Sinkhole Attack Detection Scheme using Neighbors’ Information for LEAP based Wireless Sensor Networks

Jae-jin Lee  
College of Information and Communication Engineering  
Sungkyunkwan University  
Suwon 440-746, Republic of Korea

Tae-ho Cho  
College of Information and Communication Engineering  
Sungkyunkwan University  
Suwon 440-746, Republic of Korea

ABSTRACT
Intrinsic resource constraints and vulnerability to a variety of malicious attacks hinders the widespread deployment of wireless sensor networks (WSNs). One of the malicious attacks is the so-called sinkhole attack where one or more compromised nodes, pretending to be closer to the base station, disseminates a false advertisement. The event reporting nodes start forwarding their reports to these compromised nodes. These compromised nodes can take control of the network traffic, eavesdrop on real communication, and forge reports that are then forwarded to the base-station. In the localized encryption and authentication protocol (LEAP) key management protocol, compromised nodes can expose the keys to the adversary. Therefore, it is crucial to detect and evict compromised nodes instead of using a key sharing approach. In this paper, a fuzzy logic system-based method to detect the compromised nodes and to prevent sinkhole attacks is proposed. Proposed method use neighbor information (i.e., number of common neighbors and their parent node information) to detect compromised nodes. Experimental results demonstrate the validity of the proposed approach in that it provides maintained safeguards and reduces communication cost.

Keywords
Wireless sensor network, Sinkhole attack, Fuzzy logic, Genetic algorithm.

1. INTRODUCTION
Wireless sensor networks (WSNs) are comprised of a base station (BS) and a large number of sensor nodes for data gathering over short range communication. The sensor nodes have limited resources and are left unattended in an open environment. Therefore, because of the limited security and unsecure wireless communication, these sensor nodes can easily be compromised, at worst rendering the WSN out of service [1]. One of the malicious attacks is the sinkhole attack [2]. In the sinkhole attack, one or more compromised nodes pretend to be closer to the base station and disseminate a false advertisement. These false advertisements are used to compromise sensor nodes. These compromised nodes effectively take control of the network traffic to eavesdrop on real communication. Furthermore, forged reports can also be forwarded to the BS. The compromised node(s) can also selectively drop legitimate reports or control the flow of traffic destined for the BS. In addition, these nodes can facilitate launches of other types of attacks as well. Localized encryption and authentication protocol (LEAP) was proposed by Zhu in 2003 to counter these attacks [3]. LEAP uses four types of keys in each of the sensor nodes instead of a single key mechanism to address different security requirements for a message exchange scheme that employs sensor nodes. Using this protocol, the compromised node is detected and verified through these four types of keys.

LEAP can be utilized against external attack detection and prevention, however, it is vulnerable to internal attack. This limitation provides motivation for the development of a detection and prevention scheme against internal attacks. In order to solve this problem, a fuzzy logic-based system against internal sinkhole attacks is proposed.

The remainder of this paper is organized as follows. Works related to sinkhole attacks and countermeasures are described in Section 2. Background information related to fuzzy logic systems and genetic algorithms is provided in Section 3. The proposed scheme is introduced in Section 4, and the experimentation results are presented in Section 5. Conclusions and future work are discussed at the end of the paper.

2. RELATED WORK
This section explains sinkhole attacks and countermeasures.

2.1 Sinkhole Attacks and countermeasures
Sinkhole attacks are exploited by impersonating the most efficient routing path to the destination BS or acting as a forwarding node to monitor traffic [2, 4-7]. Depending upon the location (e.g., close, intermediate or far away from the BS) of the sinkhole node, the extent of the damage may vary, with nodes closer to the BS being more malicious. Since the sinkhole node can monitor and exploit neighboring node traffic, a variety of attacks are possible, including selective forwarding attacks and wormhole attacks. Various methods have been proposed to defend against sinkhole attacks [7, 8]. The method using the link quality indicator (LQI), which indicates the signal strength or quality of a received packet, was proposed in 2009. The appropriate technique consists of a detector node (DN) with general sensor nodes, which has a battery that lasts longer than a normal sensor node. The DN path change request around route request (RREQ) / route reply (RREP) operates under the assumption that every node should be checked. The DN records the minimum link cost of each node to find a node that can verify the damaged RREQ message. However, there is a disadvantage in terms of the performance of this method in that it varies depending on the number and location of the DN. The wireless sensor received signal strength indicator (RSSI) strategy proposed by Varakulisriputh in 2009 takes into account the limited resources of the network. This strategy is based on the detection of the attack technique. The RSSI technique is used to measure the distance between a transmitter and a receiver in a RF signal to measure the intensity of the signal at the receiver. A variation of this technique involves placing an extra monitor (EM) for measuring the RSSI value nodes.
between nodes. The initial base station in the route setting process sends a hello message to all sensor nodes, and the EM node monitors the traffic of all of the nodes in response to the message. The base station passes the data to match the ID of the sending node from the RSSI based sinkhole detector (RBSD), and the RBSD can be used to develop a visual geographic map (VGM) based on the response. This verifies the RREQ message to a later generation so it can be used to detect a sinkhole attack. However, there are disadvantages in that the EM sensor node placement must be configured properly, and the routing tables of all the nodes must have antennas with higher performance.

3. BACKGROUND
This section explains localized encryption and authentication protocol, fuzzy logic systems, and genetic algorithms in detail.

3.1 Localized Encryption and Authentication Protocol
LEAP key management protocol is proposed to counter sinkhole attacks for detection, prevention and limiting the damage to neighbor nodes. The packet exchanged in a sensor network may belong to different categories (e.g., control packets vs. data packets, broadcast packets vs. unicast packets, and command vs. sensor readings) for different security responses. There is no single security mechanism that is appropriate for all secure communications. LEAP uses the following four keys instead of a single key mechanism to cater to different security requirements. Each of the four types of keys and the reason for including each in LEAP is as follows:

Individual Key (IK): This unique key is possessed by each sensor and shared with the BS. The purpose of this key is to secure communication between the nodes and the BS.

Cluster Key (CK): A cluster key is shared by a node with all its neighbors for secure local broadcast messages, e.g., routing control information and message security.

Pairwise Key (PK): This pairwise key is shared with all immediate neighbors to secure communication that requires privacy or source authentication. For example, the node is used to distribute CKs to their neighbors.

Group Key (GK): This key is shared by the BS with all nodes belonging to the network. The BS encrypts a message for broadcast to the entire group.

The following are the four key generation steps used by the LEAP.

Step 1: The sensor node u is assigned a GK, \( K_u \) (initial key), \( K_m \) (the base station’s Master Key) based on its location relative to the base station.

Step 2: The base station generates an IK and \( K_m \) for each node through a random function.

Step 3: The sensor nodes u and v generate a neighbor PK through a random function.

Step 4: Sensor node u generates CK from the PK and a random function for each neighboring node.

The LEAP can detect and respond to attacks, such as Sybil and HELLO Flood. External attacks can be prevented through the newly added node. However, it is difficult to detect and respond to inside attacks because an existing key is exposed. In this paper, a fuzzy logic system based method to detect the compromised nodes and to prevent sinkhole attacks is proposed. Proposed method use neighbor information (i.e., number of common neighbors and their parent node information) to detect compromised nodes. The assumptions and fuzzy logic system details are described in Subsections 3.1 and 3.2.

3.2 Fuzzy logic system
Fuzzy logic systems are a way of thinking based on the fuzzy sets introduced by professor L.A. Zadeh to quantitatively express the ambiguity of natural language [9, 10]. Instead of black and white, a fuzzy set belongs or does not belong to any set of each subject, and each subject can be expressed mathematically by representing the extent to which the membership function belongs to the set. The membership functions of the fuzzy logic expression of knowledge use terms with linguistic meanings that can easily be understood. Because of this design, system modifications are more easily implemented, which saves maintenance costs, therefore these strategies are used in many non-linear systems where it is difficult to understand the existing design, including wireless sensor network control [11, 12]. Fuzzy logic systems use numerical interpolation to address non-linear problems in a rule-based system. The fuzzy logic system draws an inference using logical reasoning to yield a positive argument that consists of the following three basic steps and additional steps. The first step is a fuzzy matching step that calculates the degree to which the input data matches the conditions of the fuzzy rule. The second step is calculated based on the degree of match based on the rules provided in the inference step. The third step combines all the fuzzy inference rules. In the additional defuzzification step, the information is converted to a clear conclusion based on several fuzzy rules. Membership functions and fuzzy rules and parameters have been mainly produced through iterative trial and error, or they have been designed based on the knowledge of the expert. However, in most cases, the parameters of the system are not designed to guarantee that the optimum parameters of a given system are used. Neural networks [13-17], and genetic algorithms [18-21] such as learning and research optimization techniques are used to determine the optimal parameters of the fuzzy system.

3.3 Genetic algorithm
The genetic algorithm was proposed by Holland John, and is a global optimization and search algorithm based on evolution of the natural world [22]. Genetic algorithms should be used for problems where the data structure (population) must be determined as numbers or strings. Each generation, called chromosomes or objects (individual), is calculated by a suitable fitness function, which selects the more suitable objects to constitute the next generation (selection). To configure a new generation, the algorithm combines several (two or more) individuals to create a new generation (crossover). Mutations also occur in a very small population (mutation). Through this process, the population gradually converges to the optimal solution over several generations.

3.4 GA-based fuzzy optimization (GAFO)
Parameter identification for the fuzzy system may be viewed as an optimization problem on the basis of the given criteria. Therefore, the parameter search to address the problem of identifying the fuzzy system can be thought of as the optimization of GA [22-26]. In addition, the membership functions and the fuzzy rules of the system must be taken into account in the parameter identification problem because of interdependence [18]. GAFO used for identifying the
parameters of the fuzzy model is composed of the dielectric operation unit shown in Figure 1. The genetic operations unit (GOU), the simulation unit (SU), and simulation results were fit to calculate the goodness of fit based on a fitness calculation unit (FCU).

![Fig 1: Process of GAFO](image)

When the optimization procedure is started, the oil field operations unit initiates a set of objects. A simulation is executed to evaluate the fitness of the chromosomes in the object set. The fitness calculation unit in the simulation calculates the fitness of the chromosome. The genetic evolution operation unit produces the next generation through selection, breeding and mutation operations on the basis of the calculated goodness of fit of the chromosome. Evaluation in the genetic algorithm is based on a parameter used by the fitness calculation as shown in Table 1.

**Table 1. Parameter of GA**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Size</td>
<td>100</td>
</tr>
<tr>
<td>Individual Length</td>
<td>57</td>
</tr>
<tr>
<td>Selection Method</td>
<td>Tournament selection</td>
</tr>
<tr>
<td>Crossover Method/Probability</td>
<td>Uniform crossover/0.8</td>
</tr>
<tr>
<td>Mutation Method/Probability</td>
<td>Bit-wise mutation/0.02</td>
</tr>
<tr>
<td>Terminate Condition</td>
<td>100 generations</td>
</tr>
</tbody>
</table>

3.5 Genetic representation

F Figure 2 below shows a fuzzy membership function and indicates the chromosomes of the genetic algorithm used to optimize the parameters of the fuzzy rules.

![Fig 2: Genetic representation](image)

The presence of the preceding parameters, the parameter rules, and the result of the membership functions of the fuzzy rules for each input variable label denote the output of each rule.

4. PROPOSED METHOD

LEAP can detect and respond to attacks such as Sybil and Hello-flood. However, although external attacks occur through the newly added node even when the sinkhole attack can be prevented, it is difficult to detect and respond to an attack on the inside when an existing key is exposed. In this paper, the sensor node through the RREQ message verification information of the peripheral node is used in the proposed technique for detecting damage from a sinkhole attack. In order to determine damage from a sinkhole attack, a number of common neighbors between the two nodes in the proposed method are used as the system inputs. These common neighboring nodes include a particular node including the parent node, the node, the average distance of a node, and the parent node of the common neighboring node. The input variable is a non-linear problem that cannot be explained by the sum product. To address this non-linear problem, including the inaccuracies and the uncertainty data, a fuzzy logic system (Approximate Reasoning) is used. In this section, the membership functions, rules for the input and output variables of the fuzzy logic system for the proposed method is described.

4.1 Assumptions

Static sensor nodes are randomly deployed in a sensor field, and each node does not know in advance the neighboring nodes’ information. LEAP uses a key management protocol. The attacker can impersonate a parent node (or forwarding node) of the neighbor nodes. The initial routing setup is assumed to be secure from this attack. In the initialization phase, the distance information is shared among all neighbors and the BS.

4.2 System overview

The proposed method is performed in each sensor node, and the neighboring node performs $v_j$ to verify the node that transmitted the RREQ. The following Figure 3 show the process of the proposed method.

![Fig 3: Process of proposed method](image)

The following Table 2 show the notations used in the operation of the proposed scheme of the proposed method.

**Table 2. System notations**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>System verification node $i$</td>
<td>$V_i$</td>
</tr>
<tr>
<td>A route requesting node</td>
<td>$R$</td>
</tr>
<tr>
<td>Common neighbors of $V_i$ and $R$</td>
<td>$CV_{i,R}$</td>
</tr>
<tr>
<td>Number of $CV_{i,R}$'s parent node is $k$</td>
<td>$N(CV_{i,R})$</td>
</tr>
<tr>
<td>Node $i$’s distance to the BS</td>
<td>$D_{i BS}$</td>
</tr>
</tbody>
</table>

4.3 System inputs

The following variables are input for the proposed method and are described in detail below.

- $N(CV_{i,R})$: The first input of the proposed method $N(CV_{i,R})$ is the number of common neighbors between the system verification node $V_i$ and the verified target $R$. Since the sensor node is randomly arranged, the number of sensor nodes in each cluster will be different. The proposed method generally shows high performance, but when a sufficient number of sensor nodes in each cluster show
using a genetic algorithm, and Table 3 shows some of these rules.

**Table 3. Fuzzy rules**

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Input1</th>
<th>Input2</th>
<th>Input3</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Less</td>
<td>Less</td>
<td>Medium</td>
<td>Normal</td>
</tr>
<tr>
<td>12</td>
<td>Medium</td>
<td>Medium</td>
<td>Short</td>
<td>Comp</td>
</tr>
<tr>
<td>19</td>
<td>Many</td>
<td>Less</td>
<td>Medium</td>
<td>Normal</td>
</tr>
<tr>
<td>24</td>
<td>Many</td>
<td>Many</td>
<td>Short</td>
<td>Comp</td>
</tr>
</tbody>
</table>

For fuzzy rule 1, if \( N(C_{V,R}) = \) Less, \( N(C_{V,R}^{P,R}) = \) Less, \( \sum_{i=1}^{n} |d_{r,1}^{P} - d_{r,1}^{S}| = \) Medium, then the value is determined as normal. On the other hand, if \( N(C_{V,N}) = \) Many, \( N(C_{V,R}^{P,R}) = \) Many, \( \sum_{i=1}^{n} |d_{r,1}^{P} - d_{r,1}^{S}| = \) Short, then it is determined as a false RREQ message, which is used to determine the node related to the sinkhole.

**5. PERFORMANCE EVALUATIONS**

This chapter describes the experiments used to detect the sinkhole attack performance of the proposed scheme. Table 4 below shows the parameters for the test environment.

**Table 4. Parameter of the experiments**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field size</td>
<td>500x500(m²)</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>700</td>
</tr>
<tr>
<td>Radio range</td>
<td>55(m)</td>
</tr>
<tr>
<td>Number of experiments</td>
<td>100</td>
</tr>
<tr>
<td>Number of RREQ messages</td>
<td>75</td>
</tr>
</tbody>
</table>

**5.1 Cost analysis**

Hardware must be used for sensor nodes [27]; therefore, the energy consumption in these systems must be considered. To calculate the required clock cycle times for various operations, an estimation method given in the literature [28] is employed. The proposed system employs triangular fuzzy sets as described in an earlier work [29], and there is a need for a total of 1374 clock cycles. For statistical en-route filtering (SEF), a message authentication code (MAC) using the 32bit RC5 algorithm requires a total of 372 single-bit addition operations, and the energy consumption is approximately 75 uJ [30, 31]. This translates to approximately 0.2016 uJ per 1 clock cycle consumed. Delivery of the SEF report requires 747.5 uJ to accommodate 46 bytes. On the other hand, the proposed fuzzy system has 3 inputs, 3 fuzzy sets for the input, 1 output, 2 fuzzy sets for the output and 11 fuzzy rules; these are expected to require approximately 1374 clock cycles and consume about 277uJ. The proposed method was also used to verify the RREQ, and this required a much lower rate than message passing. Therefore, the fuzzy system may be generally sufficient for handling the sensor node. In addition, the proposed method can save energy used in the neighbor node when receiving the transmitted RREQ detection (which determines if the communication node is compromised) because no additional node or BS communication is required to detect an attack.

**5.2 Experimental results**

Experiments were conducted to assess the performance of the proposed technique and to verify the RREQ messages from the sensor nodes arranged in any position. To validate a single
RREQ, each single event experiment was initiated at a desired position after verification, and the system delivered reports to the base station indicating which sensor node was detected. This was carried out a total of 100 times for accurate measurement of each event. The sensor node determined as the source of a sinkhole attack is identified through the proposed method, and the path is reset except for the node. Further, comparison the performance of the proposed method [32] to the existing sinkhole attack detection techniques. Figure 5 shows the proportion of false RREQ messages (sinkhole attack ratio or SAR) for proposed method (PM) and exists method (EM), which are, respectively, the proposed scheme and a traditional sinkhole attack detection method.

Figure 5: Compromised node detection ratio

In EM, if the number of nodes required for the local detection report is not reached due to the sinkhole attack, then the performance will decrease. Further, if the node has not detected a sinkhole attack using the proposed method with existing information from the peripheral node, the PM can have a negative impact that also causes the performance to decline. If the SAR is not more than 60%, then the performance is nearly the same as an EM SAR, which shows a maximum of 2% even when the SAR is 60% or more. Figure 6 below shows the false positive ratio (FPR) of the EM and PM.

Figure 6: False positive ratio

The FPR indicates the ratio of attacks that are falsely detected. If the EM FPR for detecting compromised nodes based on the local report is received, the FPR does not substantially increase. The FPR for PM did not increase as attacks increased because the conventional SAR detected the information in the peripheral node. The maximum FPR in PM was about 2.2%. Figure 7 below shows false negative ratio (FNRs) of the EM and PM.

Figure 7: False negative ratio

FNR represents the ratio of the failure to detect an attack. If the SAR is less than 60%, there is not a significant difference between the FNR of PM and the FNR of EM. Overall, the detection performance of the PM was within about 2% of that of the EM. However, the PM communication cost is significantly lower than that required for EM. This is because the neighboring node receiving the RREQ sent by the compromised node does not require a different communication node for detecting an attack. The following Figure 8 shows the cost of communication required for detection of the position of the damaged node in PM and EM.

Figure 8: Communication cost

The EM communication amount required for detection depends on the position of the node based on the information collected from the sensor nodes. However, the PM communication cost is based on the location of the damaged node; therefore, communication does not require any communication to the base station due to the verification of
the received RREQ by the neighbor node. This results in less energy consumption in PM than EM.

6. CONCLUSIONS

In this paper, a fuzzy system based on sinkhole attack detection techniques to detect damage from a sinkhole attack on the sensor node is proposed. LEAP is an existing technique that can respond to various attacks, including the sinkhole attack, through the key exchange method. However, attacks aimed at the key present a difficult problem. The proposed method determines whether or not damage to the input of the fuzzy system has occurred by using three pieces of information relating to neighboring nodes. The proposed method showed a detection rate of about 93% even when more than 40% of the nodes were used to minimize the messaging required for detection in order to save energy. This arrangement of nodes to detect an attack in an existing communication process is not required, but it has the advantages of using other nodes for the antenna and verification of performance. The proposed method requires less energy to run the fuzzy system than general sensor nodes, but this is not significantly different than the energy required for report delivery. In addition, a RREQ message is employed because the number of runs of the proposed method is expected to be relatively small compared to the number of delivery reports; therefore, the proposed method does not have a significant effect on the operating time of the entire network.

7. ACKNOWLEDGMENTS

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-2015R1D1A1A01059484).

8. REFERENCES


