Journal of King Saud University – Computer and Information Sciences (2018) 30, 334–348

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

An ANFIS estimator based data aggregation scheme for fault tolerant Wireless Sensor Networks

Sasmita Acharya [a,*], **C.R. Tripathy** [b]

[a] *Department of Computer Application, Veer Surendra Sai University of Technology, Burla, India*
[b] *Department of Computer Science and Engineering, Veer Surendra Sai University of Technology, Burla, India*

Received 28 July 2016; revised 9 September 2016; accepted 3 October 2016
Available online 8 October 2016

KEYWORDS

Wireless Sensor Networks;
Fault tolerance;
Data aggregation;
ANFIS;
Estimator

Abstract Wireless Sensor Networks (WSNs) are used widely in many mission critical applications like battlefield surveillance, environmental monitoring, forest fire monitoring etc. A lot of research is being done to reduce the energy consumption, enhance the network lifetime and fault tolerance capability of WSNs. This paper proposes an ANFIS estimator based data aggregation scheme called Neuro-Fuzzy Optimization Model (NFOM) for the design of fault-tolerant WSNs. The proposed scheme employs an Adaptive Neuro-Fuzzy Inference System (ANFIS) estimator for intra-cluster and inter-cluster fault detection in WSNs. The Cluster Head (CH) acts as the intra-cluster fault detection and data aggregation manager. It identifies the faulty Non-Cluster Head (NCH) nodes in a cluster by the application of the proposed ANFIS estimator. The CH then aggregates data from only the normal NCHs in that cluster and forwards it to the high-energy gateway nodes. The gateway nodes act as the inter-cluster fault detection and data aggregation manager. They proactively identify the faulty CHs by the application of the proposed ANFIS estimator and perform inter-cluster fault tolerant data aggregation. The simulation results confirm that the proposed NFOM data aggregation scheme can significantly improve the network performance as compared to other existing schemes with respect to different performance metrics.

1. Introduction

The Wireless Sensor Networks (WSNs) consist of a large number of small, low cost, limited energy and autonomous sensor nodes which are randomly deployed in different areas of interest. The sensor nodes are usually organized into different clusters in order to conserve energy in data transmission and to prolong the network lifetime. The Non-Cluster Head sensor nodes (NCHs) sense the environment for some phenomenon of interest, collect data and forward them to their Cluster Heads (CHs). The CH node performs in-network data aggregation and then forwards the aggregated data to the base

* Corresponding author.
E-mail addresses: talktosas@gmail.com (S. Acharya), crt.vssut@yahoo.com (C.R. Tripathy).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2016.10.001
1319-1578 © 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.

station through single or multiple hops depending upon the network topology. The WSN is used widely in many applications which include search and rescue operations, battlefield surveillance, environmental monitoring, forest fire monitoring, and home automation and weather monitoring.

Generally, the WSNs are designed to operate in harsh environments with the minimum human intervention. A sensor node has to rely on its limited battery power due to the limited resources in WSNs. There are many factors that may cause the failure of a WSN. A WSN may fail due to the malfunctioning of some of its components that may be hardware or software faults, faults in the network communication layer or the application layer or may be due to battery depletion etc. The fault tolerance refers to the ability of a system to perform at a desired level even in the presence of faults. A number of research papers have been published in the area of fault detection and recovery in WSNs (Javanmardi et al., 2012; Jiang, 2009; Chang et al., 2013; Xu et al., 2014). But most of these mechanisms consume lots of extra energy for fault detection and recovery and even some require additional hardware and software resources for the same. The problem arises when a cluster has a large number of faulty NCH nodes that transmit their faulty data to the CH and finally to the base station. This makes the entire network unreliable for future data transmission and in severe cases, the network may collapse.

This paper proposes an ANFIS estimator based data aggregation scheme called Neuro-Fuzzy Optimization Model (NFOM) for fault tolerant data aggregation in WSNs. The proposed scheme employs an Adaptive Neuro-Fuzzy Inference System (ANFIS) estimator for intra-cluster and inter-cluster fault detection for different fault cases. The ANFIS estimator takes some fuzzy inputs from the sensor node parameters and generates a fuzzy output by the application of fuzzy rules. The generated fuzzy output is then fed to a defuzzifier which returns a crisp output about the sensor node status. The CH acts as the intra-cluster fault detection and data aggregation manager. It identifies the faulty NCH nodes in a cluster by the application of the proposed ANFIS estimator. The CH then aggregates data from only the normal NCHs in that cluster and forwards it to the high-energy gateway nodes. The gateway nodes act as the inter-cluster fault detection and data aggregation manager. They identify the faulty CHs by the application of the proposed ANFIS estimator. The faulty cluster is then isolated and not allowed to participate in the data aggregation process. This pro-active approach prevents the network from partitioning into disjoint segments due to faulty clusters. The faulty CH is then replaced by its nearest one-hop neighbor with maximum residual energy.

The paper also compares the relative performance of the proposed NFOM scheme with Distributed Fault Detection (DFD) Jiang, 2009, Low Energy Distributed Fault Detection (LEDFD) Xu et al., 2014, Majority Voting (MV) Javanmardi et al., 2012 and Fuzzy Knowledge based Fault Tolerance (FTFK) Chang et al., 2013 through simulation with respect to different performance metrics and the simulation results are discussed. The related work is presented in Section 2. The grid cluster WSN model and the proposed fault model are presented in Section 3 as the system model. An overview of the ANFIS estimator and the proposed ANFIS estimator based data aggregation scheme NFOM, NFOM mechanism and the algorithm are presented in Section 4. The simulation results are presented and discussed in Section 5. The Section 6 presents the concluding remarks and the future scope for further research.

2. Related work

This section outlines the different distributed fault detection and recovery techniques used in WSNs and their relative advantages and disadvantages.

A routing protocol known as the Threshold sensitive Energy Efficient sensor Network protocol (TEEN) for reactive networks is discussed in Manjeshwar and Agrawal (2009). This protocol lacks mechanism to handle node failures and is not suitable for real-time applications. An intelligent sleeping mechanism (ISM) for WSNs is discussed in Hady et al. (2013) where the base station decides on which clusters to be set to sleep mode in a specific round depending on the significance of the data sent by that cluster to the base station.

The Dual Homed Routing (DHR) mechanism discussed in Jain et al. (2008) is efficient in case of Primary Cluster Head (PCH) failures but consumes a lot of energy in transmitting data twice to the PCH and the Backup Cluster Head (BCH) nodes simultaneously. The Informer Homed Routing (IHR) mechanism proposed in Qiu and et al. (2013) is an improvement over DHR in which the Non-Cluster Heads (NCHs) send data only to the Primary Cluster Heads (PCHs). A reinforcement-based Q-Learning technique for routing in WSNs is discussed in Sharma et al. (2012).

The Low Energy Adaptive Clustering Hierarchy (LEACH) presented in Kaur and Saini (2013) gives a mechanism to avoid out-of-power node failures in micro-sensor networks. This mechanism only extends the lifetime of the network but fails to reduce the data loss due to node failures. The Hybrid Energy Efficient Distributed Clustering (HEED) clustering algorithm proposed in Younis and Fahmy (2004) is an improvement over LEACH where a hybrid function is used to periodically select cluster heads based on the residual energy of a node and its proximity to its neighbors or nodal degree. Different energy efficient fault tolerance mechanisms are discussed in Jeevanandam et al. (2014), Heinzelman et al. (2002), Attia et al. (2007). A scheme to provide continuous sensor services against random node failures called R-Sentry is discussed in Yu and Zhang (2007). This scheme has the disadvantage that it incurs extra communication overhead between the sentry and the active nodes periodically. Further, the sentry nodes do not collect any data.

An analytical model for WSNs with sleeping nodes is discussed in Chiasserini and Garetto (2006). The sensor nodes of this WSN may enter into sleep mode corresponding to low power consumption or reduced operational capabilities like low battery life. A novel approach for faulty node detection in WSNs using fuzzy logic and majority voting technique is discussed in Javanmardi et al. (2012). An improved Distributed Fault Detection (DFD) scheme to check out the failed nodes in the network is discussed in Jiang (2009). The DFD scheme works by exchanging data and mutually testing among neighbor nodes in the network. But when the number of neighbor nodes is small and the node's failure ratio is high, the fault detection accuracy of a DFD scheme decreases rapidly. A Fault Tolerance Fuzzy Knowledge based control algorithm (FTFK) is discussed in Chang et al. (2013) which detects faulty communication between sensor nodes and provides Fault

Detection and Isolation (FDI) to eliminate the faulty communication behavior of nodes in WSNs. A Fuzzy Logic based Joint Intra-cluster and Inter-cluster Multi-hop Data Dissemination Approach in Large Scale WSNs is presented in Ranga et al. (2015) where a multi-criterion fuzzy logic based intra-cluster and inter-cluster multi-hop data dissemination protocol is proposed to balance among the nodes and choose more stable nodes as CHs.

A Low Energy consumption Distributed Fault Detection (LEDFD) algorithm for Wireless Sensor Networks is proposed in Xu et al. (2014) which take advantage of the temporally correlated and spatially correlated characteristics of the sensor nodes to obtain good fault detection performance and save network power. A Markov chain-based model for missing and faulty data in MICA2 sensor motes is discussed in Koushanfar and Potkonjak (2005). A fault management protocol for low-energy and efficient WSNs are discussed in Liu et al. (2013). A fuzzy approach to data aggregation to reduce power consumption in WSNs is discussed in Lazzerini et al. (2006). The approach in Volosencu and Curiac (2013) provides a technical solution to improve the efficiency in multisensor wireless network based estimation for distributed parameter systems. It implements a complex structure based on estimation algorithms with regression and autoregression using three kinds of estimators – linear, neural and ANFIS.

Different inter-actor connectivity restoration techniques for wireless sensor and actor networks are reviewed in Acharya and Tripathy (2014). An Artificial Neural Network (ANN) approach for design of reliable and fault tolerant WSNs is discussed in Acharya and Tripathy (2015) which proposes the use of an exponential Bi-directional Associative Memory (e-BAM) in order to train the network. A fuzzy knowledge based sensor node appraisal technique for fault tolerant data aggregation in WSNs is discussed in Acharya and Tripathy (2016) which uses a combination of fuzzy rules to identify the faulty nodes in the network.

3. System model

This section presents the WSN network model and the proposed fault model.

3.1. The WSN network model

A typical 3×3 grid cluster WSN is shown in Fig. 1. In a 3×3 grid cluster, the CH is located at the center with 8 number of NCH nodes in the periphery. This arrangement makes it easy for the CH to monitor the status of the NCH nodes. This 3×3 arrangement of grid cluster also facilitates easy rotation of CHs for load balancing. Also, the CH is just one-hop from all the NCH nodes in a 3×3 grid cluster. This feature greatly saves on energy and communication overhead. The WSN network model consists of a heterogeneous Wireless Sensor Network in which the nodes have varying energy.

In the WSN model, the WSN is assumed to have thousands of sensor nodes with varying energy which are grouped into a number of 3×3 grid clusters. The sensor nodes with comparatively low energy form the NCH node. In a cluster, the NCH with the highest energy becomes the CH. The CHs are rotated between different NCHs as per the TDMA schedule in order to conserve energy. The base station is assumed to be *stationary*. Fig. 1 shows a number of 3×3 grid clusters with the CH at the center and NCHs in the periphery. A record table is maintained which records the history of all local decisions during the data aggregation process. The cluster heads are selected with a probability equal to the ratio of the expected number of CHs to the total number of sensor nodes in the network.

3.2. The proposed fault model

In the proposed fault model, the fault tolerance capability of WSN is assessed by observing the network's reaction to different types of faults. In the proposed work, the NCH and CH faults are injected into the network at different rates in order to simulate different fault cases. The proposed NFOM data aggregation scheme is tested for five cases with different fault types and with different node densities and the different performance metrics are recorded. The simulation results for the proposed scheme are then compared with that of DFD [35], LEDFD [38], MV [34] and FTFK [36] techniques and the results are presented. It is assumed that the fusion center or the data aggregator is unaware about the sensor fault types in advance. Depending on the type of data collected by a sensor node, the sensor faults are categorized into the following broad types.

(a) *Fixed faults*: The sensors with this kind of fault collect data with the same readings. These data are not affected by the environment. They are of two types - *stuck-at-zero faults* and *stuck-at-one faults*. In the former, a faulty sensor node always transmits a fixed local decision '0' to the fusion center irrespective of the real observation while in the later, it always transmits a fixed local decision '1' to the fusion center irrespective of the real observation.
(b) *Random faults*: In this case, a faulty sensor node reports randomly its local decision to the fusion center irrespective of the real observation. The sensor node readings are random and uncertain.
(c) *Transient faults*: These faults may occur due to hardware features or effects of environment in the process of data collection. They can be easily corrected using the majority voting technique discussed in Javanmardi et al. (2012).
(d) *Mixed faults:* It is a combination of two or more fault types.

4. Proposed NFOM data aggregation scheme

This section gives an overview of the ANFIS estimator and the proposed NFOM data aggregation scheme. It also describes the NFOM mechanism and algorithm.

4.1. Overview of the ANFIS estimator

The proposed NFOM data aggregation scheme uses an ANFIS estimator (Volosencu and Curiac, 2013). The

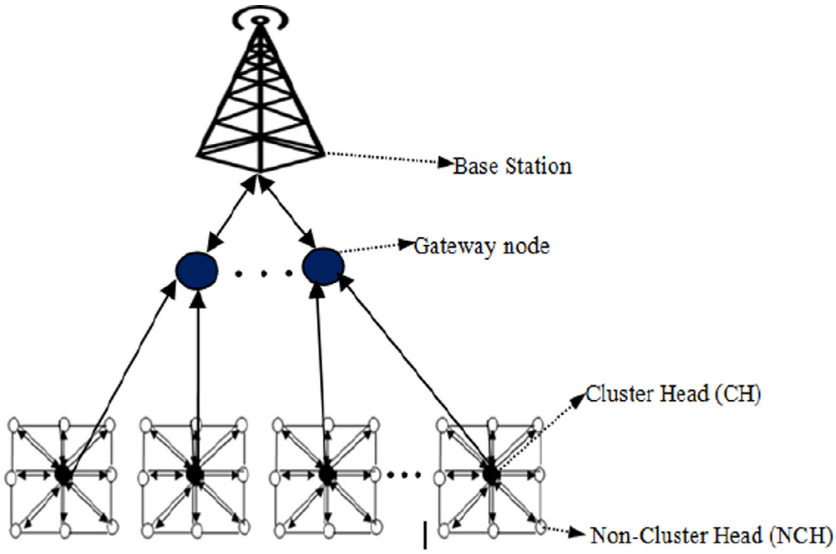

Figure 1 A 3 × 3 grid cluster WSN.

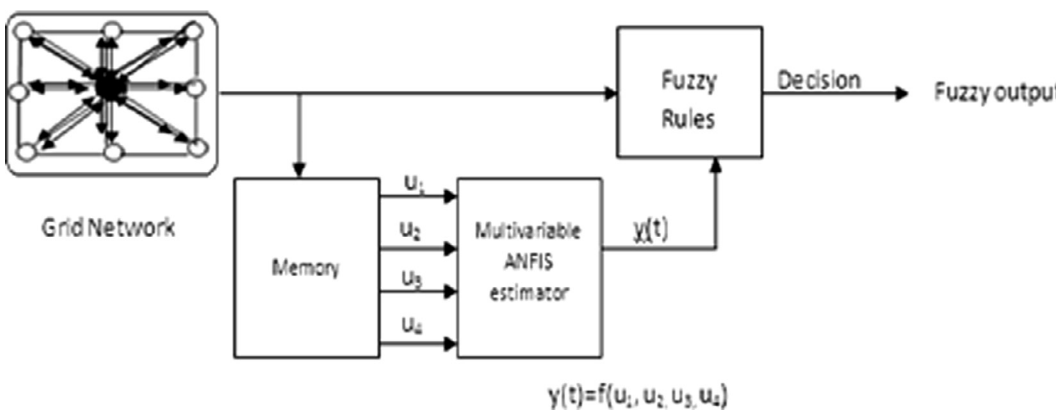

Figure 2 Estimation and detection structure of the ANFIS estimator.

estimation and detection structure of the ANFIS estimator used in the proposed model is shown in Fig. 2.

The ANFIS is a non-linear estimator defined by a function $y(t) = f(u1, u2, u3, u4)$ where$u1$, u 2, $u3$ and $u4$ are the four inputs. It uses the adaptive network-based fuzzy inference system and a set of fuzzy rules to generate the fuzzy output. It uses a combination of least-squares and back propagation gradient descent methods for training membership function parameters and modeling a given set of input or output data. The first layer is the *input layer* where the four inputs are applied. The second layer represents the *input membership function* and is called as the fuzzification layer. The neurons represent fuzzy sets and are used to determine the membership degree of the input. The activation function is represented by the membership function. The third layer represents the *fuzzy rule base*. Here, each neuron corresponds to a single rule from the set of fuzzy rules. The fourth layer represents the *output membership function* which is given by the activation function. The fifth layer represents the *defuzzification layer* with single output. The defuzzification method used for this model is the centroid method.

Table 1 Sample of fuzzy rules.

Fuzzy inputs				Fuzzy output
RNE	PDR	FR	NOR	NS
VL	P	P	P	F
L	P	P	P	F
H	G	G	G	N
M	A	A	A	N
H	VP	VP	VP	F
M	P	VP	VP	F
VL	VP	VP	VP	F
L	VP	VP	VP	F
VL	P	VP	VP	F
VL	VP	P	P	F
M	VP	VP	VP	F

4.2. Description of the proposed NFOM data aggregation scheme

This section gives an overview of the proposed NFOM data aggregation scheme. The ANFIS estimator in the proposed model monitors the status of each NCH node in a cluster

```
Input :  Network configuration, number of simulation rounds 'R', cluster threshold CT, grid size gd_size, total
         number of clusters 'n'.
Output : A fault tolerant data aggregation NFOM scheme.
         for ∀R in the network                         // for each simulation round

    Step 1: Initialize the network parameters.

           Initialize node 0 as BS and node g_t at level 't' as the gateway node.
           Set count[c_k] = 0 for cluster k = 1,2,…,n.

    Step 2: Apply the ANFIS estimator to estimate the node status of all the nodes in a cluster.

           for ∀cluster c_k ( k = 1,2,…,n)
             for ∀node n[i,j] in cluster c_k    // 'i' and 'j' represent row and column positions in the grid.
             Apply the ANFIS estimator to estimate the node status NS(n[i,j]) using fuzzy rules

    Step 3: Set the Cluster Fault Matrix (CFM) entries and forward flag (fwd_flag) values for all the nodes in
            a cluster.

           for ∀cluster c_k ( k = 1,2,…,n)
             for ∀node n[i,j] in cluster c_k

              if ( NS(n[i,j]) == 'F' )
                 then CFM(n[i,j]) = 1;                // faulty node
                      and fwd_flag(n[i,j]) = 'N';     // faulty NCH data not forwarded to CH

              else if ( NS(n[i,j]) == 'N' )
                 then CFM(n[i,j]) = 0;                // normal node
                      and fwd_flag(n[i,j]) = 'Y';     // normal NCH data forwarded to CH

    Step 4: Count the number of faulty nodes in a cluster

           for ∀cluster c_k ( k = 1,2,…,n)
             for ∀node n[i,j] in cluster c_k
               if ( CFM(n[i,j]) == 1 )
                  then count[c_k] = count[c_k] + 1;

    Step 5: Compute the Cluster Fault Index (CFI) for each cluster and forward it to the gateway node

           for ∀cluster c_k ( k = 1,2,…,n)
             CFI [c_k] = ( count[c_k] * 100 ) / gd_size;

    Step 6: Identify faulty clusters in the network by the gateway node

           for ∀cluster c_k ( k = 1,2,…,n)
             if ( CFI [c_k] < CT )                       // normal cluster
                then cfwd_flag[c_k] = 'Y';              // forward data from CH to gateway node

             else if ( CFI [c_k] ≥ CT )                  // faulty cluster
                then cfwd_flag[c_k] = 'N';              // faulty cluster data not forwarded
```

Figure 3 Proposed NFOM pseudo-code.

and assigns a node status to each node by the application of fuzzy rules. The node status value may be 0 (normal) or 1(-faulty). The node status of a 3×3 grid cluster is stored in a Cluster Fault Matrix (CFM). A CFM is generated for each cluster and the Cluster Fault Index (CFI) is calculated. If the CFI value is greater than or equal to the Cluster Threshold value (CT), then the cluster is identified as a faulty one and is isolated. The faulty cluster is not allowed to participate in the data aggregation process.

The fuzzy rules are generated using a combination of four fuzzy inputs for inter and intra-cluster criteria – residual node energy (RNE), packet delivery ratio (PDR), fault ratio (FR) and number of re-transmissions (NOR) which are denoted by $u1$, $u2$, $u3$ and $u4$. The RNE is given as the energy remaining after each simulation round. The PDR is given as the ratio of the number of data packets successfully sent to the CH to the total number of data packets. The FR is given as the ratio of the number of simulation rounds in which the NCH is found to be faulty to the total number of simulation rounds. The NOR is given by the number of times the data are transmitted again (*re-transmitted*). The RNE values vary from Very Low (VL), Low (L) and Medium (M) to High (H) and have a trapezoidal membership function. For example, if the residual energy of a NCH node is 0.05 J, it may be assigned Very Low (VL) level. A trapezoidal MF is specified by four parameters. The PDR, FR and NOR values vary from Good (G), Average (A) and Poor (P) to Very Poor (VP) and have a trapezoidal membership function assigned to them. These values are updated by the CH at the end of each simulation round and are stored in a record table. The fuzzy output is the Node Status (NS) which can take values – Normal (N) or Faulty (F) according to the fuzzy rules and has a triangular membership function assigned. The fuzzy output NS is updated periodically to the CH in intra-cluster and to the gateway node in inter-cluster data aggregation process.

The job of the NCHs is to sense the environment, collect data and forward it to their respective CHs. The CH acts as the intra-cluster fault detection and data aggregation manager. It performs three vital functions – monitor the status of each NCH in the grid cluster; pro-actively identify the faulty NCHs by the application of the proposed ANFIS estimator and perform intra-cluster fault tolerant data aggregation. The aggregated data are then forwarded by the CH to high energy nodes called the gateway nodes. The gateway nodes act as the inter-cluster fault detection and data aggregation manager. They perform three vital functions – monitor the status of each CH; pro-actively identify the faulty CHs by the application of the proposed ANFIS estimator and perform inter-cluster fault tolerant data aggregation. The gateway nodes have higher energy than the CH nodes. So, they are entrusted with the additional responsibility of replacing the faulty CH in a cluster by its nearest one-hop NCH node with the highest residual energy. This change in CH is communicated to all the NCHs in the cluster and the base station and the routing table is updated accordingly. Finally, the aggregated data are forwarded from the gateway node to the base station.

4.3. The proposed NFOM mechanism and algorithm

This section discusses the proposed NFOM mechanism and presents the NFOM algorithm.

Table 2 Simulation configuration parameters.

Parameter	Value
Network size	1200 m × 1200 m
Base station location	(100, 100) m
Number of sensor nodes	450
Number of clusters	50
Number of gateway nodes	5
Simulation time	500 s
Radio range	150 m
Energy transmitted	0.66 W
Energy received	0.395 W
Data packet size	2 KB
Packet aggregation ratio	10%

Table 3 Average FND and HNA metrics.

Algorithm	FND	HNA
DFD	200	530
LEDFD	275	600
MV	300	610
FTFK	310	640
Proposed NFOM	330	670

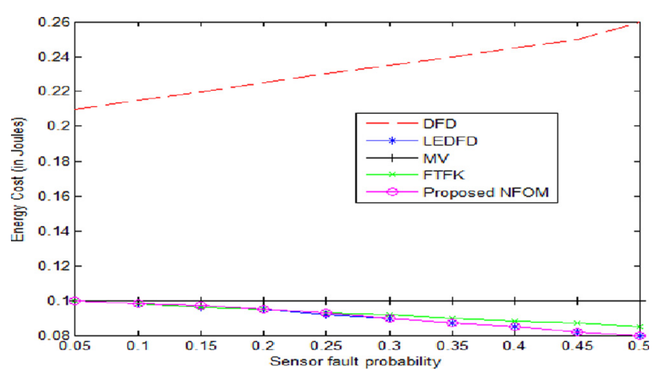

Figure 4 Comparison of energy cost versus sensor fault probability.

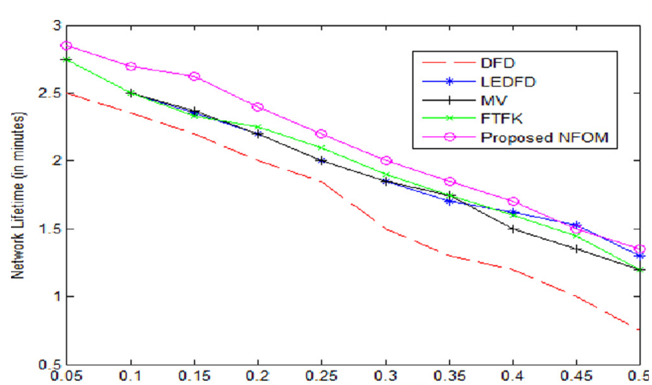

Figure 5 Comparison of network lifetime versus sensor fault probability.

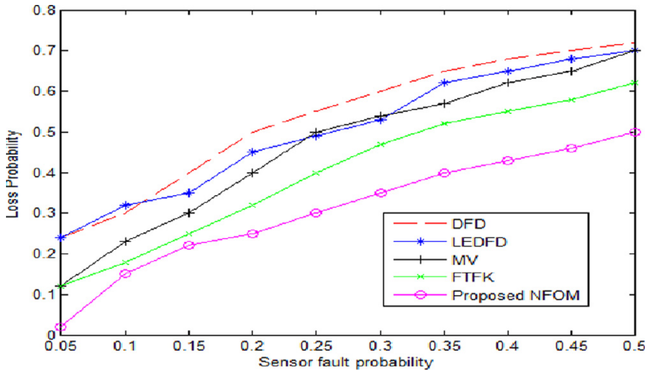

Figure 6 Comparison of loss probability versus sensor fault probability.

4.3.1. NFOM mechanism

The proposed NFOM mechanism is outlined below.

(a) At the beginning of each simulation round, the CH employs the ANFIS estimator to estimate the status of each NCH node.
(b) The NCHs are assigned node status as normal (NS = 0) or faulty (NS = 1) according to the fuzzy rules for the different intra-cluster criteria.
(c) Only the normal NCH nodes are allowed to forward their data to their respective CH. The faulty NCHs are isolated.
(d) The 3 × 3 Cluster Fault Matrix (CFM) is generated for each cluster and the Cluster Fault Index (CFI) is calculated asCFI = (No. of faulty nodes in a cluster/Total number of nodes in the cluster) * 100.

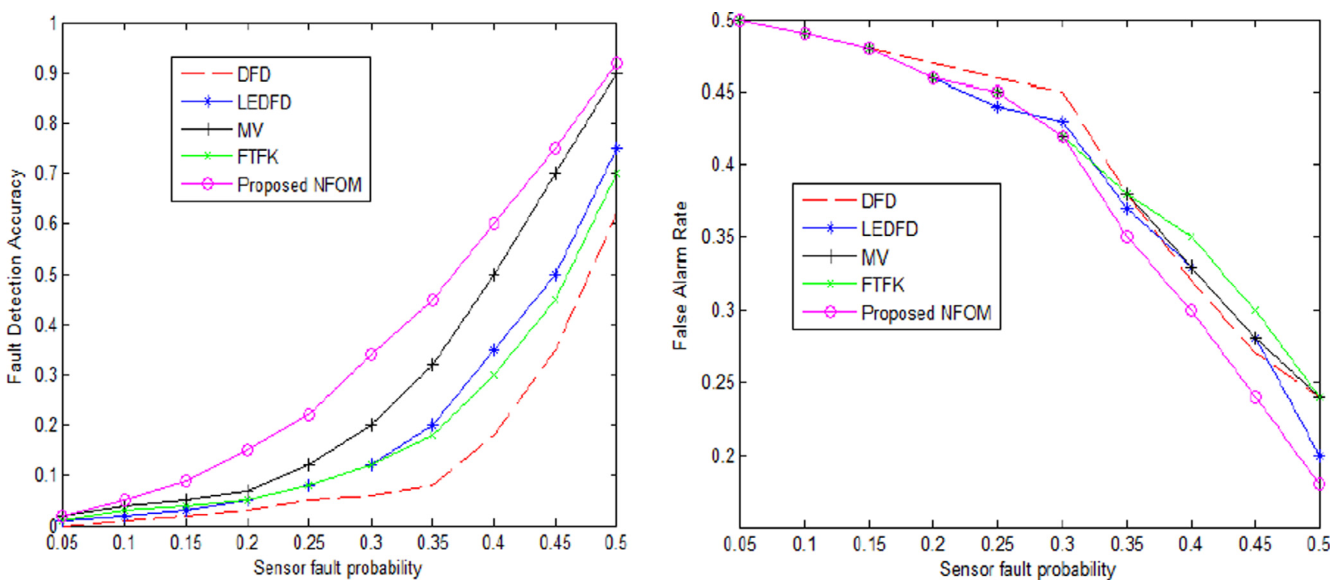

Figure 7 Comparison of FDA and FAR values for fixed faults with average node density = 7.

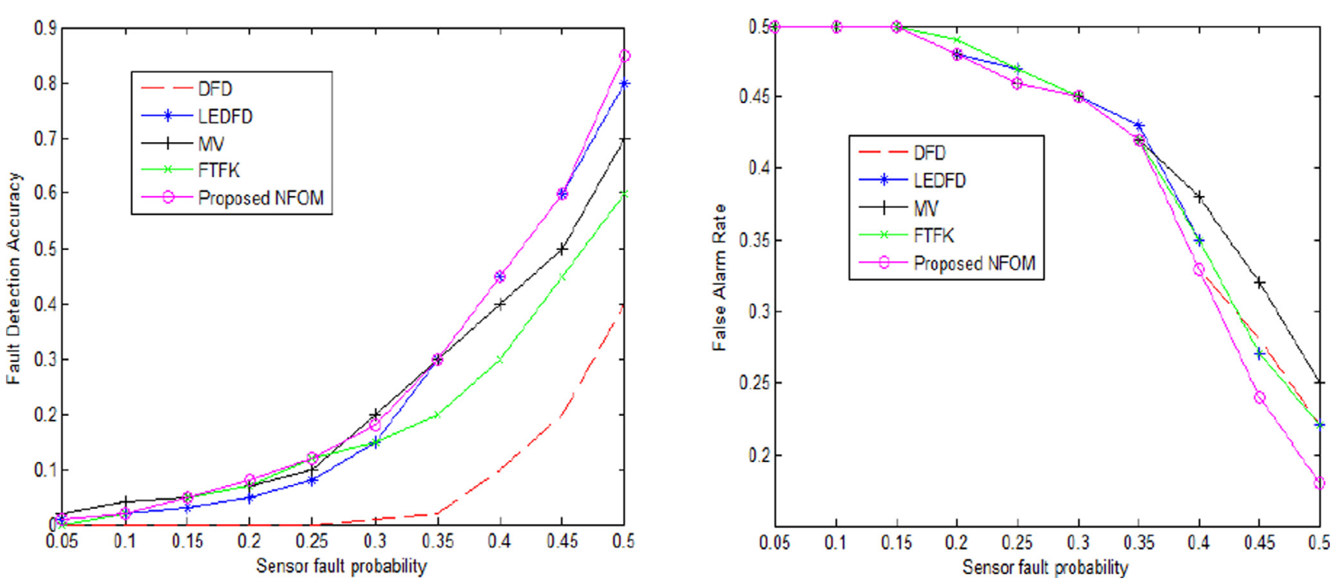

Figure 8 Comparison of FDA and FAR values for fixed faults with average node density = 10.

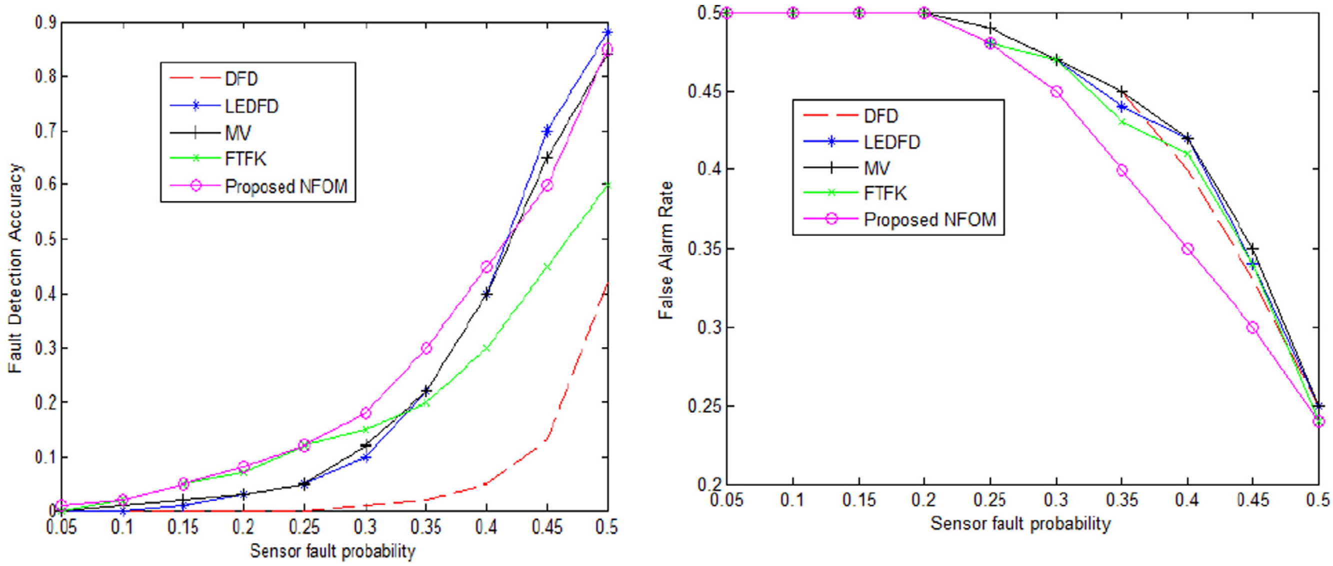

Figure 9 Comparison of FDA and FAR values for fixed faults with average node density = 20.

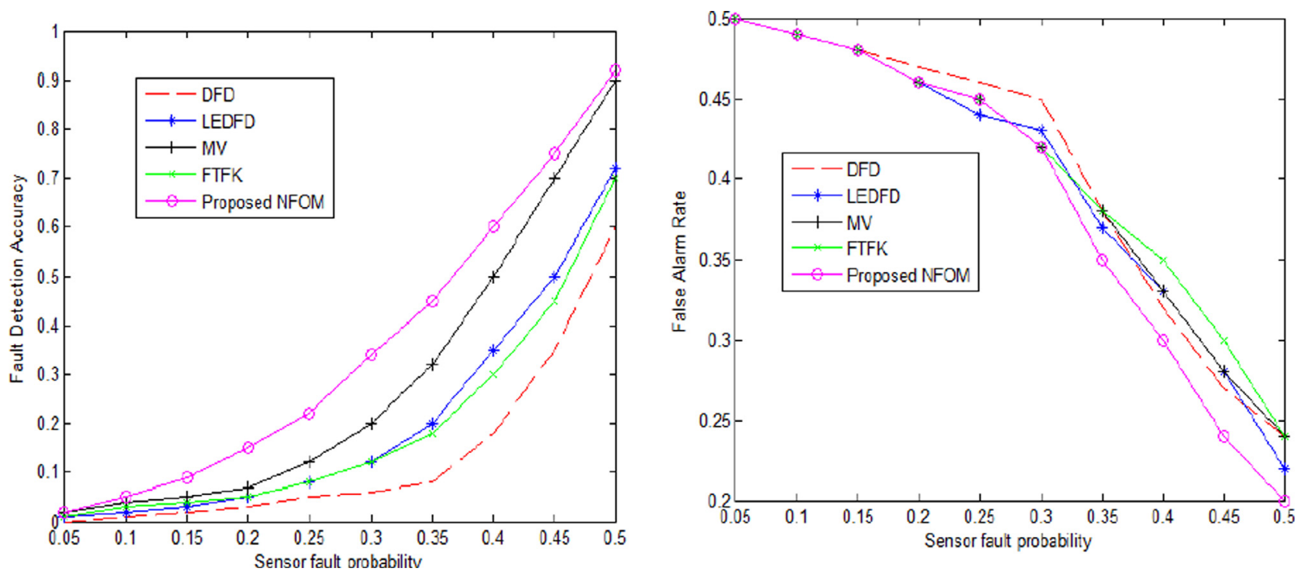

Figure 10 Comparison of FDA and FAR values for offset faults with average node density = 7.

The CFI value is forwarded from the CH to the gateway node to help in assessing the cluster as normal or faulty.

(e) If the CFI index is less than the Cluster Threshold value (CT) for any cluster, then, the cluster is considered as normal and it participates in the data aggregation process for that specific round by the gateway node.
(f) If the CFI index is greater than or equal to the CT value for any cluster, then, the cluster is considered as faulty and it does not participate in the data aggregation process for that specific round by the gateway node.
(g) The gateway node then forwards the aggregated data to the base station through single or multiple hops.

A sample CFM for a 3×3 grid cluster is given by $\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ where each entry corresponds to the node status of a NCH node in the grid cluster. Here, a '0' signifies a normal NCH node while a '1' signifies a faulty NCH node. A sample of fuzzy rules is presented in Table 1. It gives the fuzzy output Node Status (NS) corresponding to the four fuzzy inputs (RNE, PDR, FR and NOR) generated by the application of different fuzzy rules. For example, the fuzzy rule $f(VL, P, P, P) = F$ indicates that if RNE value is very low, PDR, FR and NOR are all poor, then, the node status is faulty. In a similar way, the different combinations of fuzzy inputs are assessed by the ANFIS estimator to give a fuzzy output. In the last step, the ANFIS estimator converts the

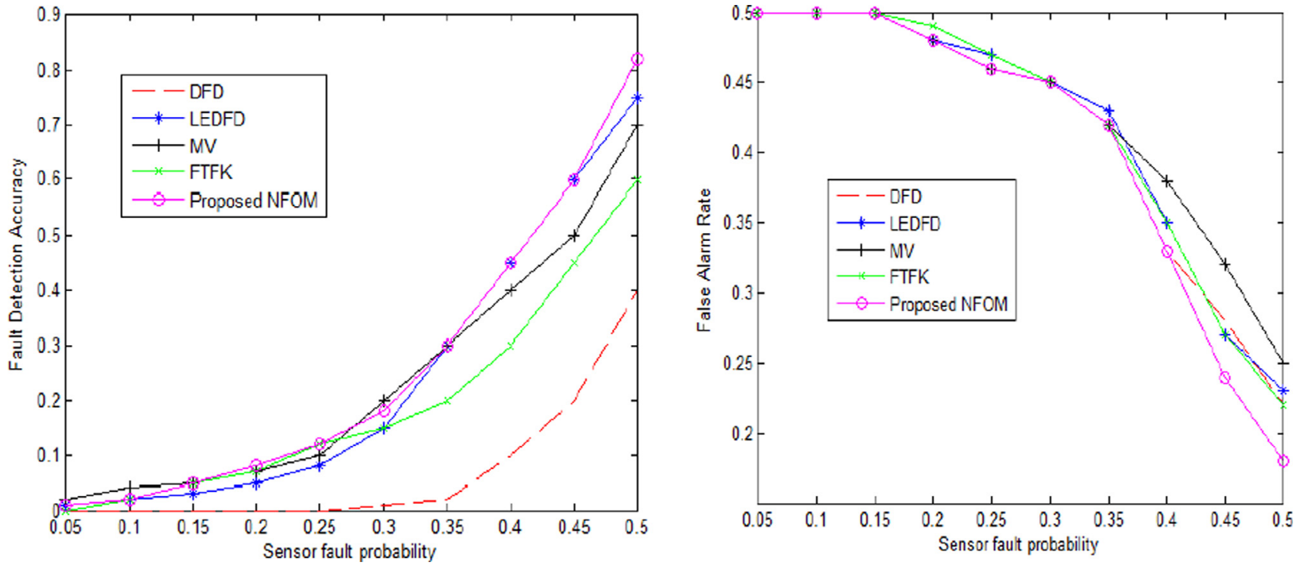

Figure 11 Comparison of FDA and FAR values for offset faults with average node density = 10.

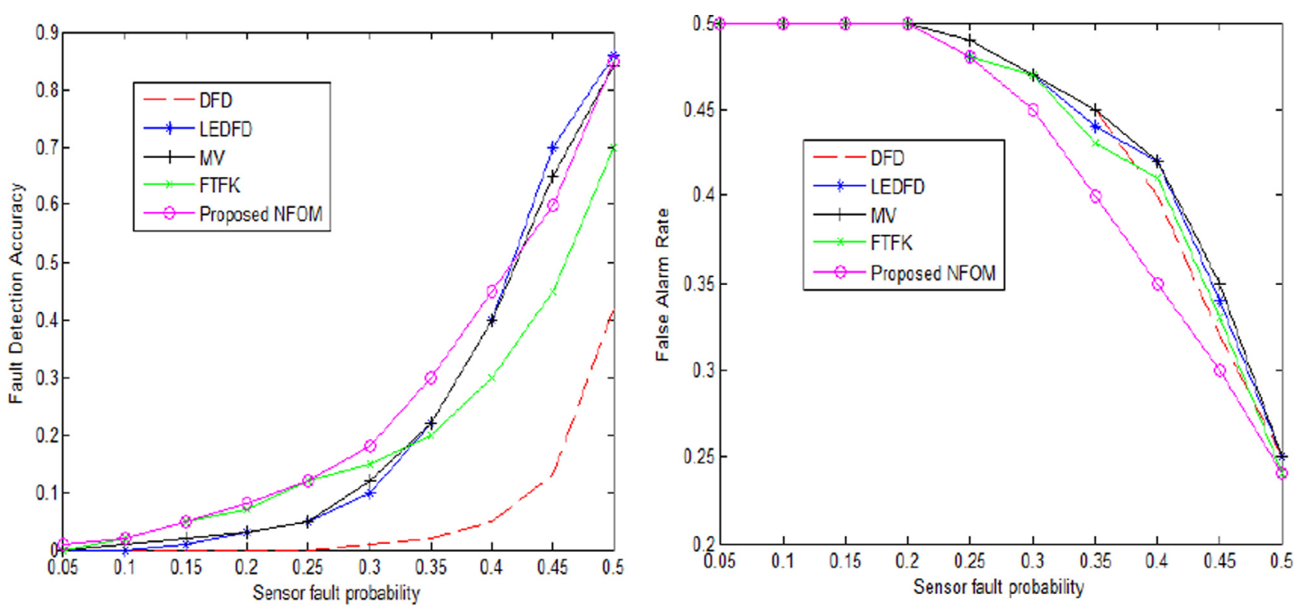

Figure 12 Comparison of FDA and FAR values for offset faults with average node density = 20.

fuzzy output to a crisp value by a defuzzification process. In the proposed NFOM data aggregation scheme, the defuzzification is done using the centroid method.

Thus, this mechanism not only conserves energy but also prevents faulty data from reaching the base station. In order to ensure coverage and connectivity through the WSN, the faulty clusters need to be replaced periodically by good ones.

4.3.2. Proposed NFOM algorithm

Fig. 3 presents the pseudo-code for the proposed NFOM algorithm. The proposed NFOM pseudo-code takes as input a network configuration with number of simulation rounds 'R', cluster threshold value CT, grid size gd_size and total number of clusters 'n' and returns a fault tolerant data aggregation NFOM scheme by repeating the following sequence of steps for each simulation round. The proposed algorithm is run at the respective Cluster Head (CH). The Step 1of the algorithm initializes node 0 as the base station BS and node g_t at level 't' as the gateway node. It also sets the counter value for the number of normal nodes in a cluster (count[c_k]) to 0 for all clusters from 1-n. The Step 2 applies the proposed ANFIS estimator to estimate the node status NS using fuzzy rules for each node in the network and for each cluster. The Step 3 checks the node status NS value and accordingly assigns the Cluster Fault Matrix (CFM) entries and forward flag (fwd_flag) values for each cluster in the network. The CFM values are stored in the respective CH. The CFM values are used to compute the Cluster Fault Index (CFI) value for a cluster as described in the pseudo-code. If the node status for a node in a cluster is estimated to be faulty ('F') by the ANFIS estimator, then

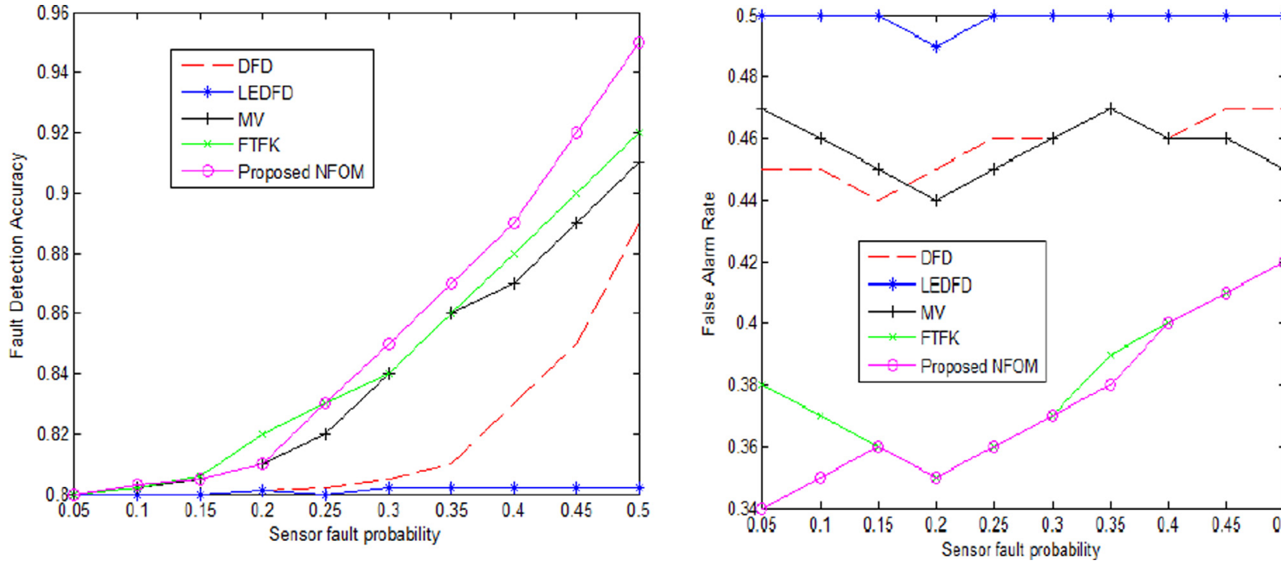

Figure 13 Comparison of FDA and FAR values for random faults with average node density = 7.

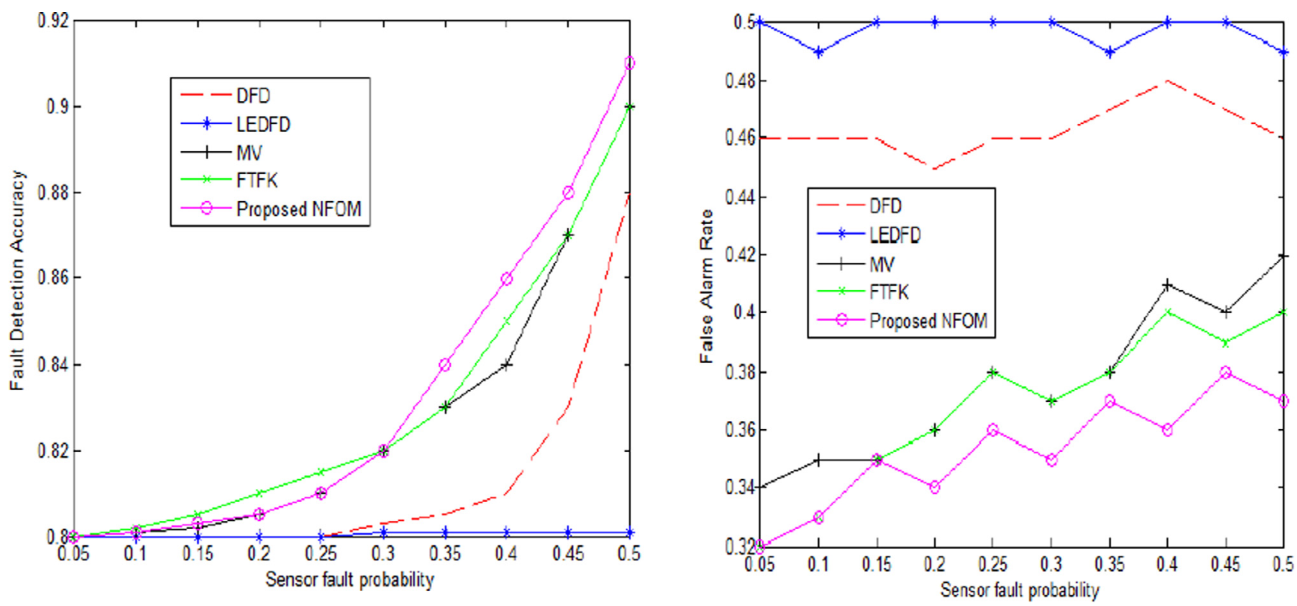

Figure 14 Comparison of FDA and FAR values for random faults with average node density = 10.

the corresponding CFM entry is set to '1' and the forward flag fwd_flag is set to 'N' in order to prevent the faulty NCH node to participate in the data aggregation process. The identified faulty NCH node is not allowed to forward its data to its respective CH. On the other hand, if the node status is estimated to be normal ('N') by the ANFIS estimator, then the corresponding CFM entry is set to '0' and the fwd_flag is set to 'Y' in order to allow the normal NCH node to forward its data to the CH. This step thus prevents the faulty data from being transmitted to the CH.

The Step 4 counts the number of faulty nodes in a cluster by checking the CFM entry for each node. This count is incremented by 1 for faulty nodes (with CFM entry '1') for each cluster in the network. The Step 5 computes the Cluster Fault Index (CFI) value for each cluster using the formula given in the pseudo-code. The CFI value is then forwarded by the respective CH to its respective gateway node to assist in assessing the cluster as normal or faulty. The Step 6 finally identifies the faulty clusters in the network by comparing their CFI value with the pre-defined cluster threshold CT value that is set at the beginning of the simulation. If for a cluster, its CFI value is less than the pre-defined CT value, then the cluster is tagged as a normal cluster and the cluster forward flag (cfwd_flag) is set to 'Y' in order to forward the cluster data through the CH to the gateway node for onward transmission. But if the CFI value is greater than or equal to the CT value for any cluster, then the cluster is identified as a faulty cluster and the cfwd_flag is set to 'N' in order to prevent the faulty cluster data from onward transmission. Thus, the proposed NFOM data aggregation scheme pro-actively isolates faulty data from the data aggregation process both at the intra-cluster and the inter-cluster levels.

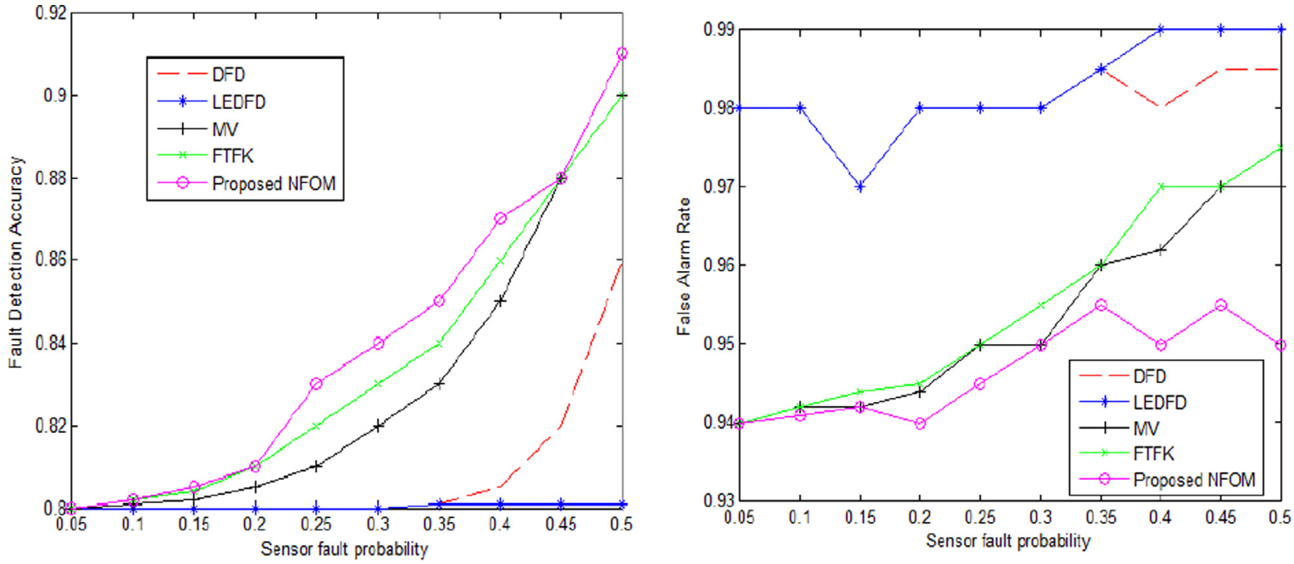

Figure 15 Comparison of FDA and FAR values for random faults with average node density = 20.

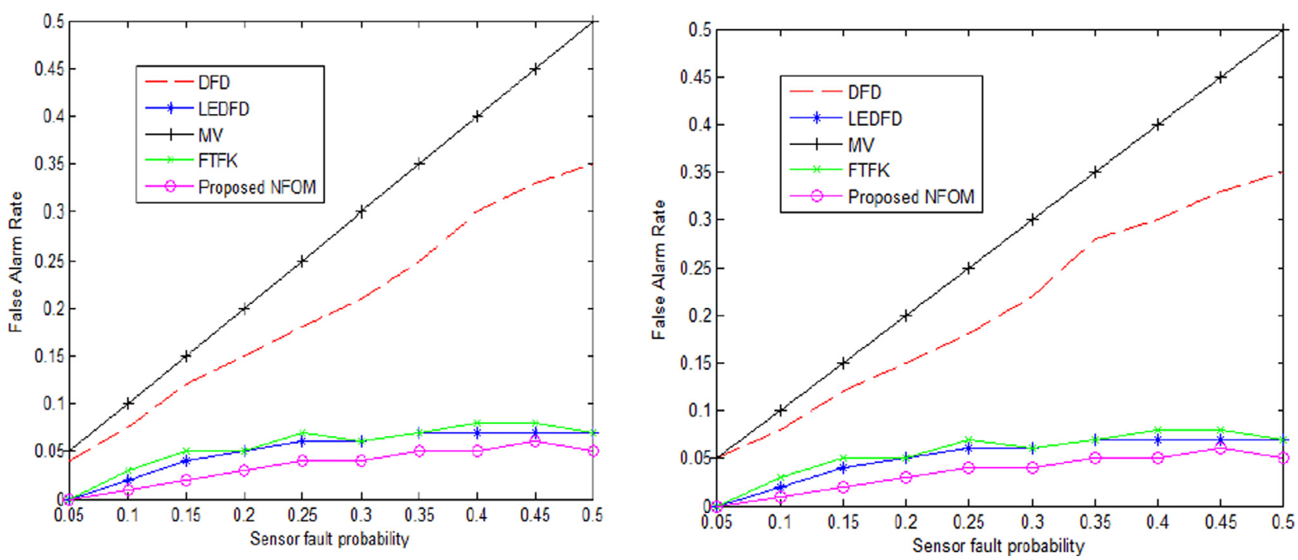

Figure 16 Comparison of FAR values for transient faults with average node densities 7 and 10.

5. Simulation results and discussion

This section presents the simulation model, ANFIS modeling of the WSN network, performance metrics and compares the performance of the proposed NFOM with those of DFD (Jiang, 2009), LEDFD (Xu et al., 2014), MV (Javanmardi et al., 2012) and FTFK (Chang et al., 2013) algorithms through simulation for five different cases. The simulation is done using MATLAB 7.8.0.

5.1. Simulation Model

In order to compare the performance of the different fault detection algorithms for varying node densities, the following five cases are considered.

Case 1: Only fixed faults at varying average node densities
Case 2: Only offset faults at varying average node densities
Case 3: Only random faults at varying average node densities
Case 4: Only transient faults at varying average node densities
Case 5: Mixed faults at varying average node densities

The configuration details for the simulation setup are given in Table 2. It specifies values for different simulation parameters like network size, number of sensor nodes, data packet size, initial energy and other information. The WSN network was tested with 450 sensor nodes which were organized into 50 clusters. The status of 50 clusters was monitored by 5 gateway nodes so that each gateway node was assigned 10 clusters. These simulation details are presented in Table 2.

The *Packet Aggregation Ratio* (PAR) is defined as the ratio of the number of nodes in which a packet gets aggregated to the total number of nodes through which the packet is transmitted to the base station. The PAR value is limited to 10% in the simulation. The base station is assumed to be *stationary* in the simulation and its location is fixed at (100,100) m.

5.2. ANFIS modeling of the WSN network

The ANFIS model was developed using MATLAB Fuzzy Logic Toolbox and it is a Sugeno-type ANFIS model. For the network operation, 70% of the data set was randomly assigned as the training set. The remaining 30% was employed for testing the performance of the ANFIS estimator. The output parameter of the WSN network depends on four input parameters, namely the Residual Node Energy (RNE), Packet Delivery Ratio (PDR), Fault Ratio (FR) and Number of Re-transmissions (NOR). The only output parameter is the Node Status (NS).

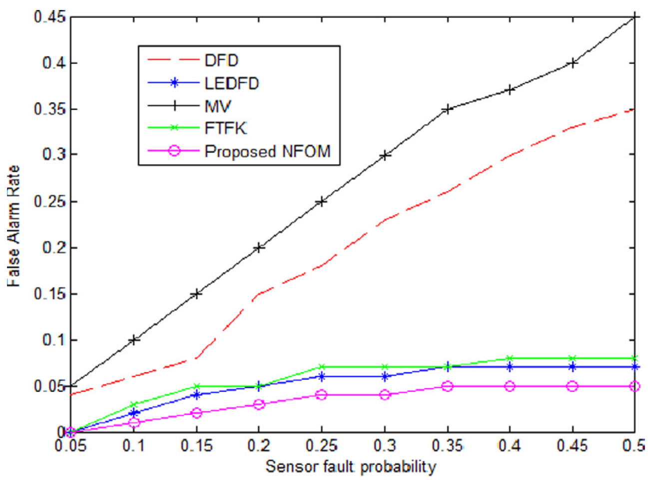

Figure 17 Comparison of FAR values for transient faults with average node density = 20.

The proposed NFOM algorithm was trained using a proper set of training data so that the outputs can be estimated based on the input–output data. The data were trained to identify the parameters of Sugeno-type fuzzy inference system based on the *hybrid algorithm* which combines the least square method and the back propagation gradient descent method. The criterions used for measuring the network performance were the network lifetime, energy cost, loss probability, Fault Detection Accuracy (FDA) and False Alarm Rate (FAR) as outlined in detail in Section 5.3.

5.3. Performance metrics

The performance of the proposed NFOM data aggregation scheme is compared with that of the DFD (Jiang, 2009), LEDFD (Xu et al., 2014), MV (Javanmardi et al., 2012) and FTFK (Chang et al., 2013) through simulation with respect to the following metrics.

(a) *First Node Dies (FND) and Half of the Nodes Alive (HNA):* FND and HNA metrics are used to estimate the network lifetime. The FND gives the simulation round at which the first node dies. It is used to estimate the network lifetime for a sparsely deployed WSN. The HNA gives the simulation round at which half of the nodes of the network die out.
(b) *Network Lifetime (in minutes):* It is the time until the network is completely partitioned due to the failure of the cluster heads in the network.
(c) *Energy Cost (in Joules):* It is the amount of energy consumed for different algorithms.
(d) *Loss Probability:* It is given by the ratio of the number of data packets dropped (d) to the sum of the number of data packets received at the base station (r) and the number of data packets dropped until the end of simulation. That is, Loss Probability = $d/(r + d)$.

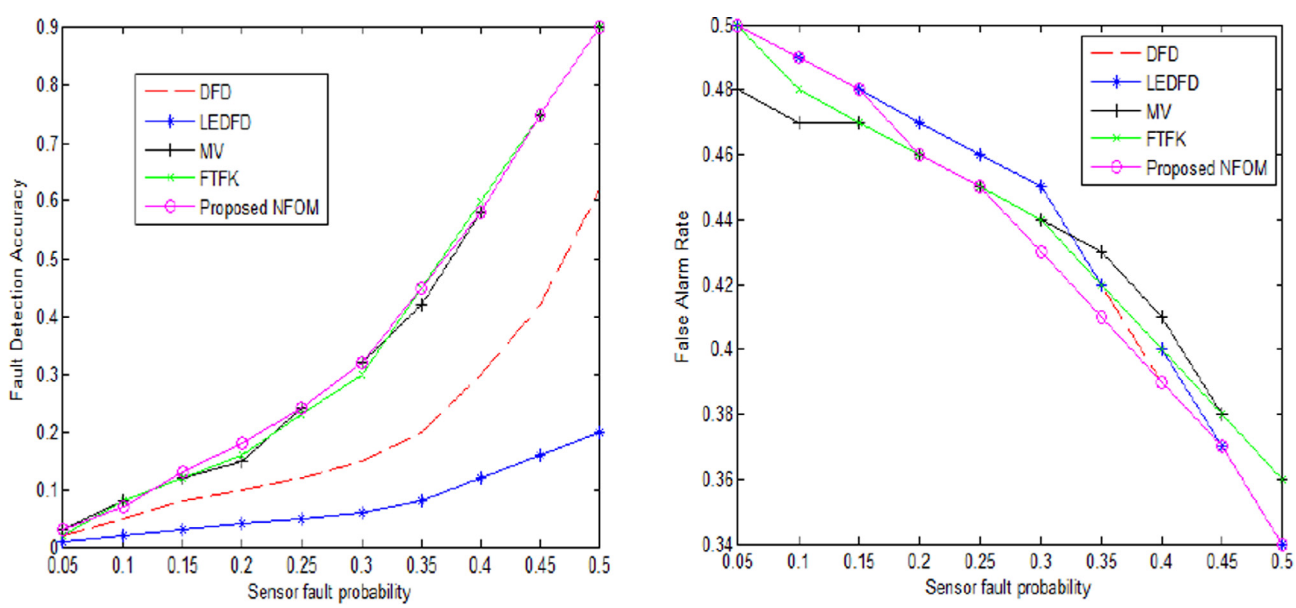

Figure 18 Comparison of FDA and FAR values for mixed faults with average node density = 7.

(e) *Fault Detection Accuracy (FDA):* It is given by the ratio of the number of correctly identified faulty nodes to the total number of actual faulty nodes. Let 'A' denote the set of actual faulty nodes and 'F' denote the set of faulty nodes detected by a fault detection algorithm. Then, FDA is given by

$$\mathrm{FDA} = |F \cap A| \div |A|$$

(f) *False Alarm Rate (FAR):* It is given by the ratio of the number of normal nodes which are mistaken to be faulty nodes to the total number of normal nodes. Let 'F' denote the set of faulty nodes detected by a fault detection algorithm, 'A' denote the set of actual faulty nodes and 'N' denote the total number of nodes in the network. Then, FAR is given by

$$\mathrm{FAR} = |F - A| \div (N - |A|)$$

5.4. Results and discussion

The simulation results for different performance metrics are presented in this section for the purpose of comparison. The performance of the four algorithms – DFD, LEDFD, MV, FTFK and the proposed NFOM are compared and the simulation results are discussed.

5.4.1. Comparison of FND and HNA

The Table 3 gives the simulation rounds for the average FND and HNA for each simulated algorithm – DFD, LEDFD, MV, FTFK and NFOM (proposed). The simulation is done on a

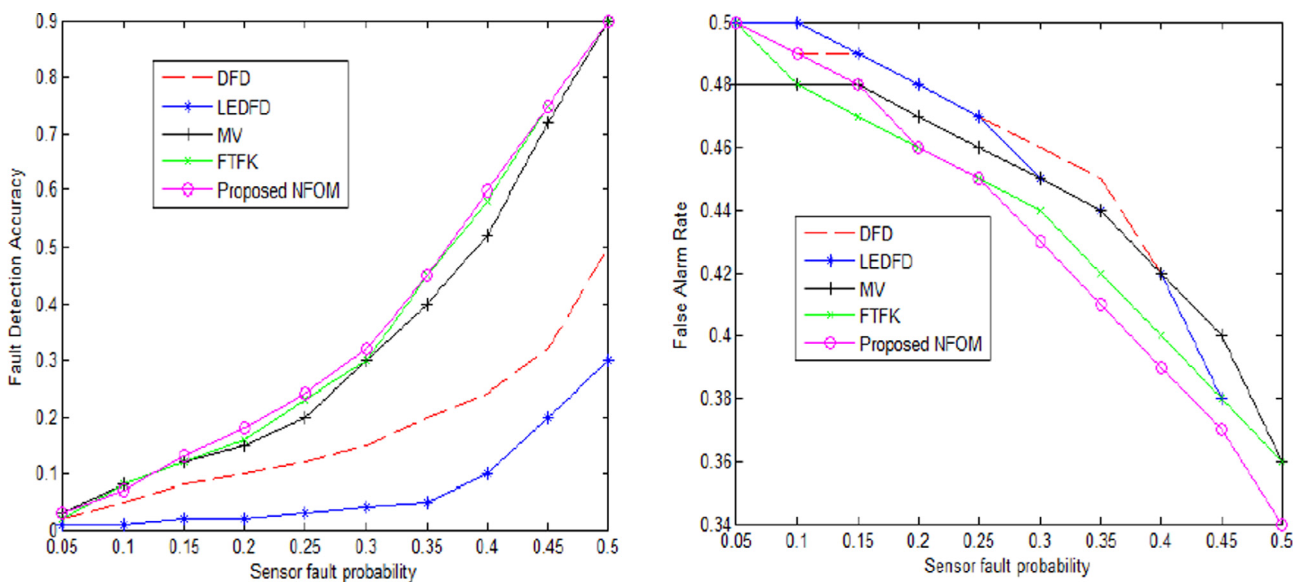

Figure 19 Comparison of FDA and FAR values for mixed faults with average node density = 10.

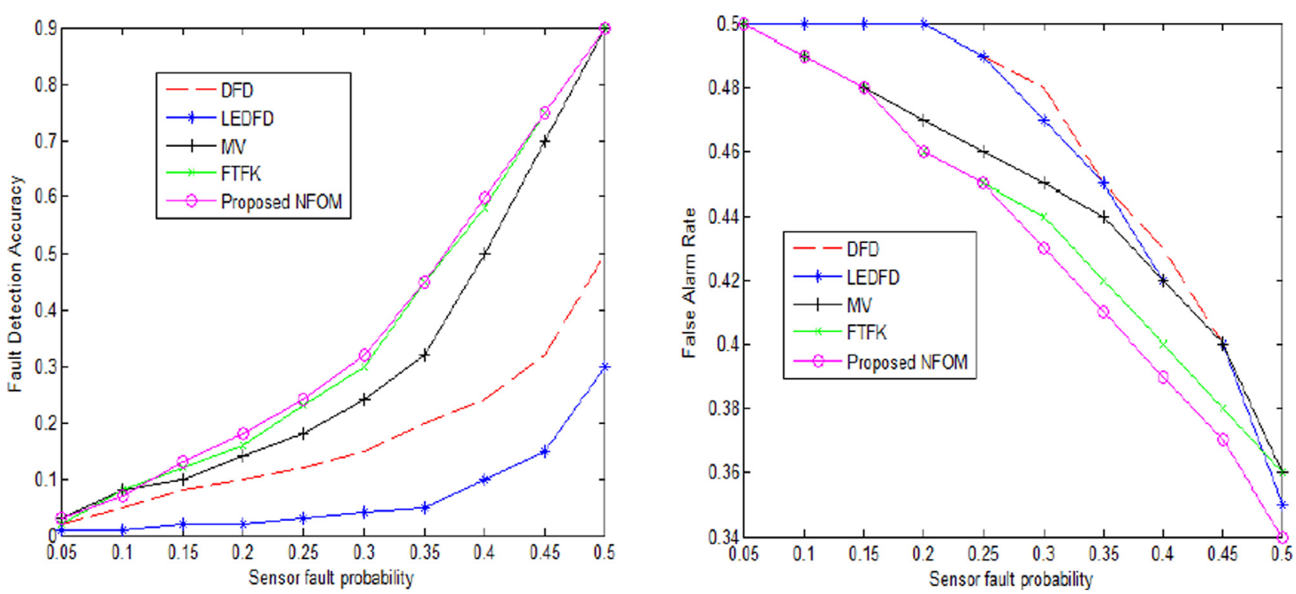

Figure 20 Comparison of FDA and FAR values for mixed faults with average node density = 20.

combination of one *stuck-at-zero*, one *offset* and one *random fault*.

It is observed from Table 3 that both the FND and the HNA values are highest for the proposed NFOM algorithm and the lowest for the DFD algorithm. The performance of the MV and the FTFK algorithms are close to that of the NFOM and the LEDFD gives better performance than DFD. The Table 3 confirms that the proposed NFOM scheme greatly enhances the network lifetime for the WSN.

5.4.2. Energy cost versus sensor fault probability

Fig. 4 depicts the energy cost in Joules for varying sensor fault probabilities for each of the simulated algorithms. The simulation is done on a combination of one *stuck-at-zero*, one *offset* and one *random fault*. It clearly shows that the energy cost is the maximum for DFD algorithm and the minimum for the proposed NFOM algorithm. The variations in energy cost are largely uniform for MV algorithm. The energy cost of FTFK algorithm matches with that of the proposed NFOM algorithm till 0.25 sensor fault probability and with gradual increase in the sensor fault probability, it becomes slightly higher than that of the proposed NFOM. The energy savings in case of the proposed NFOM algorithm owes to its use of the intelligent ANFIS estimator and longer network lifetime.

5.4.3. Network lifetime versus sensor fault probability

Fig. 5 shows the network lifetime (in minutes) for varying sensor fault probabilities for each of the simulated algorithms. The simulation is done on a combination of one *stuck-at-zero*, one *offset* and one *random fault*. It clearly shows that the network lifetime is the maximum for the proposed NFOM algorithm and the minimum for the DFD algorithm. The variations in network lifetime with varying sensor fault probabilities are almost similar for the rest of the three algorithms – LEDFD, MV and FTFK with LEDFD outperforming the other two after 0.4 sensor fault probability. The maximum network lifetime for the proposed NFOM algorithm owes to its easy identification of faulty nodes by the application of the ANFIS estimator, isolating the faulty nodes and forwarding data from only the normal nodes.

5.4.4. Loss probability versus sensor fault probability

Fig. 6 shows the loss probability for varying sensor fault probabilities for each of the simulated algorithms. The simulation is done on a combination of one *stuck-at-zero*, one *offset* and one *random fault*. It clearly shows that the loss probability is the lowest for the proposed NFOM algorithm and the highest for the DFD algorithm. The variations in loss probability with varying sensor fault probabilities are similar for the LEDFD and MV algorithms and the performance of FTFK algorithm lies in between the two extremes. The lowest loss probability for the proposed NFOM algorithm is due to its early identification of the faulty nodes by the application of the intelligent ANFIS estimator, isolating the faulty nodes and forwarding data from only the normal nodes.

5.4.5. Comparison of FDA and FAR (for Case 1)

Figs. 7–9 compare the FDA and FAR values for Case 1, that is, *fixed faults* with varying average node densities for each of the simulated algorithms. It clearly shows that in all the three figures for Case 1, the FDA value is the highest and the FAR value is the lowest for the proposed NFOM algorithm. It has also been observed that the performance of the DFD algorithm worsens with increasing average node densities. The performance of the other three algorithms lies in between the two extremes. The optimal performance of the proposed NFOM algorithm is due to its early detection of faulty nodes by the application of the intelligent ANFIS estimator and isolating the faulty nodes from the data aggregation process.

5.4.6. Comparison of FDA and FAR (for Case 2)

Figs. 10–12 compare the FDA and FAR values for Case 2, that is, *offset faults* with varying average node densities for each of the simulated algorithms. It clearly shows that in all the three figures for Case 2, the FDA value is the highest and the FAR value is the lowest for the proposed NFOM algorithm. It has also been observed that the performance of the DFD algorithm worsens with increasing average node densities. The performances of the other three algorithms lie in between the two extremes. The optimal performance of the proposed NFOM algorithm is due to its early detection of faulty nodes by the application of the intelligent ANFIS estimator and isolating the faulty nodes from the data aggregation process.

5.4.7. Comparison of FDA and FAR (for Case 3)

Figs. 13–15 compare the FDA and FAR values for Case 3, that is, *random faults* with varying average node densities for each of the simulated algorithms. It clearly shows that in all the three figures for Case 3, the FDA value is the highest and the FAR value is the lowest for the proposed NFOM algorithm. It has also been observed that the performance of the DFD and LEDFD algorithms worsen with increasing average node densities. The performance of the other two algorithms – MV and FTFK lie in between the two extremes. The optimal performance of the proposed NFOM algorithm is due to its early detection of faulty nodes by the application of the intelligent ANFIS estimator and isolating the faulty nodes from the data aggregation process.

5.4.8. Comparison of FAR (for Case 4)

Figs. 16 and 17 compare the FAR values for Case 4, that is, *transient faults* with varying average node densities for each of the simulated algorithms. The FDA values have not been given as transient faults are easily corrected by the application of MV technique (Javanmardi et al., 2012) by comparing with the readings of the neighboring sensor nodes. It clearly shows that in both Figs. 16 and 17, the FAR value is the highest for MV algorithm closely followed by the DFD algorithm and the lowest for the proposed NFOM algorithm. It has also been observed that the performance of the other two algorithms – LEDFD and FTFK are close to the proposed NFOM. The optimal performance of the proposed NFOM algorithm is due to its early detection of faulty nodes by the application of the intelligent ANFIS estimator, isolating the faulty nodes and selective forwarding of data from only the normal nodes to the base station.

5.4.9. Comparison of FDA and FAR (for Case 5)

Figs. 18–20 compare the FDA and FAR values for Case 5, that is, *mixed faults* for varying average node densities for each of the simulated algorithms. The Case 5 is a combination of one *stuck-at-zero fault* and a *random fault*. It clearly shows that in all the three figures for Case 5, the FDA value is the highest and the FAR value is the lowest for the proposed NFOM algorithm. It has also been observed that the performance of the DFD and the LEDFD algorithms worsen with increasing average node densities. The performance of the other two algorithms – MV and FTFK lie in between the two extremes. The optimal performance of the proposed NFOM algorithm is due to its early detection of faulty nodes by the application of the intelligent ANFIS estimator, isolating the faulty nodes and selective forwarding of data from only the normal nodes to the base station.

6. Conclusion and future work

This paper proposed an ANFIS estimator based data aggregation scheme called Neuro-Fuzzy Optimization Model (NFOM). The proposed scheme used an ANFIS estimator in order to identify faulty sensor nodes in a cluster and faulty clusters in the network and isolated them from the data aggregation process. The paper further investigated a combination of different fault cases and also compared the relative performance of the proposed NFOM algorithm with that of the Distributed Fault Detection (DFD) Jiang, 2009, Low Energy Distributed Fault Detection (LEDFD) Xu et al., 2014, Majority Voting (MV) Javanmardi et al., 2012 and Fuzzy Knowledge based Fault Tolerance (FTFK) Chang et al., 2013 algorithms through simulation with respect to different performance metrics. The proposed NFOM data aggregation scheme is observed to be the best in terms of energy cost, network lifetime, loss probability, fault detection accuracy and false alarm rate as compared to the other existing algorithms. The computation of the proposed NFOM algorithm for different possible positions of the base station is an interesting topic which can be taken as a future research work. Also, this work may be further extended in future for designing secure data aggregation schemes using cryptographic techniques for the design of fault-tolerant WSNs.

References

Acharya, S., Tripathy, C.R., 2014. Inter-actor connectivity restoration in wireless sensor actor networks: an overview. In: 48th Annual Convention of CSI, vol. 1. Advances in Intelligent Systems and Computing, vol. 248. Springer International Publishing Switzerland, pp. 351–360.

Acharya, S., Tripathy, C.R., 2015. An ANN approach for fault-tolerant wireless sensor networks. In: 49th Annual Convention of CSI, vol. 2. Advances in Intelligent Systems and Computing, vol. 338, Springer International Publishing Switzerland, pp. 475–483.

Acharya, S., Tripathy, C.R., 2016. A fuzzy knowledge based sensor node appraisal technique for fault tolerant data aggregation in wireless sensor networks. In: Computational Intelligence in Data Mining, vol. 2. Advances in Intelligent Systems and Computing, vol. 411, Springer India, pp. 59–69.

Attia, S.B., Cunha, A., Koubaa, A., Alves, M., 2007. Fault-Tolerance mechanisms for Zigbee wireless sensor networks, Tech. rep., IPP-HURRAY Research Group, Polytechnic Institute of Porto.

Chang, S., Chung, P., Huang, T., 2013. A fault tolerance fuzzy knowledge based control algorithm in wireless sensor networks. J. Convergence Inf. Technol. (JCIT) 8 (2).

Chiasserini, C.-F., Garetto, M., 2006. An analytical model for wireless sensor networks with sleeping nodes. IEEE Trans. Mob. Comput. 5 (12), 1706–1718.

Hady, A.A., El-kader, S.M.A., Eissa, H.S., 2013. Intelligent sleeping mechanism for wireless sensor networks. Egypt. Inf. J. 14 (2), 109–115.

Heinzelman, W.R., Chandrakasan, A.P., Balakrishnan, H., 2002. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wireless Commun. 1 (4), 660–670.

Jain, N., Vokkarane, V.M., Wang, J.P., 2008. Performance analysis of dual-homed fault-tolerant routing in wireless sensor networks. In: IEEE Conference on Technologies for Homeland Security, Waltham, MA, pp. 474–479.

Javanmardi, S., Barati, A., Dastgheib, S.J., Attarzadeh, I., 2012. A novel approach for faulty node detection with the aid of fuzzy theory and majority voting in wireless sensor networks. Int. J. Adv. Smart Sensor Network Syst. (IJASSN) 2 (4), 1–10.

Jeevanandam, C., Ramalakshmi, K., Priya I, D., 2014. A survey on energy efficient fault tolerance mechanisms in wireless sensor networks. Int. J. Emerg. Technol. Adv. Eng. 4, 330–334.

Jiang, P., 2009. A new method for node fault detection in wireless sensor networks. Open Access Sensors J. 9 (2), 1282–1294.

Kaur, A., Saini, S., 2013. Simulation of low energy adaptive clustering hierarchy protocol for wireless sensor network. Int. J. Adv. Res. Comput. Sci. Software Eng. 3 (7), 1316–1320.

Koushanfar, F., Potkonjak, M., 2005. Markov chain-based models for missing and faulty data in MICA2 sensor motes. In: IEEE, pp. 576–579.

Lazzerini, B., Marcelloni, F., Vecchio, M., Croce, S., Monaldi, E., 2006. A Fuzzy Approach to Data Aggregation to Reduce Power Consumption in Wireless Sensor Networks. IEEE, pp. 457–462.

Liu, T., Yi, S., Wang, X., 2013. A fault management protocol for low-energy and efficient wireless sensor networks. J. Inf. Hiding Multimedia Signal Process. Ubiquitous Int. 4 (1), 34–45.

Manjeshwar, A., Agrawal, D.P., 2000. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In: Proceedings of 15th IEEE International Parallel and Distributed Processing Symposium, San Francisco, CA, USA, pp. 2009–2015.

Qiu, M. et al, 2013. Informer homed routing fault tolerance mechanism for wireless sensor networks. J. Syst. Architect. 59, 260–270.

Ranga, V., Dave, M., Verma, A.K., 2015. A fuzzy logic based joint intra-cluster and inter-cluster multi-hop data dissemination approach in large scale WSNs. In: Proceedings of 2015 International Conference on Future Computational Technologies (ICFCT'2015), pp. 71–77.

Sharma, V.K., Shukla, S.S.P., Singh, V., 2012. A tailored Q-Learning for routing in wireless sensor networks. In: IEEE International Conference on Parallel Distributed and Grid Computing (PDGC), pp. 663–668.

Volosencu, C., Curiac, D., 2013. Efficiency improvement in multi-sensor wireless network based estimation algorithms for distributed parameter systems with application at the heat transfer. EURASIP J. Adv. Signal Process. 1 (1), 1–17. Springer.

Xiaolong Xu et al. "LEDFD: A Low Energy Consumption Distributed Fault Detection Algorithm for Wireless Sensor Networks", International Journal of Distributed Sensor Networks, Hindawi Publishing Corporation, 2014.

Younis, O., Fahmy, S., 2004. HEED: a hybrid, energy-efficient distributed clustering approach for ad hoc sensor networks. IEEE Trans. Mobile Comput. 3 (4), 366–379.

Yu, S.C., Zhang, Y.Y., 2007. R-Sentry: providing continuous sensor services against random node failures. In: IEEE International Conference on Dependable Systems and Networks, Edinburgh, UK, pp. 235–244.