nodes to complete and during data transmission, it is difficult to verify the watermark information for intermediate nodes. At present, the sink node can only validate the watermark information, and then increases the transmission distance of the false data in the wireless sensor network, which is not conducive to save the network energy. In addition, the digital watermarking technology is still in the initial stage of exploration, and a lot of technology is not yet mature, so the application of the digital watermarking technology also has great limitations in wireless sensor network.

h. false data filtering scheme based on time synchronization

Based on the highly synchronized time, time synchronization technology uses the time stamp to authenticate to data packet. The majority of nodes need exchange the synchronization message according to the time synchronization mechanism, and keeps time synchronization with other sensor nodes in the wireless sensor network. Therefore the scalability, stability, robustness, convergence, energy aware and characteristics determine the network time synchronization mechanism.

Literature [19] thinks when network exist a large number of compromised nodes, if assuming nodes maintain time synchronization with other nodes and the shared key between report producers and authenticator between, which provides a "delayed" certification. The disadvantage of this mechanism is that when network scales between authentication and network delay are too large, it is difficult to determine the delay bounds. In addition, the energy consumption which the neighbor nodes keep highly time synchronized is too large.

i. Filtering scheme based on mark method

Literature [20] earliest study on the node localization problem in WSN and put forward a tracking scheme based on tracing to the source of the chain marking method. In other words, each forwarding node labels the forwarded packets in accordance with certain probability, and sink node collects enough packets to reconstruct the path from the node to a source node. But the shortcomings of the scheme is that we supposes the data space is large enough, and in the

real network environment, limited packet space, the node cannot unlimited to mark packets, so the method has only theoretical significance.

Literature [21] presents a tracking scheme based on the probabilistic packet marking method. In the scheme, each data packet can be labeled up to two nodes, when a node receives the data packet, according to a certain probability to decide whether to label the package. If it is marked and checks this package marked by several nodes marked, if there have been two nodes labeled this package, covering the first marker node information labeled with information of the node, and second nodes labeled information clear, otherwise it will mark the information node to the correct position. Each labeled packets are preserved some path information, so nodes only need to collect sufficient quantities of the package, can trace back to the source node. The defect is the upstream node may be labeled markers covering the downstream node, so the distance from the cluster nodes farther nodes, the marks are collected If the probability is lower, if the node is far away from the sink node is captured and launched attacks, the sink node will need to collect a large number of packets to be able to locate.

To the above problems, the literature [22] presented the probability of packet marking scheme. In this scheme, the probability of each labeled node collected is substantially equal. A calculation function applied to all network does not exist in the practical application, so it need choose the calculation function according to different network topology [22], but the disadvantage of this scheme is the attack node from the sink node farther, sink node will collect more packets to its location. To solve this problem, the literature [23] proposed equal number of packages marking scheme. In the scheme, except one hop neighbors, the sink node needs collect approximately equal number of packets to locate. The program has better scalability, but it does not completely solve the problem of WSN attack sent false data, still need to cooperate with other programs. The scheme is extended well, but it cannot completely solve the problem of false data injection in WSN, and still need to cooperate with other programs.

Literature [24] proposed a probabilistic marking method of false data filtering and tracing. Its main idea is that according to certain probability, some forwarding node labels the data report it is not filtered out, and the sink node will add failed validation of data report to tracking set for

positioning the source node uses. If the data exceeds the pre-set limit parameter, the sink node is performed backtracking procedure, for tracing the origin of the operation, the final positioning of the source node.

Table 1: Comparison filtering scheme

| Classification | Advantages | Disadvantages |
|---|---|---|
| MAC | *Using relay node<br>*High filtration rate<br>*Saving network resources | Multiple hop to filter |
| One-way function | *Location compromised node<br>*High filter<br>*Higher accuracy | Larger overhead |
| PVFS | *Resistance to MAC attacks<br>*High filtration rate<br>*Resist change data | Waste energy |
| STEF | *Admission ticket model<br>*High filtration rate | Ignoring the transfer node can also be forged data |
| Location key | *To prevent the collusion attack<br>*GPS support<br>*Robot support | Spending too much |
| Elliptic curve key | *Random network coding<br>*Public key technology<br>*Increase security | Larger overhead |
| LEDS | *Network partitioning<br>*Higher filtration rate<br>*Non repeating unit | Hypothesis condition is high |
| DRA | *Group key updating process for broadcast<br>*Anti-collusion attack<br>*To ensure the integrity of the update | Maintenance cost is high |

| | | |
|---|---|---|
| | information to identify | |
| Interference polynomial | *Encryption and authentication<br>*Resist collusion attack<br>*Less overhead than MAC Technology | High storage and transmission overhead |
| Time node synchronization | *Node time synchronization<br>*Shared key<br>*To provide "deferred" certification | It is difficult to determine the lower bound of the delay in large networks |
| DSF | *Improved IHA<br>*hill climbing<br>*Helps to balance the network load and prolong the network *lifetime. | The key sharing degree is too low, and the authentication efficiency is low. |
| Multi path routing filtering | *Frequent communication of cooperative nodes<br>*Higher filtration rate<br>*Good fault tolerance | Spending too much |
| Authentication node delegation policy | *Do not rely on MAC certification<br>*High filtration, low overhead<br>*Can resist replay attack | "Denial of service" filter |

## VI. CONCLUSIONS

Research from the current situation, filtering scheme based on the symmetric key is considered most suitable for WSN [11-19], and table 1 is more visually after comparing the pros and cons of common false data filtering scheme.

In this technique, higher filtration efficiency is MAC authentication mechanism [1], STEF mechanism [7], DRA mechanism [13] and certification mechanism node appointed strategy[7], but more than two MAC problems of STEF mechanism still exist in. Therefore network security is good location of the key technology[9], elliptic curve key technology[10], DRA mechanism[13] interference polynomial technology[16] and network authentication

node assignment strategy[20]; high adaptability is DSF mechanism[18] and network authentication node assignment strategy[20]; large overhead is MAC authentication mechanism[1], one-way hash chain technology [5] (storage overhead is too large), PVFS mechanism [6], location of the key technology[9], elliptic curve key technology[10] and multi path routing filtering mechanism, especially elliptic curve key technology of network for large low density and mobility[10]; assumptions it is the location of the key technology of [9], time synchronization technology [17] proposed technology.

How to identify and filter false data is a challenging puzzle, and false data filtering research become an important problem in wireless sensor network security research. We need to carry out an in-depth study from the following aspects.

1) Moving target for tracking and monitoring more and more become the focus of people's attention. In addition, the characteristics of wireless sensor network transmission link fragile makes its performance unstable, and then most of the research is carried out on static network topology. Therefore the study for dynamic environment and dynamic network topology false data filtering strategy is a worthy of in-depth study.

2) The prominent characteristics of wireless sensor network are the low storage space and extremely limited energy. Existing filtering mechanism makes use of encryption and decryption operation to filter false data, which greatly increases the cost of network. To WSN, it is a very heavy burden to complete the control of the specified period. Therefore, we should more consider the low energy to solve the problem of false data filtering.

3) In order to save the energy transmission, data is usually compressed and aggregated in WSN. As a result, packet tail often is damaged. Therefore it is not suitable to this type of network for additional *T* MAC authentication mechanism after the data packet. Therefore, the study of the dependence MAC authentication filter mechanism should be the direction of further research.

With the development of false data filtering technology, false data filtering problem will be effectively solved in the wireless sensor network, which will greatly improve the security of WSN, to further promote more technology with WSN platform, to promote social and fast and good development.

## AKNOWLEDGEMENT

This work is supported by the Key Projects of Science and Technology of Hunan Province (No.2013FJ3095),The National College Students' Innovation and Entrepreneurship Training Program ([2015]41-12824), The Project Development Plan of Science and Technology of Hengyang City (No.2016KG63,2014KG63), project supported by the Hunan Province College Students' Research Learning and Innovative Experiment Plan Fund(Xiang Jiao Tong[2015]269-562), The Hunan Province College Research Project of the Teaching Reform (Xiang Jiao Tong[2015]291-562),and the Project of the Science Research of the Hunan Provincial Education Department (13C207).

## REFERENCES

[1] REN F Y, HUANG H N, LIN C. "Wireless sensor networks", Journal of Software, Vol.14, No. 7, 2003, pp.1282-1291.
[2] SU Z, LIN C, FENG F J, Ren FY. "Key Management Schemes and Protocols for Wireless sensor networks". Journal of Software, Vol.18, No. 5, 2007, pp.1218-1231.
[3] Ye, F., Luo, H., Lu, S., & Zhang, L. "Statistical en-route filtering of injected false data in sensor networks". Joint Conference of the IEEE Computer & Communications Societies, Vol.4, 2004, pp.2446-2457
[4] Zhu, S., Setia, S., Jajodia, S., & Ning, P. "An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks". Security & Privacy. Proceedings. IEEE Symposium on, 2004, pp.259-271.
[5] Yu L, Li JZ. "Grouping-based resilient statistical en-route filtering for sensor networks". Proceedings of 28th Annual Joint Conference of the IEEE Computer and Communications Societies, 2009, pp.1782-1790.
[6] F. Yang, X.H. Zhou, Q.Y. Zhang, "Multi-Dimensional Resilient Statistical En-Route Filtering in Wireless Sensor Networks", LNCS: Lecture Notes in Computer Science, 2010, pp. 130-139.
[7] K. Naresh, K.P. Pradeep, K.S. Sathish, "An Active En-route Filtering Scheme for Information Reporting in Wireless Sensor Networks", IJCSIT: International Journal of Computer Science and Information Technologies,Vol.2, No. 4,2011, 1812-1819.
[8] MA M. "Resilience of sink filtering scheme in Wireless sensor networks". Computer Communications, Vol.30, No1, 2006, pp.55-65.
[9] F. Yang, X. H. Zhou, Q. Y. "Zhang. Multi-dimensional resilient statistical en-route filtering in Wireless sensor networks". In: Proc. of LNCS'10, 2010, pp.130-139.
[10] K. Naresh, K.P. PrDadeep, K.S. Sathish. "An Active En-route Filtering Scheme for Information Reporting in Wireless sensor networks", International Journal of Computer Science and Information Technologies, Vol.2, No.4, 2011, pp.1812-1819.
[11] A. K. Bashir, S. J. Lim, C. S. Hussain, M. S. Park. "Energy Efficient In-network RFID Data Filtering Scheme in Wireless sensor networks", IEEE Sensors Journal, Vol.11, No.7,

2011, pp.7004-7021.
[12] X. Y. Yang, J. Lin, P. Moulema, W. Yu, X.W. Fu, W. Zhao. “A Novel En-route Filtering Scheme against False Data Injection Attacks in Cyber-Physical Networked Systems”. In: Proc. of ICDCS’12, 2012, pp.92-101.
[13] S.Zhu, S.Setia, S.Jajodia. “An interleaved Hop-by-Hop Authentication Scheme for Filtering of Injected False Data in Sensor Networks”, Proceeding IEEE symposium on Security and privacy, S&P’04, 2004, pp.259-271.
[14] H.Yang, S.Lu. “Commutative Cipher Based En-route Filtering in Wireless sensor networks”, Vehicular Technology Conference, VTC’04, pp.1223-1227.
[15] R.X. Lu, X.D. Lin, H.J. Zhu, X.H. Liang, X.M. Shen. “BECAN: A Bandwidth-Efficient Cooperative Authentication Scheme for Filtering Injected False Data in Wireless sensor networks”, IEEE Transactions on Parallel and Distributed Systems, Vol.23, No.1, 2012, pp. 32-43.
[16] K.Ren, W.Lou,Y.Zhang. “Providing location-aware End-to-end Data Security in Wireless sensor networks”. Proceedings of the IEEE Conference on Computing and Communicating, INFOCOM’06.2006, pp.585-598.
[17] A.Perrig, R.Szewczyk, V.Wen, D.Culler, D.Tygar. “Spins: Security Protocols for Sensor Networks”. ACM Mobile Computing and Networking, MOBICOM’02, 2002, pp. 521-534.
[18] S. Chen, A. Dunkels, F. O¨ sterlind, T. Voigt, and M. Johansson. “Time synchronization for predictable and secure data collection in Wireless sensor networks,” in Proceedings of The Sixth Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net 2007), Corfu, Greece, 2007.pp.165-172
[19] H.Wang, Q.Li. “PDF: A Public-key based False Data Filtering Scheme in Sensor Networks”, Proceedings of the International Conference on Wireless Algorithms, Systems and Applications, WASA’07, 2007, pp.129-138.
[20] Yu L, Li JZ. “Grouping-based Resilient Statistical En-route Filtering for Sensor Networks”, Proceedings of 28th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM’09), 2009, pp.1782-1790.
[21] Fan Y，Hao Y，Zhen L. “Catching’moles’in sensor networks”, Proc of the 27th International Conference on Distributed Computing Systems,Washington DC: IEEE Computer Society, 2007,pp.69-77.
[22] Xu J, Qian H, Ying W, et al. “A deployment algorithm for mobile wireless sensor networks based on the electrostatic field theory”, The International Journal on Smart Sensing and Intelligent Systems, Vol.8, No.1, 2015, pp. 516-537.
[23] Bai, Q., & Jin, C, “Image fusion and recognition based on compressed sensing theory”, International Journal on Smart Sensing & Intelligent Systems, Vol.8, No.1, 2015, pp. 159-180.
[24] Feng Y, Xue-hai Z, Shu-guang Z. “Hierarchical traceback in wireless sensor networks”, Proc of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008, pp.1-4.
[25] Qiao J, Liu S, Qi X, et al, “Transmission power control in wireless sensor networks under the minimum connected average node degree constraint”, The International Journal on Smart Sensing and Intelligent Systems, Vol.8, No. 1, 2015, pp.801-821.