DOI: http://dx.doi.org/10.26483/ijarcs.v9i2.5422

ISSN No. 0976-5697

Volume 9, No. 2, March-April 2018

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

BEHAVIORAL MODEL FOR DETECTION OF COMPROMISED NODES IN WSN

Manyam Thaile
Computer Science and Engineering, JNTUH-CEH,
Hyderabad, India

Dr. O.B.V. Ramanaiah
Computer Science and Engineering, JNTUH-CEH,
Hyderabad, India

Abstract: Node Compromise Detection (NCD) is an crucial requisite for transaction with possible attacks in regular distribution or random distribution in Wireless Sensor Networks. We designed Parameter Grouping (PG) model for compromised nodes. We extenuate burden of communication and attestation, also reduced false negatives and energy depletion when compared with available methods, like ZoneTrust. An Network Simulator version-2 based simulation was carried out and found that the PG model discovers compromised nodes effectively.

Keywords: WSN security; node compromise; AND-OR; parameter grouping; attestation.

1. INTRODUCTION

The sensor nodes deployment in WSN (Wireless Sensor Network) are in two types: random or regular and also located a large numbers. The physical status sense by sensor node depend on those parameters which we set. It changes the observed data into digital data, then operate it and communicates with other nodes or sink (base station) together with these outcomes. There are numerous applications covering by WSN which are Military, Home Security, Crop Pest Control, Civil Structure Monitoring, etc.,. A node is having less facilities for computing, storage, and node energy. The alive of network is quite low because the substituting or loading a new battery of a node is impossible. A dangerous security menace in WSN applications is node compromisation because of the weakness of the sensor network: unattended nature of the network, deficiency tamper-resistant hardware, the low computing power of nodes (incapability to run software-based security concepts like the firewall), unreliable communication, etc. Consequently an antagonist can be easily catch the sensor nodes and also well compromise them in two directions: either remotely or physically. An adversary would connect captured node to a high-end machine, disclose the personal keys, deploy the malicious code, and because of that node is compromised. An antagonist can set up numerous types of attacks [1],[2] with the help of compromised nodes for break the network functions and/or changed the sensed data or aggregated information maliciously. Hence, it is highly significant to find out and remove the compromised nodes very quickly.

The detection of compromise nodes are of two cases: reputation, attestation. First one is hold separate node trust based on the its activities [3][4][5]. The positive reports and negative reports are false in reputation. Attestation [6][7][8][9] checks the code working on the node against the available code at BS (Base Station) to find out malicious program which incurs a lot of overhead on trusted nodes. Few researchers have invented a method, that is to merge the above two methods to alleviate their disadvantages such as false positive rates and overhead (e.g, ZoneTrust).

ZoneTrust (ZT) [10] has two mechanisms: 1. Reputation, 2. Attestation. Initially ZT approach detects untrustworthy zones by applying reputation on the network, then it applies attestation on all nodes of those untrustworthy zones found by reputation to identify compromise nodes. But there is a restriction in this, that is even if any zone is chosen as untrustworthy, that doesn't mean all the nodes in that zone are compromised. There are few nodes which are compromised so it hurts to that zone is untrustworthy. This network overhead exhaust more energy of sensor nodes. To extenuate ZoneTrust drawbacks, we designed a new model called Parameter Grouping (PG). The prelude work of Parameter Grouping model is promulgated in [12] and this paper is discussed comprehensively.

1.1 Proposed Work

The PG model is to find out untrustworthy nodes depend on the nodes demeanour. After finding of unbelief nodes, then attestation is practiced against those unbelief nodes to detect and remove the compromised nodes. Figure-1 displays the methodology used for NCD.

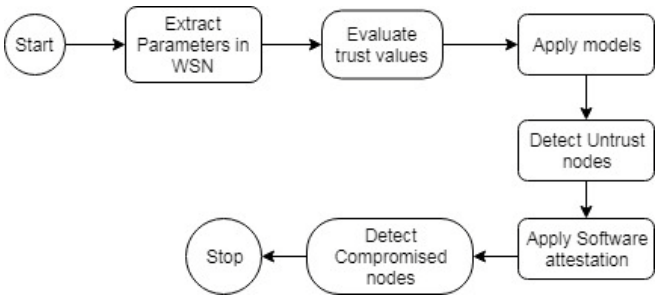

Figure 1: Methodology used for NCD

The PG model is designed based on the multiple parameters for untrustworthy nodes. We identified five parameters such as Packet Sending Rate (PSR), False Information (FI), Node Location (NL), Depletion of Node Energy (DNE), and Non-Availability of Node (NAN) to observe misconduct of the sensor nodes. The trust values are calculated depend on the information given by the sensor nodes. The information of sensor nodes is input for the five parameter DNE, PSR, FI, NL, and NAN. The models include mathematical concepts such as Entropy, Euclidean Distance and Standard Deviation (SD) etc., to calculate the trust values. These trust values are used by PG to find untrustworthy nodes. The PG model

consists of two approaches, AND-OR model and PG . The introduced AND-OR model is merging of AND model and, OR model.

The AND model calculates to true when untrustworthy condition is ($\wedge$) met by all 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. Hence it suffers from false negatives.

The OR model calculates to true when untrustworthy case is ($\vee$) satisfied by atleast one of the five parameters, then it represents that the node is untrustworthy. The hypothesis of untrustworthy nodes in this model is high compared to AND model, because here in this model atleast one parameter should satisfy the condition of untrustworthy whereas in AND model all the five parameters must satisfy the condition of untrustworthy. Both AND model and OR model have drawbacks of false negatives and overhead. To eliminate this problem we use PG.

The arrangement of sections is shown as: Surveyed literature on NCD in Section 2. The network model is discussed in section 3. An attacker model is discussed in section 4. Section 5 is given detailed concept of PG models. Section 6 discussed compromised nodes detection and revocation. Section 7 displays simulation output. Section 8 explained performance analysis. Section 9, closed with conclusion and future plan.

2. RELATED WORK

The prevention and detection are two categories to discover compromised nodes, as displayed in Figure.2.

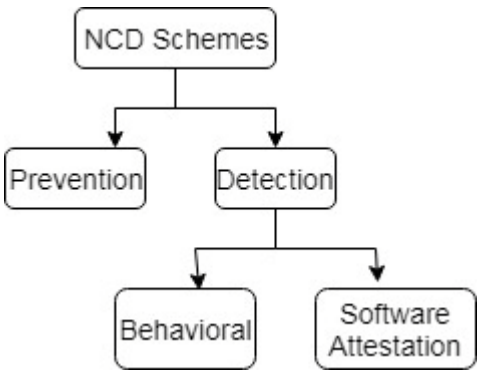

Figure 2: Classification of NCD Schemes

The first one is the first line defence for protecting sensor nodes using cryptography. Authentication and Encryption are the elementary criterion in Prevention . If the first defence is crushed, then the attacker could extract security-sensitive information (e.g., secret keys) that leads to disclose of security. The other direction is behavioral system to answered first line defence drawback. Few scholars designed different methods for the ompromise nodes depend on reputation and attestation. These methods aim to find out misbehavior of a node and checks nodes code integrity.

Reputation is to observed single nodes trustworthiness based on several parameters like packet arrival rate, packet sending rate, packet arrival time, node energy, and node location, [3,4,5] etc.,. The paper [4] introduced a reputation concept in which a Bayesian rule is used to compute single nodes trustworthiness. The paper [5] designed a method about information theory model for trustiness rating. An entropy and probability are used to calculate single nodes trustiness. However, malicious (compromise) nodes cannot be easily eliminated due to the false positives availability.

Second is to find out compromised nodes depend on the changed software code of nodes [6-9]. Generally, all attestation methods are required to each and every sensor node in that network to be attested, in practical all the nodes are not compromised. But in this case honest nodes are also part of attestation. Hence the resources of honest nodes are useless, even though they do not need to be attested.

To mitigate false positives and unnecessary attestation, ZoneTrust [10] and NodeTrust [11] have combined both the Reputation and Attestation schemes. They followed a two-step procedure: The first step is Identifying an untrustworthy zone or a node, and the second step is Software attestation for an untrustworthy zone or a node. In the step 1, misbehaving node is identified using different parameters such as packet arrival rate, packet sending rate, packet arrival time, node energy, and node location, etc, then we apply step 2. In the Step 2, After the identification of misbehaving (untrustworthy) nodes, the BS checks whether software codes of these nodes have been maliciously altered or not by performing attestation.

3. NETWORK MODEL

WSN is not in physical motion network, in which the sensor nodes positions are frozen. BS is an honest node that cannot be compromised and fixed at the middle of the network coverage area. The fixing of sensor nodes are uniform to eliminate breach in sensing and nodes operate in a hierarchical manner. The coverage of the network is split into non-intersection regions called zones. The zones are formulated depend on the Euclidean Distance. A node in every zone is randomly selected to act as Zone Head (ZH) for one iteration (for fixed interval of time). This node cannot be considered again as ZH in following iterations. For the following iteration some other node is selected as ZH. The iteration will take upto a live network.

4. ATTACKER MODEL

An attacker assumes to be an active attacker with the purpose of capturing nodes to uncover private data stored by the captured nodes. The attacker may also keep malicious code onto the trusted nodes and prepare them as compromised nodes. An attacker re-launch the nodes back into the network to keep promote attacks [15]. Therefore, to eliminate compromise nodes are very crucial. An attacker is able to compromises some of nodes in the WSN. The nodes can be disrupted by adversary with ten percent in the network.

5. PARAMETER GROUPING (PG)

Parameter Grouping model is used as the second line of defense for WSN. The behavior of a sensor node is defined as the way it operates and responses to network activities. We discuss four trust metrics to calculate trustworthiness based on euclidean distance, entropy, standard deviation and difference [5].

Entropy: It is assessed uncertainty given by information

theory [13]. The entropy trust value is defined as:

$$Trust(T_i) = \begin{cases} 1, for \quad H(p) > Th; \\ 0, for \quad H(p) < Th, \end{cases} \tag{1}$$

where H is the entropy function [13] and Th is threshold.

Euclidean Distance: it is calculated length betwixt two known places of the sensor nodes [14]. The trust value is defined as:

$$Trust(T_i) = \begin{cases} 1, d(s, p) > Th \\ 0, otherwise. \end{cases} \tag{2}$$

where $d(s, p)$ is distance between s and p nodes.

Standard Deviation (SD): it is used for measurement of false report. The trust value is defined as:

$$Trust(T_i) = \begin{cases} 1, \sigma < Th_1 \,||\, \sigma > Th_2 \\ 0, Th_1 \le \sigma \le Th_2. \end{cases} \tag{3}$$

where σ is Standard Deviation.

Difference: it is helped to calculate non-availability of time for any sensor node. The trust value is defined as:

$$Trust(T_i) = \begin{cases} 1, t_i - t_{i-1} > Th \\ 0, Otherwise. \end{cases} \tag{4}$$

Where t_i is current interval time and t_{i-1} is previous interval time. **Note:** Th is threshold, 1 means untrust and 0 means trust.

5.1 Parameters proposed for NCD

Five parameters are identified based on the regular activities of sensor node in WSN. The five parameters are defined as below:

• **PSR (Packet Sending Rate)**:

The number of packets sent out going link per unit time. The trust values is calculated for the PSR by using Entropy (E) (1,0). It is pretended that each sensor node transmit a packet to Zone Head in all intervals. The ZH is maintained packets sent by all the members of the zone. The PSR entropy value of a i^{th} node after δ number of intervals is expressed mathematically as below:

$$PSR_E(Node_i, \delta) = \frac{\sum_{k=1}^{\delta} P_k}{\delta} \times log \frac{\sum_{k=1}^{\delta} P_k}{\delta} \tag{5}$$

where $Node_i$ is i^{th} node, p_k is total of packets transmitted at k^{th} interval and δ is number of slots/intervals.

• **DNE (Depletion of Node Energy)**:

The node energy is depleted at which rate. The trust values (1, 0) can be calculated for DNE with help of the Entropy. The DNE consumed due to high packet sending rate and other reasons (drop rate, receiving rate etc). Mathematically, DNE entropy value of a i^{th} node after δ number of intervals is expressed as below:

$$DNE_E(Node_i, \delta) = \frac{\sum_{k=1}^{\delta} \Delta E_k}{\delta} \times log \frac{\sum_{k=1}^{\delta} \Delta E_k}{\delta} \tag{6}$$

where $Node_i$ is i^{th} node, ΔE_i is energy consumed at k^{th} interval and δ is number of slots/intervals.

• **NL (Node Location)**:

It refers to the physical location of a node. Euclidean Distance is helped to calculate the trust values (1, 0) for the NL. The sensor network is fixed, i.e., after deployment the locations or positions of nodes cannot change. If the position of any node is changed unusually, then that node is considered as untrustworthy. The localization methods are used to find the places of nodes [14]. The initial location of i^{th} node is $s(x, y)$, after δ number of intervals the location is $p(j, k)$. Euclidean Distance between s at (x, y) and p at (j, k) as $D(s, p)$ mathematically represented as:

$$D(Node_i, s, p) = \sqrt{(x - j)^2 + (y - k)^2} \tag{7}$$

• **FI (False Information)**:

It is depend on the rightness of the information transmitted by the node, hence it takes value either true (correct information) or false (incorrect information). SD was applied to calculate trust values (1, 0) for the False Information. Every sensor node is awaited to communicate to Zone Head (ZH) in a predefined pattern with an expected size which should be consistent with all other nodes reports. We use synthetic temperature as input for SD for measuring the temperature reports. It is defined as below:

$$\sigma(Node_k) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{8}$$

where $\bar{x}$ is mean of temperature, x_i temperature values, N is number of reports.

• **NAN (Non-Availability of Node)**:

When a node is being captured by an attacker, then the node is not available for communication by other sender nodes for a certain duration of time, therefore it considered as untrustworthy. If a node transmit data periodically to ZH, it shows its presence in the network. The trust values (1, 0) can be calculated for the NAN with help of the difference.

5.2 AND-OR Model

The motive of proposing AND-OR model is to extenuate drawbacks of ZoneTrust. The AND-OR model calculates unbelief sensor nodes depend on the binary values (1, 0) of above discussed parameters.

AND model: It identifies a node as untrustworthy when the connects of the five parameters is true. If out of five, one of the parameter is false, then that node has taken as trustworthy node. Weather monitoring applications are suitable by AND

model. The condition of AND model to be verified at i^{th} node is $C_i = (DNE_i \wedge PSR_i \wedge FI_i \wedge NL_i \wedge NAN_i)$. Node Status (NS) is defined as follows:

$$NS_i = \begin{cases} Untrust, if\ (C_i == 1) \\ Trust, else \end{cases} \quad (9)$$

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. It is suitable for military surveillance. The condition of OR model to be verified at i^{th} node is $C_i = (DNE_i \vee PSR_i \vee FI_i \vee NL_i \vee NAN_i)$. Node Status (NS) is defined as follows:

$$NS_i = \begin{cases} Untrust, if\ (C_i == 1), \\ Trust, else \end{cases} \quad (10)$$

5.3 Parameter Grouping

The primary incite of PG is to hit balance between attestation overhead and the risk of attack. The OR model has a primary advantage of low risk, whereas the AND model has the chief advantage of low 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, consumed energy and packet sending rate are inter-related as more packet sending rate results in more consumed energy. The parameters discussed ahead 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}. Mathematically, the i^{th} node is declared as *Node Untrusted* (NU) when:

$$NU_i = \begin{cases} (DNE_i \wedge PSR_i) \vee (FI_i \wedge NL_i) \vee (NAN_i) \\ or \\ (G1_i \vee G2_i \vee G3_i) \end{cases} \quad (11)$$

where G1= $(DNE_i \wedge PSR_i)$, G2= $(FI_i \wedge NL_i)$, G3= (NAN_i).

6. COMPROMISED NODES DETECTION AND REVOCATION

After finding out untrustworthy nodes by using behavioral method, that is Parameter Grouping, BS applies attestation to untrustworthy nodes to check the changing of code in it. The MD5 algorithm [16] is used for attestation, because it is suited for WSN as it obtains low computational overhead. Once the network is deployed in the beginning, the attestation of all the sensor nodes using MD5 algorithm is computed and stores resultant hash values at BS. During the process of identifying untrustworthy nodes of the network, if any untrustworthy node is found, then we calculate the hash value of that untrustworthy node using same MD5 algorithm. If both the hash values are not same, then the node is considered as compromised. If both the hash values are same, then the node is trustworthy. These steps are algorithmically indicated in the algorithm 1. After identification compromise nodes, elimination is done in two ways: 1. Manually new sensor nodes are replaced in the place of compromise nodes. 2. Compromised node code is manipulated.

Algorithm 1: Detection of compromised nodes

1: Compute hash code at the initial deployment by the BS
2: **For** Number of nodes **do**
3: Hash1[ ]=compute hash for all nodes
4: end for
5: Compute hash code for the untrusted nodes
6: **For** Number of untrusted nodes **do**
7: Hash2[ ]=compute hash for all untrusted nodes
8: end for
9: **if** (Hash1==Hash2) **then**
10: Benign nodes
11: else
12: Compromised nodes
13: end if

7. SIMULATION STUDY

In this section, simulation environment and simulation results are discussed.

7.1 Simulation Environment

We used Network Simulator version 2 (NS-2) [17], which is an open source simulator. We analysed proposed models: AND, OR and PG. The network is considered as hierarchical, simulated with 100 nodes, 10 zones, 10 ZHs and one BS.

7.2 Simulation Results

In this section, the performance of the proposed PG, AND, and OR are evaluated, these experimental results are compared to ZoneTrust.The given five metrics are helped to determine the accuracy of the proposed schemes: True Positive, False Negatives, False Positives, Recall (R), and Precision (P). The R and P values are helped to calculate the accuracy of the designed methods to find out untrustworthy nodes in WSNs. These five metrics are defined as follows.

- **True Positive (TP):** A trustworthy sensor node is noticed as a trustworthy node.
- **False Positive (FP):** A trustworthy sensor node is noticed as untrustworthy.
- **False Negative (FN):** An untrustworthy node is noticed as a trustworthy node.
- **Recall** = $\frac{TP}{TP + FN}$
- **Precision** = $\frac{TP}{TP + FP}$

The perofrmance analysis of the designed models depend on the above metrics is showed in Table 1.

Table 1: Performance analysis

Model	Nodes	Compromised Nodes	FP	FN	TP	Recall	Precision
Zone Trust	100	10	90	0	100	1	0.52
AND	100	10	0	2	100	0.98	1

OR	100	10	10	0	100	1	0.92
PG	100	10	0	0	100	1	1

Table 2: Untrustworthy Nodes (UN) Analysis

Model	Nodes	Compromised	UN	Really compromised	Attestation
Zone Trust	100	10	90	10	90
AND	100	10	10	10	10
OR	100	10	20	10	20
PG	100	10	10	10	10

The performance analysis of the ZT, AND, OR, and PG models is indicated in the Table 2, and same is plotted as graphs as shown in Figures 3 and 4.

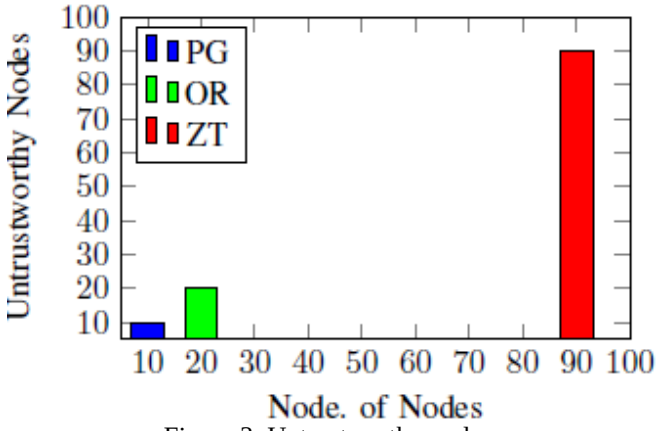

Figure 3: Untrustworthy nodes

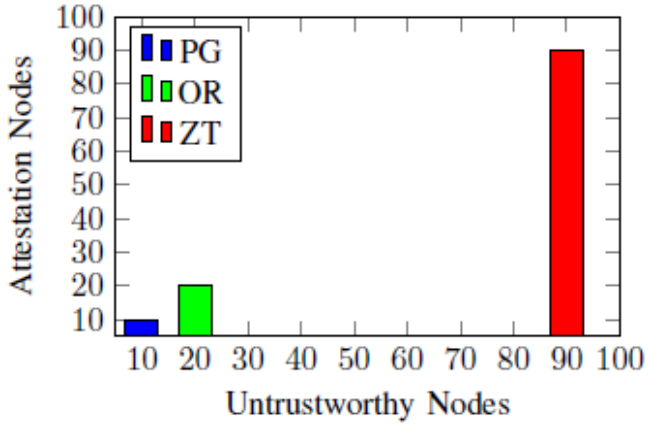

Figure 4**:** Attestation analysis

8. PARAMETER GROUPING PERFORMANCE ANALYSIS

The PG overhead of communication and attestation are analyzed.

Communication Overhead:

The Communication Overhead is defined as the number of node-trust reports that are sent or forwarded by the Zone Heads in the network. Let k be the number of zones, which are identified as z_1, z_2, z_3, z_4,.... z_k. The number of members (nodes) of i^{th} zone (z_i) is m_i (i=1,2,...k). The BS gets at most k (Number of Zone Heads=Number of Zones) number of node true reports in each time slot. The $O(\sqrt{N})$ is the mean hop length between two any chosen nodes [10], where N is all nodes in the sensor network. It is to be noted that the BS is a designated sensor node placed inside (at the center) the network. Hence the mean hop length between a ZH and the BS is $O(\sqrt{N})$. Therefore, $O(k\sqrt{N})$ is the mean number of node-trust reports per time slot. The communication overhead of PG model is much more less than when compared with ZoneTrust because PG sends only untrustworthy reports to BS whenever untrustworthy reports are available.

Attestation Overhead:

Let us assume that $O(1)$ is the attestation (by BS) overhead per node. In the worst case if one node per zone is compromised then ZoneTrust concept declares every zone as untrustworthy as per the procedure [10]. Hence it is necessary to attest all nodes of the network. Hence, the entire overhead is $O(N)$, where N is all sensor nodes. The worst case attestation of PG model is $O(Q)$=($O(Q) << O(N)$), by reason of BS should be launched only on untrustworthy nodes but not on the across zone. where Q is represented the all unbelief nodes.

9. CONCLUSION AND FUTURE PLAN

The Parameter Grouping is exploited for compromised node detection in a zone-wise sensor networks. It detects effectively untrustworthy sensor nodes. The overhead of attestation and communication of the proposed models are reduced when compared with ZoneTrust, hence automatically node lifetime also increased. The false positives and the false negatives rates are also minimal. Parameter Grouping model simulated and results showed that it detects quickly untrustworthy nodes.

In future work, how to analysed compromised nodes intelligently by using the Parameter Grouping.

REFERENCES

[1] B. Parno, A. Perrig, and V.D. Gligor, "Distributed Detection of Node Replication Attacks in Sensor Networks," May 2005, Proc. IEEE Symp. Security and Privacy (S&P), pp. 49-63.

[2] Mr. Manish M Patel, Dr. Akshai Aggarwal, "Security Attacks in Wireless Sensor Networks: A Survey," 2013, International Conference on Intelligent Systems and Signal Processing (ISSP), pp. 329 - 333.

[3] F. Li and J. Wu, "€œMobility Reduces Uncertainty in MANET,"? May 2007, Proc. IEEE International Conference on Computer Communications, pp. 1946-1954.

[4] S. Ganeriwal and M. Srivastava, "Reputation-Based Framework for High Integrity Sensor Networks," Oct. 2004, Proc. ACM Workshop Security of Ad Hoc and Sensor Networks (SASN), pp. 66-77.

[5] Y. Sun, Z. Han, W. Yu, and K. Liu, "A Trust Evaluation Framework in Distributed Networks: Vulnerability Analysis and Defense against Attacks," Apr. 2006, Proc. IEEE INFOCOM, pp. 1-13.

[6] A. Seshadri, A. Perrig, L. van Doorn, and P. Khosla, "SWATT:SoftWare-Based Attestation for Embedded Devices,"? May 2004, Proc. IEEE Symp. Security and Privacy

(S & P), pp. 272 -282.

[7] T. Abuhmed, N. Nyamaa, and D. Nyang, "Software-Based Remote Code Attestation in Wireless Sensor Network," December-2009, Proc. of IEEE Global Telecommunications Conference, pp. 1-8.

[8] T. Park and K. G. Shin, "Soft Tamper-Proofing via Program Integrity Verification in Wireless Sensor Networks," vol. 4, no. 3, May/June 2005, IEEE Trans.Mobile Computing, pp. 297-309.

[9] Y. Yang, X. Wang, S. Zhu, and G. Cao, "Distributed Software-Based Attestation for Node Compromise Detection in Sensor Networks," Oct. 2007, Proc. IEEE 26th International Symp. Reliable Distributed Systems (SRDS), 219-230.

[10] 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," vol. 9, no. 4, July/August 2012, IEEE Transactions on Dependable and Secure Computing, pp. 494-511.

[11] Manyam Thaile, and O.B.V Ramanaiah, "Node Compromise Detection Based on NodeTrust in Wireless Sensor Networks,"Jan, 2016, An IEEE International Conference on Computer Communication and Informatics (ICCCI),Coimbatore, INDIA, pp. 193-197.

[12] Manyam Thaile, and O.B.V Ramanaiah, "Node Compromise Detection Based on Parameter Grouping in Wireless Sensor Networks," ISBN: 978-1-61208-493-0, July. 24-28, 2016, SECURWARE 2016:The Tenth International Conference on Emerging Security Information, Systems and Technologies, Nice-France, pp. 14-20.

[13] Thomas M. Cover, Joy A. Thomas, "Elements of Information Theory," Second Edition, 2006 , Published by John Wiley & Sons, Inc.

[14] H. Rashid, A. K. Turuk, "Localization of Wireless Sensor Networks Using a Single Anchor Node", 2013, Wireless Personal communications (Springer), 72(2).

[15] G. Padmavathi and M. D. Shanmugapriya, "A survey of attacks, security mechanisms and challenges in wireless sensor networks," vol. 4, no. 1, Aug. 2009, Int. J. Comput. Sci., pp. 1-9.

[16] R. Rivest, "The MD5 Message-Digest Algorithm," RFC:1321, Network Working Group, 1992, MIT Laboratory for Computer Science and RSA Data Security Inc.

[17] Teerawat Issariyakul, Ekram Hossain, "Introduction to Network Simulator NS2", 2008 , Springer Science, DOI: 10.1007/978-0-387-71760-9.