Detecting Node Replication Attack In Wireless Sensor Networks Using Distributed Routing

[1]Roshini K [1, 2]P Hasitha Reddy, [3]A V S M Adiseshu, [4]S Lavanya Reddy

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor, [4]Associate Professor
[1,2,3,4] Department of Computer Science & Engineering, Sree Dattha Institute of Engineering & Science

Abstract –Current sensor nodes lack hardware support and are often deployed in such environments where they are vulnerable to capture and compromise by an adversary. A serious consequence of node compromise is that once an adversary has obtained the credentials of a sensor node, it can secretly insert replicas of that node at strategic locations within the network. These replicas can be used to launch a variety of insidious and hard-to-detect attacks on the sensor applications and the underline networking protocols. Security in sensor network is, therefore, a particularly challenging task. This paper discusses the current state of the art in security mechanism for WSN. We present a novel distributed approach called localized multicast for detecting node replication attacks. The two variants of the localized multicast approach are analyzed those are: (1) Single deterministic cell: SDC, (2) Parallel – Multiple Probabilistic Cell: P-MPC, nodes which as their name suggests differ in the number of cells to which a location claim is mapped and the manner in which the cells are selected. We evaluate the performance and success rate of these approaches both theoretically and via simulation.

Keywords-Wireless Network Security, Distributed Protocol,Node Replication Attack.

I. INTRODUCTION

A sensor network typically consists of hundreds or even thousands, of small, low-cost nodes distributed over a wide area. The nodes are expected to function in an unsupervised fashion even if new nodes are added or old nodes disappear (e.g., due to power loss or accidental damage). While some networks include a central location for data collection, many operate in an entirely distributed manner, allowing the operators to retrieve aggregated data from any of the nodes in the network. Furthermore, data collection may only occur at irregular intervals. For example, many military applications strive to avoid any centralized and fixed points of failure. Instead, data is collected by mobile units (e.g., unmanned aerial units, foot soldiers, etc.) that access the sensor network at unpredictable locations and utilize the first sensor node they encounter as a conduit for the information accumulated by the network. Since these networks often operate in an unsupervised fashion for long periods of time, the aim is to detect a node replication attack soon after it occurs. If we wait until the next data collection cycle, the adversary has time to use its presence in the network to corrupt data or otherwise subvert the network's intended purpose. Here, the adversary cannot readily create new IDs for nodes. There are several techniques to prevent the adversary from deploying nodes with arbitrary IDs, Randomized Multicast, distributes location claims to a randomly selected set of witness nodes. The Birthday Paradox predicts that a collision will occur with high probability if the adversary attempts to replicate a node, Line-Selected Multicast, exploits the routing topology of the network to select witnesses for a node's location and utilizes geometric probability to detect replicated nodes.

II. SYSTEM ANALYSIS AND SYSTEM DESIGN

A. Existing System

Centralized approaches are already available to detect node replicas. In a centralized approach for detecting node replication, when a new node joins the network, it broadcasts a signed message (referred to as a location claim) containing its location and identity to its neighbors. One or more of its neighbors then forward this location claim to a central trusted party (e.g., the base station). With location information for all the nodes in the network, the central party can easily detect any pair of nodes with the same identity but at different locations. However, this solution is vulnerable to a single-of-point failure. If the base station is compromised or the path to the base station is blocked, adversaries can add an arbitrary number of replicas into the network without being detected. Hence, a distributed solution is desirable.

Distributed approaches for detecting node replications are based on storing a node's location information at one or more witness nodes in the network. When a new node joins the network, its location claim is forwarded to the corresponding witness nodes. If any witness receives two different location claims for the same node identity (ID), it will have detected the existence of a replica and can take appropriate actions to revoke the node's credentials. The basic challenge of any distributed protocol in detecting node replicas is to minimize communication and per node memory costs while ensuring that the adversary cannot defeat the protocol. A protocol that deterministically maps a node's ID to a unique witness node would minimize both communication costs and memory requirements per node, but would not offer enough security because the adversary would need to compromise just a single witness node in order to be able to introduce a replica without being detected.

B. Proposed System

We present a novel distributed protocol for detecting node replication attacks that takes a different approach for selecting witnesses for a node. In our approach, which we call Localized Multicast, the witness nodes for a node identity are randomly selected from the nodes that are located within a geographically limited region (referred to as a cell). Our approach first deterministically maps a node's ID to one or more cells, and then uses randomization within the cell(s) to increase the resilience and security of the scheme. One major advantage of our approach is that the probability of detecting node replicas is much higher than existing solutions. We describe and analyze two variants of the Localized Multicast approach: Single Deterministic Cell (SDC) and Parallel Multiple Probabilistic Cells (P-MPC), which as their name suggests differ in the number of cells to which a location claim is mapped and the manner in which the cells are selected.

C. System Design

We consider a sensor network with a large number of low-cost nodes distributed over a wide area. In our approach, we assume the existence of a trusted base station, and the sensor network is considered to be a geographic grid, each unit of which is called a cell. Sensors are distributed uniformly in the network. New sensors may be added into the network regularly to replace old ones. Each node is assigned a unique identity and a pair of identity-based public and private keys by an offline Trust Authority (TA). In identity-based signature schemes like [6], the private key is generated by signing its public key (usually a hash on its unique identity) with a master secret held only by the TA. In other words, to generate a new identity-based key pair, cooperation from the TA is a must.

Therefore, we assume that adversaries cannot easily create sensors with new identities in the sense that they cannot generate the private keys corresponding to the identities claimed and thus fail to prove themselves to the neighbors during the authentication of the location claims. It requires that, when a node is added into the network, it needs to generate a location claim and broadcast the claim to its neighbors. Each neighbor independently decides whether to forward the claim with a given probability. For those neighbors that plan to forward the claim, they determine the destination cell(s) according to the output of a geographic hash function , which uniquely maps the identity of the sender of the location claim to one or a few of the cells in the grid. Then, the claim is forwarded to the destination cell(s) using a geographic routing protocol such as GPSR. We assume that the major goal of adversaries is to launch node replication attacks. To achieve this goal, we assume that adversaries may launch both passive attacks (e.g., eavesdropping on network traffic) and active attacks (e.g., modifying and replaying messages or compromising sensors), and the information obtained from the former can be used to enhance the effectiveness of the latter.

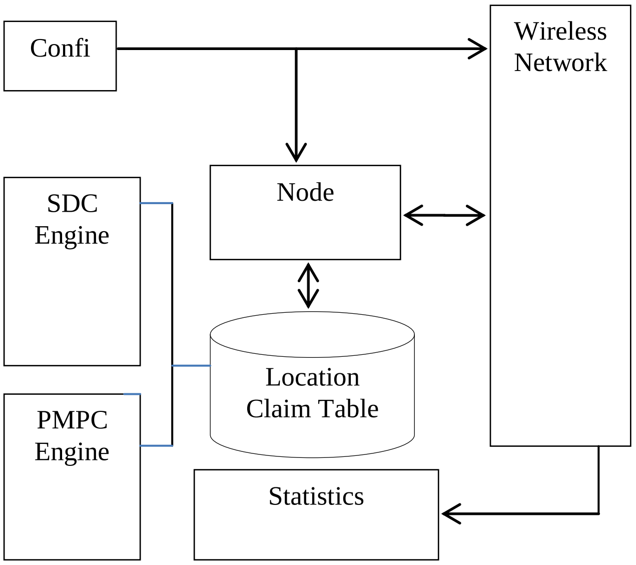

Fig 1.System Architecture

III. APPROACHES FOR DETECING NODE REPLICATION ATTACKS

We have designed two variants of the Localized Multicast approach, specifically Single Deterministic Cell (SDC) and Parallel Multiple Probabilistic Cells (P-MPC).

Single Deterministic Cell:

In the Single Deterministic Cell scheme, a geographic hash function is used to uniquely and randomly map node L's identity to one of the cells in the grid. When L broadcasts its location claim, each neighbor first verifies the plausibility of lL (e.g., based on its location and the transmission range of the sensor) and the validity of the signature in the location claim. In identity-based signature schemes, only a signature generated with the private key corresponding to the identity claimed can pass the validation process. Thus, adversaries cannot generate valid signatures unless they compromise the node with that identity. Each neighbor independently decides whether to forward the claim with a probability pf. If a neighbor plans to forward the location claim, it first needs to execute a geographic hash function to determine the destination cell, denoted as C. The location claim is then forwarded toward cell C. Once the location claim arrives at cell C, the sensor receiving the claim first verifies the validity of the signature and then checks whether cell C is indeed the cell corresponding to the identity listed in the claim message based on the geographic hash function. If both the verifications succeed, the location claim is flooded within cell C. Each node in the cell independently decides whether to store the claim with a probability ps. Note that the flooding process is executed only when the first copy of the location claim arrives at cell C and the following copies are ignored.

As a result, the number of witnesses in the cell w is s*ps on average, where s is the number of sensors in a cell. Whenever any witness receives a location claim with the same identity but a different location compared to a previously stored claim, it forwards both location claims to the base station. Then, the base station will broadcast a message within the network to revoke the replicas. Compared to the Random Multicast and Line-Selected Multicast algorithms, a major advantage of SDC is that it ensures 100 percent success rate for detecting any node replication, as long as the location claim is successfully forwarded toward cell C and stored by at least one node in the cell. An important limitation on the Random Multicast and Line-Selected Multicast algorithms is that both the communication/memory overhead and the security (in terms of the success rate of detecting node replications) of the two algorithms are tightly related to the number of witnesses (w). On the one hand, the larger w is, the higher the communication and memory overhead. On the other hand, the smaller w is, the lower the success rate of detecting node replication. In contrast, in the SDC scheme the communication cost and memory overhead are related to the number of neighbors that forward a location claim. Moreover, the randomization against potential node compromise and low memory overhead are achieved through flooding the location claim within the destination cell while storing it on only a small number o f randomly chosen nodes. Assuming that the capability of the adversary (in terms of the number of nodes that can be compromised without being detected) is limited, by appropriately choosing the cell size (s) and ps, the probability that adversaries control all the witnesses for an identity is negligible. Consequently, SDC can achieve a low communication cost by setting r to a small value, and at the same time ensure low memory overhead and good security (i.e., a high success rate of detecting node replication and high level of resilience against potential node compromise), by choosing an appropriate value for w (s and ps, actually).

Parallel Multiple Probabilistic Cells:

We assume the existence of a monitoring mechanism that can detect a node compromising operation with a certain probability. Therefore, the larger the number of nodes that an adversary attempts to compromise, the higher is the probability that the node compromising attack is detected, thereby triggering an automated protocol or human intervention for removing compromised nodes. However, in certain cases (e.g., when the number of nodes in a cell is relatively small), a determined adversary may be willing to take the risk of being detected in return for a high probability of controlling all the witness nodes for one or more identities.
Another potential risk is that a smart adversary can take advantage of the knowledge that the destination cell for a given identity is deterministic and launch a blocking attack. Informally, after compromising a small set of sensors denoted as V, the adversary can generate replicas of members in V and deploy them in such a way that all the location claims of these replicas are forwarded through members of V. In the SDC approach, all the location claims are first forwarded from the neighbors of L to a deterministic cell. Therefore, there is a high probability that these forwarding paths intersect with each other. In particular, when L and the destination cell (i.e., cell C) are far from each other, there is a high probability that all the location claims will pass through one or a small set of nodes of size y.

Therefore, the adversary only needs to compromise one or y nodes per replica so as to block the forwarding of a location claim. Hop-by-hop watch monitoring may help mitigate this attack. However, it will fail if all or most of the neighbors of an intersection point are compromised. Even worse, the adversary can insert a replica in such a way that its location claim will always be forwarded through a small set of compromised nodes. An example of blocking attack against the SDC approach is shown in Fig. 2. Cell C1 and C2 are the deterministic cells for the identity IDC1 and IDC2, respectively, and B is an area in which all the nodes have been compromised (referred to as a black hole). In this example, three replicas (i.e., L1 C1 , L2 C1 , and L3 C1) claiming the same identity that is mapped to cell C1 are added to the network sequentially, with a certain time interval between any pair of consecutive joins. In the SDC approach, nodes enroot between the replica and the deterministic cells do not store the location claim. As a result, as long as the location claims from different replicas do not arrive at the same time, forwarding nodes are not able to detect the conflicts. Finally, all the location claims are delivered to the black hole and blocked. In other words, adversaries can insert replicas without being detected. As shown in Fig. 2, two replicas (i.e., L1 C2 and L2 C2) claiming the same identity that is mapped to cell C2 are inserted into the network and their location claims are also blocked by the black hole.

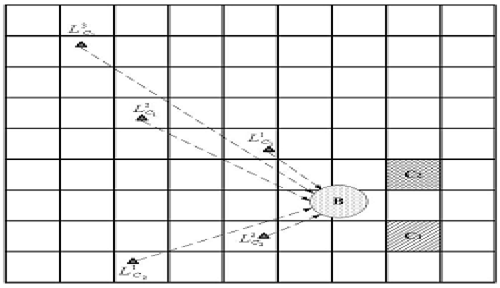

Fig 2. The Blocking Attack

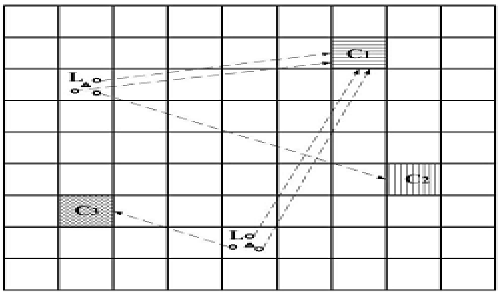

Fig 3. The P-MPC Approach

Like SDC, in the P -MPC scheme, a geographic hash function is employed to map node L's identity to the destination cells. However, instead of mapping to a single deterministic cell, in P-MPC, the location claim is mapped and forwarded to multiple deterministic cells with various probabilities. An example of P-MPC is shown in Fig. 2. When L broadcasts its location claim, each neighbor independently decides whether to forward the claim in the same way as the SDC scheme. Afterwards, each neighbor helping forward the claim first calculates the set of cells (i.e., C) to which L are mapped, based on a geographic hash function with the input of IDL. Once the location claim arrives at cell Cj, the sensor receiving it first verifies whether Cj is a member of C which can be calculated based on the geographic hash function and the identity listed in the claim message. In addition, this sensor needs to verify the validity of the signature in the location claim. If both the verifications succeed, the claim is flooded within the cell and probabilistically stored at w nodes in the same manner as in the SDC scheme. For example, in Fig. 2, there are two replicas with the same identity in the network. In this example, an identity is mapped to three cells (i.e., C1; C2; C3) with different probabilities (i.e., pc1 >pc2 >pc3). The neighbors of one replica forward the location claims to cell C1 and C2, while the neighbors of the other replica forward the location claims to cell C1 and C3. Therefore, any witness node with cell C1 can detect the node replication.

IV. ANALYSIS OF THE APPROACHES

In this section, we analyze the communication and memory overhead of both the schemes.

Communication Overhead: The comparison of the communication overhead between the two schemes shows that the communication overhead is more in PMPC then SDC as the number witness nodes are more. The communication overhead is analyzed on the basis of number of packets sent/received, to number of nodes .The overhead is less in SDC as there is a single cell which contains the witness node.

Memory Overhead: The memory required is more in PMPC as the number of cells with witness nodes are increased. Each witness node in the cell required the significant amount of memory.

V. CONCLUSION AND FUTURE ENHANCEMENT

In this paper, we proposed two variants of the Localized Multi cast approach for distributed detection of node replication attacks in wireless sensor networks. Unlike the two randomized algorithms proposed by Parno et al., our approach combines deterministic mapping (to reduce communication and storage costs) with randomization (to increase the level of resilience to node compromise). Our theoretic analysis show that, compared to Parno et al.'s algorithms, our schemes are more efficient in large-scale sensor networks, in terms of communication and memory costs. Moreover, the probability of replica detection in our approach is higher than that achieved in these two algorithms. The two schemes are compared where their communication and memory overhead is analyzed. One of our future works is to simulate the RED protocol and then have a more detailed comparison of efficiency based on empirical results.

REFERENCES

[1] H. CHOI, S. ZHU, AND T.F. LA PORTA, "SET: DETECTING NODE CLONES IN SENSOR NETWORKS," PROC. THIRD INT'L CONF. SECURITY AND PRIVACY IN COMM. NETWORKS (SECURE COMM.) 2007.

[2] M. Conti, R. Di Pietro, L.V. Mancini, and A. Mei, "A Randomized, Efficient, and Distributed Protocol for the Detection of Node Replication Attacks in Wireless Sensor Networks," Proc. ACM Mobile ad Hoc, pp. 80-89, 2007.

[3] J.R. Douceur, "The Sybil Attack," Proc. First Int'l Workshop Peer-to-Peer Systems (IPTPS '02), pp. 251-260 2002.

[4] L. Eschenauer and V.D. Gligor,"A Key-Management Scheme for Distributed Sensor Networks," Proc. Ninth ACM Conf. Computer and Comm. Security, pp. 41-47, 2002.

[5] G. Gaubatz, J. P. Kaps, and B. Sunar, "Public Key Cryptography in Sensor Networks-Revisited,"Proc. First European Workshop Security in Ad-Hoc and Sensor Networks (ESAS '04), pp.2-18, 2004.

[6] F. Hess, "Efficient Identity Based Signature Schemes Based on Pairings," Pro c. Ninth Ann. Int'l Workshop Selected A as in Cryptography (SAC '02), pp. 310-324, 2002.

BIBLIOGRAPHY

Ms. Roshini K, received her M.Tech (CSE) from Jawaharlal Nehru Technological University (JNTU) Hyderabad, Telangana**.** Her interested area is Data Mining and Computer Networks. Presently she working as an Assistant Professor in CSE Department at Sree Dattha Institute of Engineering & Science, Hyderabad, Telangana

Ms. P Hasitha Reddy received her M.Tech (CSE) from Jawaharlal Nehru Technological University (JNTU) Hyderabad, Telangana**.** Her interested area is Big Data, Mobile Computing and Human Computer Interaction. Presently she working as an Assistant Professor in CSE Department at Sree Dattha Institute of Engineering & Science, Hyderabad, Telangana.

Mr. A V S M Adiseshu received his M.Tech (CSE) from Acharya Nagarjuna University, Andhra Pradesh**.** His interested area is Big Data and Data Mining. Presently he working as an Assistant Professor in CSE Department at Sree Dattha Institute of Engineering & Science, Hyderabad, Telangana.

Ms. S. Lavanya Reddy received her M.Tech (CSE) from Jawaharlal Nehru Technological University (JNTU) Hyderabad, Telangana**.** Her interested area is Big Data and Data Mining. & Data Warehousing .Presently she working as an Associate Professor in CSE Department at **Sree** Dattha Institute of Engineering & Science, Hyderabad, Telangana.