JCST Journal of Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777771

ESTABLISHING RELIABLE IDENTITIES DETECTION SYSTEM AGAINST THE SYBIL ATTACK IN WIRELESS SENSOR NETWORK

Dr . D. Maheshwari
Associate Professor, Department of Computer Science with Data Analytics, KPR College of Arts Science and Research, Coimbatore, India
maheshwari.d@kprcas.ac.in / maheshgkrish@gmail.com

Dr. P. Kavipriya
Associate Professor, Department of Computer Science KPR College of Arts Science and Research, Coimbatore, India
kavipriya.p@kprcas.ac.in / kavipriya.rajen@gmail.com

Abstract:

Wireless Sensor Network (WSN) is considered as distributed large scale systems utilized to monitor and acquire the huge complex information for effective data management and situation handling. Due to the complex nature and node were considered as highly resource constraint. Hence, it becomes mandatory to establish the light weight security solutions to the monitoring node to prevent from various security challenges. However, WSN need a unique, distinct, and persistent discovery method in addition to establishing the security protocols against multiple threats. Especially Sybil attacks poses a high challenging threat to the network and its information. In order to mitigate the Sybil attack and its attacker. It is considered to be significant to determine the launching behaviour of the attacker. Sybil attacker generates more than one identity in the network to launch a coordinated attack. Further attacker has capability in navigating among the nodes during the detection process and it promotes the lack of network accountability. In this paper, a novel Sybil detection approach is proposed to detect the Sybil identities of Sybil attackers on employing the unsupervised classification phenomena's for collecting the Sybil nodes evolution on the proposed network topology. Simulation analysis of the proposed model proves that it is highly efficient in exploiting the Sybil Nodes in the network on compared to traditional approaches with high detection accuracy with less false positive rate and detection speed.

Index Terms: Node Replica attack, Wireless Sensor network, Unsupervised Clustering, Distributed systems, Discrete Hidden Markov Model

Introduction:

Wireless mobile nodes are capable in dynamically organizing the network topologies for sensed data transmission to base station for future processing[1]. It allows wireless devices like Bluetooth towards seamless interface with each other without any predefined infrastructure and topology. It is high suitable during natural disasters as it helps in establishing the network for data communication effortlessly. The distinctive characteristics of wireless sensor networks is considered on basis of its dynamic topology, decentralization and resource-constrained devices which is provides numerous challenges in establishing security solutions against the attacker in the network.

Node replication attacks pose a significant threat to such networks because to the lack of centralised identity management and the necessity of a different, unique, and durable identity for each node for the viability of their security measures. Two identities imply two different nodes since this creates a one-to-one mapping between an identity and an entity, which is typically assumed either implicitly or explicitly by many protocol mechanisms. A node replica attacker has a number of techniques to harm networks [2]. For instance, a node replica attacker can interfere with location-based or multipath routing by taking part in the routing and creating the appearance that different nodes are present at various places or on node-disjoint pathways..

A replica node can impair the accuracy of reputation and trust-based misbehaviour detection techniques by boosting its own reputation or trust and lowering others' reputations or trust by taking use of its virtual identities. In wireless sensor networks, a node replica attacker can modify any node to alter the traffic and packets that make up the entire aggregated reading result. Therefore, node replica, also known as sybil assaults, will significantly affect wireless ad hoc networks' ability to function normally. Finding Sybil assaults and removing them from the network is highly desirable. Using cryptographic-based authentication or trustworthy certification is the conventional strategy to thwart Sybil attacks [3], [4].

On the other hand, one of the most promising approaches for wireless ad hoc networks is localisation based on received signal strength (RSS)[5]. However, this strategy needs a mobile device with a GPS or geographical positioning tools. Hence, a novel method to identify Sybil attack identities will designed and simulated on basis of the data mining based classifiers in this work. Proposed model utilizes the RSS signal to distinguish between the normal node and Sybil nodes effectively. Initially, the entry and exit behaviour of legitimate nodes and Sybil nodes has been simulated. Second, based on the entry and exit behaviours of nodes, threshold is constructed to distinguish authentic identities from Sybil identities. Third, threshold is optimized using the RSS signal fluctuation of the nodes in the simulation. Fourth, performance evaluation of the proposed scheme is carried out using extensive simulations, and its results show that it model is capable of detecting the Sybil attacks with 90% true positives and about 10% false positives in wireless environments[6].

The remaining part of the paper is organized as follows. Section 2 represents the review analysis of related work in the Sybil attack detection. In Section 3, defines the attack detection protocol to wireless sensor network with design and simulation. In Section 4, simulation analysis and performance analysis of the proposed attack detection protocol is evaluated with respect to traditional protocols. Finally, section 5 concludes the article.

Related Works

In this section , various traditional attack detection protocol to wireless sensor network is been analyzed on basis of security and performance is as follows .

2.1. Centralized Replica detection Approach

Employment of a centralised architecture in which each node joining the network should estimate a signed message (referred to as a location claim) with its neighbours to resolve node replication attacks intuitively as the most straightforward method. This location claim is forwarded to the base station by atleast one of its neighbours. The base station detects node replication attacks if it receives several location claims with the same identity in different locations. To revoke the replicated node, it transmits a message over the entire network. 100% node replication attacks are detected by the centralised system. The centralised system, however, has some drawbacks in terms of communication and memory expenses..

2.2. Distributed Replica Detection Scheme

The distributed nature of wireless sensor networks makes distributed replica detection protocols very interesting. Particularly highly security solution employs the probabilistic key sharing scheme to detect replicated keys rather than replicated nodes in order to detect node replication attacks[7].

Parno et al.'s employs the distributed methods for node replications detection[8] utilizing two protocols, one is employed for randomised multicast and another is applied for line-selected multicast as both of protocol function similarly and each node broadcasts its location claims to base station, base station has employed to record all of the claims they receive from sensor node. A base station will swarm the network with authenticated revocation requests if it receives duplicate claims.

Each intermediate node on the propagation path with neighbour node and the random node utilizes the claim and verifies it for any duplicate location claim. If it detect the duplicates, it starts a network-wide revocation request else it transmits the claim to the subsequent base station From a broad perspective, the intersection of two lines drawn by replicated nodes from various network points can be used by the witness to identify node replication. The overall cost of communication is messages, and each node needs claims in memory.

2.3. Multicasting based Replica Detection scheme

Each node broadcasts its location claim to neighbours in the randomised multicast solution. Each neighbour transfers the location claim of with probability to nodes randomly chosen (and which are protocol parameters) using a geographical routing protocol. As a result, the set of witnesses for the node is made up of all the nodes selected by all of its neighbours. This protocol's fundamental flaw is its high communication cost. Since each node requires contacting witnesses and sending messages to them costs messages, the overall cost of the network is messages.

The line-selected multicast technique was suggested in the same study as a solution to the communication cost problem. Each intermediate node that routes messages from a neighbour to another witness is included in the collection of witnesses in this algorithm. As a result, the algorithm takes the lines formed between the witnesses into account (thus, the algorithm name). So, with just five lines per node, for example, duplicated nodes can be found with a 95% chance. In conclusion, whenever a node declares its location claim, each of its neighbours probabilistically forwards claims to a different random node [9].

Proposed Model : Replica Attack detection Approach

In this part, design of replica attack detection protocol is designed along behaviour modelling of the attack, which is as follows

3.1. Attack Model:

Node replica attacks come in two kinds. In the first, the attacker generates a new identity while discarding the one they already had, so only one of their identities is active at any given time in the network. The goal of this attack, also known as a join-and-leave or whitewashing one, is to erase any evidence of harmful behaviour in the past. The lack of accountability in the network may be encouraged by this attack. The second sort of Sybil attack, often known as a simultaneous Sybil attack, involves an attacker simultaneously using all of its identities for an attack. The goal of this assault is to disrupt the network or attempt to acquire access to or more resources than a single node in a network. The variation among the nodes is only on basis of the notion of simultaneity on their consequences.

Both sorts of node replication attack has been taken into account in the proposed method. The goal of proposed detection technique is to identify every new Sybil identity that an attacker creates, regardless of whether they intend to use it for simultaneous. As a result, a new attack categorization algorithm in this research to determine both the new Sybil identity and the whitewash identity (WID) is mentioned. Assuming that malicious nodes do not coordinate with one another and it is assumed that the attacker joins the network using a single identity. There are two methods that attackers can obtain identities. They can first create identities (for instance, by coming up with a random identification). Second, they have the ability to spoof the identities of legitimate nodes (masquerading)in the network.

3.2. Signal Strength Based Analysis

Based on their neighbourhood joining behaviour, a new legitimate node and a new replica identity can be distinguished using a behaviour clustering approach. The signal strength of the replica attacker's newly created identity will be strong enough to be distinguished from the recently joined neighbour. Analysing entry behaviour to determine the difference between a legitimate newcomer and a replica identity is high possible using signal strength computation.

$$RSS = RssL_{ref} + 10n\log\left(\frac{d}{d_{ref}}\right)$$

n is node and d is node density and dref is density Threshold and Rss Lref is signal strength of legitimate node

3.3. Classification based detection Mechanism through event analyser

Assuming that no node can move faster than the network's maximum speed. Proposed model detect the attack using detection confidence value for each node based on maximum speed limit. This node confidence value will help distinguish it from a attacking node from legitimate node . The higher confidence values indicate unusual entry of sybil node into the neighbourhood. The detector node will notify its 1-hop neighbours of the discovery of Sybil identity by sending a unique detection update packet.

Each node will consider an identity to be Sybil when it gets two or more packets indicating that identity from two different nodes. The aforementioned detection technique has two problems. First, a valid node that turns off its transmitter or other equipment in one neighborhood and turns it on in another neighborhood. The suitable solution would be that since legitimate nodes will try to preserve their identities and they will reveal their identities on each new node emerges.

Nodes publishing the legitimate nodes' identities list in the network will easily compute the existing nodes' appearances in the network. Confidence value for node is calculated as follows

Confidence value = $\int_0^\infty x[\mathrm{i}]$

x[i] is the strength the attack behaviours in the network

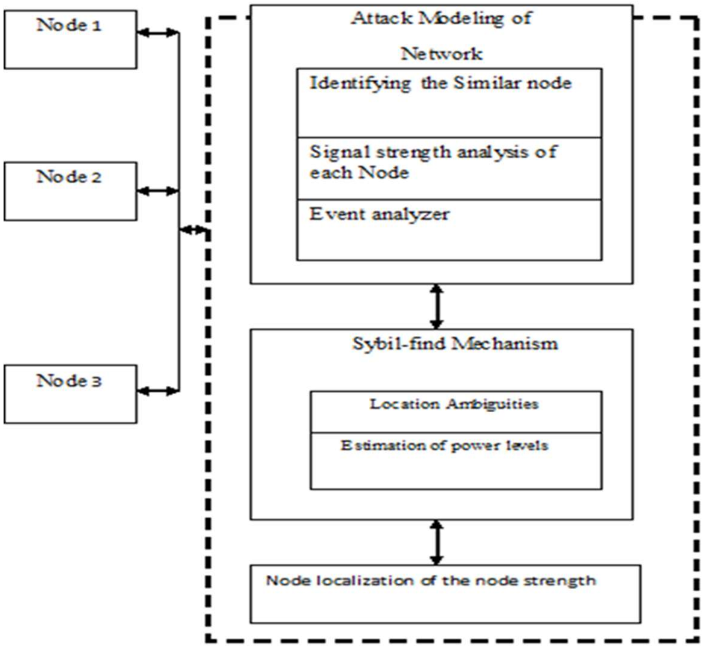
Node 1
Node 2
Node 3
Attack Modeling of Network
Identifying the Similar node
Signal strength analysis of each Node
Event analyzer
Sybil-find Mechanism
Location Ambiguities
Estimation of power levels
Node localization of the node strength

Figure 1: Attack detection Process against the replica attacks

3.4. Applying the Sybil find model for partitioning the Replicas:

By using feature ranking based behaviour-based calculations, it is possible to identify the eight features of the Sybil identities, which are most sensitive. It is possible to generate a new functional feature vector containing attack characteristics. The attack features are ranked according to its importance and relevance in detecting the attack [9]. The linear correlation coefficient is one of the most well-known and often employed metrics. The linear correlation coefficient is a feature quality measure that is well-suited for grouping or Classification of the sybil node in the network.

It should be emphasised that alternative methods can also be used to determine the importance of the features, such as symmetrical uncertainty and asymmetric dependency coefficient. Because of its widespread use and simplicity, the linear correlation coefficient is chosen in this paper.

Linear correlation Coefficient $L_c = \sum_{k=0}^{n} x^k a^{n-k}$

Where a is the node and k is value of similarity features.

We have defined five stages, which are also hidden states in HMM

- Scanning: Attacker tries to gather the information about the target system
- Enumeration: Attacker tries to find the vulnerabilities of the target system
- Access attempt: Attacker tries to obtain the access to the target system's resources.
- Denial of service : Attacker tries deny service to other users.
- Malware attempt : Attacker tries to execute attack detection code on the target system.

There were 88 rules to train the detection model. Once the rule set is mentioned, attack can be mapped with analysis of legitimate node to the alert by neighbour nodes.

Training Process of Hidden markov model

- Initialization: This step initializes the state, transition, and observation probabilities.
- Forward algorithm: This step calculates the observation probabilities based on the obtained observation sequence.
- Backward algorithm: This step calculates the state and transition probabilities based on observation probabilities and sequence.
- Re-estimation of probabilities : This step re-estimated the state, transition, and observation probabilities by iterating the above three steps number of times.

Prediction of Attacker Behavior

- As we have trained our system and stored probabilities in our database, our next step is to match the set of incoming alerts with one of our stored behavior.
- To find the closest behavior for a set of alerts, we have used Kullback Leibler Distance Calculator [6].
- The Kullback-Leibler distance (K-L) [6] is a measure of the similarity among two completely estimated probability distributions.

The Kullback-Leibler distance (K-L)

- Let $p_1(x)$ and p2(x) be two linear probability distributions. By notation, the K-L distance D (p_1, p_2) among $p_1(x)$ and $p_2(x)$ is:

  $$D(p_1, p_2) = \int p_1(x).log\frac{p_1(X)}{p_2(x)}dx$$

Basic Properties

- D (p_1, p_2) is the quantity mean of the log [p1(x)/p_2(x)], with $p_1(x)$ being the reference distribution.
- The K-L distance is considered as nonnegative. It is zero during two distributions are same.
- It is similar to determine the symmetric version of the K-L distance among p_1 and p_2:

  $$D_s(p_1, p_2) = [D(p_1, p_2) + D(p_2, p_1)] / 2$$

Algorithm 1: Attack Detection

```
Input : Sensor node N1, N2,N3..Nn
Output: Attack nodes AN1,AN2
Process
Attack Model ()
  Replica Attack()
      Signal Strength Analysis of Node = $L_{ref}$ +10nlog ($\frac{d}{d_{ref}}$)
 If (signal Strength of Node 1= Node 2)
 Consider it is replica attack
Sybil attack
 Linear Correlation C= $\sum_{k=0}^{n} x^{k} a^{n-k}$
     Node density of Node 1= Node density of Node 2
Consider it is sybil attack
```

4. Experimental results

The experiment's goal is to enable a comparison between new valid nodes' activity and new Sybil identities. The Sybils nodes will have many shared neighbours because they attach to the same physical device and transmit signals through the same node. Therefore, the identity pairs that have at least one common neighbour will be examined using our Sybil detection mechanism. We categorise the prior strategies into three groups based on the detection mechanisms: identity-based, location-based, and signal-print-based approaches.

Identity-based strategies typically reduce the production of legitimate node information, hence thwarting Sybil assaults. Location-based methods make use of the fact that each node can only be in one place at a specific moment. In the signal-print based detection approach, the detector tries to collect the attack properties of the node signals and detect the malicious claims of the node identities. Figure 2 explains the Packet delivery ratio is calculated to detect the successful data based on the data transfer rate after the Sybil attack detection.

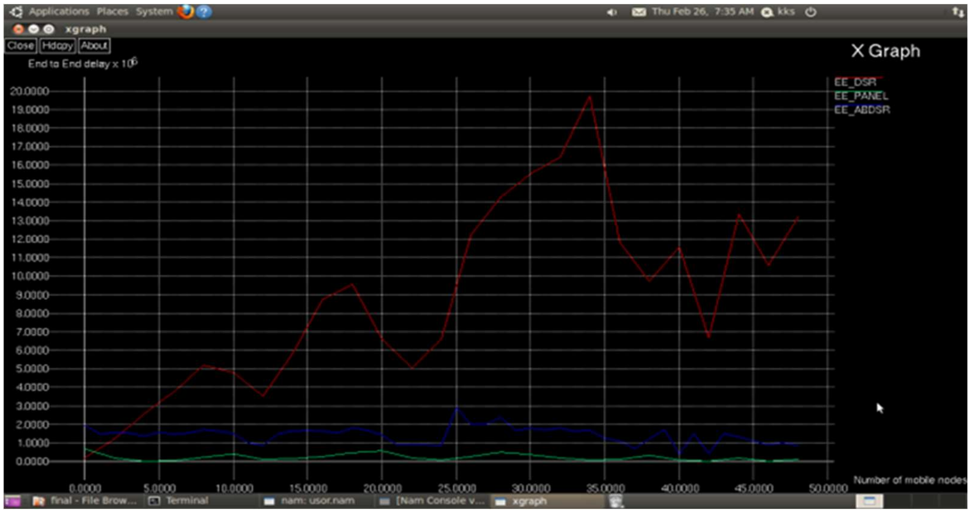

Figure 2: Packet delivery Ratio of the transmission of the packet

Routing of the transmission is carried out against the proposed solution based on the location and identity based detection for the discrete hidden markov model. Feature selection of the node determines the Sybil characteristics to remove the attacking nodes in the network. even in the attack evolution in the network, system frees the Sybil node from the transmission in order to produce the high transmission rate.

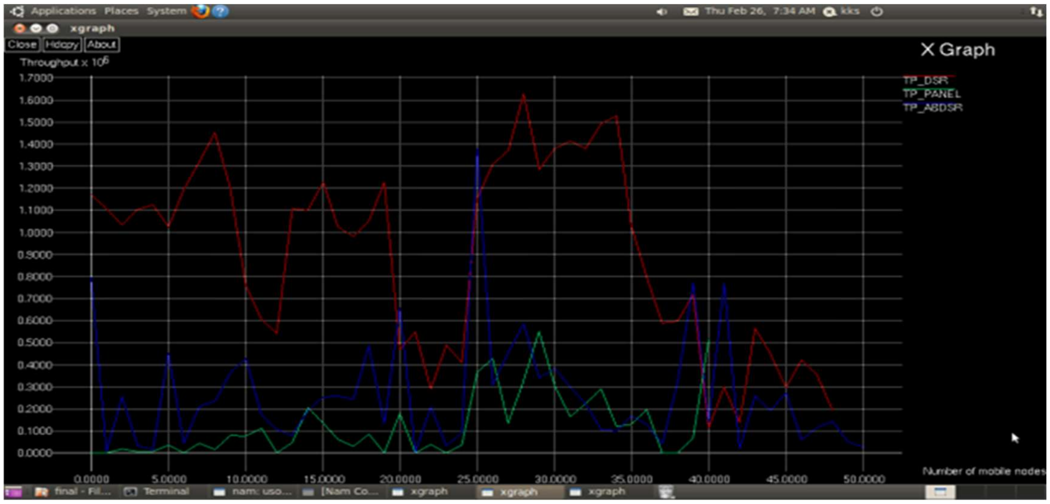

Figure 3: Throughput of the data transmission of the packets

Figure 3 explains Data transmission of the packet transfer rate is measured against the transmission route. Transmission path is attack free against the Sybil nodes and reduces the delay of the path.

Conclusion

We modelled and implemented wireless sensor network for data, which represent complex distributed systems. Dynamic characteristics of the WSN will lead to a different attack cases like sybil and DDOS which acts as resource constraint nodes, lightweight security solutions will mitigate the attack nature. Sybil attacks represent a significant danger to WSNs since each node must have a separate identity that is both unique and enduring for the security measures to work. In order to execute a coordinated attack on the network, a Sybil attacker might either construct multiple identities on a single physical device or switch identities to thwart detection, encouraging lack of responsibility in the network. Discrete hidden Markov Model scheme will be detecting the new identities of Sybil attackers through the classification phenomena's for gathering the Sybil node evolution in the network.

Reference

- S. Mathew, D. Britt, R. Giomundo, S. Upadhyaya, S. Sudit, Real-time Multistage Attack Awareness Through Enhanced Intrusion Alert Clustering, In Situation Management Workshop (SIMA 2005), MILCOM 2005, Atlantic City, NJ, October, 2005.
- Mukherjee, S., Chattopadhyay, M., Chattopadhyay, S., & Kar, P. (2018). EAER-AODV: enhanced trust model based on average encounter rate for secure routing in MANET. In S. Mukherjee (Ed.), Advanced Computing and Systems for Security (pp. 135–151). Springer.
- Kumar, E. Boopathi, and V. Thiagarasu. "Segmentation using Fuzzy Membership Functions: An Approach." IJCSE, ISSN (2017): 2347-2693.
- Yang, Gasior, Katipally,Cui, Alerts Analysis and Visualization in Network-based Intrusion Detection Systems, The Second IEEE International Conference on Information Privacy, Security, Risk and Trust (PASSAT2010), 2010, USA.
- Reddy, C. S., Yookesh, T. L., & Kumar, E. B. (2022). A Study On Convergence Analysis Of Runge-Kutta Fehlberg Method To Solve Fuzzy Delay Differential Equations. Journal of Algebraic Statistics, 13(2), 2832-2838.
- Yang, Katipally, Gasior, Cui, Multistage attack detection system for network administrators, CSIIRW -6 , 2010, USA.
- Sethuraman, P., & Kannan, N. (2017). Refined trust energy-ad hoc on-demand distance vector (ReTE-AODV) routing algorithm for secured routing in MANET. Wireless Networks, 23, 2227–2237.
- Kumar, E. Boopathi, and M. Sundaresan. "Fuzzy inference system based edge detection using fuzzy membership functions." International Journal of Computer Applications 112.4 (2015).
- CYBERSPACE: United States Faces Challenges in Addressing Global Cyber security and Governance, July 2010 .
- Reddy, A. P., & Satyanarayana, N. (2017). Energy-efficient stable multipath routing in MANET. Wireless Networks, 23, 2083–2091.
- Kashani, A. A., Ghanbari, M., & Rahmani, A. M. (2019). Improving the performance of opportunistic routing protocol using the evidence theory for VANETs in highways. IET Communications, 13, 3360–3368.
- Yookesh, T. L., et al. "Efficiency of iterative filtering method for solving Volterra fuzzy integral equations with a delay and material investigation." Materials today: Proceedings 47 (2021): 6101-6104.
- Deshmukh, M., & Kakarwal, S. (2019). Proactive neighbor knowledgebased hybrid broadcasting in MANET. International Research Journal of Engineering and Technology (IRJET), 6(07), 3668.
- Farheen, N. S., & Jain, A. (2022). Improved routing in MANET with optimized multi path routing fine tuned with hybrid modeling. Journal of King Saud University-Computer and Information Sciences, 34(6), 2443–2450..