# ENERGY PROFICIENT OUTLIER DETECTION IN WSN

**Ms. Sapna Aggarwal*, Ms. Saima Jan****
*** Assistant Professor,Jind Institute of Engg. & Technology, Jind,Haryana,**
**** Student(M.Tech), Jind Institute of Engg. & Technology, Jind, Haryana.**

***Abstract-*** *A wireless sensor network is a network made of huge number of nodes that consist of very less battery life and data processing capabilities. These microelectronics nodes are capable of measuring pressure, temperature and various other environmental conditions. In this paper we will present an algorithm of wireless sensor network having outlier detection system in the same. We will present an energy efficient algorithm with outlier detection system based on trust voting algorithm and sensor rank. Also, we will represent the result after the simulation of our work.*

***Keywords:*** *Wireless, Sensor Nodes, Outlier, Deployment, Clustering, Labeling, Outlier Identification***.**

## I. INTRODUCTION

Wireless Sensor Network (WSN) consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants and to cooperatively pass their data through the network to a main location. The more modern networks are bi-directional, enabling control over the activity of the sensors. The growth of wireless sensor networks was aggravated by armed forces applications such as battleground surveillance; these days such networks are also used in many manufacturing and end user application, such as engineering process monitoring and control, machine health monitoring, environment and habitat monitoring, healthcare applications, home automation, and traffic control. The WSN is built of nodes from a few to several hundreds or even thousands, where each node is connected to one (or sometimes several) sensors. Each such sensor network node has typically several parts: a radio transceiver with an internal antenna or connection to an external antenna, a micro-controller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery or an embedded form of energy harvesting. A sensor node size may change but its generally small in size. The cost of sensor nodes is also likewise changeable, having range from hundreds of rupees to thousands, depending on the complexity of the any sensor node that is to be deployed. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and communications bandwidth. The WSN may have bus, mesh or star topology depending upon the scenario. Also, the propagation methods among the hops of the network can be by routing or by flooding. Outlier detection refers to the method of looking for problem in data of any event related to network in our case. These anomalous patterns are often referred to as outliers, anomalies, discordant observations, exceptions, faults, defects, aberrations, noise, errors, damage, surprise, novelty, peculiarities or contaminants in different application domains. In WSNs, outliers can be defined as, “those measurements that significantly deviate from the normal pattern of sensed data” [1].

**II. TYPES OF OUTLIER** When the total data is examined as per the central data approach by any central authority then outliers can be recognized properly and can be tackled appropriately at the corresponding station. When type of data is measured the outliers can be divided as local and global outliers:

**Local Outliers**: considering the point that local outliers are known in wireless sensor network at individual sensor nodes, important task is recognizing techniques reducing communication overhead and maintaining scalability of wireless sensor network with proper determination of outliers. Many event based applications like, surveillance, vehicle following and monitoring can be done by the use of local outlier detection.

There are two variations in Local outlier identification in wireless sensor network. One variation is that historical values are used for determining the wrong or faulty value in the given sensor network. Another option is adding historical reading of their own; where the value of neighbor is taken to determine the value is proper or not i.e. the anomaly is based on the feedback from the neighbor node. When compared with the second approach the first one lags as it doesn’t provide that much accuracy and robustness in the detection of outliers.

**Global Outliers:** Global outliers are popular as they have global perspective and also they draw more attention as they focus on the complete characteristics of WSN instead of working locally like local outlier. On basis of different network architecture, different type of identification can be done on many nodes. All the data collected id transmitted to sink node in the centralized architecture. It delay the response time very much and cause a lot of communication overhead. Cluster head collect the data and identifies outlier in cluster based approach. It has better response time and energy consumption as compared to the former one. [10, 12]

## III. PROPOSED PROTOCOL

We propose a simple, static fault detection model which will improve the concept of SensorRank. In [57], the problem of determining faulty readings in a WSN without compromising detection of important events was studied. By exploring correlations between readings of sensors, a correlation network was built based on similarity between readings of two sensors. By exploring Markov Chain in the network, a mechanism for rating sensors in terms of the correlation, called SensorRank, was developed. In light of SensorRank, an efficient in-network voting algorithm, called TrustVoting, was proposed to determine faulty sensor readings. But to make SensorRank energy efficient we make use of clustering. The Concept of clustering is taken from ESC_HTN Protocol. Before we discuss about the concept of clustering we need to compute the value of Sensor Rank first.

**SENSORRANK** In the early work in the field, distances between SNs were taken into consideration when modeling the correlation of sensor readings. However, it is also possible that the readings of two geographically close SNs to have dramatically different readings. Thus, it's critical to truly capture the correlation of sensor readings rather than their distance. So a Correlation network is to be maintained for sensor readings.

The correlation network is modeled as a graph G = ($V$;$E$), where $V$ represents the SNs in the deployment region and $E$ ={($s_i$; $s_j$)|$s_i$, $s_j \in V$; dist($s_i$; $s_j$) <$R$ and $corr_{i,j}$> 0}. The weight of an edge ($s_i$; $s_j$) is assigned to be $corr_{i,j}$. Once the correlation network of sensors is constructed (and maintained), one can easily deduce the correlations among SNs. Based on the correlation network, we shall further develop an algorithm to compute SensorRank for each SN, in terms of the correlation with its neighbors, in the network. SensorRank is to represent the trustworthiness of SNs. By our design, two requirements need to be met in deriving SensorRank for each sensor.

**Requirement 1:** If a sensor has a large number of neighbors with correlated readings, the opinion of this sensor is trustworthy and thus its vote deserves more weight.

**Requirement 2:** A SN with a lot of trustworthy neighbors is also trustworthy.

These two requirements ensure that:

1. A SN which has a large number of similar neighbors to have a high rank.

2. SN which has a large number of 'good references' to have a high rank. Given a correlation network G = ($V$; $E$) derived previously, we determine SensorRank for each sensor to meet the above two requirements. Based on the above setting, we can formulate SensorRank of $S_i$ , denoted as $rank_i$, as follows:

$$p_{j,i} = \frac{corr_{i,j}}{\sum_{k \in nei(i)} corr_{i,k}}$$

$$rank_i = \sum_{s_j \in nei(i)} p_{j,i} \cdot rank_j \qquad (2)$$

where$p_{j,i}$ is the transition probability from state i to state j.

The key advantages achieved by clustering are energy efficiency, scalability and communicational efficiency. Preventing all nodes from communicating directly with BS saves energy; it also enables BS to handle more number of clusters, thus more WSNs, thereby provides scalability. Reducing the number of nodes competing for communication channel of BS provides better utilization of bandwidth and thus, enhances communicational efficiency). Cluster-based model takes less time to detect outlier nodes in sensor network. This model reduces communication overhead. Thus saves energy as well as time.

**TRUSTVOTING ALGORITHM** consists of two phases: a) self-diagnosis; and b) neighbors diagnosis phase. In the self-diagnosis phase, each sensor verifies whether the current reading of a sensor is unusual or not. Once the reading of a sensor goes through the self-diagnosis phase, this sensor can directly report the reading. Otherwise, the SN consults with its neighbors to further validate whether the current reading is faulty or not. If a reading is termined as faulty, it will be filtered out.

**Self-diagnosis Phase** When a set of SNs is queried, each sensor in the queried set performs a self-diagnosis procedure to verify whether its current reading vector is faulty or not. Once the reading vector of a SN is determined as normal, the SN does not need to enter the neighbor-diagnosis phase. To execute a self-diagnosis, each sensor $s_i$only maintains two reading vectors: i) the current reading vector atthe current time $t$ (denoted as $b_i(t)$); and ii) the last correct reading vector at a previous time $t_p$(expressed by $b_i(t_p)$).$b_i(t_p)$ records a series of readings occurred in the previous time and is used for checking whether the current reading behavior is faulty or not. If these two reading vectors are not similar, $b_i(t)$ is viewed as an unusual reading vector. Once a SN is detected an unusual reading vector, this SN will enter the neighbor-diagnosis phase Note that when $b_i(t)$ is identified as a normal vector through the neighbor-diagnosis, $b_i(t_p)$ is updated so as to reflect the current monitoring state.

**Neighbour-diagnosis Phase** If a SN $s_i$sends$b_i(t)$ to a neighbor $s_j$, $s_j$will compare $b_i(t)$ with its own current reading vector $b_j(t)$ and then give its vote with respect to $b_i(t)$. From the votes from neighbors, $s_i$has to determine whether $b_i(t)$ is faulty or not. Notice that some votes are from sensors with high SensorRank. A SN with high SensorRank has more similar neighbors to consult with and thus is more trust- worthy. Therefore, the votes from the neighbors with high SensorRank are more authoritative, whereas the votes from the neighbors with low SensorRank should cast less weight. When sensor $s_i$sends$b_i(t)$ to all its neighbors for the neighbor-diagnosis, each neighbor should return its vote after determining whether $b_i(t)$ is faulty or not. If a neighbor $s_j$considers$b_i(t)$ is not faulty by comparing the similarity of the two reading vectors (i.e., $corr_{i,j} \geq \sigma$)$s_j$will send a positive vote, denoted $vote_j(i)$, to $s_i$. Otherwise, the vote will be negative. In addition, the vote from $s_j$will be weighted by its SensorRank.

$$vote_j(i) = \begin{cases} rank_j, & corr_{i,j} \geq \sigma \\ -rank_j, & otherwise \end{cases}$$

After collecting all the votes from the neighbors, $s_i$has two classes of votes: one is positive class ($b_i(t)$ is normal) and the other is negative class ($b_i(t)$ is faulty). If the weight of the former is larger than the weight of the later, the most neighbors will view $b_i(t)$ as normal. Note that the weight of a vote represents how authoritative a vote is. It is possible that a neighbor $s_j$of$s_i$with a large SensorRank has a small correlation with $s_i$. In this case, these two SNs may not provide good judgments for each other. Therefore, each vote (i.e., $vote_j(i)$) has to be multiplied by the corresponding correlation, $corr_{i,j}$. Thus, we use the following formula to determine whether the reading is faulty or not.

$$dec_i = \sum_{s_j \in nei(i)} corr_{i,j} \cdot vote_j(i)$$

If the weight of the positive votes is more than the weight of the negative votes, $dec_i$ will be positive which means that $s_i$'s reading is normal and the current reading can be reported. Otherwise, $dec_i$ is negative, implying that the current reading of $s_i$ is faulty.

Each CH find the outlier nodes with in the cluster, it will send data to the BS. BS will aggregate the data and send the aggregated outlier data to every cluster. Now each CH has aggregated outlier data of every other cluster. So whenever there is an Inter-Cluster communicates within the network, CH will check the local aggregated outlier data. We are combining two techniques i.e. SensorRank [57] & heterogeneous node proposed Protocol. As per the result of these techniques, these two techniques are energy efficient. So we can say we will have an energy efficient system.

**The Pseudo code of Proposed Model is as Follows:**

**Step1: Start and deploy a sensor network of hundred nodes in the .m file created in MATLAB.**

**Step 2: Next step is creation of clusters as per the proposed methodology with constraint of immobile sensor nodes with our protocol.**

3 phases of **ESC-HTN** protocol:
1. Number of nodes (heterogeneous with normal and advanced nodes) with fixed processing and energy are deployed.
2. STARTING OF FIRST ROUND, FIRST IS ADVERTISEMENT

A) cluster head election is performed on the basis of energy of node taken as criteria and. Also energy used for intra cluster communication and transmission is one tenth of the normal transmission energy used from cluster head to base station
B) status of being a cluster head is advertised to the nearby nodes using one tenth transmission energy level
C) non cluster nodes listen to the medium to decide their cluster membership on the basis of signal strength.
3. SECOND IS SETUP PHASE,
A) nodes broadcast their membership status
B) cluster listen to medium and create a time division multiple access based dynamic time slots for them.
4. DATA TRANSMISSION IS THIRD STEP OF PHASE
A) nodes sleep until its TDMA slot and transmit during its slot
B) cluster head collect data aggregates and compresses and pass it to sink with energy normal transmission level.
5. Completion of single iteration
6. Phase works until completion of number of rounds.

**Step 3: calculation of rank of each sensor node using Sensor Rank [57].**

**Step 4: applying trust voting algorithm to the network created up to step 4 consisting of two steps:**

**a. Current reading vector is confirmed using self-Diagnosis phase.**

**b. In the second phase i.e. Neighbor diagnosis phase the vote of nearby nodes are given weightage as per their sensor rank i.e. higher rank higher weightage.**

**Step 5: For each vote association, $asso_{i,j}$, following formula is used to determine whether the reading is faulty or not.**

$$dec_i = \sum_{s_j \in nei(i)} Asso_{i,j} \cdot vote_j(i)$$

**if $dec_i$= +ve, node's reading is normal.**

**Otherwise, *dec* =-ve, implying that the current reading of node is faulty.[57]**

**Step 6: Collection of outlier data within the cluster using CH, it will send data to the base station located away from the network taken as assumption in the simulation. Aggregated data from the BS can be broadcasted to every cluster head, thereby recognising outliers simply and efficiently.**

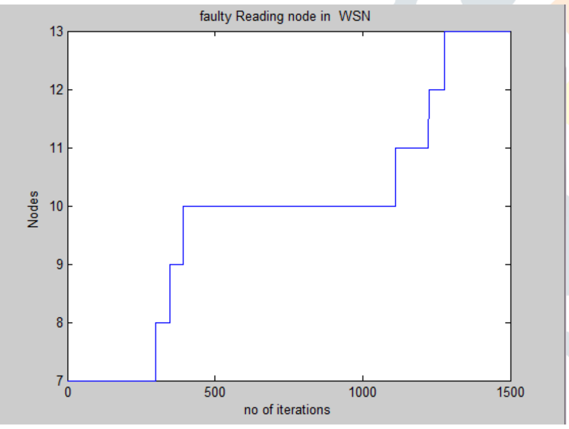
**Step 7: Stop**

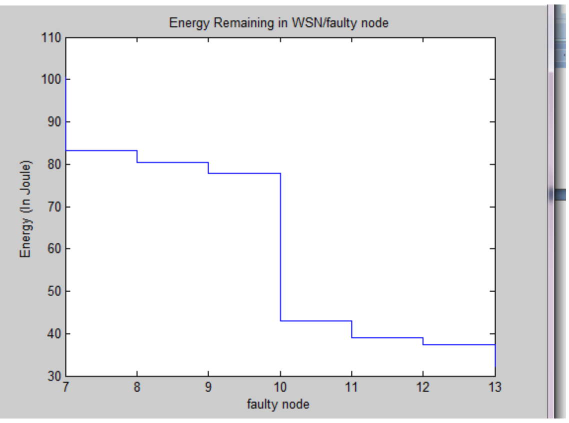
## IV RESULTS

Parameters employed in Simulation

| Parameter | Value |
|---|---|
| Field Size | 100m X100m |
| No. of Nodes | 100 |
| Probability of cluster | 0.2 |
| Initial Energy of sensor node | 100 J |
| MinReading | 1 |
| maxReading | 10 |
| $E_{fs}$ | 10 J/bit/m$^2$ |
| $E_{mp}$ | 0.0013 J/bit/m$^4$ |

Based on these parameters we will carry out the simulations. These parameters are taken after studying different research papers used in Wireless sensor network. Figure 1,2 and 3 showing Faulty reading Nodes detected in total number of iteration of simulation.





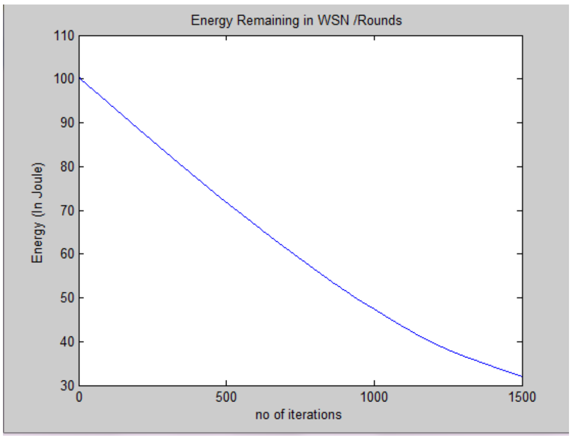


Figure 1 A Fig 1 no of faulty reading nodes in WSN

B Energy remaining in WSN/Faulty Nodes

1c Energy remaining in WSN/No of Iteration

JETIR

## V. CONCLUSION & FUTURE SCOPE

In this paper we presented review the basics of wireless sensor network and outlier in the wireless sensor network. We also presented various features of outliers like their types, how they are identified, various sources and degree of outliers. At last, we represented various challenges in detection of outlier in wireless sensor network. More study can be carried out as review of types of outliers as further classification of local and global outliers. And, various algorithms can be implemented for detection of outliers in the wireless networks.

## VI. REFERENCES

[1] Chandola, V., Banerjee, A. and Kumar, V., “Outlier detection: a survey”, Technical Report, University of Minnesota, 2007.

[2] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, “Distributed anomaly detection in wireless sensor networks”, in Communication systems, 2006. ICCS 2006. 10th IEEE Singapore International Conference on, pp. 1 –5, October 2006.

[3] http://www.americanlaboratory.com/913-Technical-Articles/156961-Statistical-Outliers-in-the-Laboratory-Setting/

[4] S. Rajasegarar, J. C. Bezdek, C. Leckie, and M. Palaniswami, “Elliptical anomalies in wireless sensor networks,” ACM Trans. Sen. Netw., vol. 6, pp. 7:1–7:28, January 2010.

[5] D. J. Hill, B. S. Minsker, and E. Amir, “Real-time bayesian anomaly detection for environmental sensor data”, in proceedings of the 32nd conference of IAHR, 2011.

[6] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, “Online outlier detection in sensor data using non-parametric models”, in proceedings of the 32nd international conference on Very large data bases, VLDB ’06, pp. 187–198, 2006.

[7] L. B. Oliveira, E. Habib, H. C. Wong, A. C. Ferreira, M. A. Vilaa and A. A. Loureiro, “Security of cluster-based communication protocols for wireless sensor networks” In 4th IEEE International Conference on Networking (ICN05), volume Lecture Notes in Computer Science, pages 449-458, Washington, DC, USA, 2005.

[8] W. Heinzelman, A. Chandrakasan and H. Balakrishnan,"Energy-Efficient Communication Protocol for Wireless Microsensor Networks", , January 2000.

[9] W. Heinzelman, “Application-specific protocol architectures for wireless networks”, Ph.D. thesis, Massachusetts Institute of Technology, 2000.

[10] Y. Zhang, N. Meratnia, and P. Havinga, “Outlier detection techniques for wireless sensor networks: A survey,” IEEE Communications Surveys & Tutorials, vol. 12, no. 2, pp. 159–170, 2010

[11] Z. Yang, N. Meratnia, and P. Havinga, “An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine”, in Intelligent Sensors, Sensor Networks and Information. Processing, 2008, ISSNIP 2008. International Conference on, pp. 151 –156, December 2008.

[12] T. Kavitha, A. Chandra, “Wireless networks: a comparison and classification based on outlier detection methods “in CSEA 2012, vol. 4, special issue 1; 2013