American-Eurasian Journal of Scientific Research 10 (3): 141-147, 2015
ISSN 1818-6785

DOI: 10.5829/idosi.aejsr.2015.10.3.22257

Fault Detection in Data Aggregation for Data Accuracy in WSN

Lithiya Sara Babu and N. Uma Maheshwari

Department of CSE, Anna University, Chennai, India

Abstract: WSNs consist of thousands of nodes to sense the data. A secure data sensing in wireless sensor network is a challenging task because of complexity and high overhead. To overcome all the issues we introduce an efficient method. Here introduce a combination of two detectors in the network in order to detect the fault nodes. Combination of detectors may lead to more accurate, less energy consumption and also increase the network lifetime. Here we use Naive Bayes detection technique and Fuzzy logic system for the detection. In Naive Bayes end to end transmission time will be used the fault node detection. During fuzzy logic system, set a fuzzy rule according to the parameters and find the fault node. Our technique may improve the throughput and packet delivery ratio with minimum packet drop and less energy consumption. It efficiently detects the fault nodes by the help of detectors and we got a secure data aggregation process throughout the network.

Key words: Data aggregation · Naive Bayes detector · Fuzzy logic

INTRODUCTION

Wireless sensor networks have tremendous utilization in the past two decades and seen different advances. Starting from, mining, weather, even battle operations and petroleum exploration, all of these require sensor applications. The entire user wants to do is to gather the data sent by the sensors and with certain analysis mine meaningful information from them. Usually sensor applications involve many sensors deployed together. These sensors form a network and collaborate with each other to gather data and send it to the base station. The base station acts as the control center where the data from the sensors are gathered for further analysis and processing. In wireless network consisting of spatially distributed nodes which use sensors to monitor physical or environmental conditions.

Intrusion detection is an important field of research, because it is not possible to set up a system with no vulnerabilities. For intrusion detection one main factor is that is to find out the attacks from a large quantity of communication activities. In different machine learning (ML) algorithms like Neural Network, Support Vector Machine and Data mining are employed in detecting intrusion activities both known and unknown from the large quantity of complex and dynamic datasets. To differentiate standard behaviors from strange behavior by examining the dataset which is a list of tasks created by the operating system that is registered into a file in historical sorted order. Researches with data mining as the chief factor to find out newly occurred intrusion in the network.

The implementation of the Naïve Bayes module for detecting faulty data that is flexible to the needs of sensor nodes. It also offers support for reducing the energy consumption of the real sensor nodes using predicted data instead of the sensed one. Sensor nodes can choose how much time should be committed to the detection task; the more the time required by the algorithm the more is the accuracy of the detection. On the other hand, lowering the response time allows the network to be more reactive to the many changes. The algorithm is able to find the best trade-off between response time and detection accuracy.

Related Work: This section briefly describes some of the works related to fault data detection in wireless sensor network in various areas.

Event Detection for Application Specific: Different publication researches are dedicated to detect events in a particular application such as vehicle tracking, military surveillance and medical diagnosis. Keally *et al*. [1]

Corresponding Author: Lithiya Sara Babu, Department of CSE, Anna University, Chennai, India.

proposed an event detection framework to fulfill user specific requirements mostly on object tracking. In this work he explores the sensing capability of nodes firstly, to perform collaboration between nodes to meet the required accuracy based on effective user demands and secondly, to change detection capabilities based on runtime observation adaptively.

Hill *et al*. [2] proposed their experiment in predicting possible events, based on analyzing and monitoring a received stream of data sensed by thousands of sensors in different field. He introduced an effective infrastructure for analyzing event detection by real time monitoring in order to detect possible failures. This framework uses fourtiers, including, early event warning tier, sensor publisher tier and ontology tier to address the challenge such as maintaining a long history of events, fast response time and combining reported events.

Middleware's for Event Detection: Some of the researches concentrate on devising middleware to facilitate applications for efficiently reporting the detected events.

In TinyDB an event based query is provided for event detection applications. This type of query is triggered when there is an explicit event has happened. It is based on the sensed value the particular event will raise a suspend and the query will be executed. In order to use this ability the programmer should write an element to introduce the signals and the event. The distinct events can be further used in queries when they are required.

For application specific Impala provide an event based programming model. It will be assigned a specific middleware agent called fault filter that is to fulfill the model requirements of programming. The fault filter agent is accountable for dispatching and capturing detected events to other middleware agents as well as applications.

Uncertainty in Event Detection: Heinzelman *et al*. [3] has introduced MiLAN a proactive service oriented WSN middleware. The interesting aspects of the middleware are the capability of switching between sensors with different sensing accuracy. It is able to hold different Quality of Service in mixed nodes. Applications executing in above the middleware layer are motorized by the capability to identify their accuracy needs based on sates of their application. In common, uncertainty in event detection contains range of wider issues than QoS. MiLAN is a significant research in vigorously handling accuracy in sensing but it does not deal with issues such as false positive detection, event warning loss and aggregating uncertainties.

An event detection service (DSWare) has defined in S. Li *et al*. using a data centric approach. It supports CEs for detecting faults in a sensor network with heterogeneous nodes. By submitting an SQL-like statement to a group of specified nodes an application program can be register events. To address CEs, sub-event sets are defined in the statement. This sub-event consists of several parameters, such as a minimum confidence value and a confidence function for detecting it. To deal with uncertainty DSWare uses confidence functions. For this a confidence function takes rate of sub-events, as an input parameter, in a Boolean data type format and computes a numeric value showing how likely the CE has happened. So that in DSWare aggregates the reported events along with the path to sink, it is not applicable to a clustered network. Node level abstraction in interpreting the sensed values is not provided.

In this paper, we devised an effective detection model for the network. Here a combination of detectors are used, (Naive Bayes and Fuzzy logic) used to find the fault node. Naive Bayes was implemented according to the Probability Density Functions (PDFs) obtained by MLE of exponential distribution for both normal and faulty end-to-end transmission time. In fuzzy logic, it is to classify the sensor nodes into normal node and worst node based on the threshold value. So the combination of these two detectors may lead to more accuracy in the network.

System Architecture: In a common Wireless Sensor Network architecture, the measurement nodes are deployed to calculate measurements such as temperature, heat, or even resolved in oxygen. The nodes are part of a wireless sensor network administered by the gateway that governs network aspects such as data security and authenticity. From the gateway it collects the measured data from each and every node and sends it through Ethernet, to a host controller.

Cluster-based network developments and proposals have been designed to build a network for just one type of node, where all nodes in a network can communicate with any other nodes in their coverage area. The sensor nodes are divided into clusters and each cluster consists of cluster head which acts as an aggregator. Data's in the network will be periodically collected and aggregated by the aggregator. This node can forward the aggregated data to the base station through other aggregator node. Due to the low deployment cost requirement of wireless sensor networks,

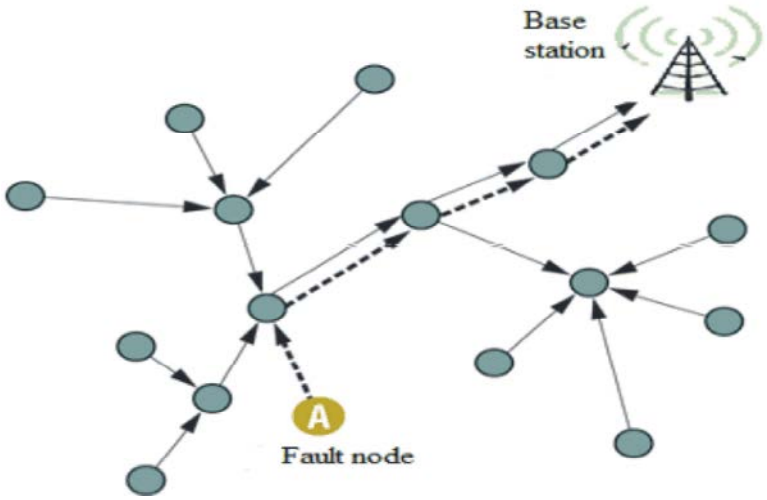


Fig. 1: Data aggregation in a wireless sensor network

sensor nodes consist of several resource constraints and simple hardware. It may also a challenging task to give an efficient solution to data aggregation. Among these constraints, "to save battery power" is the most limiting factor in designing wireless sensor network protocols. Data aggregation protocols aim to combine and summarize data packets of several sensor nodes so that the amount of data transmission is reduced.

An example for data aggregation scheme is presented in Figure 1, where a group of sensor nodes collects information from a region. In different architectures clustering provide many benefits. Reference shows the most important features of cluster-based architectures over ad hoc and sensor networks. Clustered wireless sensor networks may have two important advantages, one consist of the clustered wireless sensor networks are capable of extending the nodes' sleep times by allowing cluster heads to coordinate and optimize the activities. The next one is that reducing the volume of inter-node communication by localizing data transmission within the formed clusters and decreasing the number of transmissions to the sink node [4-11].

Detection of faulty sensor nodes can be obtained by two methods, i.e. self-detection and active-detection. In self-detection, sensor nodes are to periodically monitor along with their residual energy and identify the potential failure of the node. We consider the battery depletion as a main cause of node sudden death in the network. A node fails because when its energy drops below the threshold value. When a common node is fails due to energy depletion, it sends a message to the cell manager that it will be going to sleep mode according to energy below the threshold value.Self-detection is measured as a local computational process of sensor nodes and requires less network communication to preserve the node energy. In addition to this it may also reduce the response delay of the management system towards the potential failure of sensor nodes.

In fault management system to detect the node effectively, we use a working an active detection mode. In this process, the message about updating node may have residual battery and it may apply to track the existence of sensor nodes. For active detection, asks its cell members on a regular basis to send their updates to the cell manager. Such as the cell manager sends get messages to the corresponding common nodes on a regular basis and in return nodes send their updates. This may called as in-cell update cycle. When the cell manager does not receive an update from any node, then it sends an instant message to the node acquiring about its status. If the cell manager does not receive the acknowledgement in a given time, it then passes this information to the remaining nodes in the cell and declares as node faulty. Cell managers concentrate on its cell members and only inform the group manager for further usage,it improves network performance of its small region has been in a critical level.

Proposed Work: In the proposed work a combination of Naive Bayes and Fuzzy rule based system to detect the fault node. We obtain a network with less data redundancy, energy efficient and accurate data across the network. When any set of actions, attempt to compromise with the security attributes such as confidentiality, repudiation, availability and integrity of resources, then these actions are said to be the faults and detection

of such faults is known as a fault detection system. The basic functionality of this system depends only on three main modules such as data collection, detection and response modules. The data collection module is responsible for collecting data from various data sources such as system audit data, network traffic data, etc. Detection module is responsible for analysis of collected data. While detecting fault if the detection module detects any suspicious activity in the network, then it initiates response bythe response module. We will discuss a combination of naïve Bayes and fuzzy logic based proposed fault detection system. So we can improve the accuracy of the system.

Naive Bayes Classifier: Naive Bayes is fast, easy to implement with simple structure and effective. Its also useful for high dimensional data as the probability of each feature is estimated independently [4]. That is based on applying Bayesian theorem with well-built (naive) independence assumption. A naïve Bayesian classifier assumes that the absence or presence ofthe given feature is unrelated to the absence or presence of any other feature, given the particular class variable.Let C be the class of an observation X. In order to predict the class of the observation X by using the Bayes rule, the highest probability should be found.

$$P(C|X) = \frac{P(C)\,P(X/C)}{P(X)} \quad (1)$$

In Naive Bayes classifier the assumption that the features X1, X2,...Xn are conditionally independent of each other in a given class. In Classification problems P(C/X) is sufficient to predict the most probable classes given a test observation.CNBD was implemented according to the Probability Density Functions (PDFs) obtained by MLE of exponential distribution for both normal and faulty end-to-end transmission time. In the faulty and normal PDF from one node are compared. It can be seen that faulty transmissions would have a caused longer transmission time compared with normal transmission. To evaluate the performance of these methods, Scenes Hit Rate, False Alarm Rate and Hit Ratio are also compared.

Naive Bayes achieved the best result of Scenes Hit Rate in the cases with congestions. Ratio of suspicious nodes matched the theoretical faulty nodes to the total scenes is called scenes hit rate. To increase the Naïve Bayes classification accuracy it may use to establish the conditional probability of fault data. For good in heavy traffic congestion cases and light traffic congestion cases, the fault detection technique may use.Hit Ratio is the ratio of properly detected faulty transmission time over the total faulty transmission time. When the faulty nodes increased in the congested scenes the Hit Ratio decreased. It was because the longer transmission time was mostly caused by the network congestion and the long-lasting transmission time of different path became a small percent of the reason of the longest transmission time. The conditional probability of faulty transmission time rate was similar to the conditional probability of normal transmission time rate. For advanced probability the data from faulty network was classified to be data from normal network. So that the faulty nodes increased by the network, the Hit Ratio decreased. False Alarm Rate is the rate of total suspicious nodes against misclassified suspicious nodes. The congested scenes suffered in all methods of high false alarm rate because both faulty nodes and congestions caused the similar longer transmission time. It was not easy to have an accurate classification in all phases. In the congested scenes faulty node may increase so that the false alarm rate decreased. It was because the higher the possibility to be detected and more the faulty nodes. End-to-end packet transmission time is possible to be used for fault detection of the WSN.

Fault Detection Using Fuzzy Logic: In WSN applications entails that fault detection is crucially important. WSN's have been utilized and developed and within a multitude of security platforms, military As a result, the detection of faults can be vital to the successful operation of the WSN and can be fundamental to the decisions made on the data received. Sensors are deploying into highly sensitive application areas that need accurate data recording. This paper emphasized on proposed fuzzy logic based intrusion detection system in wireless sensor networks. This system, anomaly-based intrusion detection makes use of effective rules identified in accordance with the designed strategy that the data mined effectively. Fuzzy logic is used in intrusion detection because it is able to deal with uncertainty and complexity which is derived from human reasoning. To provide good detection results for specified and familiar attacks, the leading advantage of anomaly-based detection techniques is their ability to detect unseen data and unfamiliar intrusion occurrences. The generated fuzzy rules of the proposed strategy can be able to provide better classification rate in detecting the intrusion behavior. With the help of fuzzy variables or linguistic terms, intrusion detection features can be viewed easily

No. of Packets	Energy, E	Result
Less	Less	Normal
Less	High	Normal
High	High	Normal
High	Less	Worst

Fig. 2: Fuzzy Rules

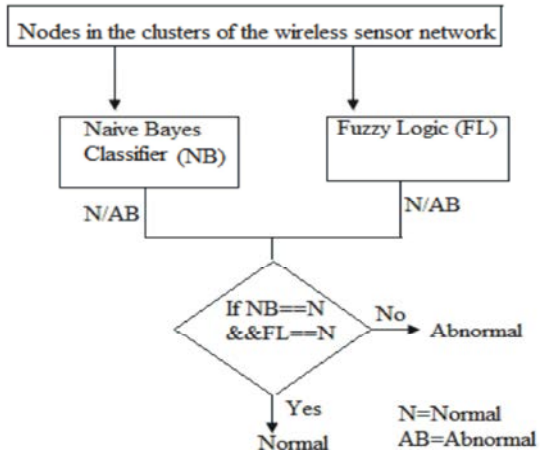


Fig. 3: Detection Model

and decision of normal and abnormal activity in the network is based on its fuzziness nature that can be identified the degree of maliciousness of a node instead of yes or no conditions.

Fuzzifier, inference engine and defuzzifier are the parts of fuzzy systems. Fuzzification process maps the input crisp values to the corresponding fuzzy sets. And it will be assigns a degree of membership functions for fuzzy set. These values will be processed in inference engine, which consist of rule base and various methods for inferring the rules. Simply rule base is a sequence of IF-THEN rules that relate to the fuzzy variables with the output fuzzy variables using linguistic variables. In rule base all rules are processed in parallel manner by the inference engine. The inference rules direct the manner in which the following fuzzy sets are copied to the final fuzzy solution space.

In Fuzzy if-then rule consider the parameter number of packets and energy and by this parameter we consider the resulting possibilities are Normal Node (NN), Worst Node (WN). Here the inputs take 2 values Less and High. To find the fault node, we may set a threshold value according to the energy of packets. Decision module will be set forfault detection. If the corresponding value will be less than the threshold, it may consider fault node otherwise normal node. According to this value class label will be set for the node. From the class label we know the node may fault or not. If the class label is normal node will be normal, if class label is abnormal node will be fault. Figure 2 shows the conditions for decision making in fuzzy logic for inputs and its corresponding results.

During the process of defuzzification we may use maximum method. After decision making on the basis defuzzification, the normal and the best nodes are selected by the cluster head for data aggregation whereas the worst nodes are neglected by the cluster head. Then the cluster head transfers the aggregated data to the destination i.e., sink. Since the values of malicious and faulty sensors are not aggregated, secure data aggregation is ensured in the wireless sensor network.

Error Detection Model: Detection model detects an error in the network, so that we can improve the data accuracy and save the energy in the network. By using two detection models such as naive Bayes and fuzzy logic overall accuracy and efficiency will be improved. According to this model faulty node and normal node will be identified easily. A secure and efficient network will be obtained.

Figure 3 shows the overall detection model for fault node detection in WSN. Both the detectors will be accurately detected the fault node in the network and it may give an efficient network. To support data aggregation along with false data detection, every node will be monitored along with Naïve Bayes Detector and Fuzzy logic model. By using two detectors we can find the fault data easily and increase the accuracy of the network.

Performance Evaluation: This section describes the experimental results and performance evaluation of the proposed system.The performance of fault detection technique is evaluated through NS2 simulation. For this we develop a protocol called Dynamic Energy Aware Routing Protocol (DEAR) to improve the accuracy. The performance of DEAR technique is compared with the Dynamic Source Routing Protocol (DSR).

Simulation Model and Parameters: In the simulation, the number of nodes is kept as 40. The nodes are arranged in 1150 X axis distance and 800 in Y axis distance for 60 seconds of simulation time. Here using wireless channel type with Omni antenna. Initial energy will be set as 70 joules. 802_11 Mac type is used for connection and CMUPriQueue is the interface queue type that used in the network. Maximum packet in queue is 300. Source and destination nodes can be selected as any of the 40 nodes in the network during the run time. The simulated traffic is in Constant Bit Rate (CBR).

Performance Metrics: The performance is evaluated by metrics like Average packet delivery ratio, Packet drop, Energy, Bandwidth, False positive, Misdetect.

RESULTS

In our experiment, vary the rate as 100,200, 300,400 and 500 Figure 4 gives the packet delivery ratio when the rate is increased.

Here in X axis denotes the rate in kb and Y axis denotes the delivery ratio. It shows our proposed DEAR technique achieves a good delivery ratio when compared to DSR.

Figure 5(a),(b) gives the packet drop when the rate is increased. Here in X axis denotes the rate in kb and Y axis denotes the packets. It shows our proposed DEAR has lower packet drop than DSR.

Figure 6 gives the energy consumption when the rate is increased. Here in X axis denotes the rate in kb and Y axis denotes the energy in joules. It shows our proposed DEAR has less energy consumption than DSR.

Figure 7 shows the false positive rate of our DEAR technique and DSR. From the figure, we can observe that our DEAR scheme attains low false positive rate, when compared with DSR scheme, since it accurately detects the intrusion.

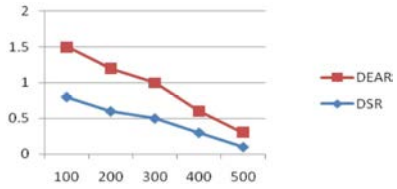
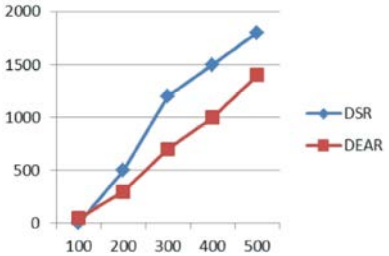
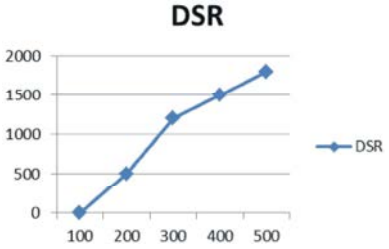


Fig. 4: Rate Vs Delivery Ratio



(a)



(b)

Fig. 5(a), (b): Rate Vs Packet Drop

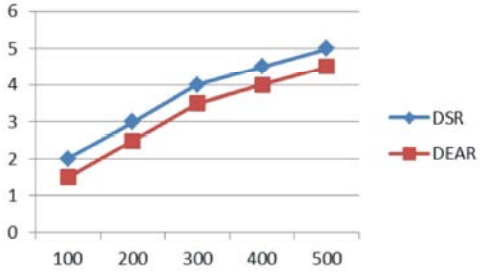


Fig. 6: Rate Vs Energy

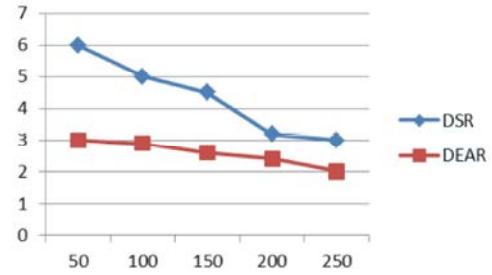


Fig. 7: Rate Vs False positive

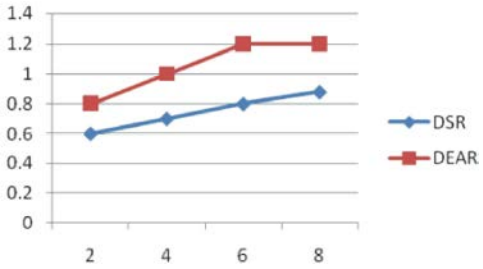


Fig. 8: Attackers Vs Bandwidth

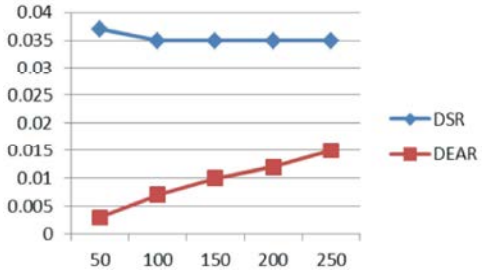


Fig. 9: Rate Vs Misdetect

Figure 8 gives the received bandwidth for normal legitimate users when varying the number of attackers. It shows that the bandwidth obtained from normal users is more in the case of DEAR when compared with DSR.

Figure 9 shows the misdetection ratio of DEAR technique and DSR. From the figure, we can see that the misdetection ratio is significantly less in our DEAR scheme when compared with DSR scheme, since it accurately detects the intrusion.

CONCLUSION

In this study, we have developed an error detection model which consists of the combination of Naïve Bayes detector and fuzzy logic. By this error detection model we can find the worst node and normal node. Since the values of malicious and faulty values are not aggregated during the sensing, so that secure data aggregation is ensured in the wireless sensor network. Our technique is improved the data accuracy of data during transmission and network lifetime is improved by avoiding unwanted data transmission.

REFERENCES

1. Keally, M., G. Zhou and G. Xing, 2010. Watchdog: Confident event detection in heterogeneous sensor networks,Real-Time and EmbeddedTechnology and Applications Symposium, IEEE, pp: 279-288.
2. Hill, M., M. Campbell, Y.C. Chang and V. Iyengar, 2008. Event detection in sensor networks for modern oil fields, in Proceedings of thesecond international conference on Distributed eventbased systems, ser. DEBS ’08. New York, NY, USA: ACM, pp: 95-102.
3. Heinzelman, W.B., A.L. Murphy, H.S. Carvalho and M.A. Perillo, 2004. Middleware to support sensor network applications, IEEE Network, 18(1): 6-14.
. Shanmugam, B. and N.B. Idris, 2006. Anomaly Intrusion Detection based on Fuzzy Logic and Data Mining, In Proceedings of the Postgraduate Annual Research Seminar, Malaysia.
4. Wahengbam, M. and N. Marchang, 2012. Intrusion detection in manet using fuzzy logic, 3rd IEEE Science (NCETACS), ISBN: 978-1-4577-0749-0, 189-192.
5. Abadeh, M.S., H. Mohamadi and J. Habibi, 2011. Design and analysis of genetic fuzzy systems for intrusion detection in computer networks, Expert Systems with Applications, 38: 7067-7075.
6. Greiner Russel and Wei Zhou, 2002. Structural Extension to Logistic Regression: Discriminative Parameter Learning of Belief Net Classifiers, In 13th international conference on uncertainty in artifi- cial intelligence.
7. Pernestal Anna, Mattias Nyberg and Bo Wahlberg, 2006. A Bayesian Approach to Fault Isolation with Application to Diesel Engine Diagnosis, In Proceedings of 17th International Workshop on Principles of Diagnosis (DX 06), pp: 211-218.
8. Madden, S.R., M.J. Franklin, J.M. Hellerstein and W. Hong, 2005. Tinydb: an acquisitional query processing system for sensor networks, ACM Trans. Database Syst., 30: 122-173.
9. Liu, T. and M. Martonosi, 2003. Impala: a middleware system for managing autonomic, parallel sensor systems, SIGPLAN Not., 38: 107-118.
10. Li, S., S.H. Son and J.A. Stankovic, 2003. Event detection services using data service middleware in distributed sensor networks, in Proceedingsof the 2nd international conference on Information processing in sensor networks, ser. IPSN’03. Berlin, Heidelberg: Springer-Verlag, pp: 502-517.
11. De Paola, A., G. Lo Re, F. Milazzo and M. Ortolani, 2013. QoS-aware fault detection in wireless sensor networks, International Journal of Distributed Sensor Networks, pp: 1-12.