DETECTING SYBIL ATTACKS USING HETEROGENEOUS TOPOLOGIES IN STATIC WIRELESS SENSOR NETWORK

SOHAIL ABBAS¹, MUHAMMAD HAQDAD² AND SALMA BEGUM¹, MUHAMMAD ZAHID KHAN²

¹Department of Computer Science, University of Sharjah, College of Science, Sharjah, UAE
sabbas@sharjah.ac.ae

²Department of Computer Science & IT, University of Malakand, KPK, Pakistan
haqdad5050@gmail.com; salmaswat2012@gmail.com; mzahidkhan@uom.edu.pk

ABSTRACT

Wireless Sensor Network (WSN) is composed of few to several hundred nodes that coordinate to perform a specific action. Data is propagated in multihop fashion from sources to sink(s). Security is an important issue in WSNs, especially when they are used to protect or monitor critical situations. The WSNs require a unique identity per node in order to function properly. However, an attack called Sybil attack violates this one-to-one mapping of identity and node; hence, a Sybil attacker controls multiple identities on a single physical node. The Sybil attacks are serious threats to the correct functioning of WSNs. Thus, effective mechanisms for security of WSNs became vital. Various techniques have been proposed in the literature that use received signal strength in order to thwart and detect Sybil attacks. For coping with signal fluctuation and improved accuracy, the proposed techniques incorporate extra hardware, such as GPS or directional antennae, or communication overhead, such as periodic beacons and/or node collaboration. In this paper, we propose a detection mechanism for Sybil attacks without using the extra hardware or imposing overhead while trying to achieve the same level of detection accuracy. We start off by proposing a pre-planned nodes deployment strategy, as a first line of defence, to prevent these attacks. The resulted topology will be Sybil-free that will also help in improving the detection of these attacks. We also propose a signal strength based detection mechanism for the counteraction of these attacks in static WSNs. Our proposed technique does not require using any sort of extra hardware, like directional antennae or GPS. The simulation based evaluation of the scheme demonstrates good detection accuracy, i.e. high true positives with low or no false positives.


1. INTRODUCTION

A Wireless Sensor Network (WSN) is a type of network constituted by a large number of spatially distributed, battery enabled tiny devices, called sensors; where each sensor is capable of detecting physical phenomena, like heat, light, pressure, etc. Sensors are mostly resource constraint and the protocols designed for them usually focus on energy efficiency. Each sensor node conveys the sensed information to the sink node or base station for further processing. WSNs are formed in self-organized manner constructing ad hoc topologies that do not depend on any pre-installed infrastructure. Hence, these networks are deployed in environments where there is no pre-installed network infrastructure or there is hard to install one. Usually the sensor nodes are used to collectively undertake a mission or collectively observe real-world environments, for instance flood detection, natural resources conservation, manufacturing productivity improvement, habitat monitoring, homeland security, battle fields monitoring, etc [1, 2].

Today, WSNs play an important role in the countries critical infrastructure. These networks are an integral part of such infrastructures. For example, smart cities [3], smart grid, smart homes [4], intelligent transportation systems [5] are considered temporary infrastructure systems that connect the world more than ever perceived. The entire vision of the idea of this infrastructure is based on a concept called Internet of Things (IoT).
Sybil attacks. In these schemes, a Sybil attacker can also be disturbed by detection systems that use reputation and trust to give false impression of being in more than one place. Sybil attackers disrupt the geographic process in order to manipulate the final outcome. Various protocols proposed for WSNs are based on voting. A Sybil attacker can employ its virtual identities to disrupt the aggregated outcome. Various protocols proposed for WSNs are based on voting. A Sybil attacker disrupts the voting process by rigging in the voting process in order to manipulate the final outcome. Some of the protocols use geographic based routing. Sybil attackers disrupt the geographic based routing by utilizing their virtual identities to give false impression of being in more than places simultaneously in the network. In misbehavior detection systems that use reputation and trust to counteract selfish nodes can also be disturbed by Sybil attacks. In these schemes, a Sybil attacker can utilize its virtual identities in order to increase its own reputation or trust and also maliciously defame the benign nodes. Moreover, if an identity of a Sybil attacker detected and blocked from the network due its bad reputation or trust, it can evade the detection by discarding the current identity and creating a new one thereby whitewashing all the bad reputation [11].

In order to tackle the issue of Sybil attacks, various approaches have been used in the current literature. For instance, cryptographic based authentication [12], resource testing, and position verification. The cryptographic based schemes are not quite suitable of WSNs because of their costly implementation and maintenance whereas resource testing schemes are usually based on unrealistic assumption regarding the resources owned by a single node. These issues will be discussed in Section 2. Position verification on the other hand is considered to be a viable solution in WSNs in which received signal strength (RSS) is used to detect and counteract Sybil attacks [13]. RSS based schemes are considered more promising because of being lightweight and it is intrinsic property of the nodes. However, the main issue in RSS based detectors is its low detection accuracy. Various authors tried to improve the detection by using directional antennae and periodic broadcast of beacon messages for location updates [14]. Since RSS fluctuates with time that reduces the overall detection accuracy, some authors use node collaboration to improve the detection, such as [15, 16]. In collaboration, nodes share the RSS gathered from the 1-hop neighbors for improved detection. The main problems in the collaborative detection are the increased overhead and the establishment of trust among nodes. In this work we tried to eliminate the use of the above mentioned techniques while achieving almost the same level of accuracy. More specifically, we propose a novel Sybil attack detection scheme based on RSS without using extra hardware such as, GPS or directional antennae, generating periodic beacons, or relying on node collaboration. We devise a Sybil-free topology in which nodes will be deployed pre-planned such that the distance between any two nodes in every neighborhood should not be the same. So, each neighbor will be heterogeneously deployed apart from every other neighbor. We conducted real world testbed of Java Sunspot sensors [17] in order to determine the real fluctuation that exists in RSS. The fluctuation helped us in tuning the threshold for node sparseness. After the deployment phase, RSS will be used to detect Sybil nodes. The detector module will analyze the
RSS and will detect Sybil attacks if messages are received from more than one identity while transmitting from the same location. The advantage of our proposed technique is that it does not need collaboration among nodes. The simulation results show that our proposed scheme produces high true positives and low or none false positives.

The rest of the paper is organized as follows. In Section 2, we briefly review the current proposed schemes for Sybil attack detection. In Section 3, the design rationale of the proposed work is explained in detail. In Section 4, we present the simulation based performance evaluation of our scheme. The paper is concluded in Section 6 where we also discuss our future work.

2. RELATED WORK

Various techniques have been proposed in the existing literature for Sybil attack detection. The more efficient among those is cryptographic based authentication [12]. In these techniques, each identity is bound by a digital certificate which is obtained from a centralized trusted third party (TTP). This is an efficient Sybil attack prevention technique; however, it suffers from various problems. One of the main problems in these schemes is that they rely on centralized or semi-centralized TTP which make it unsuitable choice for WSNs. They required heavy computation for cryptographic operations which again make it not appropriate for WSNs due to their resource constraint devices. Other issues include, certificate issuance, revocation, management, and its distribution. The TTP must be available all to time to all the nodes in the network. This final requirement makes it a challenging task in order to implement this technique in WSNs. Another category of schemes proposed for Sybil attack detection is resource testing. However, these schemes are based on unrealistic assumption of resources, such as hardware, being limited. Some of the resources, such as hardware, are very cheap now-a-days and attackers can easily get them. For example, installing multiple network interface cards, adding processing power or memory to the existing systems is not a big issue now-a-days. A promising category of solutions is the position verification based Sybil attack detection mechanisms. In these schemes, each identity is assumed to be bound by a distinct location, i.e. messages received from two identities belonging to the same location would imply Sybil attack. Researchers use received signal strength (RSS) measurements (instead of GPS) for position verification [13] in WSNs. The RSS based schemes are considered to be more promising because of being lightweight and it is intrinsic property of the nodes. Some of them are given below.

Demirbas et al. [16] used Zhong’s algorithm [18] of localization for the detection of Sybil nodes in WSN. The scheme relied on sharing of the ratios of RSSI. At least four nodes were required to collaborate to detect and localize Sybil attackers.

Jiangtao et al. [19] proposed Sybil attack detection mechanism for clustered WSN using RSSI. Both cluster heads and cluster members used RSSI to verify the status of each other for possible Sybil attack detection. The scheme needed node cooperation in order to detect Sybil cluster heads.

Shaohe et al. [15] proposed a Cooperative RSSI based Sybil Detection (CRSD) scheme for WSN. Each node collects RSS and calculates distances to other nodes and further shares the distances with its neighbors for the attack detection. Based on this information each node constructs suspected group of those identities having the least difference in RSSI. After a time period if the RSSI difference of two nodes still persists to be less than the threshold, the nodes are considered as Sybil ones.

Suen et al. [20] proposed a scheme for Sybil attack detection. The author assumed that each node must equip with GPS and directional antennae. The scheme used node cooperation and triangulation [21] to detect and localize the Sybil identities. The author assumed trusted nodes in the network.

M. A. Jan et al. [22] proposed a detection mechanism for Sybil detection in centralized clustered WSNs in order to conduct Sybil-free cluster head selection process. The scheme used collaboration of at least two high energy nodes for the analysis of the neighbors’ RSS.

N. Alsaeedi et al. [23] proposed Sybil attack detection scheme for clustered WSN in which network is divided into group of nodes called clusters administered by a cluster head. The
detection is based on position and trust level of nodes; however, it has not been specified how the position will be calculated, such as using GPS or some other localization mechanism. The author assumed only a single hop scenario, no multi-hop communication was considered.

M. A. Jan et al. [24] proposed Sybil attack detection scheme for WSN, specifically for wildfire monitoring. The author assumed semi-centralized clustered based WSN. The detection scheme worked in two phases. In the first phase attacker are detected using high energy nodes and in the second phase if still Sybil nodes are created then it will be detected by the two base stations.

P. Bhatia et al. [25] proposed a detection mechanism for evil-twin attack, also called simple Sybil attack. In this attack, it is assumed that the Sybil attacker is using only one stolen identity. The scheme used four-sector directional antennas for the fake identity detection. A localization algorithm, called HPB (Hyperbolic Position Bounding) [26], was used to localize the attacker.

As we discussed, RSS is an attractive choice to be used for Sybil attack detection; despite its few meters detection inaccuracy. The main attempt of authors in the literature is to increase the detection accuracy which is basically caused by the RSS fluctuation. For example, authors [20] and [25] used directional antennae whereas [15, 16], [19, 20], [22] employed node collaboration without discussing the issue of trust establishment among nodes. Similarly, in [23] it is not clear how the author gathered location related data but they used one-hop communication (not multi-hop). Overall, the above schemes produced an overhead in terms of communication and/or extra hardware for the detection accuracy improvement. In the next section, we propose a Sybil attack detection mechanism that avoids the above mentioned overheads while achieving the same level of detection in WSNs.

3. THE PROPOSED TECHNIQUE

In this section, we explain our design rationale of our proposed scheme. Our proposed design consists of two steps.

Step-I: We follow the pre-planned sensor nodes deployment technique in which we propose that nodes should be deployed in such a way that they be heterogeneously fall apart. This implies that all distances in the two hop region will be significantly different. By significant, we mean that the locations of two transmitters should be fairly distinguishable to an arbitrary receiver. As shown in Figure 1, the distance $d$ between node $A$ and each of its neighbor, i.e. $B$, $C$, $D$, and $F$, should be different. The $\gamma$ is the random number and its value falls in the range of $[\alpha, \beta]$ and $\alpha \neq \beta$.

Since, there is considerable amount of fluctuation in the RSS that obscure the degree of distinguishability among nodes and hence, low detection accuracy. For this purpose, the value of $\alpha$ will be carefully adopted because it will help in improving the degree of distinguishability. The distance $d$ across any two hops should be at least as different as $\alpha$. In Section 4, we will discuss the relationship of $\alpha$ with the RSS fluctuation.

Step-II: After node deployment, we need to propose our RSS based Sybil attack detection scheme. We assume that each node transmits with homogenous transmit power, i.e. each node transmits data at constant power. Like other wireless networks, in WSNs too, the RSS is treated as distance estimator.

Mostly RSS based systems are based on a radio communication model which states that the received power roughly decays with the $m^{th}$ power of the distance, i.e.
\[ R_s \propto \frac{P_t}{d^m} \]  

Eq-1

where \( P_r \) is the received power observed at the receiver, \( P_t \) is the transmit power at the transmitter and \( d \) is the distance between the receiver and the transmitter. The value \( m \) is referred to as the path loss exponent and its value depends on the environment, i.e. for outdoor Line-of-Sight (LoS) conditions its value is considered to be 2 and for indoor environments, its values is considered to be almost 4. If the transmit power is known, the receiver may calculate the distance.

The RSS received indicates the distance between the transmitter and the receiver. Since, we deployed nodes well apart from one another; each identity may be tagged with or represented by its distinct distance. More formally we can formulate this as follows.

Let \( B \) and \( C \) are two nodes lying at distance \( d_1 \) and \( d_2 \) from node \( A \), while \( d_1 \neq d_2 \), as shown in Figure 1. And suppose \( R_j \) be the RSS of node \( j \) received at node \( i \), node \( A \) can detect the attack according to Eq-2.

\[
\text{Detection} = \begin{cases} 
| R_A^B - R_A^C | = 0 & \text{Attack} \\
| R_A^B - R_A^C | \neq 0 & \text{Normal} 
\end{cases}
\]

Eq-2

Figure 1: Topology Of Nodes Having Different Distances

Each node will capture the RSS values along with the time of reception from its neighboring nodes and store those values in a table. The data will be furnished in the table like <Address, Rss-List <time, rss>>, as shown in Table 1. Each node will record the RSS values not only for the direct frames; but also for the overheard frames.

The number of linked records, shown in Table 1, may be adjusted according to the memory requirement. We have used \( \alpha \) to be 5 throughout our simulations. Since each node saves the record only for its 1-hop neighbors, the size of the table does not grow and also does not create burden on resource constraint sensor nodes.

The RSS received indicates the distance between the transmitter and the receiver. Since, we deployed nodes well apart from one another; each identity may be tagged with or represented by its distinct distance. More formally we can formulate this as follows.

Figure 2: (A) Free Range Sensor (B) Base Station
In order to analyze the actual variation in RSS, we conduct a real-world testbed using Sun Microsystems newly developed Sunspot sensors [17]. We used two types of Sunspot sensors, a base station that is directly connected to a laptop or desktop through a wired link and a free range Sunspot sensor, as shown in Figure 2. The sensor wirelessly transmits data directly to the base station while being static or mobile. We configure the sensor nodes to transmit packets to the base station after each 100 ms through using a radiogram connection. The base station on the other hand is configured to receive and record the RSS into a log file along with their reception times. In order to analyze the variation in RSS, we collected 1000 samples of RSS from Sunspot base station at three different positions, i.e. when sensor and base station were 1, 15, and 33 feet apart. The data distribution along with their descriptive statistics is given in the Appendix.

It is evident from the results in the Appendix that the RSS at 33feet (boundary) produces high variation, which is considered as worst case scenario or upper bound. More specifically, the variation at the boundary is 2.24 dbm, as shown in Table C-3. Using this high variation, it is crucial to determine the ensued distance from 2.24 dbm signal strength. To do this, we extend Eq-1, and follow the formulation from our previous work [10], we can obtain the expression for the received power at distance \( d \), i.e. \( P_d \) where \( d > d_0 \).

\[
P_d \text{ (dbm)} = P_{d_0} \text{ (dbm)} - 10 \log_{10} \left( \frac{d}{d_0} \right)^m \quad \text{Eq-3}
\]

where \( d_0 \) is called the reference distance, we know from our Sun Spot experiment Table C-1, \( d_0 \) is 1 foot (0.30 meter) and \( P_{d_0} \) is approximately -17 (dbm) from Table C-1. We will use Eq-3 in order to find out the received signal strength at the boundary, i.e. where \( d = 10 \) meters, as shown below.

\[
P_{10} = (-17) - \left[ 20 \log_{10} \left( \frac{10}{0.30} \right) \right]
\]

\[
P_{10} = (-17) - [20 \log_{10}(33.33)]
\]

\[
P_{10} = (-17) - [20 \times 1.52]
\]

\[
P_{10} = -47.4 \text{ dbm}
\]

Now our aim is to find out the distance for variation \( d_v \) resulted from \( \pm 2.24 \) dbm variation. The received power of 2.24 dbm variation in worst case can be computed as

\[
P_{d_v} - P_{10} = -2.24 \text{ dbm}
\]

\[
P_{d_v} = -2.24 - 47.4
\]

\[
P_{d_v} = -49.64 \text{ dbm}
\]

It is worth mentioning that 2.24 dbm can be added or subtracted to the original RSS but in worst case we added its value to the RSS. The distance \( d_v \) can be computed using Eq-3 as follows.

\[
d_v = 10 \left( \frac{(-P_{d_v} + P_{d_0}) + 20 \log_{10}(d_0)}{20} \right)
\]

\[
d_v = 10 \left( \frac{(-49.64 + 17) + 20 \log_{10}(0.30)}{20} \right)
\]

\[
d_v = 10^{(1.132)}
\]

\[
d_v = 13.5 \text{ m}
\]

In essence, we can realize from the above calculations and discussion that for 95% confidence interval (CI), the RSS can fall in the range of \([-45.16, -49.64]\), which is approximately two standard deviations or \( \pm 2.24 \) dbm. So the distance produced from this variation (in physical space) will be approximately 3.5 meters, i.e. \((d_b - d_v)\), where \( d_b \) is the mapping distance at the boundary, i.e. 33 ft or 10 meters. The value of \( \alpha \) should be greater than that of 3.5 meters.

5. SIMULATION SETUP

We use Network Simulator NS-2.35 to implement our scheme and evaluate it by using the parameters
listed in Table 2. To find out the detection accuracy of our scheme in different scenarios, is the target of this simulation study. Various attributes of the network like number of network connections, sparseness, network area, etc. have been considered for the evaluation of our scheme. The simulation results have been calculated as an average of 30 different simulation runs or random scenarios.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>1000m², 2000m², 3000m²</td>
</tr>
<tr>
<td>Pause time</td>
<td>600</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>42</td>
</tr>
<tr>
<td>Simulation time</td>
<td>600s</td>
</tr>
<tr>
<td>Maximum connections</td>
<td>25</td>
</tr>
<tr>
<td>No. of IDs per Sybil Node</td>
<td>5</td>
</tr>
<tr>
<td>Application</td>
<td>CBR</td>
</tr>
<tr>
<td>Packet size</td>
<td>64 Bytes</td>
</tr>
<tr>
<td>Transmission Range</td>
<td>250m</td>
</tr>
<tr>
<td>Topology</td>
<td>Grid</td>
</tr>
<tr>
<td>Attack</td>
<td>Sybil</td>
</tr>
</tbody>
</table>

Table 2: Simulation Parameters

5.1 Metrics

To find out the detection accuracy of our scheme in different environments we use two main metrics, i.e. True Positive Rate (TPR) and False Positive Rate (FPR). TP means a malicious node is truly detected and FP false positive means a legitimate node is incorrectly detected as a malicious one, i.e.

True Positive Rate (TPR) = detected Sybil identities/ total Sybil identities

False Positive Rate (FPR) = incorrectly detected legitimate identity/total good identity

5.2 Attack Implementation

Mostly in simulation based evaluation, implementation details are kept secret; however, we believe that it will not only promote clarity of the rationale but also guide beginner researchers about implementation process.

In NS-2, to create Sybil attackers that create and take on new identities during the live and continuous simulation is a difficult task. We adopt a work around of this by binding \( n \) number of nodes together to represent one node having \( n - 1 \) identities (because NS-2 starts node ids from 0). In order to imitate them as one node, we set up all the delays among them to be zero. In our simulation, we bound 5 nodes together; initially identity 1 is up and the other 4 nodes are off. We achieve this effect in the NS-2 by using the node “off” and “on” options in the \texttt{command()} method of the mobile node class.

![Identity Up Sequence Of Sybil Node](image)

Figure 3: Identity Up Sequence Of Sybil Node

Each node (or identity) other than the first one will awake itself after its timer expires and will be awake till the end of simulation, as shown in Figure 3. In the given figure, shaded circles represent up identities. After the first identity, \( Id_2 \) is up and added to attacker’s new identity list after \( t_1 \) timer expires, similarly for \( Id_3 \) is added after \( t_2 \) and so on.

5.3 Results Analysis

5.3.1 The effect of network sparseness:

We evaluate our scheme for different distances among nodes, i.e. node sparseness. As shown in Figure 4(a), when nodes are deployed very close to each other, they produce high false positives. This may be due to the fluctuation in the RSS. With the increase in sparseness, false positives decreases and reach to zero when sparseness becomes 6 meters. Similarly, the true positives also increase with increase in distance between...
nodes. In order to check for the optimal network dimension, i.e. the size of x-axis and y-axis. The optimal dimension is 150 as shown at which the system produces zero false positives and greater than 90% true positives, shown in Figure 4(b).

5.3.2 The effect of simulation time and connections:

In order to check the proposed scheme for convergence time, i.e. how long does it take to detect all Sybil attackers, we conducted this experiment, the result of which are shown in Figure 4. We relate this scenario with various numbers of connections because more connections mean more RSS’s in the network; hence, detection will be faster. It can be seen in Figure 5(a) that greater number of connections converges faster, i.e. at around 400 seconds all nodes were detected as compared to fewer number of connections which converges latter. As shown in Figure 5(b), number of connections has no effect on the false positives. The reason behind the zero false positives is that we have selected the optimal values for other parameters, such as network sparseness.

5.3.3 The effect of simulation time and area:

In this experiment, our aim is to check the performance of our scheme for different network areas and simulation time. As shown in 6(a), it can be observed that the 1K x 1K size network converges faster than the other two areas. At about 400 second simulation times, all the Sybil identities are detected. One reason behind this is that in larger areas (with constant node density) routes are longer and takes longer to traverse from source to destination, as a result the RSS are delayed. Here again, there is no effect of network area on false positives is found, as shown in Figure 6(b).

(A)      (B)

Figure 5(A) TPR And (B) FPR With Different Simulation Time And Connections
6. CONCLUSION AND FUTURE WORK

In WSNs, almost all the proposed protocols presume and rely on a single identity representing a node. Sybil attackers violate this identity-node connotation thereby fabricating fake identities on a single node. We have shown the proposed techniques for Sybil attack detection along with their pros and cons. The main drawback was that these schemes incorporate extra hardware and/or cause substantial communication overhead. In this paper, we proposed an RSS based Sybil attack detection technique for static WSNs. First, we proposed the strategy to deploy sensor nodes in pre-planned fashion such that no two distances are same at the 2-hop scope. Second, we setup a detection mechanism that enabled each node to effectively detect Sybil identities. For the node deployment we needed the real maximum degree of sparseness which we obtained from Java Sunspot sensors testbed where we analyzed the real fluctuation of RSS. Finally, we implemented and evaluated the performance of our scheme using NS-2. The results obtained showed that our scheme performed well in terms of detection accuracy.

One of the limitation of our scheme is that we used homogenous transmit power at each transmitter; however, it is important to design solutions for the heterogeneous transmit powers. In our future work, we will adapt and extend our scheme to cover heterogeneous transmit powers.
REFERENCES


APPENDIX

Figure C-1: Data Distribution For 1 Foot Distance

Table C-1: Descriptive Stat For 1 Foot Distance

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSS_P_1</td>
<td>1000</td>
<td>-16.9180</td>
<td>.02408</td>
<td>.76148</td>
</tr>
</tbody>
</table>

Figure C-2: Data Distribution For 15 Feet Distance
Table C-2: Descriptive Stat For 15 Foot Distance

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>RSS_P_2</td>
</tr>
</tbody>
</table>

Figure C-3: Data Distribution For 33 Feet Distance

Table C-3: Descriptive Stat For 33 Foot Distance

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>RSS_P_3</td>
</tr>
</tbody>
</table>