Ilkogretim Online - Elementary Education Online, 2020; Vol 19 (Issue 4): pp. 7025-7038
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2020.04.765107

Outlier Detection Techniques in WSN

T P Rani , S Susila Sakthy Sri Sai Ram
Engineering College, Chennai

Abstract

Wireless sensor network is an emerging technology and gaining importance because of its nature of sensing and diverse applications in which it is deployed. Just as there is an increase in applications, likewise erroneous data transmission in Wireless Sensor Network is inevitable. It is because sensor nodes may be deployed in harsh unattended environments. It may be disturbed by environmental factors or malicious nodes. During transmission of data, due to noise or external intervention, the data may turn into erroneous data. Such deviated data is an outlier. As the sensor nodes are used in crucial applications like industrial monitoring system or a patient monitoring systems, an erroneous data should be avoided. In this paper we propose to detect outliers based on Dempster-Shafer Theory. The results obtained are compared using NS2 and JTOSSIM simulators. By detecting outlier nodes, the resilience against node capture can be greatly enhanced as well. Comparative studies of challenges of sensor network and outlier detection methods are provided so that readers can have a rudimentary analysis of the security issues.

Index- Termsoutlier, erroneous data, wireless sensor networks, Dempster-Shafer theory.

I. Introduction

Wireless sensor network (WSN)[1] consists of a number of autonomous sensor nodes that work in cooperation with each other to perform event monitoring, data collection and filtering. Recent years WSN has gained its importance because of its wide range of applications and also due to the evolution in MEMS technology that has initiated sensor manufacture in terms of microscopic scale. The applications of WSN are not limited to Structural monitoring, Bio-habitat monitoring, Industrial monitoring, Disaster management, Military surveillance, home or building security system. There are various types of sensors ranging from normal temperature and humidity sensors to the tactile sensors. These sensors are incorporated into a sensor node along with a memory, transceiver and a control unit to form a sensor node. The sensor node monitors the data for which it is designated. The data from the sensor nodes are collected by the common nodes such as a sink node, the cluster head or the base station. The sensor nodes during data collection, normally performs data aggregation to minimize communication cost. The main challenge srelated to WSN [1] are discussed in the following section.

- Limited storage: Sensor nodes are smaller in size and hence have limited memory space. Most of the applications require more number of sensor nodes to be deployed for performing the various operations related to sensing the environment. The number of operations related to sensor nodes should be limited in order to limit the memory size.
- Low energy: Sensor nodes rely on battery life as the main source of energy. Sensor nodes are mostly battery operated and in most applications one time sensors are used. If battery gets depleted, the nodes would die. However, few applications use external energy source such as solar energy. This becomes difficult for all applications.
- Limited operations and communication: As sensor nodes have very less energy in the form of battery life, the number of operations and the communication exchanges between the nodes has to be limited.
- Deployment: In some applications like military surveillance, environment monitoring system and bio- habitat monitoring systems, sensor node may be deployed in unattended environments and are subject to anonymity.

In spite of the challenges with the use of sensors in WSN, another significant issue to be considered is to provide security for communication. As sensor nodes are deployed in distributed and hostile environment, there are various security threats[2] [3].The main security threats and issues related to WSN are discussed in the followingsection. The security attacks related to sensor nodes and WSN is classified as shown in Figure 1.1.

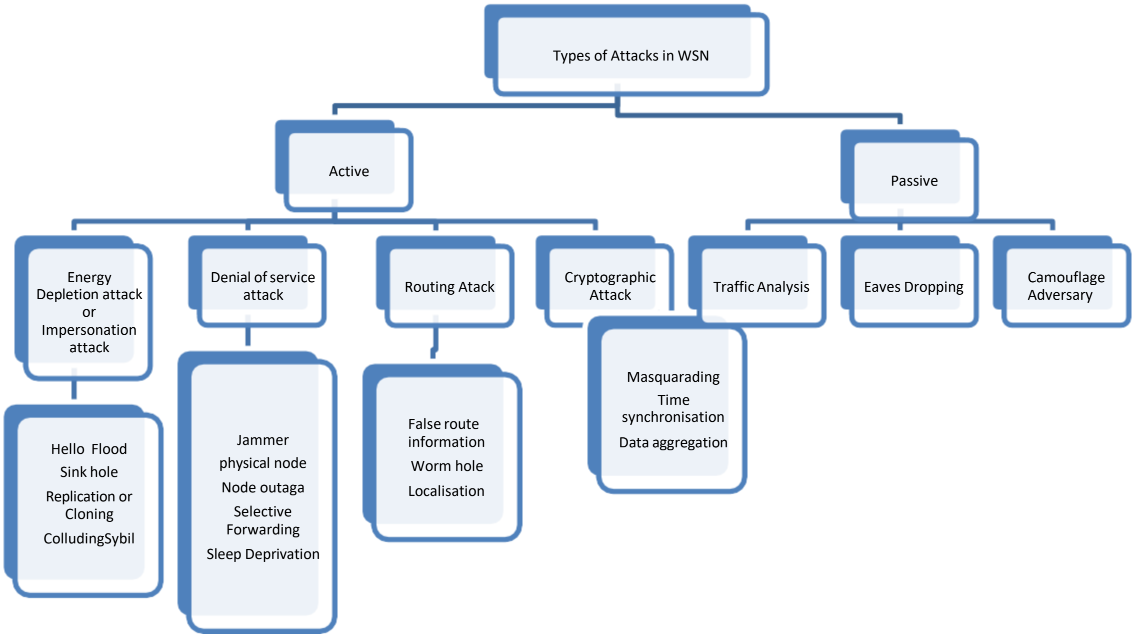


Fig 1.1 Types of security attacks in WSN

Energy depletion attack: The energy can be depleted on its own due to the nature of application or an external attacker can continuously probe the system to cause energy depletion. The different forms of attacks that cause energy depletion in sensor nodes are flooding attack [4], sink hole attack [5], clone attack or replication attack [6], [7], colluding [8]and Sybil [9] attacks.

Cryptographic attacks: Cryptography is one of the techniques that can be adapted to overcome the intruders during data transmission [10]. The public key cryptography is commonly used in WSN [11]. In some of the applications the private keys [11] used for encryption and decryption are implanted in sensor nodes. These keys once hacked cannot be regenerated after deployment [12]. Those nodes become useless because of the exposure of the private key and should be removed from network. Masquerading attack, Time synchronization attack and Data aggregation attack are some of the cryptographic attacks [11].

Other attacks related to WSN are Routing attack, Impersonation attack, Denial of service attack (DOS) [3], [13]. In Routing attacks, the malicious nodes may forward packet or routing information to a genuine node. The data that is routed to a genuine node may be faulty. False route information attack, Worm hole attack [14], [15], Localization attack are some of the routing attacks. In Impersonation attack [3], the genuine nodes may be impersonated with the aid of the credential details obtained from it. Some of the attacks that belong to this category also belong to the category of energy depletion attack. It may then play any role as in replication attack, colluding attack, spoofing attack or masquerading attack. If there is a delay in responding by the nodes or if it is reluctant to work, it may lead to Denial of service attack (DoS) [3]. The foremost type of DoS attack is the physical node attack and jamming attack [16],[17]. Node Outage [3], Selective forwarding [18] and Sleep deprivation attacks [19]are some examples of DoS. As in any networks, confidentiality, availability, authenticity and integrity are primary goals of providing security for WSN. This paper focuses on providing data integrity during data aggregation in WSN.

II. Related Work

Sensor nodes have limited computation and communication capabilities. The sink node aggregates the data from a group of sensor nodes and the sensor nodes would report the sink node in case of any abnormality in situation. Due to its limitations of power and resources there is a possibility of erroneous data either in the way it is reported by the sensors or the way it is collected by the sink node. Additionally, as the sensor nodes are deployed in rigid environments the data transferred may be affected by noise or any malicious user or node may tamper and try to modify the data. The erroneous data which deviates from normal data is saidto be an outlier.

In some of the applications sensor nodes are used to monitor for an occurrence of an event. For instance, Forest fire monitoring system, the nodes report the sink node if there is any abnormal change in temperature for the occurrence of the Forest fire event. The event detection is a central issue in wireless sensor networks. The process of event detection has been implemented using empirical distribution based on evidence theory. The abnormality is recognized based on the data reported by the sensor nodes. The accuracy of the eventdetection is completely dependent on the accuracy of the reported data. This requires mechanisms to detect outlier data present in the reported data to determine the accuracy of the reported data by eliminating the outliers.

III. Proposed system

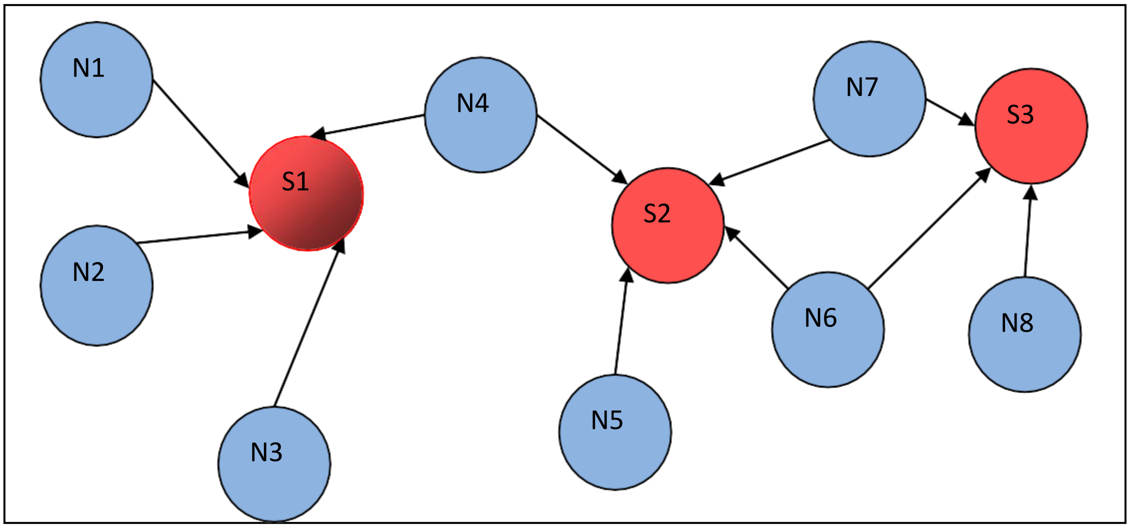
This paper focuses on error as an outlier source. If such a deviated data is transmitted and data aggregation is performed resultant data may not be accurate. In this paper, we propose a method based on Dempster-Shafer (D-S) Theory [36] for outlier detection in WSN based on Bayesian Belief Network Models [33]. It belongs to the category of Classification based approach [37]. It is a distributed prediction mechanism. It is based on mathematical theory of evidence. It combines evidence collected from other sources and conclude ata degree of belief.

The D-S theory was the evidence theory introduced by Dempster in 1960s and was later modified by Shafer, hence the name. In the D-S theory, for a data set, the degree of belief [36] or also called as mass is represented as belief function. Belief is a hypothesis that is constituted by the sum of masses of all set enclosed by it i.e. sum of masses of all subsets of hypothesis.

The proposed algorithm is based on Dempster-Shafer theory which is a mathematical theory of evidence. The algorithm has been implemented in the nodes in a distributed way. The Base station or the cluster head which collects data from sensor nodes performs the calculations in determining outliers. A node which repeatedly reports, (threshold value in the system is 6) with the erroneous or outlier data is designated as outlier node. Once the outlier node is detected, it can be removed from the list of nodes to be communicated by the cluster head or the base station. The information about the outlier nodes would be broadcasted to every other node and the outlier node would be blacklisted from communication. This would protect WSN from further malicious communications which would improve the accuracy of sensed data.

IV. System Architecture

The system architecture is shown in Fig 4.1.S1, S2 and S3 are the nodes which serve as the sink node. The sink node is a normal sensor node with additional functionality as data gathering and thereby data aggregation, in addition to the work of monitoring the temperature. Hence there is a possibility of decrease in energy level due to additional functionalities; there is a possibility of decrease in life of a node. Hence the responsibilities of a sink node are periodically shared among the sensor nodes. In fig N1, N2 etc are the temperature sensor nodes. The sink nodes perform the data aggregation and detect outliers from the data collected from the temperature sensor nodes. The data aggregation may be finally carried out at the base station after some level of data aggregation at the sink node.



V. ALGORITHM

For the obtained values from the sensors the base station or the cluster head performs thefollowing steps.

1. Calculate the pivot p which is the average$_1$of the data D. $\boldsymbol{p} \sum^{n} \boldsymbol{D}$
2. Calculate the difference d of each D and pivot p.
   a. $\boldsymbol{d}[\boldsymbol{i}] = \boldsymbol{D}[\boldsymbol{i}] - \boldsymbol{p}$
3. Calculate Pivot/difference pd for each D.
   a. $\boldsymbol{d}[\boldsymbol{i}] = \boldsymbol{p}/\boldsymbol{d}[\boldsymbol{i}]$
4. Calculate sum s of pd
   1 a. $\boldsymbol{s} = \sum^{n} \boldsymbol{pd}$
5. Calculate Mass m for each D.
   a. $\boldsymbol{m}[\boldsymbol{i}] = \boldsymbol{pd}/\boldsymbol{s}$
6. Calculate Belief bel.
   a. $\boldsymbol{bel} = \sum \boldsymbol{m}$
7. Calculate Plausibility
   a. $\boldsymbol{pl} = \mathbf{1} - \boldsymbol{bel}.$
8. Find the whether there is any $pl_i>0.5$, if $pl_i<0.5$, the node is not an outlier, if yes declare the node asoutlier and black list the node.

Consider an example with a forest fire monitoring system. The system uses temperature sensors. When there is no abnormal temperature the data obtained from sensors are 37, 38, 38, 38 and 37. All the sensors have more or less the same value. If a forest fire occurs, then the data collected from sensors can be 107, 105, 106,107 and 107. But consider the data obtained as 37, 38, 16, 38 and 37. It is understood that 16 is the varying data. We calculate the plausibility for the data. The pivot p is 33.2. The differences between data and pivot are 3.8, 4.8, 17.2, 4.8 and 3.8. After the Pivot/Difference pd is found (tabulated in table) s is found a summation of pd, after pd is calculated is 35.66. The Plausibility Calculation as per algorithm is tabulated in table 6.1.

Node	Pivot/ Difference Pd	Mass m	Belief Bel	Plausibility pl
1	8.74	0.22	0.22	0.34
2	6.92	0.18	0.18	0.3
3	1.93	0.05	0.05	0.7
4	6.92	0.18	0.18	0.3
5	8.74	0.22	0.22	0.34
2 or 3 or 4 or 5	1.35	0.03	0.66	0.78
1 or 3 or 4 or 5	1.35	0.03	0.66	0.78
1 or 2 or 4 or 5	1.26	0.03	0.7	0.82
1 or 2 or 3 or 5	1.06	0.27	1.07	0.95
1 or 2 or 3 or 4	1.26	0.03	0.7	0.82

Table 5.1 Plausibility Calculation

In the table 6.1, bel (2 or 3 or 4 or 5) is calculated as ∑m (2, 3, 4 and 5).

pl (1) is calculated as 1-bel (2 or 3 or 4 or 5)

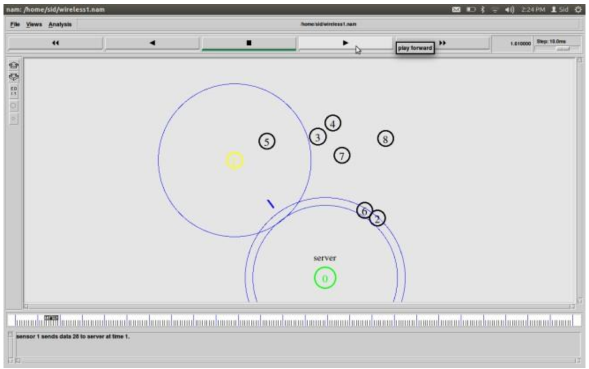
pl(2 or 3 or 4 or 5) is calculated as 1-∑bel (2,3,4,5) Similarly all the values are tabulated.

We found the difference between the obtained plausibility and found that the third node produces outlier data. Further data from the third node which is black listed will be ignored by the base station. Broadcasting revocation message is not required as it save energy and the application requirement is satisfied to obtain the genuine data.

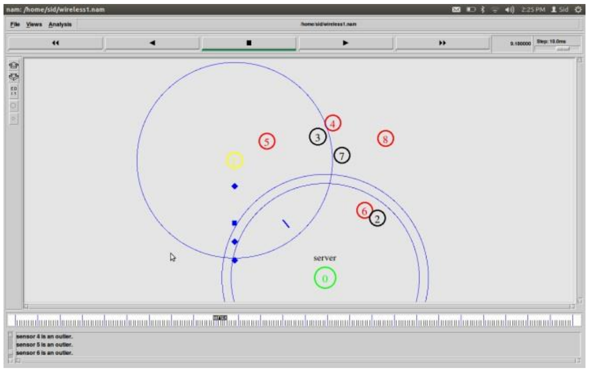
VI. Performance Evaluation

[1] NS2 Implementation:

The system is simulated using ns2 . It is an event driven simulation tool that has proved useful in studying the dynamic nature of communication networks. Ns2 2.29 is a network simulator that is used in the system and it is added with a mannasim patch , so that it adheres to the functionality of a WSN. Leach Protocol is used to get a hierarchical sensor network. Cluster head election is based on the nodes remaining energy level in the nodes and its proximity with the neighboring nodes in Leach protocol. The cluster head serves as the sink node to collect data from the lower level nodes. The nodes are created using random way point model. The cluster head collects data from the nodes. For evaluating the results obtained by the system, artificial data and real-time data are considered. The real-time data is used so that we obtain real-world conditions. But in real-time data, real malicious data is difficult. Hence synthetic or artificial data is also used to simulate the normal and abnormal conditions. Data transmission and Outlier detection is depicted in Fig 6.1.1 and 6.1.2.



sensor 1 sends data 28 to server at time 1.



sensor 4 is an outlier.
sensor 5 is an outlier.
sensor 6 is an outlier.

Fig 6.1.1Data transmission

Fig 6.1.2 Outlier detection

A degree of damage D is used in the system. It is defined as the difference between abnormal and normal data. For example if the correct data is C and a malicious data is M, D=m-C. The system is evaluated using different D values. The obtained results are plotted as receive operating characteristic (ROC) curves. ROCshows the probability of detection rate and false alarm rate for varying threshold D. The ROC curves for anomaly detection are shown in Fig 6.1.3.

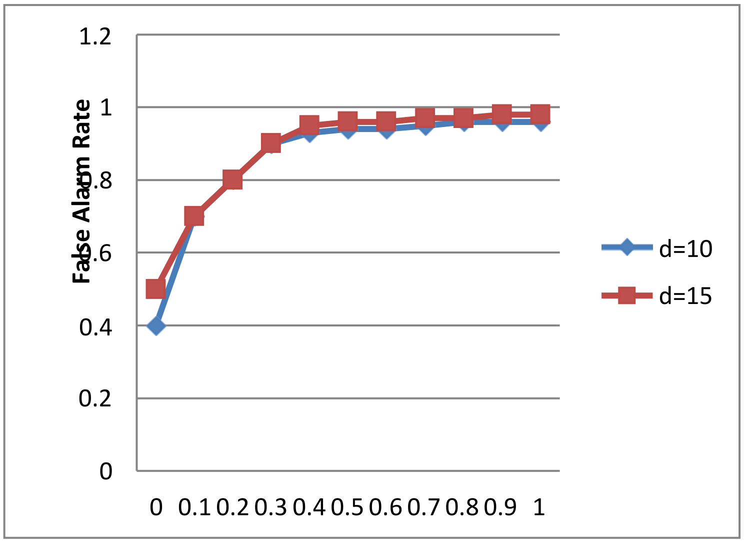


Fig 6.1.3 ROC curves for Outlier detection

[2] Implementation in J Tossim

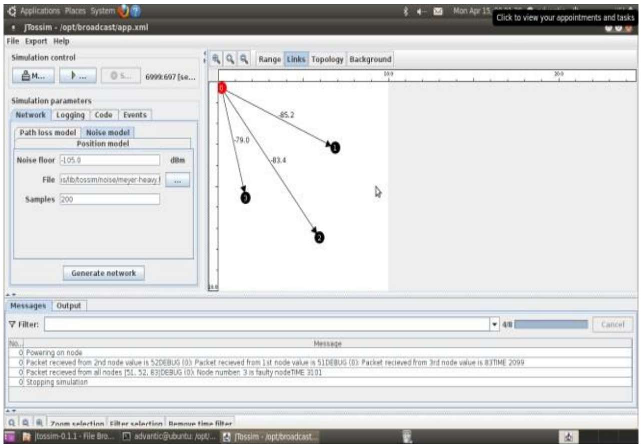
The system uses Mica2 motes in Tiny OS and evaluated the overhead in terms of ROM size and RAM size. We use 8 Mica2 sensor nodes connected to a laptop, which will act as the base station. The sensor nodes transmit the value of temperature collected and base station executes the algorithm. The memory requirements for algorithm execution are shown in Table 6.2.1.

ROM Size	1204 Bytes
RAM Size	32 Bytes

Table 6.2.1

The algorithm is executed using J Tossim. The compiler is NesC in TinyOS environment. The debug channel used is Tree routing. Default radio parameters of J

Tossim are used. It is same as in a Tossim simulator. The standard noise trace of J Tossim is used. We deployed the size of node area as 100 x 100m. We set the number ofnodes as 80 and random deployment. Here the multiple sensor nodes were plotted in different area. Every node is capable of communicate with neighboring sink nodes. Each sink node has a maximum of seven nodes communicating to it. Leach protocol is used to implement clusters. For saving the life of sensor nodes cluster head will send wake up alarm when to collect data. Then all the nodes will collect the environment temperature. After that the cluster head collects data from the node. The values are analyzed and outliers are identified. Fig
6.2.1 and Fig 6.2.2 show the nodes and its data transmission to the base station in J Tossim.



Noise floor -105.0 dBm

File ıs/lib/tossim/noise/meyer-heavy.t

Samples 200

```
No.  Message
0    Powering on node
0    Packet recieved from 2nd node value is 52DEBUG (0): Packet recieved from 1st node value is 51DEBUG (0): Packet recieved from 3rd node value is 83TIME 2099
0    Packet recieved from all nodes [51, 52, 83]DEBUG (0): Node numben 3 is faulty nodeTIME 3101
0    Stopping simulation
```

Fig 6.2.1 Nodes with the Sink nod

The time for execution of the system in both ns2 and J Tossim is tabulated in Table 6.2.2.And a graph is shown inFig 6.2.3

No. of nodes	Execution Time NS2 ms	Execution Time JTOSSIM Ms
1000	0.000	0.000
2000	0.010	0.012
3000	0.020	0.018
4000	0.020	0.019
5000	0.020	0.020
6000	0.020	0.022
7000	0.030	0.028
9000	0.040	0.035
15000	0.060	0.056
20000	0.100	0.068
25000	0.110	0.090
40000	0.170	0.095

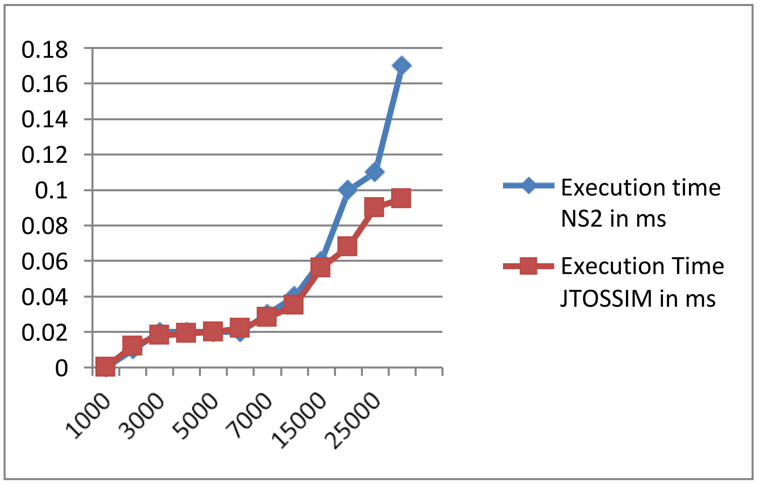


Fig 6.2.3

[3] Calculation of Overhead involved

Communication Overhead: It is the cost involved in transmitting data from the sensor node to the base station or the cluster head. Though the clusters may perform computation locally and identify outliers, it need to be communicated the base station.

Storage Overhead: The values of data and final plausibility for all nodes need to be maintained by the cluster head or the base station.

Computation Overhead: As per plausibility calculation, the overhead is directly proportional to the number of nodes.

Overhead	Using Base station	Using Cluster head
Communication Overhead	O(n)	O(nxm)
Storage Overhead	O(n)	O(nxm)
Computation Overhead	O(n^2)	O(n^2)

Table 6.3.1

Table 6.3.1 presents the details of overhead, where n is the number of sensor nodes and m is the number ofclusters**.**

I. Conclusion and Future Enhancement:

In this paper wireless sensor network security issues and challenges are studied and an overviewis presented. A survey on available outlier detection methods in WSN is tabulated. The proposed system has been implemented for outlier detection using Dempster-Shafer Theory. Simulation of the wireless sensing environment and analysis of the data to identify outliersis done using NS2 and J Tossim simulators and comparedtheir performance.

The system works well if there are more than half of genuine nodes or nodes rendering their services correctly. D-S theory can be modified and used to enhance the system to provide accurate results even in inconsistencies of data or in the presence of noise [45]. Such a system can be adapted to render appropriate services even when there are large discrepancies in data obtained.

II. REFERENCES

[1] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, “Wireless Sensor Network Survey”, International Journal of Computer and Telecommunications Networking, vol.52, no. 12, pp. 2292-2330, Aug, 2008.

[2] Raja Waseem Anwar, Majid Bakhtiari, Anazida Zainal, Abdul Hanan Abdullah and Kashif Naseer Qureshi, “Security Issues and Attacks in Wireless Sensor Network”, World Applied Sciences Journal, vol. 30, no. 10, pp. 1224-1227, 2014.

[3] Kai Xing, Shyaam Sundhar Rajamadam Srinivasan, Manny Rivera, Jiang Li, Xiuzhen Cheng, “Attacks and Countermeasures in Sensor Networks: A Survey”, Network

Security, Springer US, pp. 251-272, 2010.

[4] Virendra Pal Singh, Sweta Jain and Jyoti Singhai, "Hello Flood Attack and its Countermeasures in Wireless Sensor Networks", IJCSI International Journal of Computer Science Issues, vol. 7, no. 11, May 2010.

[5] H. Shafiei, A. Khonsari, H. Derakhshi, P. Mousavi, "Detection and mitigation of sinkhole attacks in wireless sensor networks", Journal of Computer and System Sciences, Special Issue on Wireless NetworkIntrusion, vol. 80, no.3, pp. 644–653, May, 2014.

[6] Chia-Mu Yu, Yao-Tung Tsou and Chun-Shien Lu, "Localized Algorithms for Detection of Node Replication Attacks in Mobile Sensor Networks", IEEE Transactions on Information Forensics and Security, vol. 8, no. 5, May, 2013.

[7] Xiang-yi Chen, Li-xia Meng and Yong-zhao Zhan "Detecting and Defending against Replication Attacksin Wireless Sensor Network", International Journal of Distributed Sensor Networks, May, 2013.

[8] Xingfu Wang, Lei Qian and Haiqing Jiang , "Tolerant Majority-Colluding Attacks for Secure Localization in Wireless Sensor Networks" , in Proc. of the International conference on Wireless Communications, Networking and Mobile Computing, Sept. 2009, pp. 24-26.

[9] James Newsome, Elaine Shi, Dawn Song, Andrian Perrig, "The Sybil attack networks: analysis & defenses", In Proc. Of ACM International symposium on Information processing in sensor networks 2004, pp. 259-268.

[10] Fang Liu & Xiuzhen Cheng, Dechang Chen, "Insider Attacker detection in Wireless Sensor Networks", in Proc. of International Conference on IEEE Compter and Communications societies, May, 2007, pp. 6- 12.

[11] Alka P.Sawlikar, Dr.Z.J.Khan, Dr.S.G. Akojwar, "Analysis of Different Cryptographic and Encryption Techniques using Message Authentication Code (MAC) in Wireless Sensor Networks (WSN)", International Journal of Engineering Science Invention, vol. 2, no. 7, pp. 09-13, July 2013.

[12] Hangyang Dai and Hongbing Xu, "Key Pre- distribution Approach in Wireless Sensor Networks Using LU Matrix", IEEE Sensors Journal, vol.10, no.8, pp. 1399-1409, Aug 2010.

[13] Xin Miao, Kebin Liu, Yuan He, Yunhao Liu Dimitris Papadias, "Agnostic Diagnosis: Discovering Silent Failures in Wireless Sensor Networks", in Proc. of International Conference on IEEE Infocom 2011, pp. 1548-1556.

[14] Moutushi Singh, Rupayan Das, "A Survey of Different Techniques for Detection of Wormhole Attack in Wireless Sensor Network", International Journal of Scientific & Engineering Research, vol.3, no.10, Oct,2012.

[15] Garcia-Otero, Poblacion-Hernandez, "Detection of wormhole attacks in wireless sensor networks using range-free localization", in Proc. Of IEEE International

workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD) Sept, 2012, pp. 21-25, 17-19.

[16] Mingyan Li, Iordanis Koutsopoulos, Radha Poovendran, “Optimal Jamming Attack Strategies and Network Defense Policies in Wireless Sensor Networks”, IEEE transactions on Mobile Computing, vol. 9, no. 8, pp. 1119-1133 August 2010.

[17] Justin Raj S.S, D.Thilagavathy, “Security Threats and Jamming Attacks of Multichannel Wireless sensor Networks”, International Journal of P2P Network Trends and Technology (IJPTT), vol.2, no.1, pp.27-31, 2012.

[18] Leela Krishna Bysani, Ashok Kumar Turuk, “A Survey On Selective Forwarding Attack in Wireless Sensor Networks”, in Proc. of IEEE International Conference on Devices and Communications (ICDe Com) Feb.2011, pp. 978-981.