International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 3 | ISSN : 2456-3307

A Study on Big Data Analysis on WSN for Different Data

Lovepreet Singh, Er. Dinesh Kumar

Giani Zail Singh Campus College of Engineering & Technology, Bathinda, Punjab, India

ABSTRACT

Big data represent a new era in data exploration and utilization and covers various industrial applications, for example, professionalizing business intelligence operation in the automobile industry, solving routing and scheduling problems in transportation systems , improving the performance of supply chains by minimizing the negative effect of demand uncertainties , providing security for buildings and physical infrastructure in home surveillance and security systems. In this paper the different problems are studied from the previous work these problems are like routing and scheduling problems are occurred during the processing of big data on WSN. The most important problem in the design of the routing algorithm of WSN is the maximization of the network lifetime and balancing the energy consumption of sensor nodes. In the future we propose a Real-Time Big Data Gathering Algorithm with high Energy Efficient Distributed (HEED) network.

Keywords: HEED, Big Data, WSN , LEACH

I. INTRODUCTION

Big data represent a new era in data exploration and utilization and covers various industrial applications, for example, professionalizing business intelligence operation in the automobile industry, solving routing and scheduling problems in transportation systems , improving the performance of supply chains by minimizing the negative effect of demand uncertainties , providing security for buildings and physical infrastructure in home surveillance and security systems [1], and analyzing supply chains with radio-frequency identification technology from both the risk and benefit perspectives . Big data can also be used to analyze risks in industrial operations, particularly in product manufacturing. Many manufacturing enterprises have strict requirements on equipment working conditions and environment conditions for high quality products, such as chip fabrication plants, pharmaceutical factories, and food factories.

In the product manufacturing process of the aforementioned enterprises, the working condition parameters of manufacturing equipment and environment data need to be gathered in real time. Abnormal information is extracted from these data for the risk analysis of product manufacturing to ensure the normal operation of a production system [2]. However, a technical challenge exists in gathering real-time big data in various environments. This limitation on real-time data collection can be overcome by wireless sensor networks (WSN).

WSN has become an important technological support for gathering big data, such as temperature, humidity, equipment working condition, health information, and electricity consumption, particularly for data collection and transmission in indoor environments. Real-time data can be gathered by using smart sensors, including atmospheric sensors, thermometric sensors, humidity sensors, and accelerometers. The volume of data gathered by these sensors may reach the order of petabytes according to a report of ORCALE [3]. In the industrial field, big data denote the enormous volume of various real-time data that are gathered, managed, processed, and analyzed for industrial operations.

However, a major challenge for WSN is ensuring that real-time data can be transmitted to the data center. Sensor nodes require enough energy to relay the data gathered by many surrounding sensors [4]. Therefore, energy is one of the most important indicators in WSN and energy consumption should be managed well to maximize network lifetime. To solve the aforementioned problems, an energy-efficient routing algorithm for WSN needs to be designed to gather big data in real time.

Many routing algorithms for WSN have been reported to prolong network lifetime. Some routing algorithms are proposed for the universal environment. A centralized clustering algorithm called the low-energy adaptive clustering hierarchy (LEACH). The disadvantage of centralized algorithms is that each sensor node must transmit its location and residual energy information to a base station (BS)[5].

What is big data?

Big data refers to the dynamic, large and disparate volumes of data being created by people, tools and machines; it requires new, innovative and scalable technology to collect, host and analytically process the vast amount of data gathered in order to derive real-time business insights that relate to consumers, risk, profit, performance, productivity management and enhanced shareholder value[1-2].

Big data includes information garnered from social media, data from internet-enabled devices (including smartphones and tablets), machine data, video and voice recordings, and the continued preservation and logging of structured and unstructured data. It is typically characterized by the four "V's":

- Volume: the amount of data being created is vast compared to traditional data sources
- Variety: data comes from different sources and is being created by machines as well as people
- Velocity: data is being generated extremely fast — a process that never stops, even while we sleep[2]
- Veracity: big data is sourced from many different places, as a result you need to test the veracity/quality of the data Evolving technology has brought data analysis out of IT backrooms, and extended the potential of using data-driven results into every facet of an organization.

However, while advances in software and hardware have enabled the age of big data, technology is not the only consideration. Companies need to take a holistic view that recognizes that success is built upon the integration of people, process, technology and data; this means being able to incorporate data into their business routines, their strategy and their daily operations [4]. Organizations must understand what insights they need in order to make good strategic and operational decisions. The first part of the challenge is sorting through all of the available data to identify trends and correlations that will drive beneficial changes in business behavior. The next step is enriching this organizational information with that from sources outside the enterprise; this will include familiar big data sources, such as those created and stored online. In a business environment that constantly and rapidly changes, future prediction becomes more important than the simple visualization of historical or current perspectives[6]. For effective future prediction, data analysis using statistical and predictive modeling techniques may be applied to enhance and support the organization's business strategy. The collection and aggregation of big data, and other information from outside the enterprise, enables the business to develop their own analytic capacity and capability, which for many years has only been available to a few larger organizations.

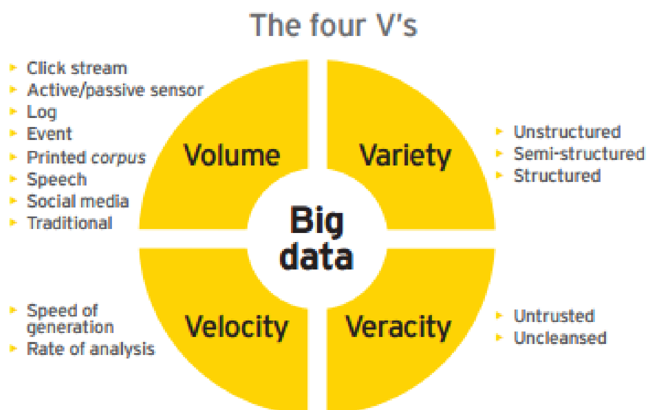

Figure 1. Big data with four V's[2]

II. BIG DATA APPLIED: IT'S A GAME CHANGER

Risk management faces new demands and challenges. In response to the crisis, regulators are requiring more detailed data and increasingly sophisticated reports. Banks are expected to conduct regular and comprehensive bottom-up stress tests for a number of scenarios across all asset classes. Recent, highly publicised 'rogue trader' and money laundering scandals have prompted further industry calls for improved risk monitoring and modeling[6]. Big Data technologies present fresh opportunities to address these challenges. Vast, comprehensive and near real-time data has the potential to improve monitoring of risk (while reducing noise-to-signal ratios), risk coverage, and the stability and predictive power of risk models. In a number of key domains - particularly operational and compliance risk - Big Data technologies will allow the development of models that will support everyday Risk Officer Decision-making. Able to process enormous amounts of data in fast timeframes, the technologies can also accommodate new requirements for scenario stress tests at the trade, counterparty and portfolio levels. The majority of benefits - and challenges - offered by Big Data stem from its massive volume and variety (fig.1.2). However, different risk domains stand to benefit from Big Data technologies in diverse ways. Big Data can be targeted to an organization's particular needs - whether they are for greater volume, variety, velocity or veracity - and strategically applied to enhance different risk domains [6].

Figure 2: Big Data stem from its massive volume and variety[4]

III. BIG DATA AND ANALYTICS

Big data poses both opportunities and challenges for businesses. In order to extract value from big data, it must be processed and analyzed in a timely manner, and the results need to be available in such a way as to be able to effect positive change or influence business decisions. The effectiveness also relies on an organization having the right combination of people, process and technology. By pure definition, analytics is the discovery and communication of meaningful patterns in data — but for business, analytics should be viewed as the extensive use of data, statistical and quantitative analysis, using explanatory and predictive models to drive fact-based business management decisions and actions. Analytics helps to optimize key processes, functions and roles[7]. It can be leveraged to aggregate both internal and external data. It enables organizations to meet stakeholder reporting demands, manage massive data volumes, create market advantages, manage risk, improve controls and, ultimately, enhance organizational performance by turning information into intelligence. Analytics can identify innovative opportunities in key processes, functions and roles. It creates a catalyst for innovation and change — and by challenging the status quo, it can help to create new possibilities for the business and its customers. Sophisticated techniques can allow companies to discover root causes, analyze micro segments of their markets, transform processes and make accurate predictions about future events or customers' propensity to buy, churn or engage[1][3][7].

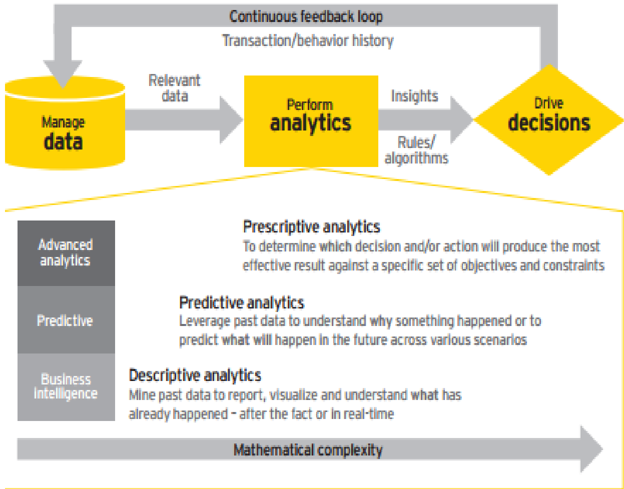

Figure 3: Big data and analytics [3]

It is no longer enough for companies to simply understand current process or operations with a view on improving what already exists, when there is now the capacity to question if a process is relevant to the business, or whether there is a new way of solving a particular issue. The key driver for innovation within organizations is to constantly challenge existing practices rather than consistently accept the same. Most organizations have complex and fragmented architecture landscapes that make the cohesive collation and dissemination of data difficult. New analytic solutions are playing an important role in enabling an effective Intelligent Enterprise (IE). An IE helps to create a single view across your organization by utilizing a combination of standard reporting and data visualization [6]:

- Data from multiple source systems is cleansed, normalized and collated
- External feeds can be gathered from the latest research, best practice guidelines, benchmarks and other online repositories
- Use of enhanced visualization techniques, benchmarking indexes and dashboards can inform management and consumers via smartphones, laptops, tablets, etc., in-house or remotely All companies need to start thinking about collecting and using relevant big data.

Data-driven decisions can reduce inefficiency between the business, legal and IT, optimize existing information assets and address disconnects between different functions of an organization. However, it is worth noting that the best data and the most advanced analytical tools and techniques mean nothing if they are not being leveraged by people who are asking the right questions. Big data, emerging storage technology platforms and the latest analytical algorithms are enablers to business success — not a guarantee of it[8].

IV. BUSINESS INTELLIGENCE AND DATA MINING

A. Business Intelligence and the Enabling Technologies

The term BI has been in existence for a long time. The oldest relevant study of BI is probably [1], which defines BI systems as, in simple terms, automatic data retrieving and processing systems that can help make intelligent decisions based on various data sources. Another even simpler definition of BI systems: "getting data in and getting data out." Such intelligent systems are highly related to the later development on decision support systems (DSSs) in the 1970s [1]. Companies can make a careful use of such systems to enhance operations decisions making. The concept is briefly illustrated in Fig. 3. Some famous industrial applications of BI can be found in the airline industry for revenue management; in the automobile industry by standardizing the processes which led to reduction in various costs ; routing problems in transportation networks ; minimizing the impact of uncertainties in supply chain systems [2]. All these applications assume data are accessible on a real-time basis, which is a technological challenge in BI system designs. This limitation on real-time data collection has been improved by advances in IT, particularly the wireless sensor technology such as RFID [3].

This enables traditional BI systems to migrate to pervasive BI systems. RFID is considered as the "the most exciting and fastest-growing technology in terms of scope of application in the next generation of BI" . On the one hand, the RFID technology is a superb channel for coordination in industrial systems especially at item-level [3]. Therefore, operational aspects such as risk assessment and inventory management are potential improvement areas with the collected data. On the other hand, the item level applications are in fact technologically constrained so the contribution of this type of technology is limited from the BI systems' point of view. Nevertheless, this and similar sensor networks can form a vital part of the overall BI system [3].

Recent development on the IoTs also facilitates handling of an immense size of real-time dataset. With a proper middleware, objects with RFID devices attached can be considered as IoTs [4]. The major distinction between IoTs and other traditional sensor networks (including the RFID network) is that they are Internet-enabled which means the objects are configured in an Internet environment. In other words, such interactions would improve the ability to control activities of the objects by adding "intelligence" via better communication. Despite the fact that it is still an infant technology, applications can already be found in environmental monitoring, inventory management, food supply chains, transportation, and so on [9]. IoT-based and also RFID-based networks, however, are subject to an intrinsic constraint, which is created by the heterogeneous nature of sensors [4], especially when the networks need to handle large datasets. It is therefore not surprising that some relevant studies linked to the agent technology have been proposed in order to resolve this issue. With a proper system design, I_oT_s can be integrated to BI systems for many industrial applications.

With the real-time information being collected, risk of management, planning, and control activities can definitely be reduced. Another highly related development is the cloud service. This is also an emerging technology which can be used to store data in

remote locations. It is not a co-incidence that both big data and cloud technology emerge almost concurrently, as they do share some similar characteristics. The cloud service provides a channel to store and process a lot of datasets, which are originated from various locations and then data analytics such as data mining, clustering, and so on, can take place somewhere else without the physical connection with the data collection sites. In addition, a cloud service could even provide a platform to reduce computation effects for the above data analytics subject to the cloud infrastructure. The Apache Hadoop ecosystem is a nice example that many companies have adopted in their cloud systems. In other words, the cloud can act as more than a data warehouse to improve business agility via the Hadoop architecture. Successful deployment of the technology can be found in the telecommunication industry and smart grid [10], for example. The above development on cloud services is again a consequence of the evolution of the Internet so subsequent discussion will focus on higher level Web-based BI systems than just cloud services. Web-based BI systems have their obvious advantages: cheaper cost, shorter time to collect data, and a single platform for data collection and a master database may be achievable. Nevertheless, such service adds concerns on security issues. Although nowadays virtually no information systems or technologies can exempt from risk and vulnerability, what is worrying here is that cloud specific vulnerability is not clearly known and thus not well managed [4-3]. In addition, mastering one piece of data always has its challenges [9]. In other words, attention should be paid in designing Web-based BI systems.
The key enabling technologies to BI. They are, however, never the synonyms to BI, and can only be part of the data collection platforms in any BI systems. In addition, the development is limited to a high level system design, which on its own is not a real limitation but the implication is that we are still far from real-life mass-scale BI applications, not mentioning if that is achievable or not.

V. LITERATURE SURVEY

Kan Yu et.al. [2014] have studied a Wireless technologies have been increasingly applied in industrial automation systems due to flexible installation, mobility, and cost reduction. Unlike traditional wireless sensor networks (WSNs), industrial wireless sensor networks (IWSNs), when expanding from wireless monitoring to wireless control, have more stringent requirements on reliability, real-time performance, and robustness in a number of industrial applications. In this paper, they explained the primary challenges of designing appropriate routing protocols and present a reliable real-time flooding-based routing protocol for IWSNs (REALFLOW). Instead of traditional routing tables, related node lists are generated in a simple distributed manner, serving for packet forwarding. Performance evaluations via simulations verify that significant improvements of reliability, real-time performance, and network recovery time can be achieved by REALFLOW, compared with traditional routing protocols.[1]

Abdullah I. Alhasanat et.al. [2015] have studied Energy consumption is an essential concern to Wireless Sensor Networks (WSNs).The major cause of the energy consumption in WSNs is due to the data aggregation. An effective way to perform such a task is accomplished by using clustering. In clustering, nodes are grouped into clusters where a number of nodes, called cluster heads, are responsible for gathering data from other nodes, aggregate them and transmit them to the Base Station (BS). In this paper they produced a new algorithm which focused on reducing the transmission bath between sensor nodes and cluster heads. A proper utilization and reserving of the available power resources is achieved with this technique compared to the well-known LEACH_C algorithm.[2]

Ms.R.S.Mahathi et.al.[2016] have studied the sensor nodes energy plays a major role in Wireless sensor networks. In most of the cases the failure of a sensor node occurs due to lack of energy thus the lifetime of a nodes are limited. Hence, the energy consumption should be managed well to maximize the network lifetime. There are various number of protocols and algorithms were proposed to decrease the consumption of energy. Here, a real-time big data gathering algorithm (RTBDG) is used for gathering the data in real-time. In RTBDG algorithm, sensor nodes can screen collected data from the sensors and then clustering data transmission structure is established based on the received signal strength indicator(RSSI)and residual energy information.

Panneer Selvam G. et.al.[2016] have studied the reduction of battery draining in the wireless sensor network using big data technique, and avoiding the failure of nodes in wireless sensor network, also to avoid repeated transmission and energy wastage in network, data collision through the repeated transmission. 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 pre-request and transmitted to the server.

Jie Cheng et.al.[2016] have studied Data gathering in sensor networks is required to be efficient, adaptable and robust. Recently, compressive sensing (CS) based data gathering shows promise in meeting these requirements. Existing CS-based data gathering solutions require that a transform that best sparsifies the sensor readings should be used in order to reduce the amount of data traffic in the network as much as possible. As a result, it is very likely that different transforms have to be determined for varied sensor networks, which seriously affects the adaptability of CS-based schemes. In addition, the existing schemes result in significant errors when the sampling rate of sensor data is low (equivalent to the case of high packet loss rate) because CS inherently requires that the number of measurements should exceed a certain threshold. This paper presents STCDG, an efficient data gathering scheme based on matrix completion.

Xuejun Ding et.al.[2016] have studied, typical residence, office, and manufacturing environments were chosen. The signal transmission characteristics of an indoor WSN were obtained by analyzing the test data. According to these characteristics, a real-time big data gathering (RTBDG) algorithm based on an indoor WSN is proposed 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 not only efficiently uses the limited energy of network nodes but also balances the energy consumption of all nodes.

VI. PROBLEM FORMULATION

In the research work big data processing on wireless sensor network different problems are faced that are given below:

1. There is routing and scheduling problems are occurred during the processing of big data on WSN.
2. The most important problem in the design of the routing algorithm of WSN is the maximization of the network lifetime and balancing the energy consumption of sensor nodes.
3. Transmission failures or deadline misses may result in disturbances to the process, degradation of the overall control performance, and even more serious economic losses or human safety problems.
4. The highly nonlinear multi- modal design problems and the NP-hard scheduling problems.
5. Another Problem is the data Optimization problem.

VII. OBJECTIVES OF WORK

The main objective of the big data processing is to solve the above problems . The main objectives are given below:

1. To propose a Real-Time Big Data Gathering Algorithm with high Energy Efficient Distributed (HEED) network.
2. To solve the energy problem and scheduling problems with energy and scheduling algorithm.
3. To design the WSN for big data with BS, CH and calculate the different parameters.
4. Compare the results being obtained.

VIII. METHODOLOGY

Sensor nodes are randomly distributed in the sensing field. In this network, the nodes are static and fixed. The sensor nodes sense the information and then send to the server. If the source node sends the packet, it will send through the intermediate node. The nodes are communicates only within the communication range. So, we have to find the node's communication range. The configuration parameter used for configure the nodes are tabulated as follows:

PARAMETER TYPE	PARAMETER VALUE
Simulation Time	60ms
Simulation Area	1500×1000m
Number of nodes	10,20,30…,50
Path loss model	Two Ray Ground
Antenna type	Omni Antenna
Mobility model	Random Way Point
Energy model	Energy Model
MAC protocol	802.11
Transmission Range	250m
Traffic model	CBR

The data from the sensor nodes are gathered by using Real-Time Big Data Gathering algorithm based on the received signal strength indicator (RSSI) and residual energy information. The sensor nodes are deployed in a field and sensor networks should have following properties:

- After deployment of sensor nodes on a field, the nodes must be immobile.
- Sensor nodes are homogeneous.
- All nodes can receive the signal which was sent from the base station(BS).
- Sensor nodes send data with fixed transmission power.

The propagation channel is symmetrical

- ✓ In RTBDG algorithm sensor nodes should screen the gathered data according to requirement of Risk Analysis. The screening process is explained below:
- ✓ Initially the data collected from the sensor nodes with the normal reference range is established.
- ✓ The data should be collected within some regular interval time.
- ✓ The data collected by sensor nodes are compared with the normal reference value. If these data are below the normal reference value then it is stored at a regular time interval.

In other case, if any abnormalities both the abnormal data and the stored data will be transmitted to Base Station (BS). d. If the stored data reaches the upper limit value of capacity then all the data are transmitted to the base station by means of established routes.

IX. CONCLUSION

Data gathering in sensor networks is required to be efficient, adaptable and robust. Recently, compressive sensing (CS) based data gathering shows promise in meeting these requirements. Existing CS-based data gathering solutions require that a transform that best sparsifies the sensor readings should be used in order to reduce the amount of data traffic in the network as much as possible. In this paper I have reviewed different researchers work and faced different problems . These problems are resolved with the help of different algorithms and WSN.

X. REFERENCES

[1]. Kan Yu et.al. "REALFLOW: Reliable Real-Time Flooding-Based Routing Protocol for Industrial Wireless Sensor Networks" Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2014.

[2]. Abdullah I. Alhasanat et.al. "Data Gathering in Wireless Sensor Networks Using Intermediate Nodes" International Journal of Computer Networks & Communications (IJCNC) Vol.7, No.1, January 2015.

[3]. Ms.R.S.Mahathi et.al. "A Real-time Big Data Gathering Algorithm in Wireless Sensor Networks" IJIEST ISSN (2455-8494) Special Issue Vol.01, No. 01, May 2016.

[4]. Panneer Selvam G. et.al. "Energy Efficiency in Wireless Sensor Network for Risk Analysis using RMS Algorithm" I J C T A, 9(9), 2016

[5]. Jie Cheng et.al. "STCDG: An Efficient Data Gathering Algorithm Based on Matrix Completion for Wireless Sensor Networks" IEEE TRANSACTION ON WIRELESS COMMUNICATIONS-2016.

[6]. Xuejun Ding et.al. "A Real-Time Big Data Gathering Algorithm Based on Indoor Wireless Sensor Networks for Risk Analysis of Industrial Operations" IEEE Transactions on Industrial Informatics-2016.

[7]. W. Shen, T. Zhang, F. Barac, M. Gidlund, "PriorityMAC: a priorityenhanced MAC protocol for critical traffic in industrial wireless sensor and actuator networks," IEEE Transactions on Industrial Informatics, vol.10, no. 1, pp. 824-835, 2014.

[8]. B. C. Villaverde, S. Rea, D. Pesch, “InRout–A QoS aware route selection algorithm for industrial wireless sensor networks,” Ad Hoc Networks, vol. 10, no. 3, pp. 458-478, 2012.

[9]. K. Yu, Z. Pang, M. Gidlund, J. Akerberg, M. Björkman, “REALFLOW: Reliable real-time flooding-based routing protocol for industrial wireless sensor networks,” International Journal of Distributed Sensor Networks,vol. 2014, Article ID 936379, 17 pages, 2014.

[10]. A. A. Kumar S., K. Øvsthus, L. M. Kristensen, “An industrial perspective on wireless sensor networks - a survey of requirements, protocols, and challenges,” IEEE Communications Surveys & Tutorials, vol. 16, no. 3,pp. 1391-1412, 2014

[11]. A. Maskooki, C. Soh, E. Gunawan, K. Low, “Adaptive routing for dynamic on-body wireless sensor networks,” IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 2, pp. 549-558, 2015.

[12]. N. Marchenko, T. Andre, G. Brandner, W. Masood, C. Bettstetter, “An experimental study of selective cooperative relaying in industrial wireless sensor networks,” IEEE Transactions on Industrial Informatics, vol. 10, no. 3, pp. 1806-1816, 2014.

[13]. R. Abrishambaf, S. N. Bayindir, M. Hashemipour, “Energy analysis of routing protocols in wireless sensor networks for industrial applications,” Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, vol. 226, no. 5, pp. 678-684, 2012.

[14]. W. Shen, T. Zhang, F. Barac, M. Gidlund, “PriorityMAC: a priority-enhanced MAC protocol for critical traffic in industrial wireless sensor and actuator networks,” IEEE Transactions on Industrial Informatics, vol. 10, no. 1, pp. 824-835, 2014.