Fault Tolerant Clustering Approaches in Wireless Sensor Network for Landslide Area Monitoring

Rehana raj T

Amrita University
Amritapuri Campus
College of Engineering
rehna@amritapuri.amrita.edu

Maneesha V Ramesh

Amrita University
Amritapuri Campus
College of Engineering
maneesha@amritapuri.amrita.edu

Sangeeth k

Amrita University
Amritapuri Campus
College of Engineering
sangeethk@amritapuri.amrita.edu

Abstract— **The installation or deployment of a wireless sensor network (WSN) in a real world application is prone to innumerable failures such as software or hardware malfunctioning, environmental hazards, radio interference, battery exhaustion, etc. In a safety critical application such as landslide prediction, fault tolerant approaches have to be followed to ensure the availability of sensor data at the analysis station during a critical situation.**

We propose a fault tolerant and energy efficient clustering approach which organizes the whole network into smaller cluster and subcluster groups enabling a considerable reduction of communication and processing overhead. Subcluster formation also gives the possibility to skillfully deal with sensor nodes, node leader, and cluster head failures. We also propose a fault tolerant approach that uses a matrix based error approximation method for providing the approximate sensor data of the failed node. The approximate code prediction takes into consideration various geological aspects of the problem.

Keywords - WSN,Clustering,Fault tolerance,Energy Efficiency

I. INTRODUCTION

In the rainfall induced landslide[1] scenario, the sensor nodes are distributed in different locations which are categorized into three zones based on a probabilistic analysis of the geological factors in each zones. Any movement or ground acceleration can be predicted only if the geological properties like pore water pressure and seepage forces exceed a threshold value. In the proposed hierarchical architecture, the lowest level nodes consist of battery powered, wireless sensor nodes containing processing and RF communication modules. The geological data that are measured for the particular application are pore water pressure, ground vibration, soil moisture, tilt or acceleration and strain on the particular sensor column into which these analog sensors are placed and buried under the ground. The data are sampled at the sensor nodes at various sampling rate. The proposed sensor network is also time driven that is suitable for periodic data monitoring. Sensor nodes periodically sample the environmental data and transmit the data at constant time intervals to the aggregating node.

II. CLUSTERING TECHNIQUES

The WSN is formed by a number of small sensor nodes (SN), cluster heads (CH), node leaders (NL), beacon node (BN) and base-station (BS). All wireless nodes physically connected to the measuring sensor columns, will be self elected as sensor nodes (SN), and the unconnected wireless nodes acts as Cluster Heads (CH) as shown in figure 1.
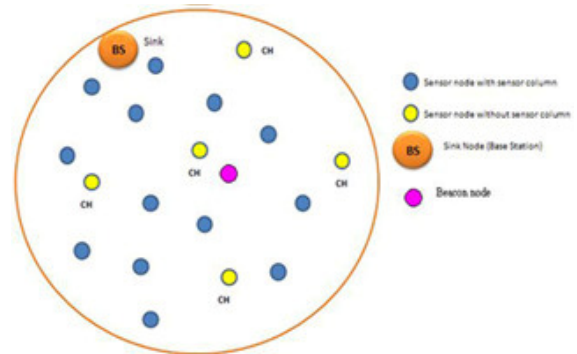

Figure 1. Distribution of wireless nodes in the sensor network

The wireless sensor nodes are connected physically to the sensor column, which contains pore pressure transducers and dielectric moisture sensors for measuring soil pore pressure, strain gauges, tiltmeters and geophones for measuring the earth movements .In many wireless sensor network applications, the sensor nodes in different regions can collaborate and aggregate their data to give more accurate reports about their local regions. This will save from communication overhead and energy loss. The sensor nodes deployed in the wide area, will form many cluster groups for efficient network organization , where each cluster group contains sensor nodes (SN) in majority, one cluster head (CH) and one node leader (NL).The consensus algorithm that is executing at the node leader (NL) will gather and aggregate the sensor data from other sensor nodes(SN) in the same cluster group.

The cluster head (CH) will then forward the aggregated data coming from the node leader, to the base station either directly or through other cluster heads. The energy consumption in cluster heads is more when compared to other sensor nodes, since they transmit data

over long distances. So we assign energy abundant nodes to act as cluster heads. The clustering reduces channel contention and packet collisions, resulting in better network throughput under high load.

III. CONSIDERATIONS IN THE LANDSLIDE MONITORING

The sensor columns will be deployed in geologically specific locations for parameter measurements. The sensor outputs from the sensor column are connected to the 8 channel ADC in wireless sensor node. When the output terminals of tilt meters and accelerometers use two channels of ADC, all other sensors uses only single channel. Based on consultation with geological scientists [1] [2], we assume the landslide deployment region as a triangular hill, where the entire deployment region is divided into three zones as ZONE1, ZONE2 and ZONE3 by considering the priority of the geological parameters. Pore water pressure builds up rapidly at ZONE1, which is at the lower portion of the hill. Therefore, sensor columns with more ratios of pore pressure transducers are placed here as shown in figure 2.
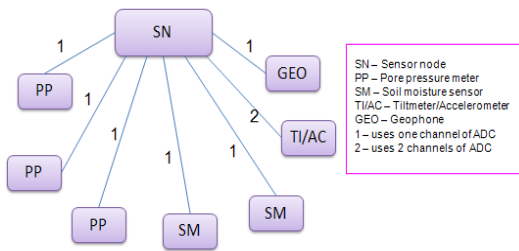
Figure 2. Possible Sensor distribution in ZONE1

ZONE2 lies in the middle portion of the entire hill, where we propose to distribute the sensors equally as shown in figure 3.
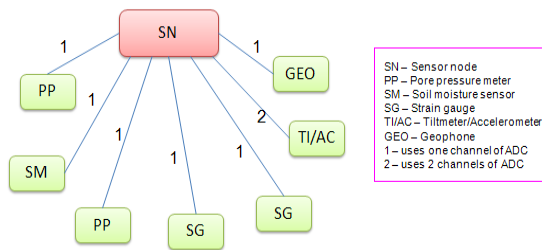
Figure 3. Possible Sensor distribution in ZONE2

Assuming that landslide movement will be rapid at the top, then more number of strain gages and tilt meters can be distributed in the sensor column that is to be deployed in ZONE 3 as shown in figure 4.
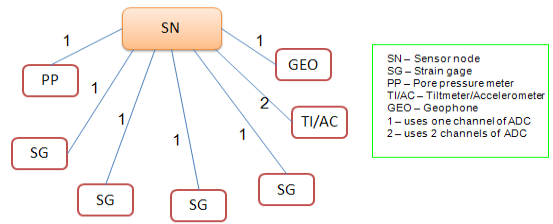
Figure 4. Possible Sensor distribution in ZONE3

Entire wireless sensor network contains low level sensor nodes (SN) to sample the geological data from the local region, cluster heads (CH) that organizes the sensor nodes in its cluster group, beacon node (BN) that will broadcast the time synchronization signal through the cluster heads to the cluster members periodically and the Base station (BS). The beacon node is placed in the network such that it hears maximum number of cluster heads and also communicates with the base station (BS).

The frequency of the crystal varies from node to node, because of manufacturing defects or change in temperature. So to synchronize the clocks in the motes, tine synchronization is needed. Initially when all the sensor nodes are ON, every node will broadcast at a fixed transmission range to find its the neighbors. On receiving an acknowledgement from the neighbor nodes in different times, each node will create a 'neighbor hood list' after ranking it based on the time stamp. Each sensor node will save its 'neighbor hood list' and which will be loaded on to 'Beacon node' and to the 'Base Station for replication. Figure 5 shows the sensor distribution and cluster formation in a landslide scenario.
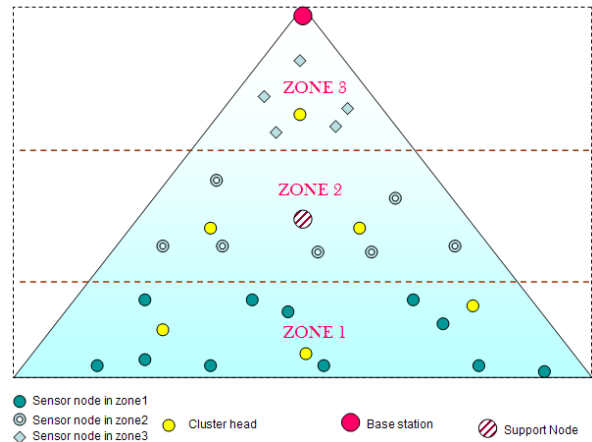
Figure 5. Distribution of sensor nodes in the landslide site

Basic Assumptions in this approach:

- Priority of sensors will be different in the three ZONES.
- Number of Cluster heads (CH) is known.
- No of sensor nodes (SN) is known.

A. Cluster Head Election

Dominating Set based Algorithm for determining the number of cluster heads

In a graph G (V, E), V is set of vertices and E is the set of edges formed from pairs of distinct vertices in V. A subset S of sensor nodes is called a *dominating set* of G, if every node in V - S is connected to some node in S. The minimum number of nodes in a dominating set of G is called domination number of G as seen in figure 6.
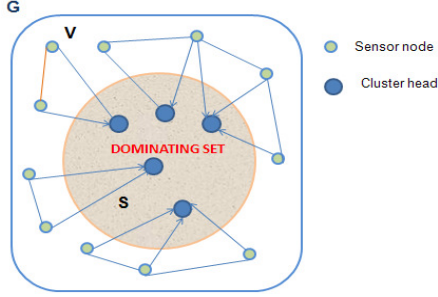

Figure 6. Dominating set based cluster head determination

Domination set problem:

Given a graph G and a positive integer k, Is Domination number (γ) of G is less than or equal to k . i.e., $\gamma(G) \le k$.

In [3], Cockayne et. al. had shown that the domination problem is NP Complete. However, we present an algorithm for the domination set problem to find the cluster heads in a cluster group. Beacon node lies outside the dominating set and is adjacent to maximum number of cluster heads (CH).Beacon node of G is a node v, where $v \in V - S$, such that v is connected to maximum number of nodes in S.

Algorithm 1: Dominating Set based Algorithm for fixing Cluster heads

- Input the sensor network graph G through its adjacency matrix.
- Create a vector containing the row sums .
- Choose a vertex, *v*, of the highest degree in the graph G and put v into the dominating set, S.
- Remove all the edges incident to v in the graph and update the row sums.
- If G has no edges, STOP. Go to step 5. Else go to step 3 (G has no edges, if all the entries of the adjacency matrix of G are zeros).
- The remaining vertices of G form a dominating set of G

Complexity of the algorithm.1:

Let there be n vertices. The algorithm needs $n(n-1)$ comparisons to find the degrees and at most n-1 comparisons to

find the highest degree, so that $n^2 - 1$ operations are required. Then the removal of edges requires at most $2n$ assignments and the updating of the row sums require $2n$ assignments, so that the complexity of this algorithm can be calculated as follows:

$$\sum_{n=1}^{n} n^2 - 1 + 4n = \frac{n(n+1)(2n+1)}{6} - n + 4\frac{n(n+1)}{2} = O(n^3)$$

B. Cluster formation.

Initialization - Adjacency algorithm to form clusters

Base station initiates the cluster heads (CH) to from their cluster groups by sending a broadcast packet. The cluster heads send a broadcast at its maximum transmission range. All nodes that hears the broadcast generates a node database containing NODE ID, Energy Level of the node Neighbor hood list, Received Signal Strength Indication(RSSI), Number of other CH's sharing the same node.

All sensor nodes hearing the broadcast from cluster head will send an acknowledgement message tagged with the node data base. Cluster heads decides the members of the group on receipt of the acknowledgement messages. After cluster groups are formed, the time synchronization signals will be broadcasted from the beacon node to the cluster members through its respective cluster heads. The cluster members start sampling at the same instant. Each cluster head determines its frame size and time slot size according to the number of nodes attached to it. This will help in the dynamic change of frame size and slot size, which will reduce a redundant delay that normally occurs in a TDMA (Time Division Multiple Access) channel access. Using TDMA scheme, each sensor node will send the corresponding 'Node database' to the cluster heads from which the cluster head can find out the sensor nodes in the overlapping zone between the cluster heads.

The cluster formation schematic of a landslide monitoring area is as shown in figure7,
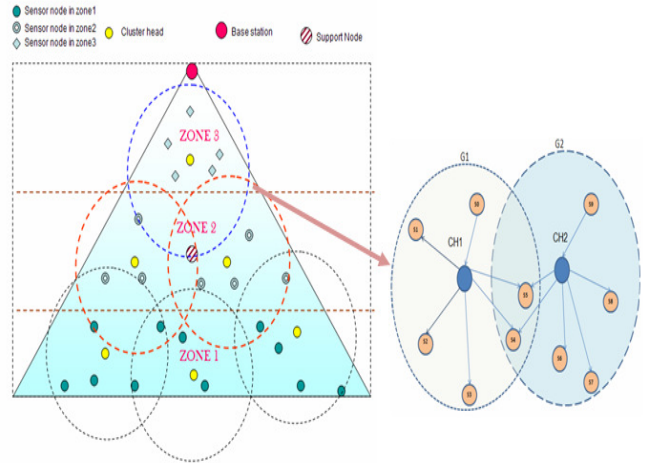

Figure 7. Cluster formation in landslide monitoring area

Cluster Formation using Adjacency matrix

- Determine cluster heads from dominating set S of the graph G = (V, E), cluster heads.

- Enter an adjacency matrix A =[a $_{ij}$] such that,

  [a $_{ij}$] = 1 if G has an edge (i, j)

  = 0 else

- Choose a cluster head and put all its neighbors in its cluster. The neighbors of a node correspond to 'ones' in the respective row of the adjacency matrix.

- If there is no more cluster heads, STOP. Else, Go to the next cluster head and go to step.3

Example: In the graph G(V,E) in figure 8 , path between the vertices are adjacent and it is represented as   '1' in the adjacency matrix as shown in figure 9.
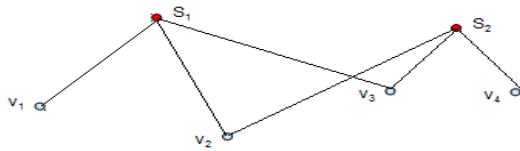

Figure 8.   Graphical representation of the sensor node and cluster head connectivity

The sensor nodes are denoted by v1, v2, v3, v4 and cluster heads by s1 and s2.The row sum gives the degree of each vertex and the highest degree vertices would be the cluster heads.

$$A = \begin{array}{c} \\ v1 \\ v2 \\ v3 \\ v4 \\ s1 \\ s2 \end{array} \begin{array}{c} \begin{array}{cccccc} v1 & v2 & v3 & v4 & s1 & s2 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{array} \right] \end{array} \begin{array}{c} RowSum \\ 1 \\ 2 \\ 2 \\ 1 \\ 3 \\ 3 \end{array}$$

Figure 9.   Adjacency Matrix of the sensor node connectivity graph

By knowing the adjacent vertices (sensor nodes) to the highest degree vertex (CH), cluster groups can be formed as shown in figure 10.
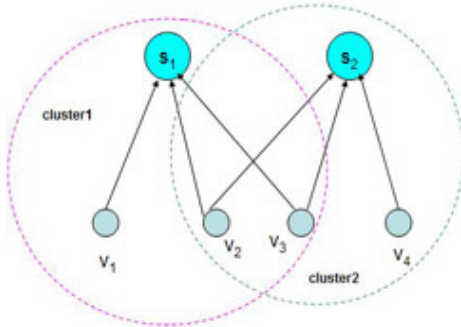

Figure 10. Cluster formation from adjacency matrix

C.   Subclusters for optimal clustering.

The sensor nodes in the intersection of two or more cluster groups will form a subcluster of the respective clusters. The intersection sub clusters are recognized from node database. Sensor nodes lying in the range of more than one or more CH's in the same range, falls within the intersection zone as shown in figure11.
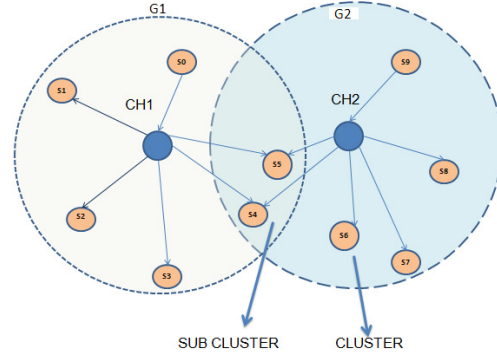

Figure 11. Intersection sub cluster

We know that each sensor node gathers different geological data from the field at different sampling rates. The information is sent in packet format for all node to node, nodes to cluster head and cluster head to cluster head communications. The energy consumed for transmission and reception of packets vary with distance of communication, although the energy consumed is fixed in mote electronics [4]. Proposed clustering techniques aim at reducing the amount of energy spent in transmitting and receiving the data packets, by aggregating the data packets in the 'Node leader' .The sensor node having minimum power consumption and which is closer to the cluster head will be elected as the node leader (NL). In a case where more than one node has minimum power consumption, it is obvious to choose the node which is closer to the cluster head as node leader. The aggregated data would be sent to the cluster head from the Node Leader, thereby conserving the energy for transmission and processing.

D.   Partitioning of clusters to sub clusters using graph partitioning based algorithm

The optimal clustering [5] will associate the sensor nodes to cluster heads such that the network lifetime with respect to energy is maximized. As a fault tolerant and optimal measure, we are considering the intersection zone as a sub cluster where the respective node leader will act as the cluster head. In a single cluster group, if cluster count is normal, the node leader will aggregate the data values from the cluster members and forward to cluster head to reduce the transmission overhead.

If the cluster count is more than the energy handling capacity of the cluster head, then partition the whole cluster into two sub clusters SC(1) and SC(2), where SC(1) uses the cluster head (CH) as the leader and SC (2) uses the Node

4

Leader, to coordinate and aggregate the data values from the respective cluster groups. The second sub cluster SC(2) will then forward the aggregated data to the cluster head which lies in the sub cluster SC(1) as shown in figure12.
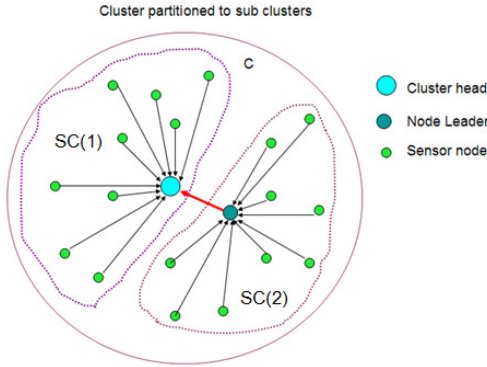

Figure 12. Cluster partitioning

The maximum energy capacity of the node leader to handle aggregation of data packets from the sensor and its transmission nodes is represented by W_{nl}. The energy consumption of the sensor node due to processing and communication is represented by $w(v_i)$. To form the sub cluster SC(2),where with node leader (NL) is the head, each sensor nodes will be added one by one to the sub cluster SC(2) until the maximum energy capacity ,W_{nl} of the node leader is met. The rest of sensor nodes which are not added to node leader cluster group are assigned to the sub cluster (SC1), with cluster head as the leader of sub cluster.

Algorithm 2: Energy efficient algorithm for cluster Partitioning

- Put the node leader into Sub cluster-2 and the cluster head into the Sub cluster-1 as leaders.

- Initially, *weight* = 0; W_{nl} = w (node leader);

  For every vertex Vi of the graph,

  weight = weight + w (V i)

  If **weight ≤ Wnl**, then put Vi to sub cluster-2

  Else go to next vertex.

- Put the remaining nodes into the sub cluster-1.

V. FAULT TOLERANCE IN CLUSTER GROUPS

To minimize energy consumption in the sensor data processing and its transmission, the node leader (NLoz) at the intersection zone aggregates the data from the nodes inside the intersection group. The node leader (NLoz) could forward the aggregated values to the cluster heads in the corresponding cluster groups i.e., the NLij can transmit it's data to either cluster heads, CH1 or CH2, whenever it hears 'free' status from any of them as shown in the figure 13.
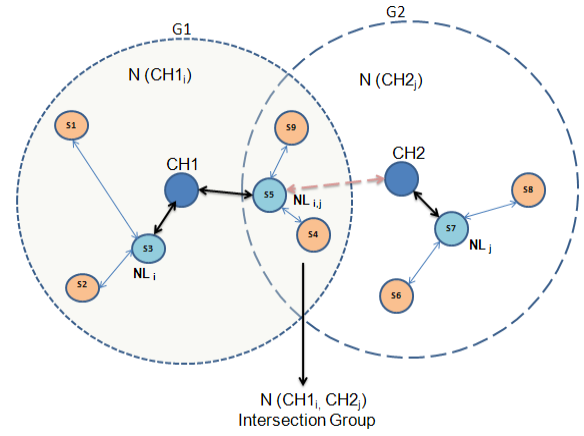

Figure 13. Data aggregation and transmission inside intersection subcluster

The node leader shows "busy" status while executing the consensus algorithm, which means that the node leader is not accessing its cluster head and the corresponding cluster head is "free". The "busy" status message from the node leader of the cluster group is forwarded to the node leader in the intersection group , thereby initiating the transmission of aggregated sensor node values from the intersection node leader to the respective "free" cluster head. We know that cluster heads are nodes having more hardware capabilities for processing and communication. Cluster head failures may occur due to reasons like radio interference, dislocation of sensor nodes, battery exhaustion etc caused by hardware faults, software faults or environmental conditions.

A. Fault tolerance of sensor nodes

Individual sensor nodes give a 'faulty' status (node off, power null, incorrect value) whenever it encounters a fault or when it gives incorrect data values. The sensor node failures make the sensor data unreachable to calculate the critical threshold value, for predicting the event. Considering the time taken for fault repair in a safety critical application like landslide prediction, we propose a fault tolerant approach that uses a 'Matrix based error approximation method' for providing the approximate sensor data closely related to the sensor data from the failed node. Even if the sensor node fails, it is possible to give an approximately predicted sensor data from the same region, in the same sensor reading time interval. Thus the sensor information from the failed sensor node can be approximated from the geographically nearest sensor using following conditions,

- If the nearest sensor is in same cluster, the instruction can be given for cluster head to approximate the data. The information from the nearest sensor will be send as the information of 'missing node' to the beacon node.

- If the nearest sensor is in different cluster, the base station or beacon node should give instruction to cluster head for the approximation of data. The information of the nearest sensor will be collected from the respective cluster heads and send to beacon node.

5

In this case theoretical values of the failed sensor can be predicted using XOR approximation matrix based on following assumptions:

- Each sensor node generates a code of length, L= (N.n), where N is the number of sensors and n is the number of bits, to the respective CH's.

- The sensors in all the three zones are prioritized

- The binary outputs from each sensor type should differ only by a prefix number of bits. The bit difference is unique for each sensor in each zone

The fractional weights w1, w2, w3, w4 and w5 are assigned in the increasing order to the bit difference and the priority of each sensor in each zone is illustrated in figure 14 .

WEIGHTS	Zone1	Zone 2	Zone 3
Priority1 (w1=0.5)	PP	PP	SG
Priority2 (w2·=0.2)	MS	MS	TI
Priority3 (w3=0.15)	SG	SG	GE
Priority4 (w4 =0.1)	TI	TI	PP
Priority5 (w5=0.05)	GE	GE	MS

W1+W2+W3+W4+W5=1
PP = Pore pressure transducer
MS= moisture sensor
SG = strain gauge
TI = Tiltmeter
GE = Geophone

Figure 14. Priority list of parameters in each zone

Matrix approximation algorithm for the missing data of failed sensor

Step.1. The data from a sensor node corresponds to a 5 x 16 matrix as shown in figure15.

$$D = \begin{bmatrix} P_{1,1} & P_{1,2} & ....P_{1,j} \\ M_{2,1} & M_{2,2} & ....M_{2,j} \\ S_{3,1} & S_{3,2} & S_{3,3} & ....S_{3,j} \\ T_{4,1} & T_{4,2} & T_{4,3} & T_{4,4} & ....T_{4,j} \\ G_{5,1} & G_{5,2} & G_{5,3} & G_{5,4} & ....G_{5,j} \end{bmatrix}$$

Figure 15. Data Matrix from the sensor node

Each row of the data matrix is a 16 bit code for each five sensors, where P is the pore pressure value, M is the moisture content, S is the strain gauge, T is the tilt meter and G is the geophone. Assume $\sum_{i=1}^{5} w_i = 1$, where w is the fractional weight assigned to each sensor in a particular zone. For highest priority parameter, highest weight will be assigned and only least bit difference is allowed.

Step.2. If the sensor v is failed at a time t. Then let D_0 be the latest 5 x 16 data matrix of the 'failed node' just before the failure at the previous time, t-1. Let $D_{i=1\ to\ k}$ be the current data matrices of the k neighboring sensors obtained from the neighbor hood list of each cluster head, at time t. Among the neighbors of the failed nodes, we choose a code with desirable properties and obtain its data matrix.

Step.3. Use the XOR approximation to predict the values for the 'missing data matrix' D_0 of the faulty node. Let A be a difference matrix obtained by XORing the data matrix of faulty node D_0 obtained at time t-1, with any nearest sensor node D_i

$$A_i = D_0 \oplus D_i$$

Consider two 2 x 3 matrices D_0 and D_1 shown in figure 16.The row values in the 'data matrix' correspond to parameter values. By XOR operation between the values of faulty node and a nearest node gives the difference matrix representing the difference in the bits as '1' and identical bits as '0'.
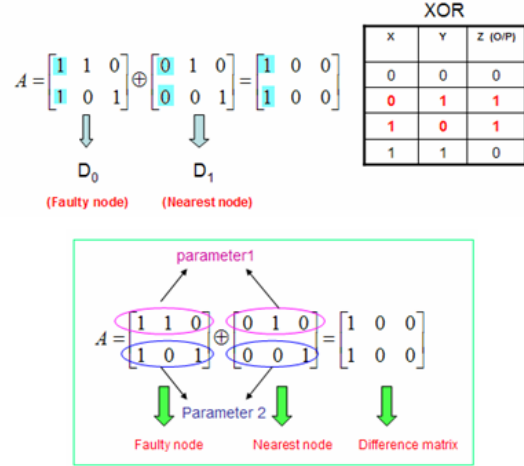

Figure 16. XOR operation of faulty node and a nearest node

Step.4. To find the error approximation matrix, let B = *A x W*, where A is an *m x n* matrix and W is an n x m matrix as shown in figure 17.

$$B = \begin{bmatrix} a_{ij} \end{bmatrix}_{mxn} \times \begin{bmatrix} w_1 w_2 w_3 w_4 w_5 \\ w_1 w_2 w_3 w_4 w_5 \\ w_1 w_2 w_3 w_4 w_5 \\ \vdots........................ \\ \vdots................... \\ ....\vdots................ \\ w_1 w_2 w_3 w_4 w_5 \end{bmatrix} = \begin{bmatrix} b_{ij} \end{bmatrix}$$

Figure 17. Error approximation matrix

Step.5. Consider sum of diagonal elements of the error approximation matrix for obtaining the accurate weight of the matrix.

$$E_{D_i} = \sum_{i,j} b_{ii}$$

Step.6. Repeat above steps to get corresponding error approximation data matrices for all nearest 'k' sensor nodes. Choose a matrix D_i such that E_{D_i} is minimum. Thus we can predict the value of missed code from the nearby sensors also by considering the geological aspects.

*Step.7.*The expected value of the code of the failed sensor can be computes as follows:

Previous value of the failed sensor + Δ= Expected value

Δ is a measure on incremental change in the code.i.e, the approximate code can be taken as the exact value of failed sensor for the purpose of prediction of landslides, subject to certain error which depends on the approximation algorithm.

B. Fault tolerance of Cluster Heads and Node leaders

In case of *single cluster head failure* [9], the respective node leader will forward the latest aggregated data to the node leader in the sub cluster of the intersection zone. The resultant data at the intersection node leader is assigned to the nearby 'FREE' cluster head of overlapping cluster. In figure 18, the aggregated values at sensor node S(5) is,

$$V(S_5) = V_{agg}(NL_i) \oplus V_{agg}(NL_{ij})$$

The aggregated value V(S₅) at the intersection zone node leader would be forwarded to the cluster head in the consecutive overlapping cluster group, when it shows the status 'FREE'.

CH1 Failed : The aggregated data value at NL at intersection zone is, $V(S_5) = V_{agg}(NL_i) \oplus V_{agg}(NL_{ij})$
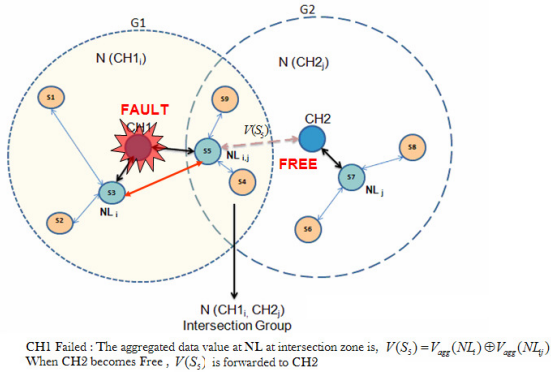When CH2 becomes Free , $V(S_5)$ is forwarded to CH2

Figure 18. Single cluster head failure

Similarly in the case of multiple cluster head failures [10], the node leaders of the same cluster will forward the data to intersection zone node leader ,which acts as fault tolerant node providing nearly faithful backup replication of data to the beacon node as shown in figure19.

Aggregated value at $V(S_5) = V_{agg}(NL_i) \oplus V_{agg}(NL_{ij}) \oplus V_{agg}(NL_j)$
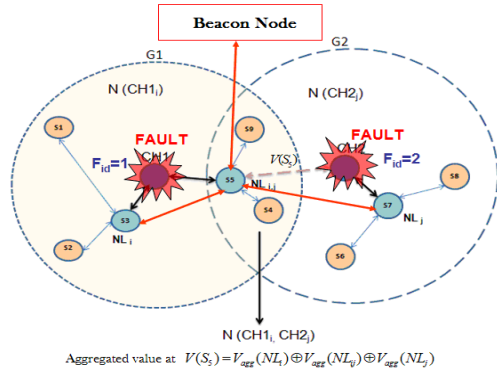
When the node leaders are failed, the respective cluster heads can take over the work and at the same time node leader election algorithm will be executed at the cluster head / beacon node to find new node leader so that cluster head can be released from the aggregation. If the new node leader is found inefficient to handle the number of sensor nodes, then the cluster partitioning algorithm is to be executed to form sub clusters for better performance.

VI. CONCLUSIONS AND FUTURE WORK

Here we introduced some energy saving clustering techniques, to compute the data gathered from sensor nodes in the overlapping zones in cluster group areas. Efforts will be made to predict the failed data based on Fuzzy control system. If control unit and base station is made into fuzzy control system, then the uncertainties or inaccuracies at the prediction of the theoretical value of the missed sensor can be compensated by fuzzifying their respective variables. For example, the predicted value of pore pressure can have a membership value of 0.9 i.e., 90% accuracy.The other considerations in the future are to reduce the energy consumed per bit transmission, to study how the Link quality and RSSI can be measured in a energy efficient way , and to develop an energy efficient hybrid channel access scheme for inter cluster and intra cluster communication. Also we are in the process of evaluating the theoretical aspects and the algorithms mentioned in the paper by simulation.

REFERENCES:

[1] Mathai.John,Sankar.G,Muraleedharan.C,Private communication,February 2007.

[2] Dr. H. M. Iyer,Private communication, Formerly of U.S. Geological Survey May 2007.

[3] E.J.Cockayne,O.Favaron,C.Payan & A.Thomason,Contributions to the theory of domination,independence and redundancy in graphs,Discrete Math. 333 1981249□258.

[4] Wang.WeiZhao., Song. Wen-Zhan., Li. Xiang-Yang., Moaveni-Nejad.Kousha.,Optimal Cluster Association in Two-Tiered Wireless Sensor Networks.

[5] Xu. Ya.,Bien. Solomon.,Mori. Yutaka.,Heidemann. John.,Estrin. Deborah.,Topology Control Protocols to Conserve Energy in Wireless Ad Hoc Networks, Department of Computer Science, UCLA.18.

[6] Younis. Ossama.,Krunz. Marwan., Ramasubramanian. Srinivasan., University of Arizona, Node Clustering in Wireless Sensor Networks, Recent Developments and Deployment Challenges.

[7] Schurgers. Curt., Srivastava. B. Mani.,Energy E_cient Routing In Wireless Sensor Networks, Networked & Embedded Systems Lab (NESL), Electrical Engineering Department, University of California at Los Angeles (UCLA), CA .

[8] Iverson. M. Richard.,Landslide triggering by rainfall in_ltration , U.S. Geological Survey, Vancouver, Washington.

[9] Beigl. Michael., Vogt. Harald., Moreira S_ de Souza. Luciana., A Survey on Fault Tolerance in Wireless Sensor Networks, SAP Research, Braunschweig, Germany. .

[10] Guha. Sudipto., Khuller. Samir., Approximation algorithms for connected Dominating Sets ,Dept of computer science, University of Marylan.

[11] Hauspie. Micha□el.,Localized probabilistic and dominating set based algorithm for efficient information dissemination in ad hoc networks, Amandine Panier INRIA Futurs,POPS research group B^at. M3, France.19