NOVEL NODE COMPROMISE DETECTION APPROACHES FOR WIRELESS SENSOR NETWORK USING PARAMETER GROUPING AND INTELLIGENT GRADIENT MODELS

A

Synopsis

Submitted

in the partial fulfillment of the requirements for

the award of the degree of

DOCTOR OF PHILOSOPHY

By

MANYAM T

[1303PH0602]

DIRECTORATE OF RESEARCH AND DEVELOPMENT,

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD,

KUKATPALLY, HYDERABAD - 500 085, TELANGANA, INDIA.

Contents

1 Introduction

A Wireless Sensor Network (WSN) will have a large number of Sensor Nodes (SNs) to sense various physical parameters such as pressure, temperature, sound, etc [1]. The SNs in a particular location work together to collect and report the observations to a central monitoring system directly or indirectly. This central trusted system in a WSN, usually known as Base Station (BS), acts as a sink that can communicate with the SNs and with the external world. The sensor network information is connected to the digital world of computer machines to make informed decisions in order to accomplish a common application-specific task. The typical application areas are health care, battle-field surveillance, forest fire detection, etc [2]. Generally WSN applications are unattended after the deployment. Hence, an attacker can physically capture and compromise sensor nodes, and launch a variety of attacks through the compromised nodes. Node Compromise (NC) is a serious security threat for a WSN and that could undermine normal sensor network operations. Hence, it is necessary to detect compromised nodes and revoke them as soon as possible. The impact of compromised nodes are of in two types: Independent Compromised Node (ICN), it means compromise node will not effect its neighbours. Other one is Dependent Compromised Node (DCN), compromise node effects its neighbouring sensor nodes.

2 Motivation

To detect the independent compromised nodes, anomaly-based is one of the available method in the literature. Anomaly-based detection method uses defined normal profiles and detects abnormal deviations from normal behavior. The behavior is closely monitored based on certain parameters (features) to detect any deviation from normal behaviour. Existing research works use features like packet rate, node location, energy of a node, etc. All these works are based on only one parameter for anomaly detection. A single feature is not sufficient to decide whether a node is compromised or not. Use of single feature leads to false alerts. Hence, we have been motivated to use multiple features of sensor networks to detect any deviation in the network to mitigate false alerts. We have proposed AND, OR and PG models based on different parameters, namely, Packet Sending Rate (PSR), Node Location (NL), Depletion of Node Energy (DNE), False Information (FI) and Non-Availability of Node (NAN).

There is limited research work available for detection of DCN. DCN will be more devastating than ICN. To quantify the damage, Intelligent Models (IMs) based on probabilistic concepts are available in literature. To mitigate the false alerts, IM is augmented with Binary-Pattern-based PG Model.

3 Problem Statement

ZoneTrust (ZT) [3] is one of the existing works for independent compromise node detection. The concept of ZT is to identify untrust-

worthy zones and then apply software attestation on them to detect compromise nodes. ZoneTrust has the demerits of high attestation overhead as well as false reports.

In case of dependent compromised node, a compromised node will be exploited by an attacker to attack its immediate neighbours for further compromise. The intensity of the attack will vary based on the physical separation of the attacker from Base Station (BS): the more farther from the BS, the more intensive the attack will be. IM concepts are applied to quantify the probability of a node compromise. This model identifies the probabilistic compromised nodes which may not be really compromised nodes. Hence this model suffers from the false reports.

The problem statement is as follows: *It is proposed to mitigate the high attestation overhead and false positives by using multiple features of sensor nodes. These features include Packet Sending Rate, Depletion of Node Energy, False Information, Node Location and Non-Availability of Node. New NCD models, namely, AND, OR and PG [5] are proposed based on these features.*

To extend the proposed PG-based NCD model for dependent compromise node detection, IM concepts proposed by Xiangqian Chen et al [4] are used. Hence the above problem statement is extended as follows:

To detect a dependent compromise node by estimating the probabilities of node compromise using

- *Intelligent Modelling (A)*
- *Node behaviour (i.e multiple feature-based PG Model) (B)*

The effective (cumulative) node compromise probability is computed as weighted average of the above two compromise probability values (A&B). The nodes whose effective node compromise probability is greater than a threshold are considered as untrustworthy. Software attestation is applied on these untrustworthy nodes to decide the truly compromised nodes for subsequent necessary action such as revocation.

4 Research Objectives

The objectives of this research work are as follows:

- To study the existing NCD models based on the criteria, false alarms and software attestation overhead.
- To propose new models with reduced false alarms and software attestation overhead.
- To extend the proposed NCD models for dependent compromise node attacks with the help of IM concepts.

5 Research Contributions

The contributions of this research work are as follows:

- New parameters (features) of sensor nodes are identified for mitigating the false alarms and software attestation overhead of single feature-based NCD schemes such as ZT. The identified parameters are PSR, DNE, NL, FI, and NAN.

AND Model is proposed based on the conjunction of all these five parameters. In other words, a sensor node is declared as untrustworthy if it satisfies all the five parameters simultaneously. Hence, a sensor node is not identified as untrustworthy even if any one parameter does not hold good.

On the same line, another model called, OR, is proposed based on the disjunction of the five parameters. That is a sensor node is identified as untrustworthy if any one parameter is valid with respect to that node.

To overcome the demerits of both AND Model (false negative rate) and OR Model (false positive rate), a new holistic model called, Parameter Grouping (PG), is proposed. The five parameters are divided into three groups based on their relevance. Each group with more than one parameter is evaluated based on the conjunction of the parameters, i.e., a group is considered as valid if the conjunction of its member parameters is true. PG Model reports a sensor node as untrustworthy if the disjunction of all the three groups is true.

- The above PG Model for NCD is proposed to deal with ICN type, where node compromise of sensor node is limited to itself. But in reality, attacks will be usually of the DCN type, where a compromised node will impact all the nodes of a network through its neighbours.

  A research work named IM based on DCN type is available in literature proposed by Xiangqian Chen et al. There are two contributions by these authors, namely, uniform model and

gradient model. In uniform model, node compromise probability is same across the entire network irrespective of its position, whereas in gradient model, the compromise probability of node will increase as its distance from BS increases, i.e, the attacker will target the nodes which are farther from the BS rather than those close to it.

Proposed PG Model is utilized in conjunction with the above intelligent model to estimate the compromise probability of a node more accurately based on its behaviour as well as its position from the BS. Two contributions based on this work are made [6]: Intelligent Uniform Model with BP (IUMBP) and Intelligent Gradient Model with BP (IGMBP). Through simulative work, its observed that extended proposed models peform better than the IM models.

6 Proposed ICN Models

The main idea of the proposed models is to detect suspect nodes (untrustworthy nodes) which are likely placed as compromised nodes in zones. In these zones, the network operator performs software attestation against suspect sensor nodes only, leading to the detection and revocation of the compromised nodes. We show analytically and through NS-2 [7] based simulation experiments that the proposed models provide effective and with little overhead.

6.1 AND Model

The AND model evaluates to true when untrustworthy condition is met by all the five parameters. If any one of the five parameters is not satisfied the untrustworthy condition then AND model evaluates to false and concludes that the node is trustworthy. This may not be true as other four parameters are true. The condition of AND model to be verified at i^{th} node is C_i=$(PSR_i \wedge DNE_i \wedge FI_i \wedge NL_i \wedge NAN_i)$.

Some (not all) parameters are not satisfied at some nodes though they are already compromised. But the AND model does not detect these compromised nodes because all the parameters are not satisfying at those compromised nodes. Obviously, this model increases the vulnerability of the network for attacks (i.e., High risk). This model reduces attestation overhead when untrustworthy nodes are less. It suffers from false negatives.

6.2 OR-Model

It decides a node as untrustworthy when the disjunction of the five parameters is true that means atleast one parameter must be true. If all the parameters are false, then only a node is declared as trustworthy node. The condition of OR model to be verified at i^{th}node is C_i=$(PSR_i \vee DNE_i \vee FI_i \vee NL_i \vee NAN_i)$.

OR model increases the number of nodes to be applied the software attestation to decide whether they are really compromised or not (Even if one parameter is satisfied by a node, it calls for software attestation for compromised node detection). This increases the software attestation overhead for OR model. OR model has low

risk but it suffers from the false positives.

6.3 Parameter Grouping Model

The primary goal of PG is to meet balance between attestation overhead and the risk of false reports. The OR model has a primary advantage of low risk, whereas the AND model has the primary advantage of low attestation overhead. To keep the merits of both, it is required to combine them. It is to group the parameters based on some criteria (inter-related). For instance, PSR and DNE are inter-related as more packet sending rate results in more consumed energy. The parameters discussed earlier are divided into three groups, namely, G1, G2, and G3, where G1={ Depletion of Node Energy, Packet Sending Rate}, G2={False Information, Node Location}, G3={Non-Availability}. The condition of PG model to be verified at i^{th}node is C_i=$(DNE_i \wedge PSR_i) \vee (FI_i \wedge NL_i) \vee (NAN_i)$.

With PG Model false alarms (namely, false negatives and false positives) are minimized substantially. This means that the sensor nodes for software attestation are reduced drastically.

7 Proposed DCN Models

The main idea of this model is to estimate affected compromise node with respect to its immediate compromised nodes. Through extensive NS-2 [7] based simulative work and analytical study, it is proved that proposed extended IM models give better performance.

The IM model estimates the compromise probability of a node by using its neighbour's information. The node's compromise proba-

bility increases if it has a more number of compromised neighbour nodes. When a node collects information from its neighbours, a compromised node is likely to give false reply regarding the number of its neighbouring nodes. If it reports less value, the compromise probability increases; else the compromise probability decreases. The false higher compromise probability leads to false positives. Similarly the false lower compromise probability results in false negatives. Summarily, IM model suffers from false reports.

To resolve the issue of false reports, the IM model is extended with BP. To detect a dependent compromise node by estimating the probabilities of node compromise using

- (P) Intelligent Modelling
- (B) Node behaviour (i.e multiple feature-based PG Model).

The effective (cumulative) node compromise probability is computed as weighted average of the above two compromise probability values (P&B). The nodes whose effective node compromise probability is greater than a threshold are considered as untrustworthy. Software attestation is applied on these untrustworthy nodes to decide the truly compromised nodes for subsequent necessary action such as the revocation of compromised nodes. Weighted average probability of a node as given in Equation-7.1.

$$WP = \frac{P + B}{2} \qquad \ldots (7.1)$$

- WP= Weighted average Probability of a node.
- P =Node compromise probability computation using IM model.

- B=Node behaviour (i.e multiple feature-based PG Model).

Proposed PG Model is utilized in conjunction with the above IM to estimate the compromise probability of a node more accurately based on its behaviour as well as its position from the base station. Two contributions based on this work are made: Intelligent Uniform Model with Binary Pattern (IUMBP) and Intelligent Gradient Model with Binary Pattern (IGMBP).

7.1 Intelligent Uniform Model with BP

For some WSN applications such as environmental and health monitoring, it is reasonable to assume the uniform probability of node compromise. This is because sensor nodes are usually deployed within a limited area. Besides, the attacker can pick any sensor node for compromise irrespective of its location and/or its distance from the BS. In other words, the compromise probability is not sensitive to the location of the victim node for certain applications as mentioned above.

7.2 Intelligent Gradient Model with BP

For some WSN applications such as battlefield monitoring a uniform model may not be appropriate because nodes close to the enemy controlled area may have a greater likelihood of compromising than nodes away from the enemy controlled area. The Compromise probability of a node will vary as a parameter of its distance from the BS, high for remotely placed nodes and low for closely

placed nodes. This notion is supported by the Intelligent Gradient Model (IGM). The distinction between a uniform model and a gradient model is that the position of a sensor node may influence the node compromise probability in the last model, while it doesn't make a difference in the past model.

8 Conclusion

A thorough and detailed survey is carried out on the existing NCD approaches. Most of the work is based on usage of single parameter for detection of untrustworthy nodes. This results in false alarms and unnecessary attestation overhead for detecting the compromised nodes.

We have identified multiple (five) parameters for identifying untrustworthy nodes based on these parameters, we have proposed new NCD models, namely, AND, OR, and PG.

Proposed models are extended to work for detection of DCN attacks utilizing the IM concepts. We have two extended IM models namely, IUM with BP and IGM with BP. These are most applicable for WSN applications like military surveilance and forest fire detection. Through extensive simulative work and analytical study, it is proved that proposed extended IM models perform much better than IM models in terms of false alarms and attestation overhead.

9 Bibliography

[1] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: scalable coordination in sensor networks," *MobiCom'99: Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, 1999, pp.263–270.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks : a survey," *Computer Networks*, March 2002, pp.393–422.

[3] Jun-Won Ho, Matthew Wright, and Sajal K. Das, "ZoneTrust: Fast Zone-Based Node Compromise Detection and Revocation in Wireless Sensor Networks Using Sequential Hypothesis Testing," *IEEE Transactionson Dependable and Secure Computing*, Vol. 9 no. 4, July/August 2012, pp.494-511.

[4] Xiangqian Chen, Kia Makki, Kang Yen, and Niki Pissinou, "Node Compromise Modeling and its Applications in Sensor Networks," *IEEE Symposium on Computers and Communications*, Aveiro, Portugal, July 2007, pp.575-582.

[5] Manyam thaile, O.B.V.Ramanaiah, **"Behavioral Model for Detection of Compromised Nodes in WSN"**, International Journal of Advanced Research in Computer Science (IJARCS), Volume 9 No.2, April 2018, ISSN No. 0976-5697, pp. 1-6.

[6] Manyam thaile, O.B.V.Ramanaiah, **"Enhanced Intelligent Model for NCD in Wireless Sensor Networks"**, Transac-

tions on Networks and Communications (TNC), Vol.6 No.5, October 2018, pp.92-100.