AN ENHANCEMENT OF CLUSTER-BASED FALSE DATA FILTERING SCHEME THROUGH DYNAMIC SECURITY SELECTION IN WIRELESS SENSOR NETWORKS

Jungsub Ahn and Taeho Cho

Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea

ABSTRACT

Today, wireless sensor networks (WSNs) are applied to various industries such as building automation, medical, security, intelligent agriculture, and disaster monitoring. A WSN consists of hundreds to thousands of tiny sensor nodes that perform monitoring tasks. A small sensor node has a limited amount of internal memory and energy resources. Sensor nodes are used to detect a variety of data in specific environmental areas. As a result, WSN should be energy efficient. Sensor nodes are vulnerable to false report injection attacks because they are deployed in an open environment. A false report injection attack consumes the limited energy of a node more quickly and confuses the user. CFFS has been proposed to prevent such an attack using a method of en-route filtering false reports by dividing nodes into clusters. However, the CFFS scheme is vulnerable for repeated false report injection attacks. In this paper, we propose an approach to prolong the WSN lifetime by adjusting the dynamic security threshold value and using a fuzzy logic-based key redistribution selection of cluster head nodes. The proposed method increases the detection rate for repeated false report injection attacks by adding the additional key distribution phase in the existing method. The experimental results show that the energy efficiency of the proposed method was increased by 40.278%.

KEYWORDS


1. INTRODUCTION

A wireless sensor network (WSN) consists of a number of sensor nodes and several base stations (BS) and transmits data through cooperation between the nodes. Sensor nodes in a WSN that sense events are useful in various industries (medical, industrial, military, Internet of Things, etc.) [1-3]. A sensor node is composed of a CPU for data processing, memory for storage, a battery for operation, and a transceiver for data transmission to other nodes [4]. Sensors configured in the sensor node sense based on a preset algorithm. In general, a tiny sensor node has limited resources because it consists of small devices. In addition, it is difficult to replace WSNs after they are deployed. Therefore, sensor nodes should be energy-efficient. After the nodes are deployed, they are exposed to a hazardous region. For this reason, node defense against a variety of application attacks is also considered and should be cluster-based to efficiently utilize the energy of the sensor nodes. Our approach attempts to communicate with the Cluster-based False Data Filtering Scheme (CFFS) [5], which configures nodes into clusters to defend against well-known false report injection attacks and transmit reports securely to the BS. This method allows the network administrator to set the security strength by setting a pre-security threshold. However, it does not consider repeat attacks of false report injection attacks and wastes energy on nodes with fixed security thresholds.
We propose a method to prolong the lifetime of a network and improve the detection probability of false data reports through additional key distribution phase and cluster-based dynamic threshold management. The proposed method improves the lifetime of the network through the selection method of fuzzy rule-based additional key distributions and dynamic security threshold values according to the situation. Experimental results show that the proposed method increases the energy efficiency by an average of 40.278% compared to the existing method, and the filtering probability increases by 30.919% on average.

The paper is outlined as follows. In Section 2, we describe related research, false report injection, cluster-based false data, and the fuzzy rule-based system. In Section 3, the proposed method is described in detail. In Section 4, we compare the performance of CFFS and the proposed method. Finally, Section 5 is the conclusion.

2. RELATED WORKS

This section introduces the false report injection attack that is a threat to the sensor network and describes the cluster-based false data filtering scheme, which is a network security protocol.

2.1. False Report Injection Attack

Figure 1 shows a false report injection attack situation. The sensor field can be accessed by anyone and the node is easily damaged by an adversary. The adversary collects deployed sensor nodes in the WSN and can easily paralyze the network using the limited resource characteristics of the sensor nodes. Well-known attack methods are as follows. 1) Capture the content of the sensor nodes. 2) Generate false reports using random data. 3) Inject the generated false report into the network. 4) False reports consume unnecessary energy through communication between nodes. 5) The incorrect report reaches the BS and informs the network manager of the wrong
event, which confuses monitoring. In particular, this attack is dangerous because repeated attacks cause the network to become paralyzed. Research such as statistical en-route filtering scheme (SEF), probabilistic voting-based filtering schemes (PVFSs), interleaved hop-by-hop authentication (IHA), and CFFS using en-route filtering methods have been proposed to defend this attack [6-8]. The SEF scheme divides the key to be distributed to the nodes into partitions. As a result, the message authentication code (MAC) is included in the report to enhance security. PVFS probabilistically filters out false positive and false negative reports using the MAC. The IHA scheme verifies the report between the nodes and the BS via pairwise MAC. In CFFS, nodes are clustered to improve energy efficiency and filter false reports through cluster head node validation.

2.2. Cluster-based False Data Filtering Scheme

![CFFS Overview](image)

Liu, Zhixiong, et al. proposed CFFS for efficient use of energy of nodes [5]. Since the nodes are organized into clusters, the load of the CH is shared by the nodes in the cluster [9]. This reduces the communication overhead of one forwarding node and efficiently uses the energy of the nodes. It also improves network scalability and error detection accuracy, as shown in Figure 2. The CFFS scheme consists of 5 phases as follows: pre-deployment and bootstrapping, distributed key assignment, report generation, en-route filtering and sink verification. In the first step, CFFS divides nodes into the cluster and calculates the burden value according to the number of nodes distributed in the cluster. Then, the CH nodes calculate the number of keys to be transmitted to the upper node according to the burden value and transmits the calculated number of keys.

When this information is transmitted to the BS, the key assignment phase proceeds. In the distributed key assignment phase, all cluster head nodes distribute the keys according to the burden value in their clusters up to the BS nodes. In the above step, the distribution of keys is determined according to the key distribution threshold value. When an event occurs after key distribution, a report is generated from a CH node, which is a delegate node in the cluster. The report is generated as follows, including the MACs that interact with the member nodes.
R :\{e; S_1, S_2, \ldots , S_i; M_1, M_2, \ldots , M_t\}

If the node receives an event from the CH, it generates the MAC by encrypting the contents of the event using the key of the node. The report includes event contents, information of nodes (S_1 to S_t), and MACs (M_1 to M_t) generated at the corresponding node. As indicated above, the number of MAC included in the report depends on the pre-established security threshold. The more MACs in the report, the higher the security, but the larger the report size, the bigger the report transmission energy. The generated security report is transmitted to the BS. During the process, the intermediate nodes that have received the report use the MAC included in the report to judge the false report as pseudo-code as follows.

0. If there are less than t \{S_v, M_v\} tuples in R, CH_i drops R;
1. If the t node IDs \{S_v, 1 \leq v \leq t\} do not belong to the same cluster, CH_j drops R;
2. If CH_i has one key K \in \{K_v, 1 \leq v \leq t\}, it re-computes M = K(e) and checks whether the corresponding M_v is the same as M. It drops R if they are not the same.
3. If an in-cluster node S_u has one key K \in \{K_v, 1 \leq v \leq t\}, CH_j sends (e, S_v, M_v) to S_u. S_u verifies M_v as in step 3 and sends the result to CH_i.
4. If CH_i receives more than one failed verification results from in-cluster nodes during a time period η, it drops R.
5. Otherwise, CH_i sends R to its upstream node.

Pseudo 1. En-route Filtering Process

Pseudo 1 represents the en-route filtration process. If the forwarding CH node has the same key, it checks the corresponding MAC of the report. If there is no abnormality, it transfers to the next upstream node and discards the report if there is any abnormality. If the report is sent to the base station, the base station examines all the MACs, and if there is no abnormality, it is transmitted to the user using the external communication network.

2.3. Fuzzy Logic System

A fuzzy logic system generates a specific rule to represent situations with inaccurate states in a multidimensional manner [10]. The fuzzy system makes inferences about the ambiguous situation through rules. It is difficult to construct an adaptive system because the environment of a wireless sensor network is difficult to express in a mathematical model and continuously changes. Since the existing system has a static structure, it is not suitable for a system that expresses the environment flexibly depending on the situation. It is difficult to determine appropriate settings such as the nature of the attack, energy, and distribution, especially if the network administrator is not constantly monitoring the network environment. Fuzzy systems effectively solve these problems through inference systems [11]. A fuzzy system consists of rules and membership functions for making decisions on various input values.

3. PROPOSED SCHEME

In this section, we discuss the problems of the existing method and explain the assumptions and suggestions for the proposed method in detail. We also describe a fuzzy rule system for determining key redistribution applied to the proposed method.

3.1. Problem Statement
CFFS catch false report injection attacks by cooperative verification between cluster head nodes. CFFS not only prolongs the lifetime of the nodes with the pre-detection function but also filters false reports to prevent false alarms. However, if the adversary repeatedly attempts to attack, energy is quickly exhausted to the cluster until it is filtered. In addition, the unsuitable security thresholds creating unnecessary energy consumption in a secure area also results in wasted energy in the nodes. Therefore, the report size should be adjusted to suit your network environment. In the proposed scheme, a new security boundary value is derived and set using the environment information of the cluster.

3.2. Assumption

In CFFS, it is assumed that all nodes are distributed at random locations as high density. The BS collects the information from all nodes for fuzzy system [12]. Routing paths of nodes are constructed by a Directed Diffusion and Hill Climb method [20-21]. We assume that when an event occurs in a cluster, all the member nodes in a cluster are able to sense it simultaneously. We assume that the report will not be attacked during the route setup process after the deployment phase. Deployed nodes are considered to satisfy the minimum-security requirement. The BS knows the situation of each cluster through verification messages. We modeled nodes based on the Micaz mote [13]. The BS has the ability to operate the fuzzy system and assumes that the energy is infinite.

3.3 Detailed Proposed Scheme

![Proposed Scheme Overview](Figure 3 Proposed Scheme Overview)

Figure 3 shows the scenario of the proposed scheme. The proposed method provides two advantages over the CFFS. The first advantage is to perform early repeat false injection attack detections by applying the additional key distribution method using the fuzzy system. If an adversary compromises a node and injects it into the network, the report will be transmitted until it encounters a CH node with a key that matches the key inside the compromised node cluster. In the worst case, it reaches the BS. In this case, all nodes on the routing path from the point of attack to the BS are exposed to the risk of energy exhaustion. However, in the proposed scheme, the forwarding node that detected the false report sends the key redistribution request message to
the BS, like Circle 1 in Figure 3. The BS receiving the request determines whether to distribute the additional key through the fuzzy system. If the output is key distribution "ON", it transmits the additional key to the next hop of the damaged node using the stored cluster route information. If not, the current state is maintained. The details of the fuzzy system will be discussed in Section 3.4. Therefore, the proposed scheme filters on the next hop of the compromised cluster through the additional distributed key. A Saved power consumption of the proposed scheme expressed as follows:

\[
\text{Saved Energy (} \mu J \text{)} = \sum_{i=1}^{NA} E_{\text{trans}} \times H_{SB} + (E_{\text{rec}} + E_{\text{ver}}) \times (H_{SB} - 1)
\]

NA is the number of Repeat False Report Injection Attacks. \(E_{\text{trans}}\) refers to the transmission energy consumed at the node. \(H_{SB}\) means the number of hops up to the BS from the location (source node) of the generated report. \(E_{\text{rec}}\) refers to the reception energy that occurred at the next hop node (upstream neighbor node). \(E_{\text{ver}}\) means the energy consumed during the verification process. The verification probabilities depend on the deployment environment of the nodes. \(E_{\text{rec}}\) and \(E_{\text{ver}}\) multiply by one less than the hop count because BS is not included in the network node energy. For example, if the Micaz mote nodes are deployed in the network and a 30byte report is generated at a distance of 21 hops from the BS, \(E_{\text{trans}}\) is 16.25 \(\mu J\) * 30 bytes and \(H_{SB}\) is 20. The energy consumed by the nodes will be described in detail in Chapter 4. In the above equation, one is reduced in \(H_{SB}\) because filtering a false report can be done at the upstream neighbor node of the report generation node. Assuming that the probability of nodes verifying the report is 30%, the proposed scheme saves up to 17,325 \(\mu J\) about one event.

The second advantage is node energy saving in a secure region where no attack occurs. As shown in the upper part of Figure 3, the sensor node normally detects the event, generates a report, and transmits it to the BS. However, the report size depends on the security threshold preset by the network administrator. In the proposed scheme, a new security threshold value for the cluster environment is derived using the information of the cluster and transmitted to the corresponding cluster node to provide different security strengths for each cluster. In other words, the proposed method manage energy efficiently according to the situation in each region. Moreover, that reduces transmission energy by adjusting the size of the report by the new threshold value according to the situation through the above steps and improves the detection ratio of false reports by improving the security in the dangerous area.

### 3.4. Fuzzy-Rule System for Proposed Scheme

The proposed method uses the fuzzy rule-based scheme to determine whether to redistribute keys, thereby saving not only early detection of false reports but also energy consumption via forwarding nodes. The BS periodically monitors the network status and decides whether to redistribute keys based on the three parameters. Depending on the nature of the fuzzy input, the decision to redistribute the key depends on the input value being monitored. The nodes selected for key redistribution store additional keys to increase the security of compromised nodes.

- False Traffic Ratio (FTR) = \{ Very_Low (VL), Low (L) , Mid (M), High (H) \}
- Density (DE) = \{ Low (L) , Mid (M), High (H) \}
- Residual Energy (RE) = \{ Low (L) , Mid (M), High (H) \}
- Additional Key Distribution (AKD) = \{ ON, OFF \}

False Traffic Ratio (FTR): The higher the FTR value, the more likely it is that a compromised node will perform repeated attacks. This should be considered, as report damage may occur due to radio interference.
Density (DEN): Density is the number of members in a cluster. This means that the CH is internally replaced by the LEACH mechanism without distributing the additional key, thereby eliminating the damaged node [9]. Therefore, additional key distribution processes should be considered depending on the density.

Residual Energy (RE): Residual energy is the residual energy of CH node. The proposed scheme has additional communication overhead for data broadcast. If the remaining amount of CH is low, the node needs more energy saving, so it will not be performed by the node.

The fuzzy system uses the center-of-gravity method for defuzzification and the Mamdani-type inference method [14-15]. The proposed method may not be performed even if the continuously monitored FTR changes according to the known node distribution knowledge. As shown in Figure 5, there are three membership functions in the proposed scheme. The parameters in each membership function were not optimized. Fuzzy optimization is outside the scope of this paper because this paper has a goal to construct a dynamic system that determines key redistribution according to the network environment. We plan to optimize the membership functions using a
genetic algorithm (GA) later [16]. (a) FTR is calculated by the BS and based on the ratio of false reports to normal reports. (b) The density is determined by the knowledge of the routing path that the nodes are made after deployment. (c) The residual energy is determined by the energy of the node being monitored periodically. The output of the proposed fuzzy system is set to "ON" or "OFF" depending on the input variables contribution as shown in Figure 4. Table 1 shows some of the rules of the proposed fuzzy system.

Table 1. Key re-distribution fuzzy if-then rules of the proposed scheme.

<table>
<thead>
<tr>
<th>Rule No.</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RE</td>
<td>DE</td>
</tr>
<tr>
<td>0</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>1</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>8</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>9</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>23</td>
<td>M</td>
<td>H</td>
</tr>
<tr>
<td>24</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>31</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td>32</td>
<td>H</td>
<td>H</td>
</tr>
</tbody>
</table>

4. PERFORMANCE EVALUATION

In this section, CFFS and the proposed scheme are compared in various aspects and the experimental environment parameters are introduced. We evaluated the network energy consumption and filtering probability according to the attack ratio and show the energy efficiency improvement rate according to the initial security threshold setting.

4.1. Experimental Environment Conditions

Table 2. Environment Parameters for the Experiment.

<table>
<thead>
<tr>
<th>parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Environment</td>
<td></td>
</tr>
<tr>
<td>Field Size</td>
<td>300 m × 300 m</td>
</tr>
<tr>
<td>Number of Nodes</td>
<td>3,000</td>
</tr>
<tr>
<td>Cluster Head Nodes</td>
<td>100</td>
</tr>
<tr>
<td>Number of Events (Occur Randomly)</td>
<td>3,000</td>
</tr>
<tr>
<td>Node Transmit Range</td>
<td>50 - 75 m</td>
</tr>
<tr>
<td>Cluster Maximum size</td>
<td>30m × 30m</td>
</tr>
<tr>
<td>Transmit Size</td>
<td></td>
</tr>
<tr>
<td>Report Size</td>
<td>20 – 30 byte</td>
</tr>
<tr>
<td>MAC Size</td>
<td>1 byte</td>
</tr>
<tr>
<td>Verification Report Size</td>
<td>10 byte</td>
</tr>
<tr>
<td>Energy Consumption</td>
<td></td>
</tr>
<tr>
<td>Transmit</td>
<td>16.25μJ (per 1 byte)</td>
</tr>
<tr>
<td>Receive</td>
<td>12.5μJ (per 1 byte)</td>
</tr>
</tbody>
</table>
We constructed the node parameters based on the Micaz node model specification [13]. The experimental network field size is 300 × 300 m. A total of 3,000 nodes were deployed at random locations, and 100 of them consisted of CH nodes. The total number of keys used in the experiment is 200, and the System Threshold is 2. In addition, the various security thresholds are constructed to evaluate the performance of the proposed scheme. The energy required for transmission is 16.25 μJ per byte and the energy required for reception is 12.5 μJ per byte. The energy required to validate the MAC is 75 μJ [17-19]. The report size is 20-30 bytes and the MAC size is 1 byte. Events occur at random locations and false reporting injection attacks occur on randomly damaged nodes according to FTR. In this simulation, packets are not lost during the transmission phase.

4.2. Experimental Results and Analysis

![Figure 6 Energy Efficiency Versus False Traffic Ratio](image-url)

Figure 6 shows the increase in energy efficiency versus the false report ratio. In the case of a secure environment where no attack occurs, there was a 0.719% greater energy savings than in existing CFFS. This is because the security threshold is loosely controlled according to the secure region. Therefore, even if an event occurs, nodes save energy by reducing the size of the report. The reason why the difference between the proposed scheme and energy consumption is small when the false data is 0% is that initial threshold value is small. This is covered in Figure 7. The
higher the attack rate, the higher the energy efficiency in this proposed scheme. By attempting to distribute additional keys, it effectively defends against DoS attacks and by detecting false reports, the cluster lifetime also increases. In the proposed method, the energy efficiency improved by 40.278% on average compared to the existing method when roughly 3000 reports were generated at random positions.

Figure 7 Energy Efficiency Versus Initial Security Threshold

Figure 7 shows the energy efficiency based on initial security thresholds in a safe area with a report size of 20. The existing CFFS preserves the initial security threshold value, so incorrect security threshold designation in a safe area results in the rapid energy consumption of the node. The proposed scheme increases energy efficiency as the initial security threshold value increases more than CFFS. As shown above, when the initial threshold value is set to 10 cases, the energy efficiency is improved by 3.598%.

Figure 8 Filtering Performance Versus False Traffic Ratio
Figure 8 shows that the filtering efficiency increases with the false report ratio. The proposed method shows that the probability of filtering increases by 30.919% on average compared with the existing CFFS method. The proposed scheme distributes the additional key of the damaged node to the cluster (known as the dangerous zone), which enables early filtering.

5. CONCLUSIONS

WSNs have limited resources and should be energy-efficient. Liu, Zhixiong, et al. proposed a CFFS for filtering false data for security reasons. However, this scheme shows that it is vulnerable to repetitive false report injection attacks, and it is impossible to manage the security strength of each cluster. In this paper, we proposed to decide whether proper key redistribution is based a fuzzy rule system using residual energy, node density, and false traffic ratio in WSN. The proposed scheme is appropriate as an alternative for the false report repeated injection attack. We have experimentally verified that the proposed method improves energy efficiency by an average of 40.278% over the existing method. However, in the proposed scheme, the membership function applied in the fuzzy system is not efficient. In addition, it is possible to the defense of only within a limited range of attacks. In other words, the proposed method is not considered when all of the nodes in a cluster were compromised. We plan to optimize the membership functions of the fuzzy system using GA. We will solve the above problems through the membership functions and context-aware architecture techniques.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. NRF-2018R1D1A1B07048961)

REFERENCES


AUTHORS

Jung Sub Ahn received the B.S. degree in computer engineering from Kyunil University in 2016 and now doing Ph.D. degree in Department of Electrical and Computer Engineering from Sungkyunkwan University, Republic of Korea. His research interests include wireless sensor network security, modelling & simulation, IoT security

Tea Ho Cho received a Ph.D. degree in Electrical and Computer Engineering from the University of Arizona, USA, in 1993, and B.S. and M.S. degrees in Electrical and Computer Engineering from Sungkyunkwan University, Republic of Korea, and the University of Alabama, USA, respectively. He is currently a Professor in the College of Software at Sungkyunkwan University, Korea.