A NOVEL APPROACH FOR DETECTING COMPROMISED NODES IN WIRELESS SENSOR NETWORKS

P. Vijayalakshmi¹, D. Somasundareswari² and V. Duraisamy¹

¹Hindusthan College of Engineering and Technology Coimbatore, India, E-mail: vijihicet@gmail.com
²Adithya Institute of Technology, Coimbatore, India, E-mail: dsomasundareswari@yahoo.com

Abstract: Increased interest in the field of wireless sensor networks has proved that wireless sensor networks can have a broad variety of applications. Current applications of wireless sensor networks are in the fields of medical care, battlefield monitoring, environment monitoring, surveillance and disaster prevention. Many of these applications require that the sensor network be deployed in an area that is hostile, inaccessible and mission critical. However, since sensor nodes are inexpensive devices, which could be easily compromised and controlled by an adversary, the compromised nodes could report false sensed results and degrade the reliability of the whole network. Therefore, how to identify these compromised nodes in a wireless sensor network is a very important security issue. In this paper we propose a novel approach using genetic algorithm to logically identify the anomalous behavior of compromised nodes in an efficient and effective way.

Keywords: Wireless sensor network, security, identify compromised nodes, Genetic algorithm

I. INTRODUCTION

Wireless sensor networking has been subject to extensive research efforts in recent years, and has been well recognized as a ubiquitous and general approach for some emerging applications ranging from the civilian domain, such as environmental monitoring, to the military domain, such as battlefield surveillance. A wireless sensor network (WSN) is usually composed of a large number of sensor nodes which use wireless links to perform distributed sensing tasks. Here, each sensor node is cheap with low battery power and computation capacity, but is equipped with sensing, data processing, and communicating components. Since a WSN is usually deployed in a harsh environment and each low-cost sensor node could be easily compromised, the security becomes a major concern. For example, when some sensor nodes are compromised and controlled by an adversary, they could behave abnormally and randomly report true/false sensing results to the sink, by which the reliability of the whole network is seriously degraded. Therefore, it will be a critical issue to develop an effective strategy for the sink to detect and identify these compromised nodes. Assume that a WSN consists of n sensor nodes, where the compromised nodes are less than the normal nodes, and each normal node can know other nodes’ trust status by adopting the similar techniques in [23], [24]. Then, a straightforward way for the sink to identify the compromised nodes is that the sink inquiries each sensor node with n–1 queries on other nodes. In such a way, after total n(n–1) queries, the sink can logically identify all compromised nodes based on these returned results. Although this approach is effective, but it is not efficient because the n(n–1) queries could consume a significant amount of the sensor nodes’ energy.

In this paper, in order to reduce the query number, we propose a novel approach to Identify Compromised Nodes using genetic algorithm in wireless sensor networks.

II. PRELIMINARIES

In this section, we describe our sensor network model, provide a brief overview on the trust assumptions, and define our design goal.

A. Network Model

We consider a simple abstraction of a WSN consisting of a fixed sink S and a number of sensor nodes $N = \{N_1, N_2, \ldots, N_n\}$, where $n \geq 3$, deployed at a remote area (around one or more hop neighborhood of the sink), as shown in Figure 1. Note that such an abstraction can be taken as a
building block of a large-scale WSN. The sink $S$ is a data collection unit, which could be a powerful workstation with plentiful resources. However, the sensor nodes $N = \{N_1, N_2, \ldots, N_n\}$ are inexpensive, low power devices which have limited resources, memory space, computation capability, and communication bandwidth. Once such a WSN is deployed and the corresponding data paths are established, the sink $S$ can inquire each sensor node $N_i \in N$ with a certain query, and then the sensor node $N_i$ reports its sensing results back to the sink $S$ over the pre-defined path. Additionally, in order to provide confidentiality to the sensing results, a pre-shared key between each $N_i \in N$ and $S$ is assumed and used to encrypt/decrypt the sensing results. We assume that the compromised nodes will be excluded from the network if the nodes resist to forward data for the other nodes. While identifying the compromised nodes under a Denial of Service (Dos) attack is not our focus, we concentrate on the case where the compromised nodes still forward data but could falsely report the sensing results.

With the passing of time, $nc$ sensor nodes are compromised by an adversary, and form compromised-set $Nc = \{N_{*1}, N_{*2}, \ldots, N_{*nc}\}$, and the rest $nd$ sensor nodes keep normal and form normal set $Nd = \{N_{†1}, N_{†2}, \ldots, N_{†nd}\}$. Then, we will have $N = Nc \cup Nd$. As in the most application scenarios, it is reasonable for us to assume the normal-set $Nd$ is larger than the compromised-set $Nc$. Then, we will have the following relations:

$$Nc < Nd$$

$$Nc + Nd = N$$

In the WSN abstraction, the sink $S$ can clearly know the network topology and properly reconfigure the data path connecting each sensor node, and each normal sensor node $N_i \in Nd$ can exchange their local information and get to know all nodes’ trust status within its communication range.

C. Design Goal

As the compromised nodes report the true / false sensing results randomly, the reliability of the whole network will be seriously impaired. Therefore, undoubtedly, it is of ultimate importance for the sink $S$ to exactly know each sensor node’s trust status. Based on the above assumptions, we design a novel algorithm to help the sink $S$ to identify the compromised nodes. Compared with the straightforward approach where $n(n-1)$ queries are required, the proposed approaches are much more efficient.

III. PROPOSED ALGORITHM

The proposed algorithm aims to distinguish the compromised nodes from the normal ones with the lowest number of queries. Genetic Algorithm (GA) has been used to solve many optimization problems. In this a initial population is randomly generated and is evaluated with a fitness function. With higher probability of fitness is being chosen to create the next generation. The size of the population is same for all the iterations and the process is iterated until a predefined threshold value is met or a maximum number of iterations are reached.

A. Overview

Genetic algorithm used to evolve simple rules for network traffic. Initially the genes should be represented as different data types like byte, integer,
float etc., and the data range of the genes should be generated, in each iteration randomly. Then the genetic algorithm uses set of rules for the network of the form

\[ \text{If } \{\text{condition}\} \text{ then } \{\text{act}\} \]

**B. Genetic Algorithm Parameters**

The general structure of the genetic algorithm for the proposed search of compromised node is given in Figure 2.

**Figure 2: Structure of Genetic Algorithm**

The fitness function will compute the best solutions from the information mentioned from the initial population. The selection operators like mutation and crossover form the most effective parts in the algorithm when they involve in operation of the each iteration. On completion of a single iteration the best individual is selected. Selection operation has two procedures. Firstly, computing fitness value; secondly, queuing it from the smallest to the biggest. Then the min fitness value is the best individual, selecting the best individual as father-individual, the selection probability of each individual is proportional to its fitness value, the selected probability is higher when the individual fitness value is bigger. The two chromosomes chosen for crossover should have at least one common gene (node), but there is no requirement that they be located at the same locus. The population undergoes mutation by an actual change or flipping of one of the genes of the candidate chromosomes, which keeping away from local optima

**IV. PERFORMANCE ANALYSIS**

The NS-2 simulator was chosen to simulate the designed algorithm for detecting compromised nodes in a wireless sensor network. To begin the simulation process, there were several parameters altered for different runs to determine what kind of effect they had on various performance measures that were taken for consideration.

Figure 3 shows that the number of messages that are transmitted using GA are more compared to conventional methods.

**Figure 3: Analysis of Number of Messages**

The smaller networks take less time than the larger networks for two main reasons. The obvious reason is that there are more compromised nodes in the larger networks, so more times is needed to find all of them. The second reason is that there is more traffic generated in the larger networks, so it takes longer for the base station to process all the requests.

Figure 4 shows the average number of packets sent by a compromised node once it has been
injected into the network. Here that the number of compromised nodes is dependent on the network size. For example, the network with 40 nodes will include 4 compromised nodes. The algorithm works efficiently since only a few packets at most are sent compared to the hundreds of valid packets.

Figure 5 shows that the delay is reduced to a considerable level on using GA.

VI. CONCLUSIONS

We have presented an algorithm to detect the compromised nodes in wireless sensor networks. The base station is alerted to claims of abnormal behavior and verifies them by checking the difference in packet transmission times of the suspected node. Simulations are conducted to demonstrate the performance metrics. We believe that a base station that has more computation and power resources would further decrease the detection time. The concept implementation has proved that the rules generated by GA have the potential capability to detect the compromised nodes effectively in wireless sensor network environment.

The purpose of the work reported in the paper is to analysis the performance metric using genetic algorithm. The future work is to extend the scope of the rules involved.

REFERENCES


