Improvement of energy efficiency and reduction of data loss using clustering in WSN

Snehal Kumar Trapsiya¹, Mr. Indr Jeet Rajput²

¹ Student, Computer Engineering, HGCE(GTU), Gujarat, India
² Asst. Prof., Computer Engineering, HGCE(GTU), Gujarat, India

ABSTRACT

WSN consist of sensor nodes of small size & having limited computational power & limited energy. Due to lower energy, sometimes it may happen that the data sensed by the nodes couldn’t reach to the sink node and data loss occurs which is also a big problem. An algorithm named FNR provides around 90% of data loss recovery with the increase in energy which uses combination of Grade diffusion algorithm and Genetic Algorithm.

In WSN, clustering approach is very important for the increase of lifetime and reduction of data loss means decreasing energy consumption and decreasing data loss. Another algorithm, DFCA provides clustering, which is more important than other clustering techniques.

By applying the DFCA in the FNR algorithm before applying the GA after the Grade diffusion process is done we can have more energy efficient and more data reliable algorithm, which helps us in many approaches for the betterment of WSN and its lifetime and data loss problem.

Keyword: - FNR, DFCA, GA

1. Introduction

A WSN is a type of wireless network consists of sensor nodes which generally sense the physical conditions, processing it and then sends it to the base station or say the place or node from where the network is connecting to the outside world. Wireless Sensor Networks (WSN) are consisting spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, pressure etc. and to cooperatively pass their data through the network to the base station. A Wireless Sensor Network is designed to detect events of interest of an application, collect data related to these events and report these data to a monitoring station (Base Station).

Wireless Sensors are small in size having limited energy & computational power. The Wireless sensors are standard measurement tools; consist of transmitters to convert signals from process control instrument into a radio transmission. The radio signal is interpreted by a receiver which then converts the wireless signal to a specific, desired output, such as an analog current or data analysis via computer software.

The WSNs are of two types. [5]
1) Flat Network.
2. Background and Existing System

In WSN, fault occurrence probability is very high compare to traditional networking. On the other hand networks maintenance and nodes replacement is impossible due to remote deployment. These features motivate researchers to make automatic fault management techniques in wireless sensor networks. As a result many fault detection and fault tolerance techniques proposed. Fault Node Recovery (FNR) algorithm is one of the fault tolerance technique. Use of FNR algorithm can results in fewer replacements of sensor nodes and more reused routing paths. And thus algorithm not only enhances WSN lifetime but also reduces the cost of replacing the sensor nodes. The FNR is based on the GD algorithm, with the goal of replacing fewer sensor nodes that are inoperative or have depleted batteries, and of reusing the maximum number of routing paths. In the FNR first the GD algorithm is applied and after that the GA is applied and thus the faulty nodes are replaced with the other ones so that the lifetime can be increased. Using the GD algorithm first, Grade Value, routing table, set of neighbor nodes, payload values for each sensor nodes etc are decided. After that the Genetic algorithm is applied and the faulty nodes are found and then they are replaced with the other healthy nodes by using the following steps.

(1) Initialization,
(2) Evaluation,
(3) Selection,
(4) Crossover,
(5) Mutation.

In the FNR, after applying the GD algorithm, the sensor nodes are detected using the sensor node detection event and after that the replacement process occurs as per the steps given above.

These steps are described below.
(1) Initialization:

In the initialization step the GA generates the Chromosomes according to the affected or the nodes which are to be replaced and its genes. The chromosome is the expected solution of the replaced nodes.

. The Fig 2 shows the Chromosome and its genes.

<table>
<thead>
<tr>
<th>9</th>
<th>7</th>
<th>10</th>
<th>81</th>
<th>23</th>
<th>57</th>
<th>34</th>
<th>46</th>
<th>66</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig 2 Chromosome and Genes [1]

(2) Evaluation

In this step, fitness value is calculated with the fitness function.

But we can’t put genes directly into the fitness function in the in the FNR algorithm. Because the genes mean whether the node should be replaced or not. The fitness function is shown (1)

\[ f_n = \frac{\max(\text{Grade})}{\sum_{i=1}^{\max(\text{Grade})} \frac{P_i \times T^{-1}}{N_i \times T^{-1} \times i^{-1}}} \]

(1)

\( N_i = \) the number of replaced sensor nodes and their grade value at \( i \)

\( P_i = \) the number of reusable routing paths from sensor nodes with their grade value at \( i \).

\( T \_N = \) total number of sensor nodes in the original WSN.

\( T \_P = \) total number of routing paths in the original WSN.

Here as WSN is looking for the most available routing path and the least number of replaced sensor nodes high fitness value is sought.

(3) Selection:

The selection step will eliminate the chromosomes with the lowest fitness values and retain the rest. Here elitism strategy is used using which keep the half of the chromosomes with better fitness values and put them in the mating pool.

(4) Crossover:

The step is used in the genetic algorithm to change the individual chromosome. One point cross over strategy is used here.
(5) Mutation:

The mutation step can introduce traits not found in the original individuals and prevents the GA from converging too fast.

Here in the FNR, after the Mutation step we find the recovered nodes using which the data loss is being decreased and energy level is increased. But here the data transmission takes place in either single hop or multi hop way.

But if we apply clustering in the FNR algorithm we can make the algorithm more energy efficient and we can use the advantage of both clustering and Fault node recovery for the betterment of WSN. For that
DFCA algorithm is used as a base of FNR algorithm means first we apply DFCA and provides clustering and after that FNR is applied on the WSN.

Why Clustering?

WSN can be of Flat base and Cluster base. Flat base is that one in which the sensor nodes send the sensed data using single hop or multi hop transmission. And the sensor nodes have limited energy and irreplaceable power sources so they can’t send the data after the exhaust of the energy. Clustering is that technique in which the sensors nodes are divided into sub areas and a node is selected as a Cluster Head (CH), and other nodes of the same cluster sends the sensed data to the CH and the CH sends it to the either Base station(BS) or the other CH. Even any node of the routing path affected or becomes faulty the data being send will not reach to the BS. Instead of that, if we apply clustering than if one of the nodes found faulty, the neighbor of that node will take place of that node and sends the data to the BS. Thus the problem of faulty node can be replaced and also we can improve the energy level and lifetime of WSN.

In [2] Fault Tolerant and Mobility aware Routing protocol for mobile WSN is explained. Also the fault tolerance and less packet loss is achieved while sending data either from cluster members to cluster head or low level cluster head to higher level cluster head. Delay also gets reduced as the cluster head sends the data to Base Station through level based architecture of nodes.

In [3] protocol for Energy efficient and energy balanced fault tolerance clustering in WSN is provided. Here also importance of clustering in the WSN is explained. And also Energy model and network model are explained. The cluster formation is based on residual energy and routing overhead with distance between sensor nodes and CH. Fault tolerant algorithm always selects new CH using either cost or payoff functions. Fault tolerant algorithm only considers a permanent failure of sensor nodes and gateway.

In [4] Energy efficient chain based cooperative routing protocol for WSN explained. In this cluster based WSN CHs are used to receive the data from the nodes which relies under that particular CH. And to transmit the data from the last CH to the BS, it sends the data to the Cluster Coordinates (CCO) assigned to it. These CCOs are that nodes which transmit the data sent through their former CH or CCO assigned as shown in Fig 5.

Balanced energy consumption is achieved by transmission of data to the intermediate nodes at all the levels. Another remarkable property is that, number of nodes can be increased without any additional cost, as all the nodes can still send data with the help of relay nodes within cluster and cluster coordinators. In [5] information for fault tolerance in dynamic cluster based WSN is provided. Also in this paper the behavior of the WSN network for both flat and dynamic cluster based and also provides the information of multiple nodes leaving the dynamic cluster based WSN. Dynamic Cluster Based WSN is a robust network to handle new nodes joining and existing nodes leaving. Here two algorithms are proposed, one for one-hop neighbor information and one for partial one-hop neighbor information.
Fig 5 Inter-cluster communication with the help of Cluster coordinators [4].

In [6] comparative study on different LEACH protocols is provided. It also explains the classification of routing protocols for WSNs. And the main purpose of designing energy efficient routing protocol is to efficiently use the energy of the network so that the network lifetime get increased.

2.1 FNR algorithm

Here the Fig 6 explains the FNR algorithm flow chart and the basic steps involved in it.

---

**Fig 6 Fault Node Recovery Flow Chart. [1]**
2.1 Distributed Fault-Tolerant Clustering Algorithm.

DFCA is a distributed fault-tolerant clustering algorithm that addresses both issues, cluster formation based on remaining energy of the gateways and fault tolerance of the WSNs owing to death of some gateways. In DFCA, the nodes select proper Cluster Head by considering a cost function which consists of residual energy of the CH, The distance between sensor nodes to the CH and distance from the CH to the base station. In cluster formation, DFCA takes care about the sensor nodes that have no CH within their communication range. It also presents a distributed run time recovery of the faulty cluster members due to sudden failure of the CH. It avoids redundant deployment of the CHs or fully re-clustering approach to recover the faulty nodes to tolerate the failure. [6]

The LEACH algorithm has the time complexity of O (mn2), very high for a large scale WSN. Then the DFCA algorithm proposed with the time complexity O (n log n) time, which is an improvement over. Thus in the DFCA following terminologies are used: [6]

1. A set of sensor nodes denoted by S={S1, S2, ..., Sn}
2. A set of gateways denoted by G = {G1, G2, ..., Gm}
3. Dist(Si, Sj) denotes the distance between two nodes Si and Sj.
4. E_{residual}(Si) denotes the remaining energy of Si.
5. Com\text{RangeCH}(Si) is the set of all those gateways, which are within the communication range (Rs) of node Si. Therefore,

\begin{align*}
\text{ComRangeCH}(Si) = \{G_j | Dist(S_i, G_j) \leq R_s \land G_j \in G\} \quad \ldots (2)
\end{align*}

For example, \{\text{ComRangeCH}(Si) = \{G_2, G_3, G_5, G_7\} means sensor node Si can be assigned to any one of the gateways G2, G3, G5 or G7.

6. Neighbour (Si) is the set of all those sensor nodes, which are within the communication range of node Si.

\begin{align*}
\text{Neighbor}(Si) = \{S_j | Dist(S_i, S_j) \leq R_s \land S_j \in \{S-Si\}\}
\end{align*}

Depending on the communication range and connectivity between the sensor nodes and gateways, also in DFCA few kinds of nodes are defined in the system as follows:

**Covered Node and Covered Set:** Covered nodes are having at least one Gateway within its communication range. And covered set is set of all covered nodes.[6]

**Uncovered Node and Uncovered Set:** Uncovered sensor has no gateway within its communication range. And uncovered set is collection of all uncovered nodes in WSN.

**Backup nodes and backup set:** Backup nodes for an uncovered sensor node Si are all the covered sensor nodes which are within communication range of Si. Backup set of an uncovered sensor node Si is the set of all backup nodes of Si. We refer this set as BackupSet (Si).[6]
Alive and Inactive Sensor node: Alive sensor nodes are those sensor nodes, which have some residual energy and can send the sensed data to the CH directly or indirectly.

Here in this thesis we apply clustering to the FNR by providing DFCA to the FNR before finding grades and by doing that we are having the advantage of both clustering and Fault node recovery algorithms.

3. Comparisons of Implemented

Table 1 Comparison of Implemented Techniques

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Title</th>
<th>Method Used</th>
<th>Advantages</th>
<th>Dis-advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fault Node Recovery Algorithm for a Wireless Sensor Network [1]</td>
<td>Combination of Grade Diffusion algorithm and Genetic algorithm</td>
<td>Reuses most routing path, provides more energy efficiency.</td>
<td>Clustering topology is not used.</td>
</tr>
<tr>
<td>2</td>
<td>Fault Tolerant and Mobility Aware Routing Protocol for Mobile Wireless Sensor Network [2]</td>
<td>Location aware(Mobility) aware routing algorithm is used</td>
<td>Fault tolerance &amp; less packet loss is achieved.</td>
<td>Mobility of cluster head is not provided for heterogeneous network.</td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>Descriptions</td>
<td>Benefits</td>
<td>Contraindications</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8</td>
<td>A Comparative Study on Advances in LEACH Routing Protocol for Wireless Sensor Networks: A survey [8]</td>
<td>Distributed Fault Tolerant by applying Clustering</td>
<td>For some amount it is energy efficient.</td>
<td>Some other protocol provides more efficient, scalable and robust clustering scheme then LEACH.</td>
</tr>
<tr>
<td>9</td>
<td>An Energy Efficient Uneven Grid Clustering based Routing Protocol for Wireless Sensor Networks [12]</td>
<td>Centralized approach and uses fixed clustering.</td>
<td>It ensures that the transmission distance for any communication in the network is less than the threshold distance of the energy consumption model. And improves energy.</td>
<td>As BS has to send messages to each CH in each round it causes extra traffic overload in network.</td>
</tr>
</tbody>
</table>
4. Proposed Work

A WSN is a type of wireless network consists of sensor nodes which generally sense the physical conditions, processing it and then sends it to the base station or say the place or node from where the network is connecting to the outside world. Wireless Sensor Networks (WSN) are consisting spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, pressure etc. and to cooperatively pass their data through the network to the base station. A Wireless Sensor Network is designed to detect events of interest of an application, collect data related to these events and report these data to a monitoring station (Base Station).

4.1 Problem definition

The overall goal behind this thesis is to improve Energy efficiency of the cluster based WSN by applying the DFCA with Fault Node Recovery algorithm. In [1] states that FNR algorithm combined with Genetic algorithm requires replacing the fewer sensor nodes and

The primary goal is to improve the energy efficiency and the fault recovery of the sensor nodes by applying the DFCA in the FNR algorithm before applying the GA after the Grade diffusion process is done.
4.2 Proposed Algorithm

Step 1: Take some number of nodes as an input.

Step 2: Apply GD algorithm to find the grade values, neighbor nodes etc.

Step 3: Apply DFCA algorithm and construct the clusters and assign CH according to that.

Step 4: Sensor node detection event takes place.

Step 5: Calculate value of Bth.

Step 6: Apply Genetic algorithm and if faulty node found then replace it with better one.

4.3 Proposed System

Fig 7.1 Proposed System
6. Implementation

6.1 Implementation Strategies

The proposed system work on matlab tool. A proposed system work on FNR algorithm. It will be provide the energy efficiency to WSN.

6.2 Tools

6.2.1 MATLAB

“MATLAB is high-level language and interactive environment used by millions of engineers and scientists worldwide. MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. It lets you explore and visualize ideas and collaborate across disciplines including signal and image processing, communications, control systems, and computational finance. MATLAB is a tool for doing numerical computations with matrices and vectors. It can also display information graphically. MATLAB can be used for a range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology. The name MATLAB stands for matrix laboratory. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects, which together represent the state-of-the-art in software for matrix computation. More than a million engineers and scientists in industry and academia use MATLAB, the language of technical computing like c, c++, JAVA.”

The MATLAB system consists of five main parts:

“The MATLAB language: MATLAB language is a high-level matrix or array language with control flow of statements, functions, data structures, input/output, and object-oriented programming features. It allows both ”programming in the small” to rapidly create quick and dirty throw-away programs, and ”programming in the large” to create complete large and complex application programs.

The MATLAB working environment: This is the set of tools and facilities that you work with as the MATLAB user or programmer. It includes facilities for managing the variables in your workspace and importing and exporting data. It also includes tools for developing, managing, debugging, and profiling M-files, MATLAB's applications.”

“Handle Graphics: This is the MATLAB graphics system. It includes high-level commands for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level commands that allow you to fully customize the appearance of graphics as well as to build complete Graphical User-Interfaces on your MATLAB applications.”

“The MATLAB mathematical function library: This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.”

“The MATLAB Application Program Interface (API): This is a library that allows you to write C and FORTRAN programs that interact with MATLAB. It include facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.”

“MATLAB has evolved over a period of years with input from many users. In university environments, it is the standard instructional tool for introductory and advanced courses in mathematics, engineering, and science. In industry, MATLAB is the tool of choice for high-productivity research, development, and analysis.”
“MATLAB features a family of application-specific solutions called toolboxes allow you to learn and apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others.”

6.2.2 Features of MATLAB

- “High-level language for numerical computation, visualization, and application development”
- “Interactive environment for iterative exploration, design, and problem solving”
- “Mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, numerical integration, and solving ordinary differential equations”
- “Built-in graphics for visualizing data and tools for creating custom plots”
- Development tools for improving code quality and maintainability and maximizing performance”
- “Tools for building applications with custom graphical interfaces”
- “Functions for integrating MATLAB based algorithms with external applications and languages such as C, Java, .NET, and Microsoft”

6.3 Base paper Implementation

![Graph of Node Calculations](image)

Fig 8 Node Calculations
Fig 9: Packet loss VS Rounds

Fig 10: Number of Cluster head Remains per Round
Fig 11 Number of Packets Received by CH per Round

Fig 12 Number of working node per Round
Here the above results shows us the different conditions and as per the conditions, behaviour of the particular nodes or cluster heads etc.

The figure 5.1 shows the Node calculations where we can get the result that from the starting to some rounds there is not that much nodes which are losing its energy and being dead but after some rounds it is directly proportional to the number of rounds.

The figure 5.2 shows us the results of the packet loss per rounds where we can see in the results that packet loss is also directly proportional to the number of rounds passes and after some of the rounds the loss of the packets can’t be affordable.

The figure 5.3 shows us the loss of the cluster heads after the starting of the data transmissions and it shows us that in a single system where we can easily see that the cluster heads are losing their energy as the number of data transmission rounds increases.

The figure 5.4 shows the number of receiving packets by cluster head per round which shows us that at starting of the data transmission and the network, the Cluster heads are forming and taking their places and so number of packets receiving are increases with respect to the number of rounds and after some rounds it receives the packets steadily.

The figure 5.5 shows us the all working nodes per rounds. Here it shows us that from starting to some number of rounds they stays steadily. But after the energy of that nodes are decreasing and number of working nodes becomes inversely proportional to the rounds or we can say that as number of rounds increases the number of working nodes decreases.

6.4 Simulation

“Here simulation of proposed method was performed to verify the method. As same as FNR, the experiment was designed based on 3-D space, using 100x100x100 units, and the scale of the coordinate axis for each dimension was set at 0 to 100. The transmission ranges of the nodes were set to 15 units. In each of these simulations, the sensor nodes were distributed uniformly over the space. There are three sensor nodes randomly distributed in 10 x 10 x 10 space, and the Euclidean distance is at least 2 units between any two sensor nodes. Therefore, there are 3000 sensor nodes in the 3-D wireless sensor network simulator, and the centre node is the sink node. The data packages were exchanged between random source/destination pairs with 90,000 event data packages. Here, at initial stage energy of each sensor node was set to 3600Ws, the actual available energy.”

“Here fig 5.6 provide us comparison of dead nodes after some rounds of transmission of data between the Original and proposed algorithms. It shows us that from 1000 to 2000 rounds the number of dead nodes is same but after that from 2000 to 7000 rounds the number of dead nodes decreases for the proposed method and are low compared to FNR algorithm so here we can have less data packet loss compared to original and thus we can easily improve our data transmission. But for around 8000 rounds no of nodes going to die are same so we can easily say that this method is better than the original one and improves data loss also. Thus it is simple to understand that for small number of rounds, proposed method is not that much effective but for larger one means for large network this proposed method is being very beneficial.”
Fig 13 Number of dead nodes per Round

Fig 14 Packets received at sink node per round
In the next fig 5.7 graph between packet received at sink node per rounds is compared between the original and proposed method, through which we can say that after the 3000 rounds, as per the graph, we can say that the original have more data received at the sink node than the proposed one but we can easily improve this by applying aggregation and make the proposed more data efficient.

In next two results we have the routing and energy efficiency information, through which we can exactly figure out the efficiency and betterment of our proposed system. In fig 6.8 path loss per round is shown. Here, the graph suggest us that as proposed method is combining the DFCA with FNR, means applying clustering, in starting path loss is more than the original one but after few round of transmission it becomes very less as it becomes stable and at that time, proposed methods’ graph is decreases continuously and so the overall path loss decreases when we apply the proposed method instead of the original one. Thus it improves the overall transmission of data and decreases overall routing path loss.

![Graph showing packet loss per round](image)

The next graph shows us the residual energy after completion of some rounds and it also shows us that at the starting of the transmission, the energy reduction of our proposed system is high as clustering and other stuffs are there but after a peak point, the reduction in the energy is very slow compared to original algorithm. So, we can say that for more transmission rounds the proposed algorithm is much better than the DFCA and FNR algorithm.
7. CONCLUSIONS

Use of FNR algorithm provides us reduction in data loss, number of sensor nodes enhanced and also provides us the energy efficiency but it does not use the clustering topology, if we apply clustering then we can make the network more energy efficient. The DFCA algorithm provides us the more energy efficiency in the WSN. So, for development of more energy efficient network with low data loss we apply DFCA before the GD phase in FNR. So, we can improve the energy efficiency & also can improve the fault tolerance by which we can easily improve the use of sensors nodes and the lifetime of sensor network. the lifetime of sensor network.
8. REFERENCES


