Grenze International Journal of Engineering and Technology, Jan Issue

Statistical Test for Distributed Cluster-based Intermittent Fault Detection in Wireless Sensor Networks

Bhabani Sankar Gouda [1], Dr. Sudhakar Das [2] and Dr. Trilochan Panigrahi[3]
[1]Research Scholar, Computer Science and Engineering, BPUT Rourkela, Rourkela, India
bhabani012@gmail.com
[2-3]NIST Institute of Science and Technology, Department of Electronic and Communication, Berhampur, India
NIT Goa, Department of Electronic and Communication, Ponda, India
Email: sudhakar.dash@nist.edu, tpanigrahi@nitgoa.ac.in

***Abstract*—Due to the regular growing number of applications, distributed WSN enhanced to enjoy very large appeal. Occasionally, a small number of distributed nodes are drop some data value or unable to transmit accurate data to the respective or specific fusion Centre or to a nearby node. In the process of distributed network's performance suffers significantly because it is not aware of the node's malfunction. Data from the defective or affected node might be treated as outliers. So that for identify the data sent by the problematic sensor node can be found using a different statistical test. In this paper, proposed a DSFDOA method that a clustering-based statistical Z-test and a self-detectable distributed fault detection system are bringing forward. Put forward in DSFDOA and distributed sensor node algorithm is perform in Python and MATLAB, and staging is measured in terms of different parameter like false data alarm rate and detection data accuracy (DDA) & (FDAR). The results initiate demonstrate that the proposed approach shown performs better than the current algorithms and when compared to the execute results of the previous algorithm.**

***Index Terms*— WSN, Fault diagnosis, K-mean, Intermittent fault.**

I. INTRODUCTION

WSN is very large network but it represent of small region in a distributed manner , independently energy sensor distributed nodes called "wireless sensor networks" (WSNs) are dispersed around an small area and used to collect data, identify unusual events, and communicate wirelessly. This new technology has become a reality because to rapid advancements in embedded micro sensing, MEMS, and wireless communication [1], [2]. In general scenario in different time sensor nodes are remotely sensing the different applications include the specific environment, data monitoring, precision statistic of agriculture, medical observation and its applications. The major objective of the sensor nodes are Sensing the node, processing the data, and communication between to specific data points based on the components that when combined in a small device, enable a wide range of these applications. The effective design and deployment of WSNs has grown in importance due to their many widely used applications.

Grenze ID: 01.GIJET.9.1.741

The key characteristics of WSNs are availability, sustainability, and consistent performance, which can be harmed by a variety of things like external noise, topology changes brought on by broken links, energy depletion, hardware failure, malicious attacks, etc. The deployment of sensor nodes in unattended regions frequently results in them becoming defective. Any environmental changes have an impact on the sensors' performance because of message corruption from impulsive noise and topological changes, as well as message loss. Different fault types, including hard faults, soft faults, transient faults, byzantine faults, and others, can occur in WSNs. Sometimes a sensor node will be silent [3, 4], or it will send the fusion centre incorrect information. Soft faults occur when a sensor node exhibits anomalous behaviour for an extended period of time [5, 6]. There should be a decrease in accuracy in the estimating or procedure for detection if the network does not recognise that sort of problem and processes the data under the assumption that the particular sensor node is trouble-free. Finding such a malfunctioning node that is producing anomalous data is thus a crucial and difficult task in WSNs.

For the purpose of locating the soft faulty node, the centralised algorithms for fault diagnosis described in the literature [7] make larger use multi-hop of data exchange among the sensor nodes. The communication process uses the majority of the energy; consequently, the algorithms do not use little energy. [8]. It is vital to build an algorithm that would utilise reduced interaction between the fault nodes discovery because the sensor nodes in distributed WSNs operate with specific small range and limited energy power with battery and typically non-renewable resource. The distributed detection strategy is suggested here in place of the centralised method to improve the energy efficiency of the detecting algorithm.

The local coordination strategy [9], one of the most crucial techniques in distributed frameworks, is the foundation of the suggested algorithm. Each sensor node works along with its neighbours to find and recognise the network's malfunctioning sensor nodes and determine the degree of node confidence interval by examining the next hop neighbour distributed sensors' data. Here, the self-determination capabilities of each sensor limit the packet transfer. The cluster approach groups comparable types of objects into groups and defines how to divide elements so as to reduce the degree of similarity and dissimilarity [3,17]. The clustering algorithms have many applications in several domains, including learning theory, visualization, artificial intelligence, neural networks, and statistics. They are highly helpful in real-world settings. The clustering method, which is based on pattern categorization under unsupervised machine learning, timer series analysis, and statistical data, is also employed in practical applications [12,18] of distributed clustering.

II. RELATED WORK

Invalid data must be protected so that the network performance doesn't suffer. Sensor malfunction or environmental noise can result in faulty data. This section discusses sensor errors and significant methods that have been put out to date to prevent them in sensor networks. The two main categories of defect detection algorithms are centralised and distributed approaches. In the centralised approach, the fusion centre gathers information from all the cluster head or centroid with other distributed sensor nodes in a specified range and determines which sensor node is malfunctioning or faulty based on factors such as loss of data packets[10-11], communication delay, absence of frequent network activity [12-13], low buffer occupancy [14], the concept of channel loading, etc. Due to flooding, each of the aforementioned strategies raised the dispersed overhead of communication and its computation on the cluster network.

More nodes powered are used by the status exchange-based strategy outlined in [15]. The amount of energy left over is used as a warning sign for node faulty. By doing this way, additional node energy is used & the node quickly enters a dead state. Each sensor node in a distributed WSN can, to some extent, detect if it is faulty or fault-free, just like a biological component. Often referred to as a self-monitoring technique and show the node capability. This method uses less energy since it sends less information to the fusion centre or its neighbours. Authors suggested a self-monitoring strategy in [16]. In this method, each sensor node updates its neighbours once per predetermined amount of time. The neighbour then recognises the node's presence and notifies the fusion centre of its status. When that time period has passed and the neighbour has not discovered any update packets, the neighbour announces that the node might not exist and transmits information to the fusion centre. The data from each sensor is reported to the cluster head, and the cluster head is analysed by the fusion center to determine the state of each node's fault in the network.

Since each sensor is aware of its neighbours, this method employs a static network. However, the strategy covered in this study is used in a dynamic clustered network where each cluster gather the information of sensor node determines its status by processing the data it receives and identifying its neighbours based on the data at each time interval. Prior to contacting the central, each node in the nearby based on the clustering and coordination approach determines its status using information from its neighbours. A nearby coordination strategy is employed in [9]. Each node in this system determines its nodes status is faulty or fault free based on the discrepancy between two data points and its own cluster sensed data & by using the median method to reading of its neighbours. Clustering increases algorithm complexity, which results in very little or extremely poor energy efficiency. Numerous nodes were deployed in the same direction out of their own interest thanks to the multilayer clustering method. The node assembles or resembles the cluster node, depending on the gateway between them [11,19], and has effectively addressed distributed WSN clustering.

III. FAULT AND NETWORK MODEL

The network model of a decentralised arbitrary network with dispersed sensors is described at the outset of this section. A fault model is then discussed, which provides a concise explanation of how the fault entered the foundational network.

In this model we consider a distributed cluster based sensor network with containing of K sensor nodes, such as Kl, K2, K3………..K_N, which are distributed at random across a rectangular terrine. Assume that the every node has some specific transmission ranges of its cluster head and all the sensor nodes are the same and uniform. In order to make things simple, we'll suppose that each sensor node Ki can link to every sensor node and located inside the peripheral of cluster Ci where cluster sensor node Ki is regarded to be present at the centre of Ci. The cluster head data points has an own transmission network range of Si is the same as the radius I of Ci.

The suggested technique takes into account any softly malfunctioning sensor nodes that are present in the dynamic distributed WSNs. Since the malfunctioning sensor faulty nodes are still functional, it is presumed that all the faulty nodes have transmitting data to their close neighbour by using one-hop neighbours. These types of node give inaccurate information to their nearby nodes because of less or more range of data noise or the node failure of any sensor node data point based on the components.

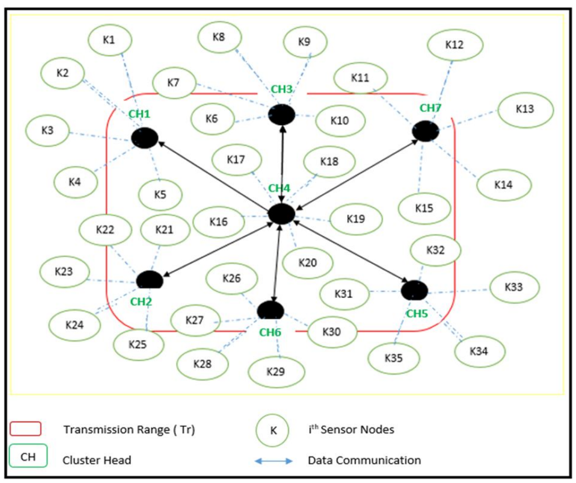

Figure 1. System model for distributed Network

Data points are divided into K clusters Sj (J = 1, 2,..., K) and associated with a representation using K-mean clustering procedures (cluster centres). In order to construct a better algorithm, the key phase of the proposed method is the division of data points into a set of represented vectors and K-mean clustering. Until a halting condition is satisfied, the mapping process is restarted starting with a starting set of cluster centres. The Lloyd iteration is used to represent the suggested distributed algorithms (based on K-mean).

Find the partition of K, where i=1, 2,..., K and Ki={X|d(X,C_i) ≤d d(X,C_j) } for all I ≠ j, then calculate the cluster head for each cluster to generate a new sensor node that set of clusters representing KC_{p+1}, given a set of cluster centres KC_p={C_j}. Find the partition of K, where i=1, 2,..., K and Ki={X|d(X,Ci) ≤ d d(X,Cj) for all I ≠ j. Then, given a set of cluster centres KCp={Cj}, calculate the cluster head for each cluster to produce a new sensor node that set of clusters representing KCp+1.

In the distributed process 7 no. of cluster head representing the network structure. All the cluster head CH1, CH2, CH3, CH5, CH6, and CH7 are connected through the main cluster or centroid CH4. Each cluster heads are sensing the connected nodes and gather the information. Based on the distance between to data point measure the distance and find out the data or node is faulty or fault free sensor node. Here sensor nodes are distributed and divided into uniform manner and find out the closest or nearest cluster head and do the communication. Each cluster head collected or connected with 5 no. of sensor nodes.

CH1 is gather the information from connected sensor nodes like K1,K2,K3,K4,K5 , CH2 cluster head collect the information from K2,K22,K23,K24 and K25, CH3 cluster head collect the information from K26,K7,K8,K9,K10 sensor nodes to fusion center, CH7 cluster head collect the information from K11,K12,K13,K14 and K15 sensor nodes, CH5 cluster head to collect the information from K31,K32,K33,K34 and K35 sensor nodes that closet to neighbor sensor nodes, CH6 collect the information from K26,K27,K28,K29 and K30 no. of sensor nodes and CH4 collect the information from assign the nodes and gather information from different cluster. CH4 cluster head is also connected or transmitted through specific nodes as K16, K17, K18, K19 and K20 respectively.

After gathered the information from sensor nodes and its cluster head compute the distance and find out faulty or not.

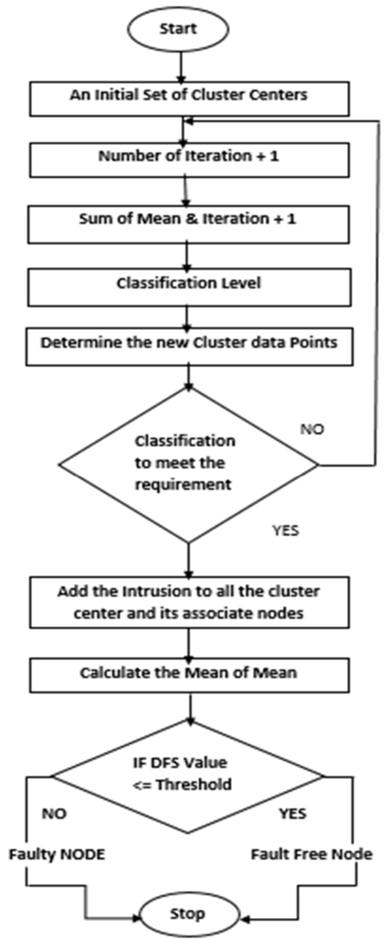

Figure 2. Flow chart of the proposed algorithm for fault finding

IV. DISTRIBUTED K-MEAN SELF FAULT DETECTION ALGORITHM

Here is a detailed description of the cluster-based k-mean self-detecting distributed algorithm. There are three different states in the DSFDOA algorithm. The following is a list of perceptive descriptions of various states.

A. Implementations states of the Proposed Algorithm

Here 3 unique states of the proposed K-mean algorithm to detect the self-cluster based faulty intermittent nodes are:

1) Data Collection state
2) Data Analysis state
3) Data Decision state

Each sensor node Ki performs a particular role at various phases. Each sensor node Ki initially extends an invitation to its neighbours by providing its own sensed data at a specific time t. Ki gathers all accessible sensed cluster data Xj from its neighbours Kj ξ Negi within its communication range (Tr) at the same time t. Then, Ki uses its neighbour's data from the analysis phase to anticipate its likely felt data interval at the regular time instant t.

Following data prediction, it uses comparison of sensed and cluster head's predicted the sense data in the data decision state to determine its status and state, based on which Ki will actively participate in cluster network functioning. Table I provides an overview of all the notations that have been utilised to describe the algorithmic processes. The following list includes the various occurrences that occur throughout the distributed self-detection algorithm's various phases.

B. Data Collection

During this stage, each sensor node Ki ξ K broadcasts the information it has sensed over its transmission or communication of cluster range (Tr). Each sensor Ki locates its surrounding nodes Kj ξ Negi using the reception data. It is expected that all of the sensor nodes in K that are present at deployment are fault free node of sensors with a fault status is 0.

C. Data Analysis

Utilizing the data collected from the surrounding nodes, each cluster sensor node Ki determines its likely probability of cluster sensed data PXi (4). In essence, the Z-score is evaluated by these mathematical equations (l)-(4). DSFDOA algorithm which is Z-test based on a statistical test in which a normal distribution can be used to approximate distribution of the test statistic under the null hypothesis. As a result, it is presumed that the cluster data sensed values collected by the various nodes is independent of one another. When the node's degree is higher, the z-test is preferred. The sensor cluster data obtained by cluster sensor node Ki's mean M is

$$M = \frac{1}{Degree(Si)} \sum_{Sj \in Negi} Xj \tag{1}$$

The standard deviation S is evaluated as

$$S = \sqrt{\frac{\sum_{Sj \in Negi}(Xj - M)^2}{Degree(Si) - 1}} \tag{2}$$

To evaluate the Z-score the following formulas are used

$$SE = \sqrt{\frac{S}{Degree(Si)}} \tag{3}$$

$$pxi = \sqrt{\frac{(Xi - M)}{SE}} \tag{4}$$

D. Data Decision

During the data decision step, each cluster sensor node Ki compares the cluster sensed data Xi with its own computed data PXi. If it falls between 2 and 3, Ki declares itself to be a good node; if not, it is labelled as a faulty sensor cluster node. The key phases of the algorithm that each individual node runs because it is purely distributed are described in the DSFDOA algorithm.

V. INTERMITTENT DYNAMIC FAULT DETECTION IN WSN WITH DISTRIBUTED DSFDOA

Data: Sensor cluster node position (XC0i, YC0i), Sensed cluster node temperature (X_i), KC: - Cluster centres, Number of sensed node (N_i), Transmission range of each centroid (Tr)

Results: Final fault status of sensor node K_i (FKi)

Step-1: K-mean Initialization State: Each sensor node cluster center K_i ε K finds their centroid and keep them in a cluster center KC_i. Set FK_i to 0 (Fault free)

TABLE I. THE ALGORITHM PARAMETER USED TO DEVELOP

Parameters	Description
K	Nodes in a Cluster of Sensors
K_i	Installed sensor node at P_i ($XC0_i$, $YC0_i$)
N	Each sensor node represents a sample of data (Total)
d	Data aspect
T_r	Transmission range of each cluster centroid K_i.
C	Set of nodes in K that contain all of the cluster sensor nodes' communications
G(K,C)	Connection between every sensor node K_i (Each cluster center)
FK_i	K_i calculates the fault state of the sensor node
$CRCD_i$	The total amount of data that was received from each cluster centroid belongs to the K_i sensor nodes
$d(K_i)$	K_i ε K is the cluster sensor node's degree
Xi	Distance is the measurement between samples from each sensed i_{th} sensor node
Threshold	The threshold level over which K_i transmits data is error-free
PX_i	K_i's anticipated computed data during interval t
KC_i	Cluster center accessible through P_i

Step-2: Analysis Phase: Set of cluster center KC_p and it perform here, the enhanced or improved set of cluster centre representatives is produced using the Lloyd iteration method.

KC_i+1 and CRCDNi=0;

For j=1…………………………….|KC_i| and K_j ε KC_i do

$CRCDN_i = CRCDN_i + X_j$;

$CRCDN_i = CRCDN_i / d(K_i)$;

$CKD_i = 0$

For j=1…………………………………|KC_i| and K_j ε KC_i do

$CKD_i = CKD_i + (X_j - CRCDN_i)^2$

$KD_i = \sqrt{CKDi/(d(Ki) - 1}$

$KE_i = \sqrt{KDi/(d(Ki)}$

$PX_{i\,=}$ (Xi-$CRCDN_i$)/Ke_i

Step-3: Data Decision State: Stop if the cluster network hasn't changed much since the prior iteration; otherwise, continue. Set P+1 -> P and go to Step-2

If $|X_i - PX_i| \geq 2$ and $|X_i - PX_i| < 3$ then

Cluster sensor node K_i is expected to be proven to be Incorrect or fault free sensor cluster node of each centroid

FK_i=0

Else

Cluster sensor node Ki is expected to be proven to be Incorrect or faulty cluster sensor node of each centroid

FK_i=1

VI. RESULT AND ANALYSIS

The MATLAB was used to implement the proposed technique (version 18). In the proposed approach MAC layer was analyzed using DCF of IEEE 802.l1a. The CSMA/CA protocol is used by each sensor node to send its data packets. A 200 x 200 square meter area was covered by 2048 sensor nodes, which were randomly dispersed to create the arbitrary network topology. The sensor nodes had fixed cluster head transmission or communication radius of 5 m and were functionally similar.

The consumption of per second 0.5 mJ, 0.10 mJ, 0.15 mJ, 0.20mJ and 0.25 mJ of energy were used for transmission, reception, and standby, respectively. Each sensor node has a set battery power of 1 J allotted to it. Each sensor node's sensed data was evenly divided. Data from broken sensor nodes have mean 50 and variance 0, while data from fault-free sensor nodes have mean 30 and variance 0. The network's performance is evaluated by adding 0.05, 0.10, 0.15, 0.20, 0.25, and 0.30 percent faulty of cluster sensor nodes, which were selected using an established in uniform distribution function.

The different parameter as like false data alarm rate (FDAR) and detection data accuracy (DDA) are used to gauge how well the algorithm is working.

The Fig. 3(a) and 3(b) depict the analysis results for find an average node degree of the sensor network are 7 and 10. The average node degree of the sensor network can be changed in regular time interval by changing the communication radius. It is discussed and shown in the both figures 3a and 3b that the performance analysis of the proposed DSFDOA algorithm is better compared to the DSFDA and JSA algorithm. When the average node

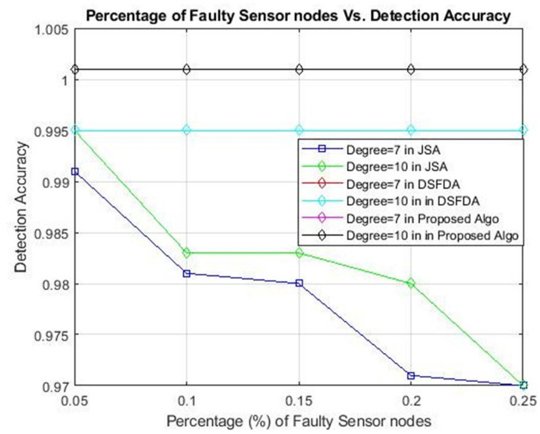

Figure 3a. % of faulty nodes Vs DDA

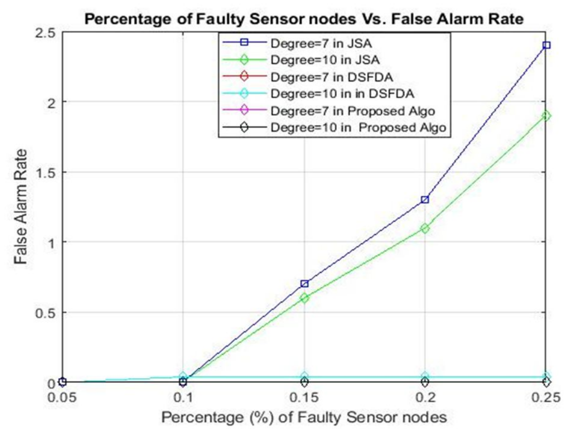

Figure 3. Performance analysis of proposed DSFDOA with JSA and DSFDA

degree of the sensor network is 10, the proposed algorithm represents DDA and FDAR are to be 1 and 0 respectively. This demonstrates shown that if an any node has consists of more degrees, it may only analyse 10 data points for the cluster based neighbours to find out the fault state and may choose not to use the remaining information. In this method, we can determine the minimal amount of information needed to analyse the fault condition and cut down on communication costs.

When the network degree is high, a node has more information to analyse its defective status. If more data are considered, the statistical z-score will be more accurate. This is reflected in the suggested statistical method's calculations for both DDA and FDAR. Additionally, reduced communication overheads are required by the distributed approach to identify the softly broken sensor node.

VII. CONCLUSION

In this research, a DSFDOA which is z-test-based distributed soft fault detection algorithm for WSNs is developed. The data from a faulty node is regarded as an outlier in the suggested approach. The sensor gathers information from nearby cluster nodes, uses z-test to analyse it, and then determines its own fault state as well as that of its neighbouring nodes. By offering high data detection precision and a low data false alarm rate, the algorithm outperforms.

REFERENCES

[1] I. Akyildiz, W. Su, Y Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Science Direct Trans. on Computer Networks, vol. 38, no. 4, pp. 393-422, 2002.

[2] I.F.Akyildiz, W. Su, YSankarasubramaniam, and E. Cayirci, "A survey on sensor networks," IEEE Trans. on Commun. Magazine, vol. 40, pp. 102-114, 2002.
[3] F. Preparata, G. Metze, and R. T. Chien, "On the connection assignment problem of diagnosable systems," IEEE Trans. on Commun. Magazine, vol. EC-16, no. 6, pp. 848-854, Dec. 1967.
[4] S. Chessa and P. Santi, "Comparison-based system-level fault diagnosis in ad hoc networks," in Reliable Distributed Systems. 2001. Proceedings. 20th IEEE Symposium on, 2001, pp. 257 -266.
[5] S. Kamei and H. Kakugawa, "A self-stabilizing distributed approximation algorithm for the minimum connected dominating set," in in the Proc. of IEEE International Parallel and Distributed Processing Symposium.2007(IPDPS 2007), June 2007, pp. I - 8.
[6] J. Y Weigang Wu, Jiannong Cao, "A self-stabilizing distributed approximation algorithm for the minimum connected dominating set," in in the Proc. of 14th International Conference on Computer Communications and Networks. 2005(ICCCN 2005), Oct 2005, pp. 165 - 170.
[7] M. Panda and P. Khilar, "An efficient fault detection algorithm in wireless sensor network," in Contemporary Computing, ser. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2011, vol. 168, pp. 279-288.
[8] D. Estrin and M. Srivastav, "Instrumenting the world with wireless sensor networks," in IEEE International conference on Acoustics. Speech. Signal Processing (ICASSP'Ol), May 2001, pp. 2033-2036.
[9] M. M. Asim and M. H. Merabati, "A fault management architecture for wireless sensor network," in Proceedings of the international conference on Wireless Communication and Mobile Computing, ser. TWCMC '08, Aug 2008, pp. 779-785.
[10] S. 1. Park, R. Vedantham, R. Sivakumar, and I. F. Akyildiz, "A scalable approach for reliable downstream data delivery in wireless sensor networks," in ACM MobiHoc Conference.2004, 2004, pp. 78-89.
[11] J. Staddon, D. Balfanz, and G. Durfee, "Efficient tracing of failed nodes in sensor networks," in ACM International workshop on wireless sensor networks and application, 2002, pp. 122-130.
[12] N. Ramanathan, K. Chang, R. Kapur, L. Girod, E. Kohler, and D. Estrin, "Sympathy for the sensor network debugger," in SenSys.2005, 2005, pp. 1-12.
[13] C. Y Wan, S. B. Eisenman, and A. T. Campbell, "Coda: Congestion detection and avoidance in sensor networks," in ACM SenSys03, 2003, pp. 266-279.
[14] Y Sankarasubramaniam, O. B. Akan, and I. F. Akyildiz, "Esrt: Event-tosink reliable transport in wireless sensor networks," in ACM MobiHoc03, 2003, pp. 177-188.
[15] Y J. Zhao, R. Govindan, and D. Estrin, "Residual energy scan for monitoring sensor networks," in IEEE WCNC.02, 2002, pp. 256-262.
[16] A. M. Mokhtar and H. M. M., "A fault management architecture for wireless sensor network," in IWCMC'08, 2008, pp. 779-785.
[17] S. Bandyopadhyay, E.J. Coyle,"An energy efficient hierarchical clustering algorithm for wireless sensor networks", IEEE INFOCOM, Volume 3, 2003, Pages 1713-1723.
[18] Manasi N. Joshi, "Parallel K - Means Algorithm on Distributed Memory Multiprocessors", Spring 2003.
[19] Y. Liu, N. Xiong, Y. Zhao, A.V. Vasilakos, J. Gao, and Y. Jia, ''Multilayer clustering routing algorithm for wireless vehicular sensor networks'', IET Commun., 2010, Vol. 4, Iss. 7, pp. 810—816.