International Journal of Engineering Research and General Science Volume 3, Issue 3, Part-2 , May-June, 2015
ISSN 2091-2730

AN OVERVIEW OF OUTLIER DETECTION IN WSN

Gaurav Goyal, Rajiv Munjal
M.tech. Student, CBS Group of Institutions, Jhajjar, Haryana;gauravgoyal.nrw@gmail.com;9953920564

Abstract- A wireless sensor network consist of large number of nodes that possesses very small battery life and data processing capabilities but these microelectronics system are capable of measuring physical and various environment related consequences like sound, pressure, motion, pollution causing agents etc. In this paper we will review the basics of wireless sensor network and outlier in the wireless sensor network. We will present various features of outliers like their types, how they are identified, various sources and degree of outliers. At last we will also present various challenges in detection of outlier in wireless sensor network.
Keywords: Wireless, Sensor Nodes, Outlier, Deployment, Clustering, Labeling, Outlier Identification**.**

I. INTRODUCTION

A wireless sensor network consist of large number of nodes that possesses very small battery life and data processing capabilities but these microelectronics system are capable of measuring physical and various environment related consequences like sound, pressure, motion, pollution causing agents etc. Wireless sensor network can be utilized in a wide variety of military applications such as war field monitoring and many more application like chemical spill prevention, heath care application, nuclear plants and traffic control etc. In surveillance applications, sensors are deployed in a certain field to detect and report events like presence, movement, or intrusion in the monitored area. Data collected by sensors are transmitted to a special node equipped with higher energy and processing capabilities called "Processing Node" (PN) or "sink". The processing node of wireless sensor network collect and compare data from various sources i.e. sensor nodes and thus extracting useful and meaningful information. In the architecture SNs are grouped into clusters controlled by a single command node. In wireless sensor network the sensor nodes are capable of doing only short distance communication which is radio based and responsible for detecting any target or event.
Every cluster has a entryway node that manages sensors in the cluster. Clusters can be formed based on many criteria such as communication range, number and type of sensors and geographical location. Sensors receive commands from and send readings to its gateway node, which processes these readings.[7,8]

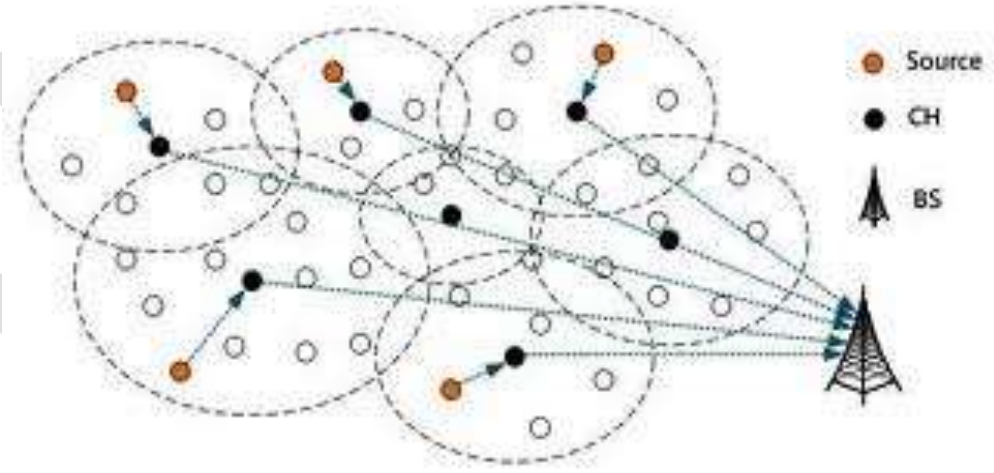

Figure 1: Sensor Network Architecture

Gateways can track events or targets using readings from sensors in any clusters as deemed by the command node. However, sensors that belong to a particular cluster are only accessible via the gateway of that cluster. Therefore, a gateway should be able to route sensor data to other gateways. Gateway nodes interface the command node with the sensor network via long haul communication links.
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]. This definition is based on the fact that in WSN

SNs are assigned to monitor the physical world and thus a pattern representing the normal behavior of sensed data may exist. Potential sources of outliers in data collected by WSNs include noise & errors, actual events, and malicious attacks.

II. TYPES OF OUTLIER

When the complete data is analyzed as per the central data approach by any central authority outliers can be identified properly and can be tackled appropriately at the corresponding station. When type of data is considered the outliers can be classified as local and global outliers:

Local Outliers: Taking the point that local outliers are recognized in wireless sensor network at individual sensor nodes, techniques for reducing communication overhead and maintaining scalability of network with proper determination of outliers is important. Many event detection applications, for example, vehicle following, surveillance and monitoring can be done using local outlier detection. Local outlier identification has two variations 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]

II. OUTLIER SOURCES AND HANDLING

There are three likely outlier sources in WSNs:

(a) Noise and errors which result in fault detection [1]

(b) Events which result in event detection [2].

(c) Malicious attacks which finally lead to intrusion detection.

Outlier handling is carried out by performing these three important steps:

Outlier labeling: Outlier labeling stands for detection of outlier from the given dataset it is performed with the help of various outlier detection algorithms.

Outlier Identification: Outlier identification deals with outlier detection as event or error or any kind of noice.

Outlier Accommodation: Once an observation is identified as a potential outlier, analysis should begin to determine whether an assignable cause can be found for the spurious result. If none of the reason can be found, a repetition can be suggested, the potential error node data should be backed up for future consequences. Robust statistical methods such as weighted least-squares regression minimize the effect of an outlier observation. Robust outlier detection techniques should be employed when the number of outliers is large, so that the resulting data distribution is not skewed, however non-robust techniques can be employed when the number of outliers is small.[11]

III. DEGREE OF OUTLIER

Outlier score recognizes the amount by which the sensor nodes reading diverges from the normal data reading. In wireless sensor network we have two scales for measuring the degree of being an outlier.

(a) Scalar: It is the outlier scale which divides the data measurement to determine whether it is normal or anomalous. This is actually zero one type of classification of data. This method neither differentiates between outliers, nor provides a ranked list of outliers.

(b) Outlier Score: In this a score is associated with outlier not only the classification of sensor reading as normal or anomalous. The score describe the degree of outlierness in the measurement of sensors.[3]

V. CHALLENGES OF OUTLIER DETECTION IN WSNs

Fetching out important information from given raw data is very tough job. [5] The complex design and data collected from sensor nodes are complex so it is difficult to determine outlier in it. Due to these reasons it is difficult to detect outlier in wireless sensor network:

- **Resource constraints**. The low quality and cheap sensor nodes have severe constraints in resources, like computational capacity, energy and communication bandwidth.

- **High communication cost.** In WSNs, radio communication consumes a big portion of energy not the computation in real. Computation cost for a sensor node is much lower than cost of radio communication [4].
- **Distributed streaming data.** Dynamic change can come in streaming data due to different streams. Additionally, the original distribution of data thus streamed cannot be known before receiving. Furthermore, direct computation of probabilities is difficult [6].
- **Dynamic network topology, frequent communication failures, mobility and heterogeneity of nodes**. A sensor network deployed in unattended environments over extended period of time is susceptible to dynamic network topology and frequent communication failures.
- **Large-scale deployment.** Deployed sensor networks can have massive size (up to hundreds or even thousands of SNs). The key challenge of traditional outlier detection techniques is to keep an extraordinary detection rate along with it keeping the rate of false alarm also was possible. This requires the construction of an accurate normal profile that represents the normal behavior of sensor data [5].
- **Identifying outlier sources.** The sensor network is expected to provide the raw data sensed from the physical world and also detect events occurred in the network. However, it is difficult to identify what has caused an outlier in sensor data due to the resource constraints and dynamic nature of 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.[7]

VI. 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.

VII. 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