EXPLORATIONS ON ENGINEERING LETTERS (EEL)
VOLUME 1, ISSUE 1 (2016):PP.173-179
SANA ACADEMIC PRESS

Big Data Gathering Concept using RSM Algorithm to Avoid Data Traffic and Battery time Conservation

Abirami S[1], Shivasankari S[2],Brindha S[3]
[1]Research Scholar, St. Peter's University, Chennai.
[2]Asst.Prof, Dept. of Computer Science & Applications, St. Peter's University, Chennai
3Asst. Prof, Dept. of Computer Science & Applications, St. Peter's University, Chennai.
rami1.techno@gmail.com

Abstract: *The era of big data has begun, and an enormous amount of real-time data is used for the risk analysis of various industrial applications. Wireless sensor networks (WSN) technology can overcome this limitation by collecting the big data generated from source nodes and transmitting them to the data center in real time. In this study, typical residence, office, and manufacturing environments were chosen. The signal transmission characteristics of WSN were obtained by analyzing the test data. According to these characteristics, a real-time big data gathering (RTBDG) algorithm is proposed along with RSM algorithm for the risk analysis of industrial operations. In this algorithm, sensor nodes can screen the data collected from the environment and equipment according to the requirements of risk analysis. Clustering data transmission structure is then established on the basis of the received signal strength indicator (RSSI) and residual energy information. Experimental results show that RTBDG and RSM not only efficiently uses the limited energy of network nodes but also balances the energy consumption of all nodes. In the near future, the algorithm will be widely applied to risk analysis in different industrial operations.*

Keywords:*Risk analysis, sensor nodes, RSM algorithm,Big data, energy consumption.*

I INTRODUCTION

Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, we must choose an alternative way to process it The first definition of Big Data comes from Merv Adrian: "Big data exceeds the reach of commonly used hardware environments and software tools to capture, manage, and process it within a tolerable elapsed time for its user population". Another good definition is given by the McKinsey Global Institute: "Big data refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse". These definitions imply that what qualifies as big data will change over time as technology advances. What was historically big data or what is big data today won't be big data tomorrow. This aspect of the big data definition is that some people find unsettling. The preceding definitions also imply that what constitutes big data can vary by industry, or even organization, if the tools and technologies in place vary greatly in capability. We should not be surprised that companies are tracking and analyzing our data. We rarely hear of theologians talking about what data can tell us about faith needs of parishioners. Big data is not a single technology but a combination of old and new technologies that helps companies gain actionable insight. Therefore, big data is the capability to manage a huge volume of disparate data, at the right speed, and within the right time frame to allow real-time analysis and reaction. Big data is typically broken down by three characteristics Volume - how much data, Velocity - how fast that data is processed, Variety - the various types of data. These characteristics are called the three Vs of Big Data and a number of vendors have added more Vs to their own definitions. Volume is the first thought that comes with big data: the big part. Some experts consider Petabytes the starting point of big data.

As we generate more and more data, we are sure this starting point will keep growing. However, volume in itself is not a perfect criterion of big data, as we feel that the other two Vs have a more direct impact. Velocity refers to the speed at which the data is being generated or the frequency with which it is delivered. Think of the stream of data coming from the highways‟ sensors in the Los Angeles area, or the video cameras in some airports that scan and process faces in a crowd. There is also the click stream data of popular e-commerce web sites. Variety is about all the different data and file types that are available. Just think about the music files in the iTunes store (about 28 million songs and over 30 billion downloads), or the movies in Netflix (over 75,000), the articles in the New York Times web site (more than 13 million starting from 1851), tweets (over 500 million every day), foursquare check-ins with geolocation data (over five million every day), and then you have all the different log files produced by any system that has a computer embedded. When you combine these three Vs, you will start to get a more complete picture of what Big Data is all about. Other authors explain the fourth V as Veracity. Most of Big Data comes from sources outside our control and therefore suffers from significant correctness or accuracy problems. Veracity represents both the credibility of the data source as well as the suitability of the data for the target audience. LinkedIn, Netflix, Facebook, Twitter, Expedia, national and local political campaigns, and dozens of other organizations are all generating enormous economic, social, and political value. Some examples of Big Data : social media text, cell phone locations, channel click information from set-top box, Big Data analysis 145 web browsing and search, product manuals, communications network events, Call Detail Records (CDRs), Radio Frequency Identification (RFID) tags, maps, traffic patterns, weather data, mainframe logs.

II RELATED WORKS

The concept Big data was implemented in the existing work. It proved the ability of implementing the bigdata concept in WSN network. It can also used for energy efficient data transmission. Here used this concept for monitoring the network .Critically send the collected data to the server for a fixed interval of time.Implemented only in monitoring network. Risk values cannot be plotted when it occurs.Concept fails in most of the WSN network.According to the report by ORACLE in , the conceptof big data is stimulating a wide range of industry sectors. Specific examples of big data generated by sensors were provided in the report. For instance, manufacturing companies usually embed sensors in their machinery for monitoring usage patterns, predicting maintenance problems, and enhancing the product quality. By studying the data streams generated by the sensors embedded in the machinery allow the manufacturers to improve their products. The numerous sensors deployed in the supply lines of utility providers generate a huge volume of data, which are consistently monitored for production quality, safety, maintenance, and so forth.

Other examples of sensors generating a bulk of the big data consist in electronic sensors monitoring mechanical and atmospheric conditions. In addition, sensors used for healthcare services (to monitor biometrics of the human body, patients' conditions, healthcare diagnoses, treatment phases, and so forth) are identified to be a rich source of big data in the report presented in .However, how to gather the sensed data from these numerous sensors in an energy-efficient manner remained beyond the scope of the report The work in presented a cloud-based federated framework for sensor services. The main objective of the workwas to enable seamless exchange of feeds from large numbers of heterogeneous sensors. Various applications using big data generated by densely distributed WSNs have also emerged in literature. In addition, in and , big data in terms of the healthcare information (e.g., blood pressure and heart rate) sensed by numerous sensors are used to realize remote medical care services. Furthermore, patients' location information areused to arrange prompt dispatch of ambulances.

Large volume of data gathered from location-sensors attached to animals enabled researchers to observe various animal habitats .Because widely and densely distributed WSNs collect various types of data, the overall data which are gathered is, indeed, overwhelming. To efficiently gather the big data generated by the densely distributed WSNs is, however, not an easy task since the WSNs may be divided into sub-networks because of the limited wireless communication range of the sensors.

In conventional research works, data gathering using the mobile sink in WSNs has been widely studied in literature. Data Mobile Ubiquitous LAN Extensions (MULEs) isthe one of the most prominent and earliest studies on the mobile sink scheme. Data MULEs follow the basic steps of all the mobile sink schemes. First, it divides sensor nodes into clusters. Second, it decides the route for patrolling each cluster. The work in assumes a simple data collection scheme whereby the mobile sink node divides sensor node into grids regardless of the sensor nodes' location, and patrols the grids by using random walk between the neighboring grids. However, this type of clustering, which is not based on the nodes' location, might result in inefficient data gathering. If there is no sensor node remaining in the cluster, patrolling the empty cluster results in waste of time and degraded efficiency. Also, patrolling based on randomness might result in unbalanced visits to clusters with different numbers of sensor nodes.

Earlier research works on sensor node clustering algorithms demonstrates that the increasing number of clusters reduces energy consumption for data transmission. Certainly, the idea holds since increasing the number of clusters decreases the cluster-sizes and shortens the transmission length. Some researchers consider that certain limitations on the number of cluster can be decided by other factors. For example, in , the limitation is the maximum acceptable latency of data collection. The authors of also defined the limitation by a node's buffer size. While these limitations are realisticassumptions, they do not considerthe energy consumption fordata requests. In our paper, we first focus on the effect of data request messages by increasing the number of clusters. Based on a simple and common data gathering model of the densely distributed WSNs, we demonstrate that the number of request messages has a noticeable impact on the energyconsumption of the sensor nodesconsumption of the sensor nodes.messages has a noticeable impact on the energy consumption of the sensor nodes.

III PRELIMINARIES

In this section, here introduce essential preliminary concepts, such as system model, and an overview of the big data gathering concept,big data implementation.

A. SYSTEM MODEL

In order to gather these data, the Wireless Sensor Networks (WSNs) are constructedwhereby the sensors relay their data to the "sink". However,in case of widely and densely distributed WSNs (e.g. inschools, urban areas, mountains, and so forth) thereare two problems in gathering the data sensed by millions ofsensors. First, the network is divided to some sub-networksbecause of the limited wireless communication range. Forexample, sensors deployed in a building may not be ableto communicate with the sensors which are distributed inthe neighboring buildings. Therefore, limited communicationrange may pose a challenge for data collection from all sensornodes. Second, the wireless transmission consumes the energyof the sensors. Even though the volume of data generated byan individual sensor is not significant, each sensor requiresa lot of energy to relay the data generated by surroundingsensors. Especially in dense WSNs, the life time of sensorswill be very short because each sensor node relays a lot ofdata generated by tremendous number of surrounding sensors.In order to solve these problems, we need an energyefficientmethod to gather huge volume of data from a large numberof sensors in the densely distributed WSNs.

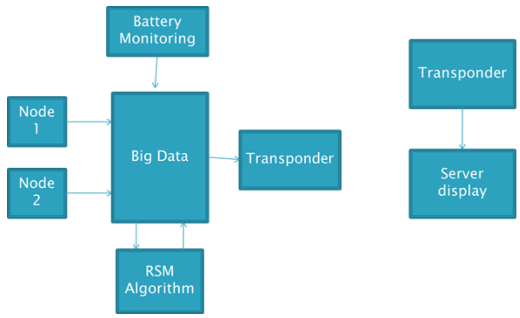

FIG 1: System Model

To achieve energy-efficient data collection in densely distributed WSNs, there have been many existing approaches.For example, the data compression technology is capableof shrinking the volume of the transmitted data. Although it is easy to be implemented, the data compression technologyrequires the nodes to be equipped with a big volumeof storage and high computational power. In addition, thetopology control technology can evaluate the best logicaltopology and reduce redundant wireless transmissions. When the redundant wireless transmissions are reduced, therequired energy for wireless transmissions can be also reduced.Furthermore flow control and routing can choose the pathwhich consists of nodes having high remaining energy. However, these technologies are not able to deal withthe divided networks problem

BIG DATA: Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data creation, search, sharing, storage, transfer, visualization, querying and information privacy.

SENSOR NODES: Wireless sensor networks (WSN), sometimes called wireless sensor and actuator networks (WSAN), are 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 a main location.

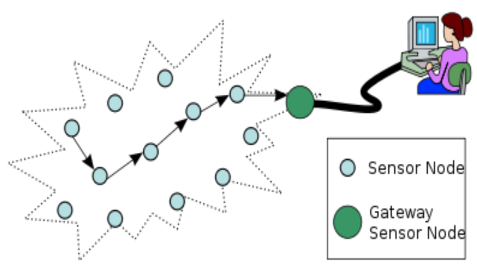

FIG 2: Sensor Nodes

TRANSPONDER:A transponder gathers signals over a range of uplink frequencies and re-transmits them on a different set of downlink frequencies to receivers on Earth, often without changing the content of the received signal or signals.

BATTERY MONITERING: Battery management system (BMS) is any electronic system that manages a rechargeable battery (cell or battery pack), such as by protecting the battery from operating outside its Safe Operating Area, monitoring its state, calculating secondary data, reporting that data, controlling its environment, authenticating it and / or balancing it.A battery pack built together with a battery management system with an external communication data.

B. CONSIDERED NETWORK MODEL

In this paper, we consider a network which consists of a mobile sink and many sensor nodes spread within a limitedfield. Every sensor node knows its location by using localization technology, and the mobile sink knows all nodes locations. Regardless of being a sink or the sensor, node has alimited communication rang Rand communication is always successful if it is within R. The mobile sink node patrolsthe cluster centroids that arecalculated to minimize energy consumption for data transmission, and collects data fromsensor nodes. Sensor nodes are equipped with a buffer memory and store sensed information until mobile sink approaches centroid. The information is transferred to the sink node by multi-hop fashion. In this paper, we assume a densely distributed WSN in a large area such as schools, urban areas ,mountains, and so forth and thus WSNs are divided into subnetworks .N sensor nodes.

C.Data Gathering Process Using Cluster Technique

After clustering, the mobile sink patrols every cluster centroid and collects the data from the nodes in the cluster. It is easy to see that delay is a main problem of using mobile sink in WSNs. This delay is the waiting time from data generation to data sending. Because the mobile sink moves relatively slowcompared with electrical communication between nodes, the mobile sink scheme causes long delay. To shorten this delay, we need to minimize total patrolling path length. Thus, inour scheme the mobile sink patrols along Traveling Salesman Problem (TSP) path of all cluster centroids. Once the mobile sink arrives at the cluster centroids, Itcollects data from sensor nodes. Directed Diffusion is one of the most famous data collection schemes in WSNs.In our method, we consider using a typical example of them ,i.e., "One Phase Pull where the mobile sink node sends data request message at the cluster centroids. When a sensor node receives a data request message from cluster k, the nod re-broadcasts the data request message and replies data to the neighboring node, which is the parent node in the data request tree of cluster k. Then, the node relays data messages to the sink.

IV IMPLEMENTATION

A. ALGORITHM

RTBDG ALGORITHM: The RTBDG algorithm proposed in this paper is designed togather big data used in the risk analysis of industrial operations.Therefore, sensor nodes should screen the gathered dataaccording to the requirements of risk analysis. The screeningprocess can be described as follows:1) The normal reference ranges of the data collected bysensor nodes are established.2) Data are collected at a regular time interval t1 .3) The data collected by sensor nodes with a normalreference value are compared. If these data are within thenormal reference ranges, the data are stored at a regular timeinterval t2 (t2 □ n □t1 , n is a natural number and can be setaccording to the requirements of risk analysis). Otherwise, boththe abnormal and stored data are transmitted to BS by theestablished routes.4) If the amount of data stored in sensor nodes reaches thepreset upper limit value of the storage capacity, all of the storeddata are also transmitted to BS by the established.

B. BIG DATA IMPLEMENTATION

Big Data implementation :There are several ways to store larger amounts of data . There is singular method to deploy a business intelligence solution to answer unique company questions, but there is an approach to take advantage of Big Data which minimizes risk and increases the likelihood of a successful outcome. Big Data projects are difficult and need know-how and experience to be successful. The implementation method of Big Data consists of 8 (or 9) steps1. begin with stakeholders and consider culture, 2. find data stewards, 3. set clear goals, 4. create the plan, 5. select the right strategy and tools, 6. establish metrics, 7.deploy the technology, 8. make big data little [CRMSearch. The Business Case for Big Data, data implementation.php], 9. design for Continuous Process Improvement (CPI).

C.Signal Transmission Characteristics

The signal transmission characteristics of indoor WSN are summarized as follows:
1) The path loss of nodes near a wall is generally greater than that away from the wall. 2) The path loss of nodes near floors and ceilings is generally greater than that away from floors and ceilings. 3) The path loss of non-line-of-sight nodes is generally greater than line-of-sight nodes.4) Two sensor nodes with a fixed distance apart generate different path losses when they are in different indoor environments such as different indoor materials, layout, and sizes, or in different locations of the same room. 5) Path loss increases with the increasing distances between nodes in a constant environment 6) Path loss between nodes significantly increases when the obstacles are added between nodes. 7) Path loss experiences significant fluctuation when a person is walking back and forth between the nodes but is close to the path loss in the constant environment after it is filtered. 8) If obstacles exist or a person is moving between nodes, the communication between the nodes can be interrupted .

D.RISK ANALYSIS EVALUATION

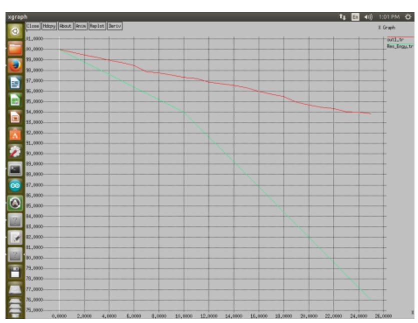

FIG 3: Residual Energy

1) Path loss increases with the increasing distances between
nodes in a constant environment. 2) Path loss will have a great fluctuation when a person is walking back and forth between nodes, but it is close to that in the constant environment after it is filtered. 3) If a person moves between nodes, the communication between nodes can be interrupted, or data packets can be lost when the path loss is large enough. 4) If a person walks back and forth between nodes, the path loss will have a decreasing fluctuation with increasing distances between nodes.

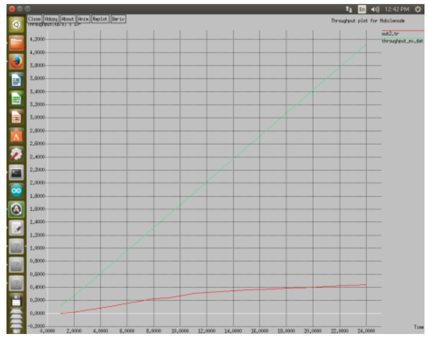

FIG 4: Throughput

V.CONCLUSION.

In Wireless Sensor Network all nodes are powered using battery source. Battery time conservation is very important in increasing the life time of the node. Continuous data transmission will result in fast draining of battery charge. In order to overcome this process here introduce a concept called big data.In this the data transmitted by the nodes are filtered. Based on the risk analysis only the value changes will be transmitted to the router or receiver end.Here propose a intelligent node which monitors all the sensor values Also monitor the level of battery and adjust the sampling rate for the sensors. Then it will monitor the values of sensor and make the decision whether the value is risk or not.If the value is found as risk the data will be transmitted to the server In other cause in order to pre request by the server data can be transmitted. Here we use RSM algorithm to support this conditions.Most of the data analysis was done by control unit in the node itself. So data transmission to server for analysis is avoided. Energy efficiency is achieved by big data analysis. And failure avoidance is maintained using RSM algorithm.

REFERENCES

[1]. I. Bisio and M. Marchese, “Efficient satellite-based sensor networks for information retrieval,” IEEE Systems Journal, vol. 2, no. 4, pp. 464–475, Dec. 2008.

[2]. I. Bisio, M. Cello, M. Davoil, and et al, “A survey of architectures and scenarios in satellite-based wireless sensor networks: System design aspects,” International Journal of Satellite Communications and Networking (IJSC), vol. 30, no. 6, 2012.

[3]. S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, “XORs in the air: Practical wireless network coding,” IEEE/ACM Transactions on Networking, vol. 16, no. 3, pp. 497–510, Jun. 2008.

[4]. K. Miyao, H. Nakayama, N. Ansari, and N. Kato, “LTRT: An efficient and reliable topology control algorithm for ad-hoc networks,” IEEE Transactions on Wireless Communications, vol. 8, no. 12, pp. 6050– 6058, Dec. 2009.

[5]. J. Zhao, Y. Qin, D. Yang, J. Duan, “Reliable graph routing in industrial wireless sensor networks,” International Journal of Distributed Sensor Networks, vol. 2013, Article ID 758217, 15 pages, 2013.

[6]. P. T. A. Quang, D. S. Kim, “Enhancing real-time delivery of gradient routing for industrial wireless sensor networks,” IEEE Transactions on Industrial Informatics, vol. 8, no. 1, pp. 61-68, 2012.

[7]. A.Tiab, L. Bouallouche-Medjkoune, “Routing in industrial wireless sensor networks: a survey,” Chinese Journal of Engineering, vol. 2014, Article ID 579874, 7 pages, 2014..

[13] Compression format for IPv6 datagrams over IEEE 802.15.4-based networks.J. Hui and P. Thubert , 2015 IEEE 802.15.4-based networks.