International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 3, Issue 4(April 2014), PP.42-49

Comparison of Anomaly Detection Techniques For Wireless Sensor Network

Sreepradhana.T, Soundarya.S
(Department Of Information Technology, Psg College Of Technology, India)
(Department Of Information Technology, Psg College Of Technology, India)

Abstract:- A wireless sensor network with a large number of sensor nodes can be used as an effective tool for gathering data in various situations. As wireless sensor networks continue to grow, so does the need for effective security mechanisms. Because sensor networks may interact with sensitive data and operate in hostile unattended environments, it is imperative that these security concerns be addressed from the beginning of the system design. The main issue in the sensor network is the data reliability. There are some factors that affect the data reliability in the sensor networks. It includes the noise and missing values, duplicated or inconsistent data. During transmission using sensor networks when the battery power is exhausted, then the probability of getting erroneous data will grow rapidly. It leads to faulty data and it affects the data reliability of the system. An efficient technique to detect anomalous data is proposed. The outlier detection at the aggregator level is done using density based technique by using Local Outlier Factor (LOF). Here enhancement over LOF is introduced. Enhanced LOF is simpler and is used to find sparse clusters.

Keywords:-Enhanced Outlier Factor, K-distance, Local Outlier Factor, Outlier detection, Reachability distance.

I. INTRODUCTION

Sensor networks consist of a huge number of small sensor nodes, which communicate wirelessly. These sensor nodescan be spread out in hard accessible areas by what newapplications fields can be pointed out. A sensor nodecombines the abilities to compute, communicate and sense. In all the applications, data is prone to attacks. The reliability of the data cannot be trusted. There may be outliers.

Outliers are those patterns that deviate from the normal pattern of the sensed data. These outliers decrease the performance and quality of the system.So, the outlier detection is necessary for wireless sensor network. At each cluster, data is aggregated and outlier detection is performed to reduce energy consumption. Data aggregation is the process of one or several sensors that collects the detection result from other sensors. The collected data must be processed by sensor to reduce transmission burden before they are transmitted to the base station or sink. After the data is aggregated, density based outlier detection is performed. The simplest version of density-based outlier detection is named as Local Outlier Factor (LOF).

In the proposed system, consensus-based outlier detection approach has been applied by enhancing the previous versions such that it addresses the problems of the existing system. The main feature of the system is to eliminate the malicious data at the cluster head level. The defined system provides mechanisms to identify the sparse anomalous clusters and also the problem of aggregator compromise.

II. NETWORK STRUCTURE

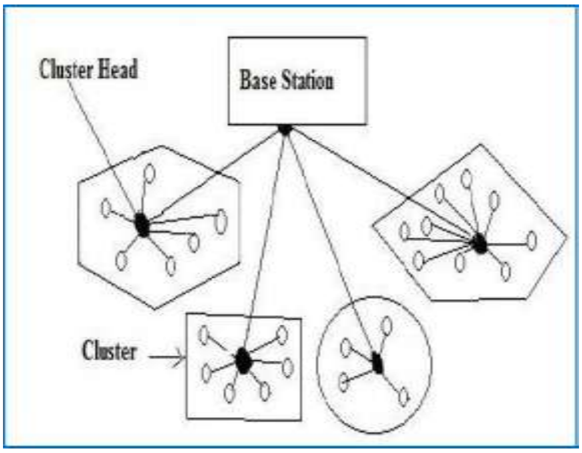

Fig 1 Cluster based mechanism in WSN

The nodes are clustered based on LEACH protocol. Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol for sensor network helps to minimize energy dissipation in sensor networks. It is a hierarchical routing algorithm for sensor networks which make clusters of the sensor nodes based on the received signal strength. LEACH forms clusters by using a distributed algorithm, where nodes make autonomous decisions without any centralized control. The advantages of this approach are that no long-distance communication with the base station is required and distributed cluster formation can be done without knowing the exact location of any of the nodes in the network. In addition, no global communication is needed to set up the clusters, and nothing is assumed about the current state of any other node during cluster formation. The goal is to achieve the global result of forming good clusters out of the nodes, purely via local decisions made autonomously by each node.

In LEACH, data fusion and aggregation are local to the cluster. Cluster heads change randomly over time to balance the energy dissipation of nodes. The node chooses a random number between 0 and 1. The node becomes a cluster head for the current round if the number is less than the following threshold:

$$T(n) = \begin{cases} \frac{p}{1-p*(r \bmod \frac{1}{p})} & \text{if } n \in G \\ 0 & \text{otheriwse} \end{cases}$$

Equation 1

The first phase is set-up phase in leach protocol where the main motive is cluster formation. For electing cluster heads residual energy and threshold are taken into account. After the clusters are formed and cluster heads are elected we move on to the next steady state phase. In the steady state phase the data transmission takes place in the network.

III. PROPOSED NETWORK

In the proposed system, consensus-based outlier detection approach has been applied by enhancing the previous versions such that it addresses the problems of the existing system. The main feature of the system is to eliminate the malicious data at the cluster head level. The defined system provides mechanisms to identify the sparse anomalous clusters and also the problem of aggregator compromise is addressed.The enhanced version of LOF involves the use of two Minpts thereby defining two different neighborhoods: (1) neighbors in computing the density and (2) neighbors in comparing the densities. In LOF, these two neighborhoods are identical due to the fact that only one Minpt is used. Hence the sparse anomalous clusters can be identified as outliers rather detected as normal data set as in LOF.

Enhanced LOF

The enhanced LOF can be calculated be calculated as follows

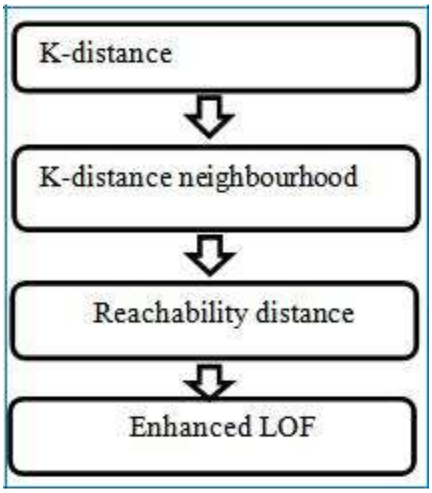

Fig 2 Enhanced LOF methodology

K-Distance

k-distance is defined as the furthest distance among the k-nearest neighbors of a data point p, where k is the minimum number of points that are used to define a neighborhood of a given data point.Consider the point p as shown in the Fig 3. Let k be 2. The 2-nearest neighbours of p are the data points O1 and O. Among O and O1, the furthest distance from p is O. As per definition, k-distance is distance between p and O.

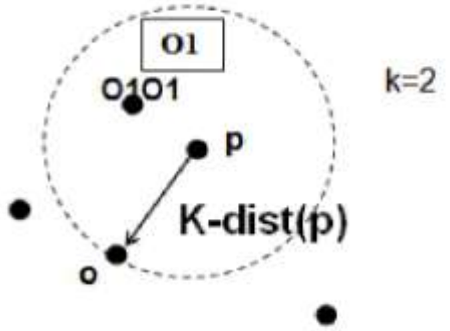

Fig 3 k-dist(p) for k=2

The k-distance of object p, is defined as the distance d(p,o) between p and an object o ϵ D such that:
(i) for at least k objects o'ϵD \ {p} it holds that d(p,o') ≤ d(p,o), and
(ii) for at most k-1 objects o'ϵD \ {p} it holds that d(p,o') < d(p,o).

K-Distance Neighborhood

The k-distance neighborhood is defined as the set of k neighbors which lie within k-distance of a point p.Consider the point p as shown in the Fig 4. Let k be 2. The 2-nearest neighbors of p are the data points O1 and O. As per definition, k-distance neighborhood of p is O and O1.

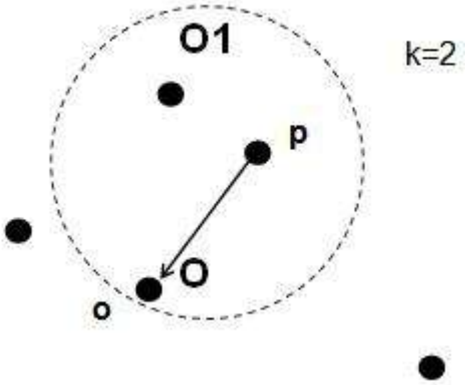

Fig 4 k-distance neighbourhood of p

The k-distance neighborhood of p contains every object, whose distance from p is not greater than the k-distance,

Nk-distance(p) = { q ϵ D\{p} | d(p, q) ≤ kdistance(p)}

Equation 2

These objects q are called the k-nearest neighbors of p.In order to detect density-based outliers, the density of the neighborhood of each object is determined which is defined by a parameter Minpts(positive integer) that specifies the minimum number of points that resides in p's neighborhood.

Reachability Distance

Reachability distance of p from o is the maximum of the radius of the neighborhood of o if p is in the neighborhood of o or the real distance from p to o.

Consider the point p, O1 and O2 as shown in the Fig 5. The neighbourhood circle of p is denoted using dotted circle and the radius denoted the k-distance. The maximum of radius of neighbourhood circle of p and the distance of p from O1 is the radius of neighbourhood circle of p. As per definition, reachability distance is the distance between p and O.

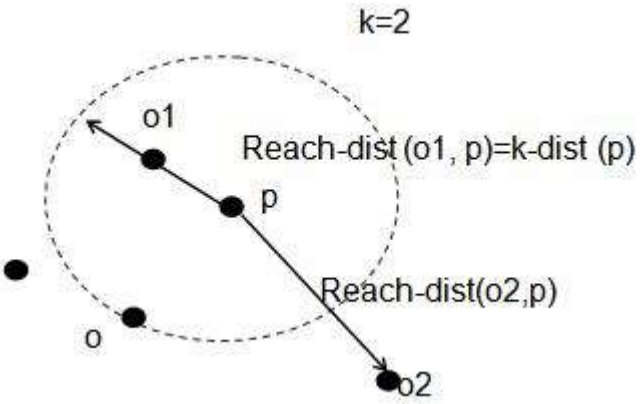

Fig 5 reach-dist(o1,p) and reach-dist(o2,p), for k=2

The reachability distance of object p with respect to object o is defined as reach-dist (p, o) = max { k-distance(o), d(p, o) } where o $\epsilon N_{MinPts(p)}$

Equation 3

Figure 5 illustrates the idea of reachability distance with k =2. Intuitively, if object p is far away from o (e.g. o2 in the figure), then the reachability distance between the two is simply their actual distance. However, if they are "sufficiently" close (e.g., o1 in the figure), the actual distance is replaced by the k-distance of p. The reason is that in so doing, the statistical fluctuations of d(p,o) for all the p's close to o can be significantly reduced. The strength of this smoothing effect can be controlled by the parameter k. The higher the value of k, the more similar are the reachability distances for objects within the same neighbourhood.

Enhanced LOF

Enhanced LOF of an object p is defined as the average of the ratio of the local reachability density of p and those k1-nearest neighbours of p.

$$\text{Enhanced LOF(p)}= \frac{\sum_{o\varepsilon N_{MinPts1 - dist(p)}} 2- \quad ()\,2- \quad ()}{N_{MinPts1 - dist(p)}(p)}$$

Equation 4

IV. PERFORMANCE ANALYSIS

Selection Of Input Parameters

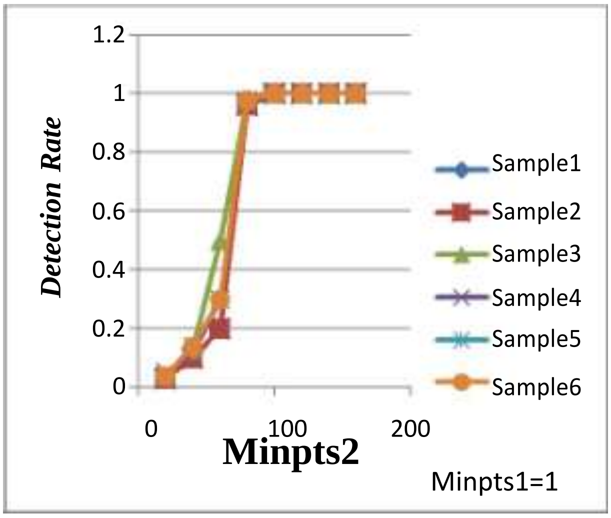

Fig 6 Selection of Minpts2

Considered various samples of sensor reading of sensor nodes with sample having 30% of anomalous value. Minpts1 is set to 30-40% of sample size, we have to determine a suitable value for Minpts2 such that detection rate is high. To identify the sparse anomalous clusters, setting of Minpts2 value is crucial as it is used to determine the density of each sensed data. Thus a plot is done by varying the Minpts2 value proportional to the Minpts1 value for various sample sensor data against detection rate. From the graph it is found that when Minpts2 value is more than 50% of Minpts1 value the detection rate is high for all samples. When it is less than 50% of Minpts1 value the detection rate is low.

Performance For Dense Clustered Data

The following graphs are plotted for various versions of LOF scheme named as follows, LOF the basic versions and LOF3 denote the enhanced version where LOF3 is based on density based outlier detection.

Detection Rate

The number of correctly detected anomalies from the total number of anomalies is said to be the detection rate of a system. The graph below depicts the detection rate of various LOF versions. It is found that the enhanced version of LOF has higher DR until 40% outliers whereas basic versions deviate after 30% outliers.

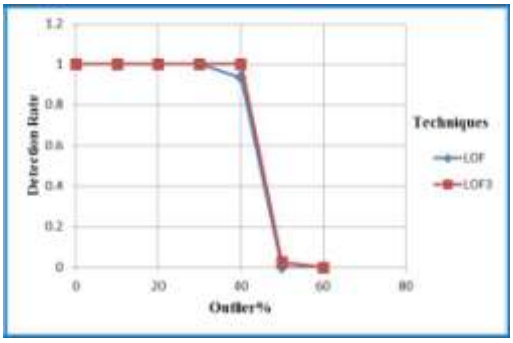

Fig 7 Comparison of Detection rate

False Alarm Rate

The graph represents the variation of false positive rate across various outlier percentages for different LOF schemes. It is found that the enhanced version of LOF with clustering produces very low FPR until 50% outliers whereas basic versions deviate after 30% to 40% outliers.

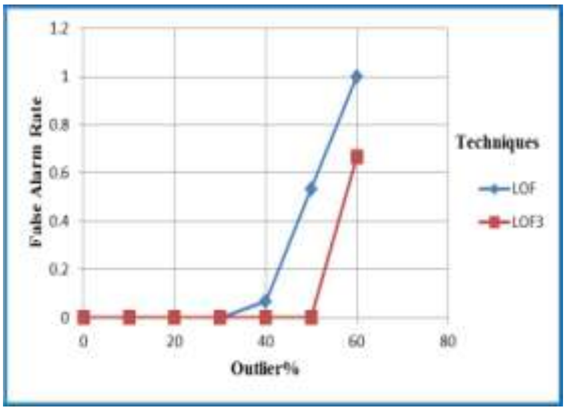

Fig 8 Comparison of False alarm rate

False Positive Rate

The following graph depicts the false positive rate of various versions of LOF. It is found that the FPR values stay low until 40% for LOF3 and until 30% for other version and finds a steep increase henceforth.

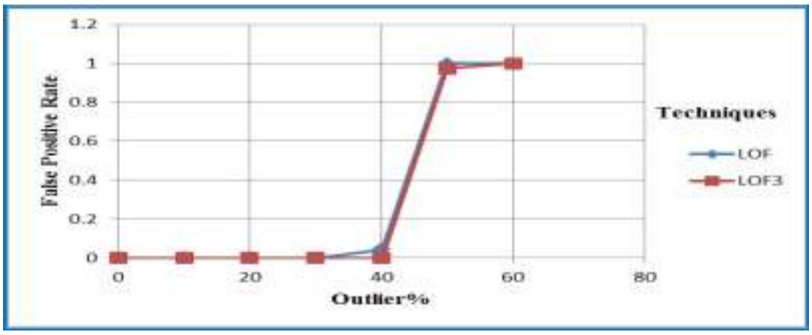

Fig 9 Comparison of False positive rate

V. PERFORMANCE FOR SPARSE CLUSTERED DATA

Parameters Detection Rate

The main feature of enhanced LOF version is their ability to detect sparse anomalous cluster. It is rightly justified from the graph that the enhanced version of LOF has higher DR until 40% outliers whereas basic versions provide mixed results after 30% anomalies.

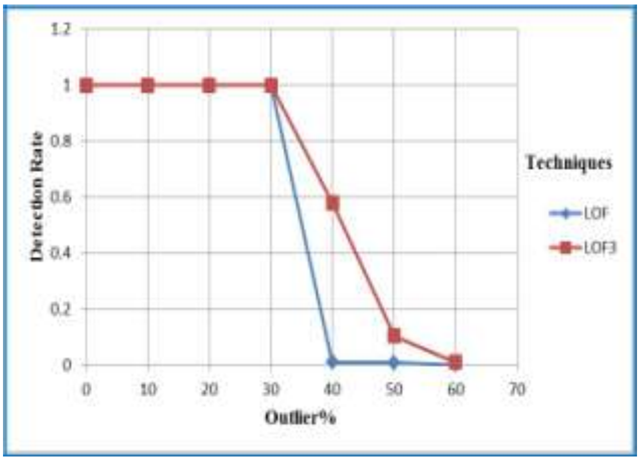

Fig 10 Comparison of Detection rate

False Alarm Rate

The graph represents the variation of false positive rate across various outlier percentages. It is found that the enhanced version of LOF with clustering produces very low FAR in comparison to other versions where FAR is very high due to the presence of sparse clusters.

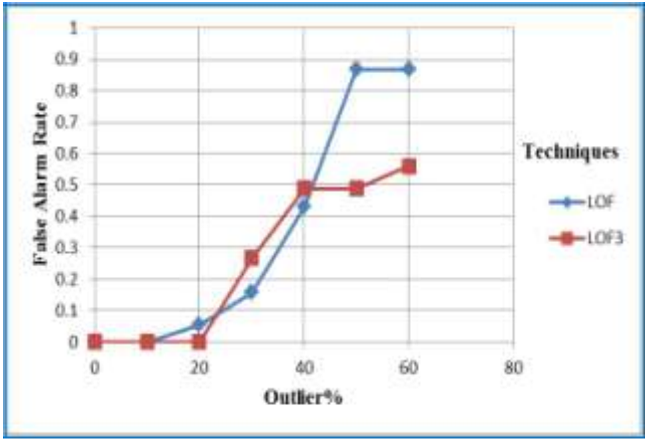

Fig 11 Comparison of False alarm rate

False Positive Rate

The variation of false positive rate against various outlier percentages is monitored for the sensor measurements under consideration. The basic versions provide very high FPR value after 30% whereas the enhanced versions provide a stable result across various outlier ranges.

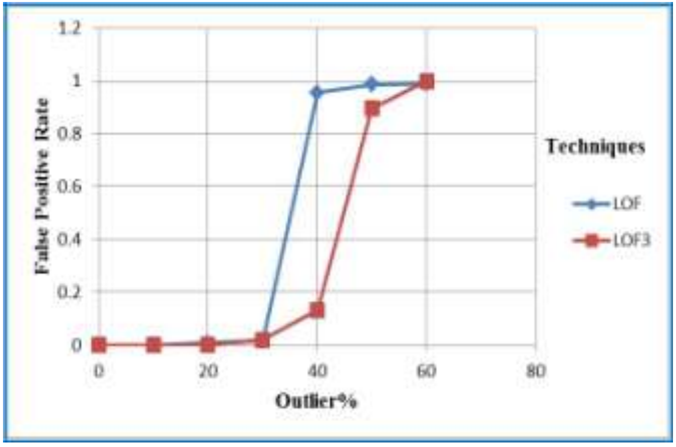

Fig 12 Comparison of False positive rate

Data Accuracy Rate

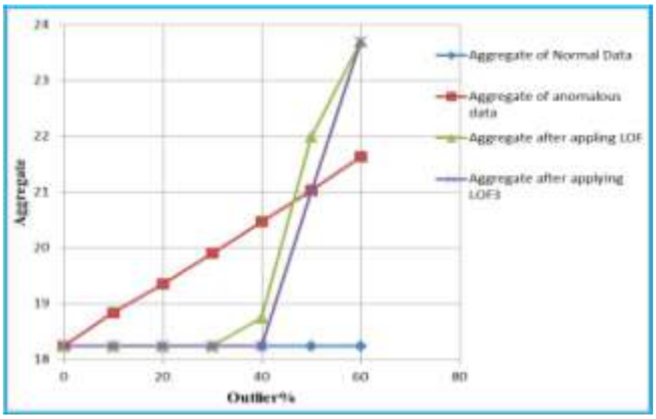

Fig 13 Data accuracy rate

The data accuracy rate, after applying outlier detection mechanism, is nearer to normal data until detection is high after which it coincides with the anomalous data.

VI. TRUST BASED FUSION DATA ASSURANCE

A single copy of valid fusion result is transmitted to the base station by the fusion node. This single transmission saves the power of the uncompromised node.Moreover, no communication is necessary between the sensors in this voting scheme. Early termination is achievable when the base station receives enough "agree" or "disagree" votes. A witness node may remain silent when it agrees with the transmitted fusion result. Only "disagree" votes need to be sent. This "silent assent" feature drastically reduces the transmission power consumption in the system. A compromised fusion node can be identified if it has been excluded by the base station during the polling process.This "traitor exclusion" is useful for further verification of the fusion result.

No forged result can be accepted by the base station unless the number of compromised nodes reaches the number of support votes that is required to verify the fusion result and these nodes collude to forge the fusion result. No forged votes will be accepted by the base station since the witness nodes are assigned trust value. Thus the reliable polling is achieved in the trust based fusion data assurance mechanism.

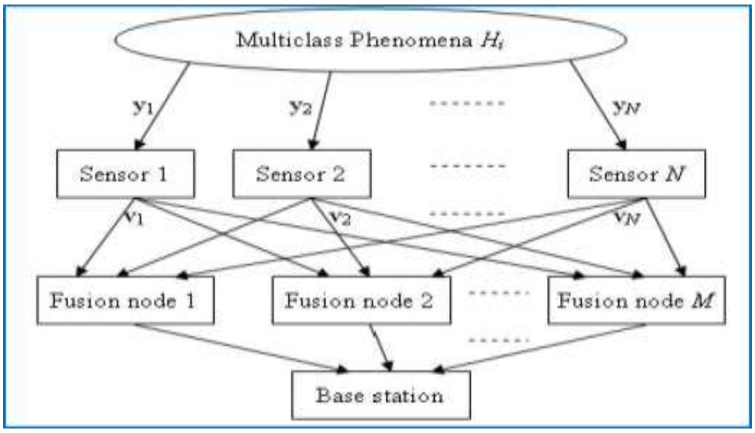

Fig.14 Trust based fusion data assurance mechanism

In secure trust based power efficient direct voting assurance, each witness node is assigned a trust value. The base station must ask the witness node whether it agrees or disagrees with the transmitted fusion result. The witness node then sends its vote to the base station. No denial-of-service attack is assumed and the vote can be clearly identified at the base station. If the transmitted fusion result is not supported by at least T witness nodes, then the base station may have to select a witness node that does not agree with the transmitted result as the next chosen node.

The base station chooses a fusion node. Other fusion nodes serve as witness nodes. Define a set of witness nodes that includes all witness nodes and let the nodes in the set be ordered based on the trust value assigned to it. The chosen node transmits its fusion result to the base station. The base station polls and sends the above fusion result to the node in the witness set by following the trust order of the witness nodes. The base station receives votes from the witness nodes and verifies its trust value; the node having high trust value will be accepted by the base station.

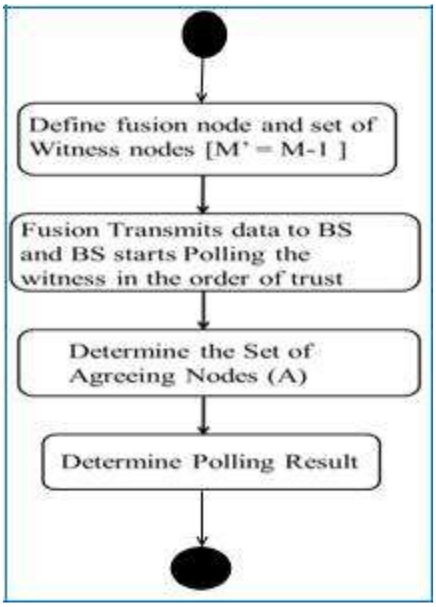
Define fusion node and set of Witness nodes [M' = M-1]
Fusion Transmits data to BS and BS starts Polling the witness in the order of trust
Determine the Set of Agreeing Nodes (A)
Determine Polling Result

Fig.15 Procedure for trust based data fusion assurance

VII. CONCLUSION

In this project, an efficient outlier detection mechanism is implemented such that anomalies are detected and eliminated from the entire system by monitoring all levels. The performance of the protocol in detecting outliers is analyzed using the detection rate, false alarm rate and false positive rate using the data collected from sensors. The observed results show that the protocol can be used to detect dense as well as sparse clusters at the cluster head level. There is also a considerable reduction in communication overhead and energy consumption due to the application of data aggregation mechanism using LEACH protocol.

REFERENCES

[1]. Anny Lai-mei Chiu, Ada Wai-chee Fu, "Enhancements on Local Outlier Detection," ideas, pp.298, Seventh International Database Engineering and Applications Symposium (IDEAS'03), 2003

[2]. Er. Barjinder Singh Kaler ,Er. Manpreet Kaur Kaler," Challenges in Wireless Sensor Networks" , RIMT-MAEC, Mandigobindgarh

[3]. KiranMaraiya, Kamal Kant, Nitin Gupta, "Wireless Sensor Network: A Review on Data Aggregation", ISSN 2229-5518,IJSER © 2011, International Journal of Scientific & Engineering Research Volume 2, Issue 4, April -2011

[4]. MarkusM. Breunig, Hans-Peter Kriegel, Raymond T. Ng, Jörg Sander.: "LOF: Identifying Density-Based Local Outliers" Proc. ACM SIGMOD 2000 Int. Conf. On Management of Data, Dalles, TX, 2000.

[5]. K.Padmanabhan , P.Kamalakkannan " A Study On Energy Efficient Routing Protocols In Wireless Sensor Networks", European Journal of Scientific Reasearch, ISSN 1450-216X vol.60 No.4(2011) © Euro Journals Publishing, Inc,2011.

[6]. Rajesh Patel, Sunil Pariyani, Vijay Ukani,"Energy and Throughput Analysis of Hierarchical Routing Protocol (LEACH) for Wireless Sensor Network", International Journal of Computer Applications (0975 – 8887), Volume 20– No.4, April 2011.

[7]. SutharshanRajasegarar, Christopher Leckie, andMarimuthuPalaniswami, University Of Melbourne, Australia., "Anomaly Detection In Wireless Sensor Networks," 1536-1284/08/ © 2008 IEEE, IEEE Wireless Communications ,August 2008.