International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 3
ISSN: 2454-4248
113 – 117

Outlier Detection in WSN based on Ranking and Clustering Technique

Ritika
Deptt. of Comp. Sci. & Engg.,
M.M.E.C, Mulana,
Ambala, India
rsharma.20sept1988@gmail.com

Amandeep Kaur
Deptt. of Comp. Sci. & Engg.,
M.M.E.C, Mulana,
Ambala, India
amandeepkaur@mmumullana.org

Dr. Rajneesh Kumar Gujral
Deptt. of Comp. Sci. & Engg.,
M.M.E.C, Mulana,
Ambala, India
drrajneeshgujral@mmumullana.org

***Abstract*—** In wireless sensor networks, the readings that deviate from the actual value of sensed data are known as outliers. The reason of outliers includes noise, errors, and malicious attacks on the network. Traditional outlier detection methods are not suitable to wireless sensor networks due to the limited resources of the wireless sensor networks. We propose outlier detection approach based on clustering and rank calculation of Sensor nodes. We applied our ideas into the SensorRankscheme and finally we achieved an efficient mechanism which is based on clustering. The cluster head detect find Faulty nodes from its own cluster and send this information to the Base Station. In this manner Base Station have information of all Faulty nodes in the network. Then Base Station can broadcast this information to all cluster heads. In this way Inter cluster outlier detection can be done. As each cluster head is associated in finding the faulty nodes, which results detection of faulty nodes in WSNs with low energy consumption. Energy consumption is affected by message communication between nodes, so our technique is efficient than traditional Sensor Rank scheme. The analysis shows that our scheme is good enough for Outlier Detection in the WSNs.

Keywords- *Outlier, Outlier Detection, Sensor rank, clustering.*

*****

I. INTRODUCTION

WSNs consist of the small size battery operated nodes that have sensing, computation and communication capabilities. For the reason of these abilities, WSN is able to monitor the environment and compute simple tasks and exchange data with Base Station (BS) and among themselves. Due to its sensing ability, WSNs is used in the wide range of applications including habitat monitoring, dense forest fire reporting, battlefield surveillance, volcanic fields monitoring and wild life supervision etc. In spite of having enormous applications WSN differs from traditional wireless networks in terms of computation power, memory and bandwidth capacity[1].

The nodes in network are resource constrained having limited battery power, computation power, less memory and transmission ranges. Once the battery of node is dead, SN is useless which may cause reducing Network lifetime, Network partition, Holes in Routing Paths, Frequent Flooding of Control Packets for Link State Maintenance and the the probability of generating erroneous data will grow rapidly [2].As the environment is unattended so SNs are more vulnerable to attacks which cause faulty and redundant information to be transmitted to sink[3]. Since actual events occurred in the physical world, e.g., forest fire, earthquake or chemical spill, cannot be accurately detected using inaccurate and incomplete data [4], it is extremely important to detect outliers for ensuring the reliability and accuracy of sensor data.

The term outlier, also known as anomaly, originally stems from the field of statistics [5]. The two classical definitions of outliers are: (Hawkins [6]): “an outlier is an observation, which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism”. (Barnett and Lewis [7]): “an outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data”.

II. RELATED WORK

Based on outlier sources, outlier detection techniques are divided into 3 as shown in figure 1.

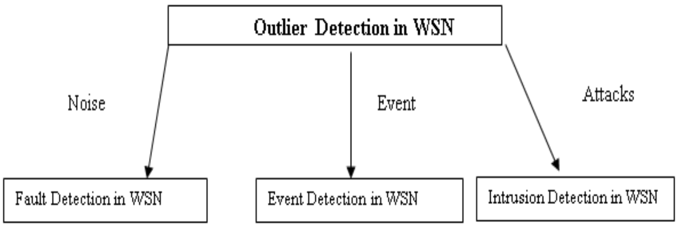

Figure 1. Outlier sources in WSNs and their corresponding detection techniques.

A. *Challenges of Outlier Detection in WSNs*

Due to the following challenges, Resource constraints, High communication cost, Dynamic network topology, frequent communication failures, mobility and heterogeneity of nodes, Large-scale deployment, Identifying outlier sources conventional outlier detection techniques might not be suitable for handing sensor data in WSNs. Thus, the main challenge faced by outlier detection techniques for WSNs is to satisfy the mining accuracy requirements while maintaining the resource consumption of WSNs to a minimum. In other words the main question is how to process as much data as possible in a decentralized and online fashion while keeping the communication overhead, memory and computational cost low [10].

B. *Techniques Designed For WSNs*

Recently, many outlier detection techniques specifically developed for WSNs have emerged. In this section, we provide a technique-based taxonomy framework to categorize these techniques.

Outlier detection techniques for WSNs can be categorized into statistical-based, nearest neighbor-based, clustering-based,

 ISSN: 2454-4248
113 – 117

classification-based, and spectral decomposition-based approaches. Statistical-based approaches are further categorized into parametric and non-parametric approaches based on how the probability distribution model is built. Gaussian-based and non-Gaussian-based approaches belong to parametric approaches and kernel-based and histogram-based approaches belong to non-parametric approaches.

III. Proposed System Model

The proposed outlier detection technique saves energy by using the concept of clustering with Sensor Rank and a simple model for the protocol implementation. We propose a simple, static fault detection model which will improve the concept of Sensor Rank. In [14], 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 Sensor Rank, was developed. In light of Sensor Rank, an efficient in-network voting algorithm, called Trust Voting, was proposed to determine faulty sensor readings. We know that Sensor Rank is a trusty worthy value that determines how the readings to two sensors are close to each other. But to make Sensor Rank energy efficient we make use of clustering based on CLEANER [15]. Later we discuss how to compute the value of Sensor Rank.

A. CLuster-based approach for ENERgy-efficiency in the WSN (CLENER)

After adjusting the transmission power, each node generates a random number (μ), which ranges from 0 to 1. Then, the node decides to become a CH by comparing μ with the T(n), which is computed by means of Equation 1. If μ is less than T(n), the node becomes a CH for the current round.

$$T(n)= \eta \frac{p}{1-p(r \bmod \frac{1}{p})} + \alpha(1 - e^{\frac{-RE^2}{2\sigma^2 RE}})........(1)$$

Where η and α are weights to give importance, the sum is exactly 1. The Residual Energy is denoted as RE, and σ means the energy variance, which is used to produce better CH candidates.

B. Sensor Rank

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. Sensor Rank 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.

Based on the above requirements, Sensor Rank of S_i, denoted as $rank_i$, as in equation 2:

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

$$rank_i = \sum_{S_j \in nei(i)} p_{j,i} \cdot rank_j \text{ ....(2)}$$

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.

With the help of clustering instead of Sensor Rank exchange their rank to each other CH collect the sensor ranks of all SNs thereby energy is saved which is our main motive.

With the help of clustering instead of Sensor Rank exchange their rank to each other CH collect the sensor ranks of all SNs thereby energy is saved which is our main motive.

C. Proposed Model :

Step1: Start

Step 2: Create a Network

Step 3: Create Clusters from network using **CLENER [15].**

a. A CH is selected from the SNs.

b. Based on last step, Non-CHs select the best CH by considering a multiple metrics and Cluster is created.

Step 4: Rank of each node is calculated using Sensor Rank [14].

Step 5: With Sensor Rank, Trust Voting algorithm [14] is used which consists of two phases:

a. Self-diagnosis: performs a self-diagnosis procedure to verify whether its current reading vector is faulty or not.

b. Neighbor diagnosis phase: the votes from the neighbors are taken. Vote with high Sensor Rank are more authoritative, whereas the votes from the neighbors with low Sensor Rank should cast less weights.

Step 6: If a neighbor with a large Sensor Rank has a small correlation node, they may not provide good judgments for each other. Therefore, each vote correlation, $corr_{i,j}$, following formula is used 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 dec_i= +ve, node's reading is normal.

Otherwise, dec_i= -ve, implying that the current reading of node is faulty.

Step 7: Collection of outlier data within the cluster using CH, it will send data to the BS.

Step 8: Aggregated data from the BS is forwarded to every cluster.

Step 9: Stop

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.

IV. SIMULATION AND PERFORMANCE ANALYSIS

Section A provides simulation scenario which introduces the system model used by presented scheme. The model is simulated in MATLAB 7.14.0.739. Section B introduces performance evaluation matrices used to analyze performance of the presented model.

A. Simulation Scenario

The presented model can be used for large WSNs, where network is divided into clusters. Initially there is a network in which nodes are distributed randomly as shown in figure 2.

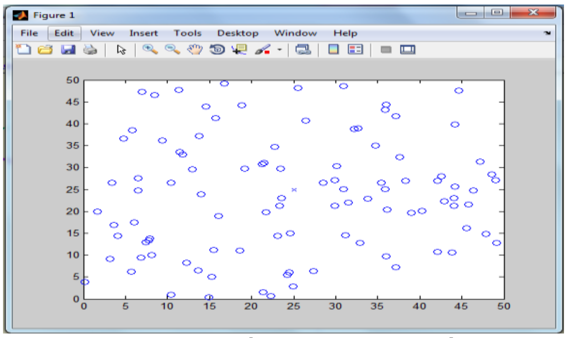

Figure 2. Network creation using 50 Nodes.

Now Clusters Heads are elected from the given network and clusters are made based on [15]. Figure 3 shows dark blue stars (*) which are marked as Cluster Heads.

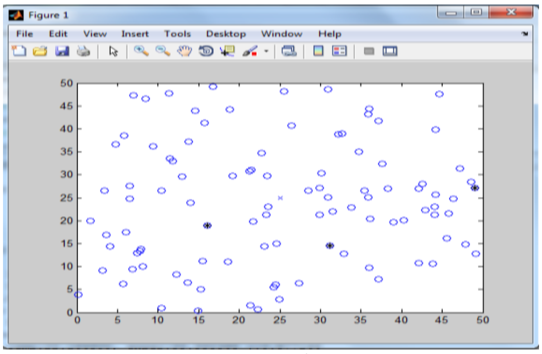

Figure 3. Election of Cluster Head

Each Normal node will elect its cluster head based on Probability which can be calculated Fuzzy Logic System using the two input variables “distance between the node & cluster head” and “Residual Energy”.

Finally figure 4 shows the surface graph for probability calculation for cluster formation.

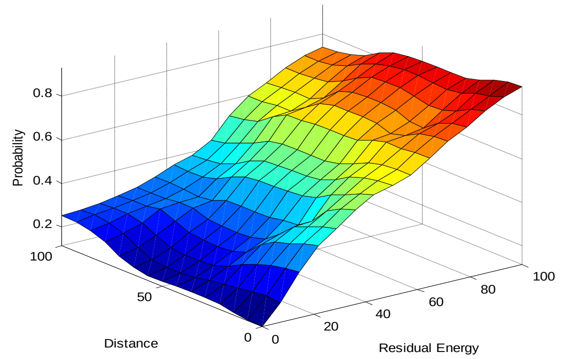

Figure 4. Surface Graph for Probability Calculation for cluster formation.

Using this Probability Calculation fuzzy logic, each normal node calculates the probability for each cluster head. The node which has the highest probability with respect to any cluster head will be the member of that cluster for cluster head in that round. In this way Cluster formation is done using [15].

Now next simulation is for Sensor Rank Calculation and Outlier Detection. Figure 5 shows the Red colored (*) Faulty Nodes found in the network in each cluster.

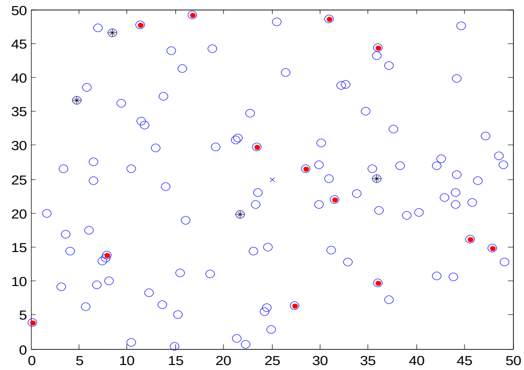

Figure 5. Faulty Nodes found in the network in each cluster.

B. Performance Evaluation

The basic parameters used for simulations are listed in Table I.

TABLE I. PARAMETERS EMPLOYED IN SIMULATION

Parameter	Value
Field Size	50m X 50m
Location of Base Station	25m X 25m
Probability of cluster	0.1
Initial Energy of sensor node	20 J
The Data packet Size	208 bits
DeltaT	10
η	0.4
A	0.6
Mindreading	1
maxReading	10
E_{fs}	10 J/bit/m^2
E_{mp}	0.0013 J/bit/m^4

ISSN: 2454-4248
113 – 117

Now Clusters Heads are elected from the given network and clusters are made based on [15]. Figure 4.2 shows dark blue stars (*) which are marked as Cluster Heads.

Based on these parameters author will carry out the simulations. These parameters are taken after studying different research papers used in Wireless sensor network.

Figure 6 shows that with the increase in no. of nodes the outlier frequency is also increasing. Also in comparison with traditional SensorRank scheme, it performs better.

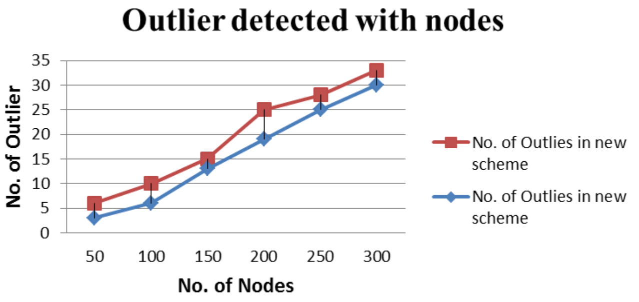

Figure 6. Outlier Detection with increasing no. of nodes

Figure 7 showing average outlier detected in each round of simulation. It shows that outlier per round is almost remains same with the increase in no. of rounds.

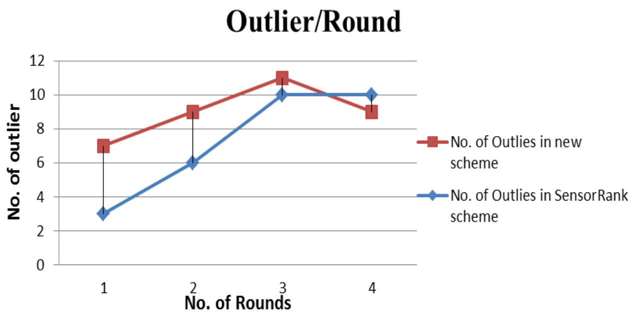

Figure 7. Outlier per Round

Finally figure 8 showing the energy consumption (joule) after completion of each round while finding the faulty node in a network. It shows that consumption of energy is very low after each round, which shows that the life time of network will increase.

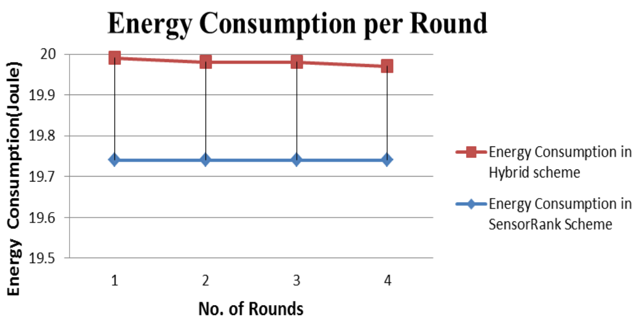

Figure 8. Energy Consumption per round

V. Conclusion and Future Scope

In this research, we have presented an efficient technique for Outlier Detection using Sensor Rank in the WSNs. In the existing Sensor Rank scheme [14], a single network is considered and rank is calculated and Faulty nodes are detected. But for a network to be good designed there should be cluster formation. Traditional Algorithms have not used clustering technique. So, an efficient Clustering technique CLEANER [15] is applied to the Sensor Rank [14] and faulty nodes are detected from each cluster. A further direction of this study will be dedicated node to which a default high energy can be given. This node is outside the cluster and performs all the calculation for the cluster due to which the overhead of cluster head can be reduced.

References

[1] Y. Zhang, "Observing the Unobservable -Distributed Online Outlier Detection in Wireless Sensor Networks", PhD thesis, University of Twente, June 2010.

[2] 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.

[3] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, "Labelled data collection for anomaly detection in wireless sensor networks," in Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2010 Sixth International Conference on, pp. 269 –274, December. 2010.

[4] 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.

[5] Peng Jiang, "A New Method for Node Fault Detection in WSNs", in proceedings of Sensors (14248220), Vol. 9 Issue 2, p1282, February, 2009.

[6] Arun Kumar K.S, Vinay Joseph Ribeiro, "A Reliable and Energy Efficient Framework for Wireless Sensor Networks", in proceedings of COMSNETS'09, pp. 604-612, 2009.

[7] Y. Zhang, N. Meratnia, and P. Havinga, "Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks," in Proc. Int. Conf. Advanced Information Networking and Applications Workshops WAINA '09, pp. 990–995, 2009.

[8] A. R. Ganguly, "Knowledge Discovery from Sensor Data", CRC Press, 2008.

[9] M. Shuai, K. Xie, G. Chen, X. Ma, and G. Song, "A kalman filter based approach for outlier detection in sensor networks", in Computer Science and Software Engineering, 2008 International Conference on, vol. 4, pp. 154 – 157, December 2008.

[10] Y. Zhang, N. Meratnia, and P. J. M. Havinga, "A taxonomy framework for unsupervised outlier detection techniques for multi-type data sets", Technical Report TR-CTIT-07-79, Centre for Telematics and Information Technology University of Twente, Enschede, November 2007.

[11] D. J. Hill, B. S. Minsker, and E. Amir, "Real-time bayesian anomaly detection for environmental sensor data", in In Proceedings of the 32nd conference of IAHR, 2007.

[12] 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, VLDB Endowment, 2006.

[13] 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.

[14] Xiang-Yan Xiao, Wen-ChihPeng and Chih-Chieh Hung," Using sensor rank for in-network detection of faulty nodes in WSNs", in proceedings of MobiDE'07, June 10, 2007.

[15] Claudio Silva, Rodrigo Costa, AdoniasPires, Denis Rosário, Eduardo Cerqueira, Kássio Machado, Augusto Neto and JóUeyama, "A Cluster-based Approach to provide Energy-Efficient in WSN", International Journal of Computer Science and Network Security, VOL.13 No.1, pp. 55-62, January 2013.

[16] K. Srinivasan and P. Levis. Rssi is under appreciated. In Proceedings of the third workshop on embedded networked sensors (EmNets), 2006.