Detection of Node Replication Attack using Distributed Multicast Approach in Wireless Sensor Networks

Anusha N. Rao[1], Rekha B S[2]

Department of Information Science & Engg, R V College of Engineering, Bangalore

[1]n.anusha.rao@gmail.com
[2]rekhakumar26@gmail.com

Abstract— **Wireless sensor network (WSN) has become a growing research area due to its tremendous number of applications which can benefit from it. WSN are often deployed in unfavourable environments where an attacker can capture various nodes and replicate them to take over the network. The defence against node replication attacks are few and suffer from high overhead with respect to computation cost and memory. To find the replicas in the network, distributed detection approach is employed. The basic challenge of this approach is to minimize the communication and per node memory costs. With the use of distributed approach the node replicas can be captured efficiently. The results are expected to compare the different variants with respect to communication cost and memory.**

***Keywords*—wireless sensor network security, node replication attack, distributed detection approach, clones, distributed multicast approach.**

I. INTRODUCTION

A wireless sensor network (WSN) is a collection of nodes organized into a cooperative network. The current advances in technology have made it possible for development sensor nodes at very low cost. The nodes are compact and inexpensive. Since their cost is low, it is feasible to deploy thousands of nodes in the area of interest. WSN are unattended and hence prone to attacks. An adversary can eavesdrop and capture nodes and access all the information stored in it. The attacker can duplicate the node and deploy them in the network to launch attack on the inside. This attack is called as node replication attack.

There are two approaches to detection of the node replication attack: centralized approach and distributed approach. In [4] centralized approach is discussed. In centralized approach all the nodes in the network transfer a list of their neighbour's location claim to a central base station. With this information the central base station can easily detect any pair of nodes with the same identity but are at different locations. The drawback of this approach is single point of failure. Also, if the nodes near the base station are compromised then this protocol will be useless. Hence distributed approach is proposed.

In distributed approach, a node's location claim is stored in one or more witness nodes. During addition of new node to a network, its location claim is sent to corresponding witness nodes. When a witness node receives location claims for the same node identity(ID) and if they have different locations then that node is said to be replicated in the WSN. Later appropriate actions should be taken to revoke the credentials of that node.

The challenge here is to reduce communication cost and memory costs while detecting the replicas using the distributed protocol. Hence a novel approach called Localized Multicast [1] is proposed. In this approach, the witness nodes for a particular node identity (ID) are selected from within a geographically region, also known as cell. The proposed approach deterministically maps a node's ID to one or more cells. Then randomly selects witness nodes in the cell for that particular node.

Two variants of the Localized Multicast approach are described [1].They are Single Deterministic Cell (SDC) and Parallel Multiple Probabilistic Cells (P-MPC). These two variants differ in the number of cells to which location claim is mapped.

The rest of the paper is organized as follows: Section 2 shows the background work done. Section 3 shows the assumptions and threat model. Section 4 describes the system architecture. Section 5 shows the results obtained by the simulation. Section 6 concludes the paper.

II. BACKGROUND STUDY

The previous solution for detection of cloned attacks relied on centralized base station approach [4]. In this solution, each sensor's location is sent to base station (BS). BS takes the responsibility of identifying the node replicas at distinct locations. BS finally revokes the replicas.

SET [2] is based on computing set operations such as intersection and union of exclusive subsets in a network. SET consists of five components: exclusive subset construction, authentication of subset covering,

distributed set computation and interleaved authentication on subset trees, and verifiable random selection. The network is divided into exclusive subsets and subset leaders (SLDR) are selected. Each SLDR is authenticated. SLDR generates report of randomly selected members and forwards report to the base station. Multiple subset trees are constructed. For each subset tree constructed, a parent SDLR aggregated reports form child SLDRs, and forwards the final report to the base station. When the base station receives these reports it verifies the report validity and detects node replicas.

The first distributed approach protocols proposed are Randomized Multicast (RM) [4] and Line-Selected Multicast (LSM) [4] are proposed. Two preliminary approaches are discussed initially. They are Node-To-Network Broadcasting and Deterministic Multicast. In Node-To-Network Broadcasting, each node uses an authenticated broadcast message to flood the network with its location information. Each node stores the location information for its neighbours and if it receives a conflicting claim, revokes the offending node. Communication cost is O (n^2). In Deterministic multicast, node's location claim is sent to a limited subset of deterministically chosen witness nodes. Each node broadcasts its location claim and its neighbours forward that claim to all the witness nodes. If the attacker replicates a node, then the witness nodes will receive two different location claims for the same node ID. This shows that these nodes are replicas and will be revoked. This communication cost is O (g ln g) where g is the number of witness nodes. In RM each node has $\sqrt{n}$ witness nodes. Thus according to Birthday Paradox, if attacker replicates node, then at least one witness node is going to receive inconsistent information about the location claims of a particular node. The communication cost is O (n^2). In LSM, line segments are drawn in the network. It uses the routing topology of the network for the selection of witness nodes. It uses geometric probabilities to detect replicated nodes. The communication cost is O ($n \sqrt{n}$).

In RED[3] protocol the witness as selected uniformly within the network. The number of witness is less compared to [4]. RED is more robust against selective node compromise and less robust against random node compromise. The overhead introduced by RED is low and almost evenly balanced among the nodes compared to LSM [4].

III. Assumptions and Threat Model

In this approach a trusted base station is assumed to be present. The network is divided into geographic cells. Sensors are distributed randomly in the network. It is assumed that an offline trusted authority manages public and private keys for each node in the network. An identity based signature scheme is used to generate private keys for each node. Cooperation from Trusted Authority is must for creating identity based key pairs. Therefore, it is assumed that the attacker cannot create sensor nodes with new identities. They fail to prove themselves to the neighbours during the authentication of the location claims because they cannot generate private keys corresponding to the identities claimed.

IV. System Architecture

The two approaches which are discussed are Single Deterministic Cell (SDC) and Parallel Multiple Probabilistic Cell (P-MPC) [1]. The system architecture shown in fig. 1 has the components which are discussed hereafter. The configuration component is used by the client to start the simulation. The client provides the number of nodes to be created and it is sent to the WSN simulator. The simulator assigns public and private keys using RSA algorithm. The client can either select SDC or P-MPC for detection of replicas. The location claim is computed using MD5 algorithm. The witness nodes store the location claim in a table. Finally, the list of node replicas are revoked shown to the clients.

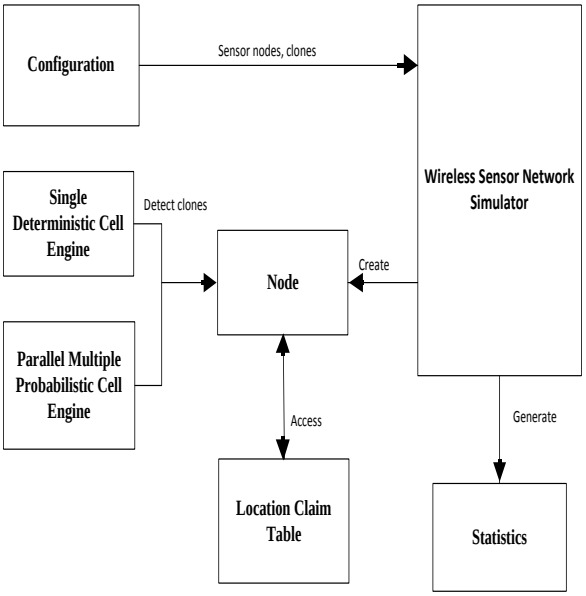

Fig. 1 System Architecture

A. *Single Deterministic Cell*

This scheme uses geographic hash function [5] to uniquely and randomly map node N's identity to one of the cells in the grid. Consider an example where a grid consist of m x n cells, a cell at m'th row and at n'th column is uniquely identified as c (where $c = m'. n + n'$). Node N is mapped to cell c.

When node N broadcasts its location claim, its neighbours first verify its location information based on its location and the transmission range. The neighbours

also check the validity of the signature of location claim. Each neighbour decides whether to forward the location claim with a probability. A destination cell is determined. The location claim is forwarded to the destination cell. When the location claim is received at the destination cell, the sensor which receives first verifies the validity of the signature. The correctness of calculation of destination cell is verified. When both verifications succeed, then the location claim is sent to few nodes randomly selected within the destination cell. The advantage of this scheme is that it ensures 100% success rate in detecting node replicas as long as location claim reaches the destination cell.

The following message types are used in the following algorithms:

1. Broadcast Location Claim (BLC1) message to neighbour in the same cell as originator.
2. Broadcast Location Claim (BLC2) message to neighbour in the same cell as target cell.
3. Location Claim message (LCM)
4. Request to send Location Claim (RLC)
5. REVOKE message to base station to inform clone detected (REVOKE)

The algorithm of SDC is shown below.

SDC Algorithm

1. **while** (true)
2. M= getMessage();
3. **if** (M is RLC)
4. Claim= formClaim(ID, xpos,ypos, RSA Private Key);
5. Create BLC1 message and send to Neighbours in the Cell.
6. **else if** (M is BLC1)
7. TargetCell = Hash(Recv ID , cell row , cell col);
8. Choose randomly to send LCM to the target Cell.
9. **else if** (M is LCM)
10. Create BLC2 message and send to Neighbours in the Cell.
11. **else if** (M is BLC2)
12. Choose randomly to store the Location Claim in the message.
13. **if** (claim already exists)
14. **if** (id in claim is same but location claim not matching)
15. Send REVOKE to base station
16. **end if**
17. **end if**
18. **end if**
19. **end while**

B. Parallel Multiple Probabilistic Cell

A potential risk is that a given adversary may be willing the jeopardy of being detected in return for high probability of controlling all the witness nodes for one or more nodes. Another risk is that an attacker can launch a black hole attack after determining the destination cell for a given node. This black hole attack is not detected by SDC scheme. This attack is shown in the fig 2.

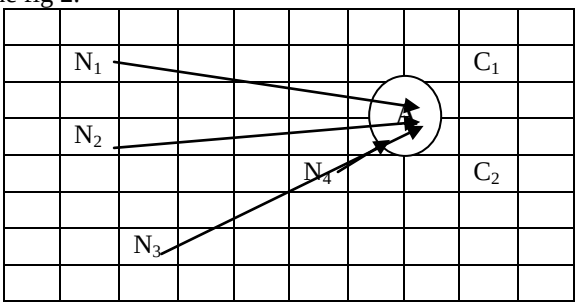

Fig 2. Black hole Attack

The network which is attacked is shown in fig. 3. There are four nodes N_1, N_2, N_3, and N4. Node N1 is mapped deterministically to cell C_1 and node N_4 deterministically mapped to cell C_2. A is an area in which all the nodes have been compromised by the adversary. The area A has got nodes which are used to reach the cells C_1 and C_2. So the location claim message of node N_1has to pass through the black hole A and never reach the cell C_1. Replicas of Node N_1 are N_2 and N_3. These are not detected by SDC[1] scheme due to the block hole. Therefore P-MPC[1] is proposed.

Similar to SDC, the P-MPC scheme uses geographic hash function [5] to map node's identity to the destination cells. The location claim is forwarded to multiple deterministic cells with various probabilities in P-MPC[1].

The algorithm of P-MPC is as follows:

P-MPC Algorithm

1. **while** (true)
2. M= getMessage();
3. **if** (M is RLC)
4. Claim= formClaim(ID, xpos,ypos, RSA Private Key);
5. Create BLC1 message and send to Neighbours in the Cell.
6. **else if** (M is BLC1)
7. TargetCell = HashTo3Cells(Recv ID , cell row , cell col);
8. Choose randomly to send LCM to 3 target Cell.
9. **else if** (M is LCM)
10. Create BLC2 message and send to Neighbours in the Cell.
11. **else if** (M is BLC2)
12. Choose randomly to store the Location Claim in the message.
13. **if** (claim already exists)
14. **if** (id in claim is same but location claim not matching)
15. Send REVOKE to base station

16. **end if**
17. **end if**
18. **end if**
19. **end while**

The P-MPC algorithm selects three cells for each node because three cells are sufficient to detect the replication node.

V. RESULTS

The replicas are detected in the network. Performance of the detection system in terms of communication cost (fig. 3) and memory overhead (fig. 4) are compared. The communication cost of the P-MPC scheme is greater than that of SDC since in the SDC scheme location claim information has to be sent to few witness nodes. The memory overhead is high in P-MPC compared to SDC since the number of witness nodes storing the location claim information is high in P-MPC.

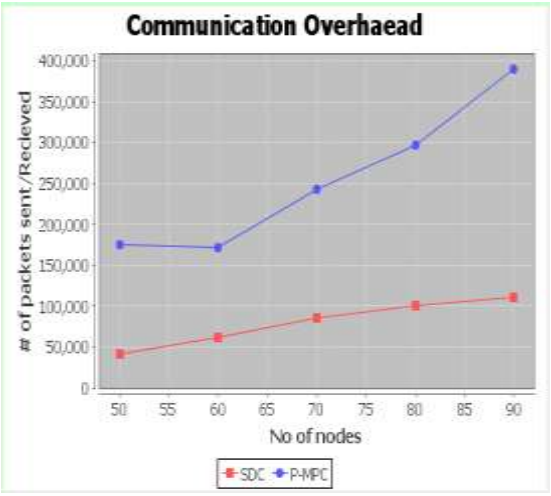

Fig. 3 Communication Overhead

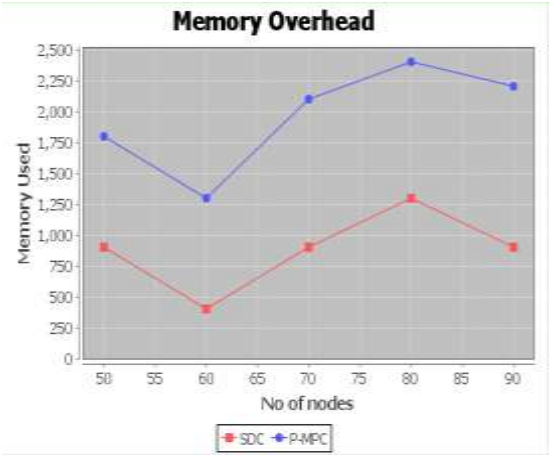

Fig. 4 Memory Overhead

VI. CONCLUSION

Two types of localized multicast approach for distributed detection of node replication attacks in wireless sensor networks are discussed. This approach combines deterministic mapping with randomization. The deterministic mapping reduces communication and storage cost. The randomization increases the level of resilience to node compromise attack. The probability of node replica detection is higher than the randomized multicast [4] and line selected multicast [4].

REFERENCES

[1] B. Zhu, S. Setia, S. Jajodia, S. Roy, "Localized Multicast: Efficient and Distributed Replica Detection in Large Scale Sensor Networks,"IEEE Transcations on Mobile Computing, Volume 9, Issue 7, pp 913-926, 2010.

[2] H. Choi, S. Zhu, T.F. La Porta, "SET: Detecting Node Clones in Sensor Networks," Third int'l Conf. Security and Privacy in Comm. Networks (SecureComm) 2007.

[3] 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," ACM MobiHoc, pp 80-89, 2007.

[4] B. Parno, A. Perrig, and V. Gligor, "Distributed Detection of Node Replication Attacks in Sensor Networks," IEEE Symp. Information Privacy, pp 49-63, 2005.

[5] S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R. Govindan, and S. Shenker, "GHT: A Geographic Hash Table for Data-Centric Storage," Proc. First ACM Int'l Workshop Wireless Sensor Networks and Applications (WSNA), pp. 78-87, 2002.