IJCSN International Journal of Computer Science and Network, Volume 5, Issue 4, August 2016
ISSN (Online): 2277-5420 www.IJCSN.org
**Impact Factor: 1.02**



# Implementation of Data Mining in Wireless Sensor Networks: An Integrated Review

**Mohd Muntjir**

Research Scholar OPJS University,
Churu, Rajasthan, India

**Abstract - Current years have witnessed the emergence of wireless sensor networks (WSNs) as a new information-gathering paradigm, in which a large number of sensors scatter over a examination field and extract data of interests by reading real-world phenomena from the physical environment. Energy consumption becomes a primary concern in a WSN, as it is crucial for the network to functionally operate for an probable period of time. The WSN's extraordinary characteristics direct us to innovative research challenges in some data mining process. Data mining is one of the most important methods by which useful patterns in data with minimal user interference are known and available information of users and analysts to make decisions relayed on their vital organizations to adopt. Data mining, as the continuance of multiple intertwined disciplines, consisting statistics, machine learning, pattern recognition, database systems, information retrieval, World-Wide Web, visualization, and lots of application domains, has made great progress in the past decade. To ensure that the advances of data mining research and technology will competently benefit the progress of science and engineering, it is important to scrutinize the challenges on data mining posed in data-intensive science and engineering and explore how to further develop the technology to assist new discoveries and advances in science and engineering. In WSNs, hierarchical network structures have the advantage of supplying scalable and energy efficient solutions.**

**Keywords -** *WSN, Partitioned WSN, Scheduling in WSN, Location-Based Scheduling, Data Mining.*

## 1. Introduction

### 1.1 Introduction to WSN

Wireless Sensor Networks (WSNs) mechanism is rather simply applicable to a variety of fields. It is based on smaller nodes, radio transceiver, and battery. The wireless sensor networks execute function in autonomous manner in the spatial field to get accurate values. Wireless Sensor Networks (WSNs) are a moderately new application in the network, which provides high quality monitoring for great geographical areas with relatively inexpensive equipment [1]. WSNs are composed of set of tiny sensor nodes, which can effectively monitor their adjacent environment. Due to the wide potential applications in battlefield surveillance, environmental monitoring, healthcare, weather forecasting, and disaster detection etc. WSNs have attracted quite attention from both academic and industrial fields in current years [2] WSN have a numerous advantages, but the available energy at each sensor nodes are treated as a restriction. Hence energy consumption is a major criterion [3].

### 1.2 Partitioned WSN

The sensor network scenario with tree based topology which having one sink node/ base station with four parent nodes. Every parent node has a link to one or more child nodes as shown in figure 1.

In the figure below Node 0 is sink node and parent node are 3, 8, 11, and 21. Their corresponding child nodes are shown through the link as shown in Fig 1. Consider power of node 11 is drain off entirely which result in partitioned WSN, node 11, 16, 17, 18, 19 and 20 are pack of sensor nodes in divided WSNs. In partitioned network sensed data from other nodes are transitory collected by earlier node of the parent node. Past mobile robot visits the parent node; MR gets the gathering information from that pervious node of parent node in partitioned system. For example consider partitioned network nodes 11, 16, 17, 18, 19 and 20.

In this node 11 is dull node. So, sensed data from 17, 18, 19 and 20 nodes are transitory collected by node 16, which is previous node of the parent node. Past mobile robot visit the node 11, MR gets the assembled information from node 16 in that partitioned network [1].

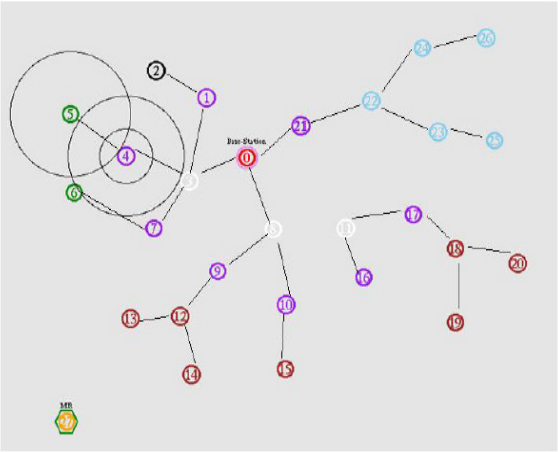

Fig.1: Example of tree based WSN

## 1.3 Scheduling in WSN

### 1.3.1 Time-Based Scheduling

(TBS) Time-based scheduling is based on the time a node becomes the status of death node and near death node. Fig. 1 shows that nodes 3, 8, and 11 are denoted as death node and the values of Death time for every node are in the order of node 8 < node 3 < node11. Allowing for above death time of parent nodes, node 8 is first lost node or going to dead node close to death node before node 3 and node 11. So, the mobile robot call node 8 first and then MR visit node 3 and node 11 for data gathering method.

### 1.3.2 Location-Based Scheduling

In most situations, the distance between the death node and the sink of the WSN is measured. Thus, location based scheduling (LBS), based on the distance amid the sink and the death node and near death node, is necessary. LBS is set to command the MR to visit the nearest the death node and near death node first to collect necessary information. The Euclidean distance formula calculates the distance between the sink and the location of the death node and near death node.

For example, distance between sink and the dead node/near dead node as follows. Fig. 2 shows that the distances of nodes 3, 8 and 11 linking the sink are in the sort of distance (3) < distance (8) < distance (11). Node 3 is little distance to the sink, when comparing distance between parent node and sink. So, the mobile robot visit node 3 first and then MR call node 8 and node 11 for data gathering process [1].

## 1.4 Data Mining

The rapid development of computer and information technology in the last twenty years has fundamentally changed roughly every field in science and engineering, transforming many disciplines from data-poor to progressively data-rich and calling for the improvement of new, data-intensive methods to conduct research in science and engineering [4].

Data mining is one of the most imperative methods by which useful patterns in data with minimal user intervention are known and accessible information of users and analysts to make decisions based on their vital organizations to adopt. Half-term data mining process involuntarily analyze large databases to find useful patterns can be applied. Data mining is a powerful innovative technology with great prospective to help companies focus on the most important material in the data they have composed about the performance of their customers and probable clients.

It determines information within the data that queries and reports can't efficiently uncover. Data mining originates its name from the Relationships between searching for esteemed information in a large database and mining a mountain for a vein of valuable element. Both procedures require either sifting through an enormous amount of substantial, or logically analytical it to find where the value dwells [10].


Fig.2. Data Mining Process

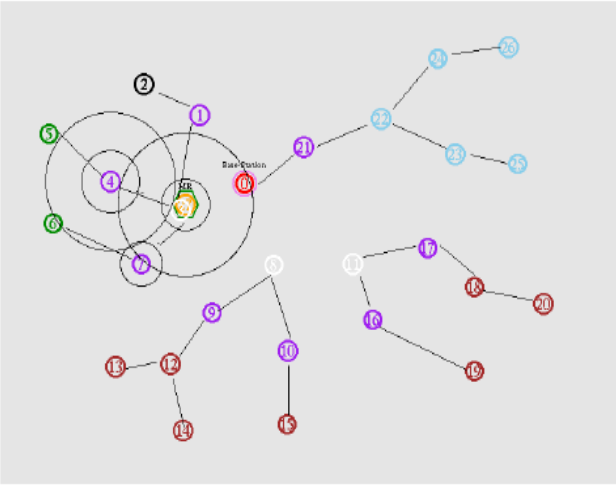

Fig.3. Example of Location Based Scheduling

Data mining is the process involves three steps are:

1. Initial excavation, 2. Construction sculpts or gain credit with the help of pattern recognition / approval, 3. Operation, Step 1: search. Usually this order with facts preparation will be done which may include data cleansing, data conversion and election the sub-set of fields is with enormous volume of variable (fields). Then according to nature analytical predictions, this order to precaution model simple or comments to recognize variables and determine the complexity of models for use in the next step requires. Step 2: Production and verification of model validity. This stage is to review different models and selecting the best model predicting the competence of the deals. Several techniques were developed to achieve this aim. And "competitive estimation of models" was called. For this purpose different models used identical data collection until be compared their competence, then the model that have the best performance, is chosen. This technique includes: Bagging, Boosting, Stacking and Meta-learning. Step 3: utilize. Last step before a model that has been selected, the work in new data until expectant`s precaution outs [6].

### 1.4.1 Cluster Analysis in Data Mining

The process of clustering the facts objects into cluster is depended on the similarities and his functionalities is data clustering. This process is much easier to collect and dissimilar sensors are placed in the nearby clusters. The Difference between the clusters is dependent on the characteristics of every cluster and they are measured by various distance functions like Manhattan distance and Euclidean distance [5].

## 2. Literature Review

**V. Karthik in 2012[1]** Data collection process is one of the significant aspects in the design consideration for future analysis in wireless sensor networks. Energy consumption by parent node increases owing to continuous forwarding of sensed data from their respective child nodes especially in the tree based topology. Once the power in the parent nodes was completely drained off, some of the child nodes get isolated/ partitioned from the sink node. The projected data collection method involves deployment of multiple mobile robots whose responsibility is to gather the data from the nodes whose energy is below the threshold value. Navigation of mobile robots to gather the data from partitioned nodes usually achieved by time and location based strategies. In projected hybrid scheduling, the navigation of mobile robots scheduled by both the combination of time and location based approaches with various region scheduling. In large network scenario, the mobile robot gets more burdens due to its extra responsibilities to visit all partitioned nodes. So the entire scenario is divided in to different regions and the deployment of multiple mobile robots is relayed on the requirements. Therefore, the efficiency of sensed data composed by the base station or sink node from partitioned/islanded WSN is improved doubly using multiple mobile robots. Through simulation below the environment of NS-2 simulator, the results from various aspects show that proposed multiple mobile robots can develop the performance of collecting the sensed data in large-scale sensing fields and also it improves the lifetime of the sensor nodes.

**M. Vijayalakshmi et al.in 2013[2]** wireless Sensor Network (WSN) is an emerging technology. WSNs usually consist of a large number of small sensor nodes with limited onboard energy supply and deployed densely in a given region for information harvesting purposes. Since the sensor devices have limited memory and power capacity, the power consumption in WSN becomes as a major issue nowadays. So that, in the proposed framework, a scheme to reduce the power consumption in WSN is introduced. Proposed framework is clustering based. Clustering and Prediction techniques, which use temporal correlation among the sensor data, supply a chance for reducing the energy consumption of continuous sensor data collection. Thus it can attain stability and prolongs network lifetime. An adaptive scheme is presented which is used to control prediction, analyze the performance trade off between reducing communication cost and prediction cost, and design algorithms to take the benefit of adaptive scheme to

enable/disable prediction operations. Localized prediction scheme is performed which takes advantages over the previous dual-prediction scheme to minimize communication and computation cost thereby reducing the energy consumption. Sleep/awake scheduling can be applied. A practical algorithm designed for data aggregation will use faster and more efficient cluster-to-cluster propagation.

**Sherin Mathew et al.in 2013[3]** lately, there has been a rapid growth in the wireless communication technique. Mobile sinks can be mounted upon urban vehicles with fixed trajectories supply the ideal infrastructure to effectively retrieve sensory data from such isolated WSN fields. Presented approach uses either single-hop transfer of data from SNs that lie within the MS's range or intense involvement of network periphery nodes in data retrieval, processing, buffering, and delivering tasks. Our projected protocol aims at reducing the overall network overhead and energy expenditure associated with the multihop data recovery process while also ensuring balanced energy consumption among SNs and prolonged network lifetime. This is attained through building cluster structures consisted of member nodes that route their measured data to their assigned cluster head (CH). CHs execute data filtering upon raw data exploiting potential spatial-temporal data idleness and frontward the filtered information to appropriate finish nodes with sufficient residual energy, located in proximity to the MS's trajectory. Simulation results specify the superior performance of our proposed algorithm to strike the appropriate performance in the power consumption and network lifetime for the wireless sensor networks.

**Laxmi Choudhary et.al. 2012[4]** with the rapid improvement of computer and information technology in the last many years, an enormous amount of facts in science and engineering has been and will continuously be produced in massive scale, either being stored in gigantic storage strategy or flowing into and out of the system in the type of data streams. Moreover, such data has been made widely available, e.g., via the Internet. Such tremendous quantity of data, in the array of tera- to peta-bytes, has fundamentally changed science and engineering, transforming several disciplines from data-poor to increasingly data-rich, and calling for new, data-intensive methods to conduct research in science and engineering. In this paper, author discuss the research challenges in science and engineering, from the data mining perspective, with a center on the following issues: (1) information network analysis, (2) discovery, custom, and understanding of patterns and knowledge, (3) stream data mining, (4) mining moving object data, RFID data, and details from sensor networks, (5) spatiotemporal and multimedia data mining, (6) mining text, Web, and other unstructured data, (7) facts cube-oriented multidimensional online analytical mining, (8) visual data mining, and (9) data mining by combination of sophisticated scientific and engineering domain knowledge.

**S. Nithyakalyani et al.in 2012[5]** Nowadays sensors are very essential for today life to monitor environment where human cannot get involved very a lot. Wireless Sensor Networks (WSN) are used in many real world applications like environmental monitoring, traffic control, trajectory monitoring. It is more demanding for sensor network to sense and collect a large amount of data, which are continuous over time, which in turn must to be forwarded to sink for further decision-making process. Clustering of sensory data act as a center job of data mining. A clustering in WSN involves selecting cluster heads and assigning cluster members (sensors) to it for proficient data relay. The constraints in power supply, limited communication, bandwidth, and storage resources are the main challenges in WSN facing today.

**Rouhollah Maghsoudi et al.in 2011[6]** Data mining knowledge in reaction to technological advances in various Rmynh, base arena is built there. Data Mining face a different situation that the data size is large and want to make a small model and not too complicated and yet the data as well as describe. The Requisite is to use data analysis to reduce the amount and the huge volume of information. One important and sensible facts in the world of machine intelligence and is robotics robots routing. Robot router has obstacle recognition and how to deal with the decision with barrier. For routing, algorithms including probabilistic methods (filtering particulate), evolutionary algorithms like as genetic, ants social and optimization particle mass, neural methods - Fuzzy, inequality of matrix method relayed on gradient methods combined sensor information, etc. There are data mining schemes in the years 2010-2008 as a technique for routing and a absolute robot has been used and still is in progress. Overview of the methods in the paper mentioned in diverse articles since 2000 has so far. Although many data mining methods comprise, but mentioned in this article with precise literature data mining will deal with the routing problem.

**Onur Tekdas et al.in 2009[7]** explore synergies among mobile robots and wireless sensor networks in environmental monitoring during a system in which robotic data mules gather measurements gathered by sensing nodes. A proof-of-concept implementation

demonstrates that this approach considerably increases the lifetime of the system by conserving energy that the sensing nodes otherwise would use for communiqué.

**R.Sivaranjini et al.in 2013[8]** Nowadays Wireless sensor networks playing vital role in all are Which is used to sense the environmental monitoring, Temperature, Soil erosion etc. Low data delivery competence and high-energy consumption are the inherent problems in Wireless Sensor Networks. Finding precise data is more difficult and also it will leads to added expensive to collect all sensor readings. Clustering and prediction techniques, which develop spatial and temporal correlation among the sensor data, provide opportunities for reducing the power consumption of continuous sensor data collection and to achieve network energy efficiency and stability. So as recommend Dynamic scheme for energy consumption and data collection in wireless sensor networks by combining adaptively enabling/disabling forecast scheme, sleep/awake method with dynamic scheme. Our framework is clustering based. A cluster head demonstrate all sensor nodes inside the region and collects data values from them. Our framework is common enough to incorporate many advanced features and show how sleep/awake scheduling can be applied, which takes our scaffold approach to designing a practical dynamic algorithm for data aggregation, it avoids the need for rampant node-to-node broadcast of aggregates, but rather it uses faster and more efficient cluster-to-cluster propagation.

**Miao Zhao et al.in 2012[9]** Recent study reveals that great benefit can be achieved for data gathering in wireless sensor networks by implying mobile collectors that gather data by short-range communications. To pursue maximum power saving at sensor nodes, intuitively, a mobile collector should traverse the transmission series of each sensor in the field such that each data sachet can be straight transmitted to the mobile collector without any relay. However, this approach may lead to considerably increased data gathering latency owed to the low moving velocity of the mobile collector. Fortunately, it is pragmatic that performing proper local aggregation via multihop transfers and then uploading the aggregated data to the mobile collector can effectively shorten data gathering latency. In such a scheme, the number of local transmission hops should not be randomly large as it may increase the energy consumption on packet relays, which would adversely affect the overall competence of mobile data gathering. Based on these observations, in this paper, author study the tradeoff between energy saving and data gathering latency in mobile data gathering by exploring a balance among the relay hop count of local data aggregation and the moving tour length of the mobile collector. Author first suggests a polling-based mobile gathering approach and formulates it into an optimization problem, named bounded relay hop mobile data gathering (BRH-MDG). Particularly, a subset of sensors will be selected as polling points that buffer locally aggregated facts and upload the data to the mobile collector when it arrives. Author then give two efficient algorithms for selecting polling points among sensors. The effectiveness of our approach is validated during extensive simulations.

**Khushboo Sharma, et al.2012[11]** in "Nearest Neighbor Classification For Wireless Sensor Data" Has used a Nearest Neighbor Classification technique to classify the data expected by sensor nodes in WSNs. This technique solves the problem of data irregularity, which occurs due to the refusal of sensed data composed by partitioned Wireless Sensor Networks and the unrelated data composed by the sensor nodes of Wireless Sensor Networks. Furthermore, a Nearest Neighbor Trajectory is used for the classification of data sensed by sensor nodes. Moreover, nearest neighbor classifier procedures the patterns only, when the demand is made to categorize a query vector in WSN. Hence, the technique works in two parts: I) predicts the classes of sensor data and II) nearest neighbor trajectory begin the training of sensor data and create a model with a finest representation.

**Cheng Chen, et al. 2011[12]** in "Overview On Data Collection Using Mobile Robots In Wireless Sensor Network" in 2011 have projected a data collecting algorithm that conducts the mobile robot to gather the sensed data from the partitioned Wireless Sensor Networks which previous used to be redundant by the base station as it could not be proficient to establish the partitioned Wireless Sensor Networks. Furthermore, due to the refusal of the sensed data of partitioned Wireless Sensor Networks by base station, troubles like data discrepancy comes into the play. Hence, to determine such issues, authors have projected two approaches for the direction finding of island Wireless Sensor Networks: I) Local based approach and II) Global based approach. Moreover, authors have also précised scheduling techniques that are essential to program the navigational system of mobile robot. Hence, the scheduling techniques are based on three things: I) time, II) location and III) dynamic moving based. Furthermore, Ns-2 simulator has been used to create the scenarios and approach with advancement.

**Bo Yuan, Member, IEEE,et.al.2007[13]** in "Optimal Robot Routing Problem In Wireless Sensor Networks" .in

2007 have proposed how the mobile robots can be used to visit given sensor nodes dispersed in Euclidean plane in order to collect data from that sensor nodes. Moreover, the authors have projected a method which can be used to solve NP-hard problem which is a unique case of Travelling Salesman Problem and Neighbourhood which in turn helps in minimizing the total travelling distance enclosed by the mobile robot while assembling the sensed data in WSN.

**Marcelo B. Soares, et.al.2007[14]** in "Hybrid Mobile Robot Navigational Strategy For Efficient Data Collection In WSN" in 2007 have proposed the Hybrid Mobile Robot Navigational approach, that is used to notice the correct location of partitioned Wireless Sensor Networks in order to conduct the mobile robot to accumulate the sensed data from that approach. Moreover, this hybrid mobile robot navigational approach works at two different layers: I) Reactive Layer and II) Planning Layer. Furthermore, the reactive layers help in assembling the representation of composed data and planning layer offers the strategy for the mobile robot in WSN. Hence, it can accumulate the data from partitioned Wireless Sensor Networks without any obstruction [14].

Table 1: Literature Survey

| Ref No | Year | Findings | Technique |
|---|---|---|---|
| 1 | 2012 | 1.Improves the lifetime of the sensor nodes 2.Improves the performance of collecting the sensed data in large-scale sensing fields | Multiple mobile robots |
| 2 | 2013 | Stability and prolongs network lifetime | A practical algorithm designed for data aggregation will use faster and more efficient cluster-to-cluster propagation. |
| 3 | 2013 | Appropriate performance in the energy consumption and network lifetime for the wireless sensor networks. | Cluster structures consisted of member nodes that route their measured data to their assigned cluster head (CH |
| 4 | 2012 | Fundamentally changed science and engineering, transforming many disciplines from data-poor to increasingly data-rich | The research challenges in science and engineering, from the data mining perspective |
| 5 | 2012 | A clustering in WSN involves selecting cluster heads and assigning cluster members (sensors) to it for efficient data relay. | Clustering in WSN |
| 6 | 2011 | Solve the routing problems | Data mining methods |
| 7 | 2009 | Significantly increases the lifetime of the system by conserving energy that the sensing nodes otherwise would use for communication. | Synergies among mobile robots |
| 8 | 2013 | Avoids the need for rampant node-to-node propagation of aggregate | Practical dynamic algorithm for data aggregation |
| 9 | 2012 | Significantly increased data gathering latency due to the low moving velocity of the mobile collector | a polling-based mobile gathering approach and formulate it into an optimization problem, named bounded relay hop mobile data gathering (BRH-MDG). |

## 4. Conclusion

Wireless sensor network technology has the prospective to enable main breakthroughs in the natural sciences by giving scientists the potential to collect high-fidelity data over large geographic regions and extended periods of time. In WSNs, since the sensor nodes are energy constrained and have limited lifetime, energy consumption of sensor nodes becomes as a major issue. Two main approaches: 1) clustering-based: sensor nodes form clusters and elect the cluster heads in such a way to improve energy efficiency, and 2) prediction based: energy-aware prediction is used to find the slight trade-off

among communication and prediction cost. Via performance evaluation, it is shown that it achieves energy efficiency even though the object arrived from any random location and moves randomly. K-Means Data Relay (K-MDR) clustering algorithm for WSN reduces the communication overhead and increases the entire network life time by reducing the number of transmission between every sensor node to sink. The K-Means Data Relay algorithm decreases the computational time and improves the performance of the network when compared to K-Means algorithm. Hybrid moving based scheduling strategy for data collection process is improved lifetime of the network. For large coverage area and more number of nodes failure in the network, multiple mobile robots used for separate partition in order to collect data.

**Acknowledgments**

I would like to gratefully and sincerely thank to The Dean of our College, Chairman of our Department for his guidance, understanding, patience, and most importantly, his friendly nature during this research paper.I would also like to thank my friends and colleagues, my supervisor and the university who provided me an efficient support to work on this atmosphere and good infrastructure. I would also like to thank to all the previous researchers who worked very hard and helped others to comprehend the subject of Wireless Sensor Networks System.

## References

[1] V. Karthik,"Region Based Scheduling With Multiple Mobile Robots for Data Collection Strategies in Wireless Sensor Networks", Vol. 2, Issue 7, July 2013

[2] M. Vijayalakshmi , V. Vanitha,"CLUSTER BASED ADAPTIVE PREDICTION SCHEME FOR ENERGY EFFICIENCY IN WIRELESS SENSOR NETWORKS", Vol 04; Special Issue; June 2013

[3] Sherin Mathew, S. Nandhini," A Novel Approach for Sensory Data Collection in Wireless Sensor Networks with Mobile Sinks", Volume 2, Issue 2, March 2013

[4] Laxmi Choudhary," CHALLENGES FOR DATA MINING", Volume 2, Issue 2 (February 2012)

[5] S. Nithyakalyani and S. Suresh Kumar,"Data Relay Clustering Algorithm for Wireless Sensor Networks: A Data Mining Approach",2012

[6] Rouhollah Maghsoudi,, Somayye Hoseini , Yaghub Heidari," Surveying Robot Routing Algorithms with Data Mining Approach", Vol .2 No.2 (2011) 284-294

[7] Onkur Tekdar and Volkan Isler University of Minnesota Jong Hyun Lim and Andreas Terzis Johns Hopkins University,"Using Mobile Robots To Harvest Data From Sensor Fields",2009.

[8] R.Sivaranjini, E.Dinesh,"An Adaptive scheme for energy consumption and data collection in Wireless Sensor Networks", Volume 2 Issue 3  March. 2013  PP.50-55

[9] Miao Zhao, Member, IEEE, and Yuanyuan Yang, Fellow, IEEE," Bounded Relay Hop Mobile Data Gathering in Wireless Sensor Networks", Vol. 61, No. 2, Febuary 2012

[10] Doug Alexander,Data Mining," http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/"

[11] Khushboo Sharma, Manisha Rajpoot, Lokesh Kumar Sharma, Department of Computer Science and Engineering, Rungta College of Engineering and Technology, Kohka Road, Kurud, Bhilai, India.

[12] Tzung-Cheng Chen, Tzung-Shi Chen, Member, IEEE, and Ping-Wen Wu.

[13] Bo Yuan, Member, IEEE, Maria Orlowska, and Shazia Sadiq, School of Information Technology and Electrical Engineering, The University of Queensland, QLD 4072, Australia, E-mail: {boyuan, maria, shazia}@itee.uq.edu.au.

[14] Marcelo B. Soares, Mario F. M. Campos, Dimas A. Dutra,V´_ctor C. da S. Campos and Guilherme A. S. Pereira

## Author Profile:

**Mohd Muntjir** is a Research Scholar in OPJS University .He has received his B.Sc in Mathematics from Choudhary Charan Singh University Meerut U.P. in 2002 and M.C.A. degree in Computer Science from Hemawati Nandan Bahuguna Garhwal University Srinagar Garhwal Uttarakhand in 2007. His research interests are Wireless Sensor Network, Data Mining, DBMS, Cloud Computing, E-Commerce and Multimedia Technology.