Detection And Recovery Of Faulty Nodes Using Clustering Optimization In Wireless Sensor Networks

Beena T1, Biju Balakrishnan2 & K. Arun Patrick3
1PG Scholar, Nehru Institute of Technology, Coimbatore
2Assistant professor(SG), 3Assistant professor, Nehru Institute of Technology, Coimbatore

Abstract: An WSN may often consist of hundreds of distributed sensors to monitor, collect and transmit information from various environments. Signals are relayed from one sensor to another sensor until it reaches the sink node which could either be mobile or a central control unit. Due to the low cost and the deployment of sensor nodes often in a harsh and uncontrolled environment, it is not uncommon that sensors would easily become faulty sensor nodes. The breaking down of sensors would inevitably affect the transmission of signals or even cause signal lost. Hence it is normal that inactive nodes miss their communication and split the network. The clustering optimization technique efficiently groups the nodes into clusters and selects a cluster head based on the maximum energy. The number of members is limited based on the communication distance and the number of cluster per network is also constrained to number of nodes in the network. The fault nodes are detected based on the signals collected from them and classified into scene fault and node fault based on NHCRF model. The data from the fault nodes are recovered by the cluster head. The secondary cluster head is chosen based on energy level and it takes the position of cluster head in the case of failure of the head. This approach allows faulty nodes recover their data by selecting neighbourhood cluster properly and increases detection accuracy of fault diagnosis and reduces time complexity.

Keywords - Clustering, Cluster head, Fault diagnosis, Secondary cluster head

1. Introduction

An Wireless Sensor Networks (WSNs) are replacing the traditional wired industrial communication systems offer several advantages including easy and fast installation and low cost maintenance and replaces the traditional wired industrial communication systems. WSN applications, such as, industrial process monitoring and control, factory automation and plant monitoring require reliability and timeliness in forwarding messages among nodes. In WSNs, transmission failures can result in missing or delaying of process or control data. And missing the process or control deadline is normally intolerable for industrial applications, as it may possibly terminate the automation or, cause problems in industrial automation, ultimately resulting in economic losses.

The sensed data should be reliably and timely transmitted to the sink node, and the re-tasking or programming data for sensor node operation, command, and query should be reliably delivered to the target nodes. The reliability, timeliness, and energy efficiency of data forwarding are crucial to ensure proper functioning of an WSN so it is also required that these networks can operate for years without replacing the device batteries. The undesirable delay as well as additional energy consumption occurs due to the varying wireless channel conditions and sensor node failures may cause network topology and connectivity changes over time, to forward a packet reliably at each hop, it may need multiple retransmissions.

A sensor node in wireless sensor network that is capable of performing some processing, gathering sensory information and communicating with other connected nodes in network. The main components of a sensor node are a power source, external memory microcontroller, transceiver, and one or more sensors. A sensor node in wireless sensor network that is capable of performing some processing, gathering sensory information and communicating with other connected nodes in network. The main components of a sensor node are a microcontroller, transceiver, external memory, power source and one or more sensors. Most transceivers operating in ideal mode have a power consumption almost equal to the power consumed in receive mode. External memory used for storing application related or personal data, and program memory used for programming the device. An important aspect in the development of a wireless sensor node is ensuring that there is always adequate energy available to power the system. The power for sensing, communicating and data processing are
consumed by sensor nodes. More energy is required for data communication than any other process. Sensors produce a measurable response to a change in a physical condition like temperature or pressure.

![Architecture of a sensor](image)

**Figure 1. Architecture of a sensor**

2. Related works

The sensor nodes in wireless sensor networks may be deployed in hostile and unattended environments. The ill-disposed environment affects the monitoring infrastructure that includes the sensor nodes and the network in addition to node failures and network partitioning. This in turn adds a new dimension to the fragility of the network topology. Such issues are found in conventional wireless networks thus, demand efficient techniques for discovering faulty behavior in such networks.[1] Traditional fault diagnosis techniques devised for multiprocessor system are not directly applicable for WSN due to their specific requirements and limitations. The fault diagnosis techniques are classified based on the correlation between sensor readings, nature of the tests and characteristics of sensor nodes and the network. Distributed self diagnosis is an important problem in wireless sensor networks (WSNs) in that each sensor node needs to learn its own fault status. The classical methods for fault finding using majority voting, median, mean, and hypothetical test based approaches which is not suitable for large scale WSNs due to large deviation in incorrect data transmission by faulty sensor nodes. [11] One of the most critical issues is the detection of corrupted readings among the huge amount of collected sensor data. Indeed, such readings could significantly affect the quality of service (QoS) of the WSN, and it is highly desirable to discard them. This issue is usually addressed through fault detection algorithms that classifies reading by exploiting spatial and temporal correlations.[3] Fault detection, isolation, and estimation are considered for a class of discrete time-varying networked edge sensing system with incomplete measurements.

The sensor nodes in wireless sensor networks may be deployed in unattended hostile environments. The ill-disposed environment affects the monitoring infrastructure that includes the sensor nodes and the network in addition to node failures and network partitioning. This in turn adds a new dimension to the fragility of the network topology. Such issues are common than those found in conventional wireless networks thus, demand efficient techniques for discovering faulty behavior in such networks.[1] Traditional fault diagnosis techniques devised for multiprocessor system are not directly applicable for WSN due to their specific requirements and limitations. The fault diagnosis techniques are classified based on the correlation between sensor readings, nature of the tests and characteristics of sensor nodes and the network. Distributed self diagnosis is an important problem in wireless sensor networks (WSNs) in that each sensor node needs to learn its own fault status. The classical methods for fault finding using majority voting, median, mean, and hypothetical test based approaches which is not suitable for large scale WSNs due to large deviation in incorrect data transmission by faulty sensor nodes. [11] One of the most critical issues is the detection of corrupted readings among the huge amount of collected sensor data. Indeed, such readings could significantly affect the quality of service (QoS) of the WSN, and it is highly desirable to discard them. This issue is usually addressed through fault detection algorithms that classifies reading by exploiting spatial and temporal correlations.[3] Fault detection, isolation, and estimation are considered for a class of discrete time-varying networked edge sensing system with incomplete measurements.

The phenomena of multiple communication delays and data missing are characterized by utilizing unified measurement model. A least-squares filter for the addressed time-varying networked sensing systems that minimizes the estimation variance is first designed, then a novel residual matching (RM) approach is developed.[5] A fault detection strategy for wireless sensor networks based on modeling a sensor node by Takagi–Sugeno–Kang (TSK) fuzzy inference system (FIS), uses the sensor measurement of a node approximated by a function of the sensor measurements of the neighboring nodes. And it also model a node by recurrent TSK-FIS (R-FIS), where the sensor measurement of the node is approximated by taking function the previously approximated value of the node and the real measurements of the neighboring nodes and itself.[6] A novel centralized...
hardware fault detection approach based on Naïve Bayes framework for a structured Wireless Sensor Network (WSN). For most WSNs, power supply is the main constraint of the network because most applications are in some situation and the sensors are with battery only. To maximize the network’s life, the Centralized Naïve Bayes Detector (CNBD) analyzes the collected end-to-end transmission time at the sink. [2]

3. Existing system

Wireless Sensors Networks (WSN) fault diagnosis problem is formulated as a pattern classification problem and introduces a newly developed algorithm, Neighborhood Hidden Conditional Random Field (NHCRF), for determining hidden states of the sensor nodes. The health condition of WSN are determined by using the NHCRF model to estimate the posterior probability of different faulty scenarios. It has relaxed the independence assumption of the Hidden-Markov model so the NHCRF model can improve the WSN fault diagnosis. To enhance, by using the concept of nearest neighbors is used when estimating dependencies the robustness and anti-noise ability of the NHCRF is enhanced. In this paper, a 200-sensor-node WSN is used to show that the proposed NHCRF method can deliver excellent and effective results for WSN-health diagnosis.[1] The contributions of this paper are three-fold. First, a new algorithm called Neighborhood Hidden Conditional Random Field which is derived for handling fault diagnosis such as locating faulty nodes in an WSN. Second, in contrast to other conventional methods that mainly detect faulty sensor nodes, our proposed NHCRF algorithm and its relaxed characteristic enables hidden states of an WSN be determined. Thus, the dependencies among sensors and transmission paths are found. And as a result, faulty scenes caused by faulty transmission path but not sensor nodes can also be detected. Third, only neighborhood dependencies rather than global dependencies are used for detecting states of hidden variables. This eliminates the effect caused by distant neighbors, and thus enhances the robustness of diagnosis significantly. Different from the previous fault diagnosis methods employing many metrics collected by extra devices monitoring the sensors in an WSN, the proposed diagnosis approach relies only on the collected signal delay data.

4. Proposed work

Some WSN with a lot of immobile node and with the limited energy and extension of many sensor nodes and their operation. Hence it is normal inactive nodes miss their communication in network, and splits the network. For avoidance split of network, a fault detection of corrupted node and recovery of data is necessary. A design techniques to maintain the cluster structure in the event of failures caused by energy drained nodes is done. Initially, node with the maximum residual energy in a cluster becomes cluster head and node with the second maximum residual energy becomes secondary cluster head. Afterwards the selection of cluster head and secondary cluster head will be based on available residual energy. Sensors provide an easy solution to those applications that are based in low maintenance and the inhospitable areas where conventional approaches prove to be impossible and very costly. Sensors are generally equipped with limited energy, data processing and communication capabilities and are usually deployed in an area of interest to monitor events and gather data about the environment. Sensor nodes are typically expected to last until their energy drains so it is vital to manage energy wisely in order to extend the life of the sensors for the duration of a particular task. Failures in sensor networks due to energy depletion may increase this often results in scenarios where a certain part of the network become energy constrained and stop operating after sometime. Sensor nodes failure may cause connectivity, data loss and in some cases network partitioning. In clustered networks, it creates holes in the network topology and disconnects the clusters, thereby causing data loss and connectivity loss. Numbers of fault tolerance solutions are

![Figure 2. Architecture for detection of fault nodes using NHCRF](image-url)
The clustering optimization is done by clustering the nodes in the network based on some constraints such as sensing range, transmission range, network size and number of nodes. The clustering information is stored in the queue and stacks in the program. The nodes are ordered based on the neighboring node distance with each node containing next hop information.

5. Cluster formation

Initially a set of sensor nodes are dispersed in the sensing area. We assume that sensor nodes know their location and the limits S and D. Algorithms for estimating geographic or logical coordinates have been explored at length in the sensor network. In the first step is to calculate Eth and Eic for every node i, N < I < 1. Eic is the total energy spent on each link of its next hop neighbours. Eth is the energy spent to communicate with the farthest next hop neighbor. Every node i has an initial energy, Einit. A flag bit called “covered flag” is used to denote whether the node is a member of any cluster or not. It is set to 0 for each node initially.

6. Determining cluster head

The node with highest energy level is chosen as a cluster head. Every cluster head sends a message cluster head status msg and Eic to its neighbors (within sensing range) and every cluster head keeps a list of its neighbor cluster heads along with its Eic. The nodes which receive Eic lesser than itself relinquishes its position as a cluster head. The cluster heads which are active send their messages to the cluster head manager outside the network. The cluster head manager has the information of the desired cluster head count. Another round of cluster head relinquishing starts if the number of cluster heads are still much more than expected, then. This time the area covered would be greater than sensing range. The area covered for cluster head relinquishing keeps increasing till the desired count is reached.

7. Adjusting number of members

Nodes send a message hello messages along with their neighbor nodes which are received by nodes within the transmission range. For example in figure (4) nodes a,b,c,d,w,x,y are within transmission range of v. After receiving the hello message, the node v calculates the distance between itself and nodes a,b,c,d,w,x,y using the coordinates from hello message. It stores the distance di and the locations in the distance table. N. Nodes within the sensing range are the neighbors of a node and the nodes w,x,y,b are neighbors of v. Among the nodes within the sensing range, first D closest neighbors are chosen as its
potential candidates for next hop. Assuming D=3, in figure (3), the closest neighbors of v are w,x,y. Among the candidates, the farthest node’s distance \( d_{\text{max}} \) is taken for the calculation of Eth. Suppose a node needs power \( E \) to transmit a message to another node who is at a distance ‘d’ away, we use the formula \( E = k d^c \) where \( k \) and \( c \) are constants for a specific wireless system.

**Figure 3. Clustering of nodes**

Usually \( 2 < c < 4 \). In our algorithm we assume \( k = 1, c = 2 \). For a node \( v \), \( d^1 * d^2_{\text{max}} = E_{\text{thv}} \), since there are \( D \) members to which a node sends message. Eic is the total energy spent on each of link of the D closest neighbors. For a node \( v \), where \( k = 1 \). Div is the distance between node i and node v. Eligibility for cluster head position based on their energies is determined after the calculation of threshold energy Eth, nodes become A node v becomes eligible for the cluster head position if its Einit > Ethv. The admissible degree D is reduced by one and then Eth is recalculated there is insufficient number of cluster heads when no nodes satisfy this condition. The lowest value that D can reach is one. Clustering is not possible when the condition Einit > Ethv is never satisfied at all, because no node can support nodes other than itself. There may also be situations where all the nodes or more number of nodes are eligible for being cluster heads and the excess cluster heads are made to relinquish their position.

**8. Tracking number of clusters**

Each and every node tracks the size \( S \) of the cluster. By sending a message to each one of them the cluster head accounts for itself and equally distributes \( S-1 \) among its next hop neighbors. The neighbours that receive the message distribute among all their neighbours except the parent and account for themselves. Until they reach a stage where the size is exhausted the messages propagated. Then the algorithms terminates if all the nodes have been covered if the size is not satisfied. The cluster is ready for operation after the cluster formation. The nodes communicate with each other for the period of network operation time.

**9. Detection of fault nodes**

Signals are relayed from one sensor to another sensor until it reaches the sink node which could either be mobile or a central control unit. Due to the low cost and the deployment of sensor nodes often in a harsh and uncontrolled environment, it is not uncommon that sensors would easily become faulty sensor nodes. The breaking down of sensors would inevitably affect the transmission of signals or even cause signal lost. Therefore, a lot of research has been done to improve the robustness of WSN fault diagnosis issue as a pattern classification problem. It uses signals collected from sensors, such as signal strength and frequency, and signal delay as features for classifying whether or not an WSN suffers from faulty sensors. The health conditions of WSNs can be determined by many attributes, e.g., package lost, radio interferences, package size, data collision on physical layer, etc. All of these measurements, however, cannot be easily obtained in real-world applications, especially for those sensors working in severe environmental conditions. The diagnosed result are classified into either node fault or scene fault.

**10. Conclusion**

The clustering optimization technique efficiently groups the nodes into clusters and selects a cluster head based on the maximum energy. The number of members is limited based on the communication distance and the number of cluster per network is also constrained to number of nodes in the network. This project is carried out with two phases. The clustering is done and the fault nodes are detected based on the signals collected and classified into scene fault and node fault. This approach allows faulty nodes detection by the cluster properly and increases detection accuracy of fault diagnosis and reduces time complexity. And it recovers the data from fault nodes by the cluster head. The secondary cluster head is chosen based on energy level and it takes the position of cluster head in the case of failure of the head. In future direction of research, various optimization approach for recovery of data can be implemented in order to improve the accuracy.
of the results. The energy efficiency can be enhanced by utilizing some improved mechanism.

11. References


