

**Peer Reviewed and Refereed Journal**

## **BIG DATA ANALYSIS FOR RISK IDENTIFICATION: A REVIEW**

**Dileep Kumar**

**Assistant Professor**

**IET, Dr R M L Avadh University**

**Faizabad, UP**

*Abstract:* - The era of big data has begun and the huge amount of real-time data has been used for various applications. Some of the challenging processes are data gathering, processing, and analysis of big data in real time. The sources of big data are widely spread across the whole world. The main sources of big data are the Wireless Networks. Wireless networks are building up of many nodes. Every node monitors the environment or some of the substantial factor and combines those data. Every nodes energy plays a major function in Wireless networks. In the majority of the cases, the failure of a node occurs due to lack of energy thus the lifetime of a node is limited. Hence, the energy utilization ought to be managed well to exploit the network lifetime. There are a various number of protocols and algorithms were proposed to decrease the consumption of energy. In the existing work, a technical challenge comes in the indoor industrial environment. Indoor wireless sensor networks (WSN) expertise can overcome this restriction by gathering the big data generated from main nodes and sink them to the Base center in real time. In the existing study a typical house based, office, and industrialized environments were selected. The signal broadcast characteristics of an indoor WSN were obtained by investing the analyzed data. In this paper different research work is reviewed.

**Keywords:** Big Data, WSN, RTBDG algorithm, CH, BS, and HEED etc.

### **LINTRODUCTION**

Big data represent a brand new era in knowledge exploration and utilization and covers numerous industrial applications, for example, professionalizing business information gathering within the industry, resolution routing and programming issues in transportation systems , up the performance of provide chains by minimizing the negative impact of demand uncertainties , providing security for buildings and physical infrastructure in home police work and security systems [1], and analyzing provide chains with radio-frequency identification technology from each the chance and profit views . Massive knowledge may also be went to analyze risks in industrial operations, significantly in product producing. Several producing enterprises have strict necessities on instrumentality operating conditions and atmosphere conditions for prime quality product, like chip fabrication plants, pharmaceutical factories, and food factories.

In the product producing method of the same enterprises, the operating condition parameters of producing instrumentality and atmosphere knowledge got to be gathered in real time. Abnormal data is extracted from these knowledge for the chance analysis of product producing to make sure the conventional operation of a production system [2]. However, a technical challenge exists in gathering period of time massive knowledge in numerous environments. This limitation on period of time knowledge assortment will be overcome by wireless detector networks (WSN). WSN has become a crucial technological support for gathering massive knowledge, like temperature, humidity, instrumentality operating condition, health data, and electricity consumption, significantly for knowledge assortment and transmission in indoor environments. Period of time knowledge will be gathered by exploitation sensible sensors, as well as atmospherically sensors, mensuration sensors, wetness sensors, and accelerometers. The degree of knowledge gathered by these sensors could reach the order of petabytes in line with a report of ORCALE [3]. Within the industrial field, massive knowledge denote the large volume of assorted period of time knowledge that area unit gathered, managed, processed, and analyzed for industrial operations.

However, a significant challenge for WSN is making certain that period of time knowledge will be transmitted to the info center. Detector nodes need enough energy to relay the info gathered by several encompassing sensors [4].

Therefore, energy is one amongst the foremost necessary indicators in WSN and energy consumption ought to be managed well to maximize network period. To resolve the same issues, associate degree energy-efficient routing algorithmic rule for WSN must be designed to assemble massive knowledge in real time.

Many routing algorithms for WSN are according to prolong network period. Some routing algorithms area unit planned for the universal atmosphere. A centralized cluster algorithmic rule known as the low-energy adaptation cluster hierarchy (LEACH). The disadvantage of centralized algorithms is that every detector node should transmit its location and residual energy data to a base station (BS)[5].

### WHAT IS BIG DATA?

Big data refers to the dynamic, giant and disparate volumes of knowledge being created by people, tools and machines; it needs new, innovative and ascendible technology to gather, host and analytically method the huge quantity of knowledge gathered so as to derive period business insights that relate to shoppers, risk, profit, performance, productivity management and increased shareowner value[1-2].

Big information includes info garnered from social media, information from internet-enabled devices (including smartphones and tablets), machine information, video and voice recordings, and therefore the continued preservation and work of structured and unstructured information. it's usually characterized by the four “V’s”:

- Volume: the quantity of information of knowledge of information being created is huge compared to ancient data sources
- Variety: information comes from completely different sources and is being created by machines further as folks
- Velocity: information is being generated extraordinarily quick — a method that never stops, even whereas we have a tendency to sleep[2]
- Veracity: huge information is sourced from many alternative places, as a result you would like to check the veracity/quality of the information the info the information Evolving technology has brought data analysis out of IT backrooms, and extended the potential of exploitation data-driven results into each side of a corporation.

However, whereas advances in package and hardware have enabled the age of massive information, technology isn't the sole thought. Firms got to take a holistic read that acknowledges that success is made upon the combination of individuals, process, technology and information; this suggests having the ability to include data into their business routines, their strategy and their daily operations [4]. Organizations should perceive what insights they have so as to create sensible strategic and operational choices. The primary a part of the challenge is sorting through the entire out there information to spot trends and correlations that may drive useful changes in business behavior. Succeeding step is enriching this structure info thereupon from sources outside the enterprise; this can embody acquainted huge information sources, like those created and hold on on-line. In an exceedingly business setting that perpetually and quickly changes, future prediction becomes a lot of necessary than the easy mental image of historical or current perspectives [6]. For effective future prediction, information analysis exploitation applied math and prophetic modeling techniques is also applied to boost and support the organization’s business strategy. The gathering and aggregation of massive information, and alternative info from outside the enterprise, allows the business to develop their own analytic capability and capability that for several years has solely been out there to a couple of larger organizations.

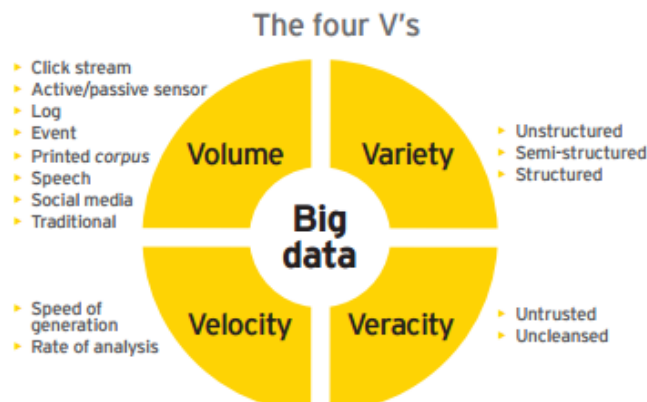
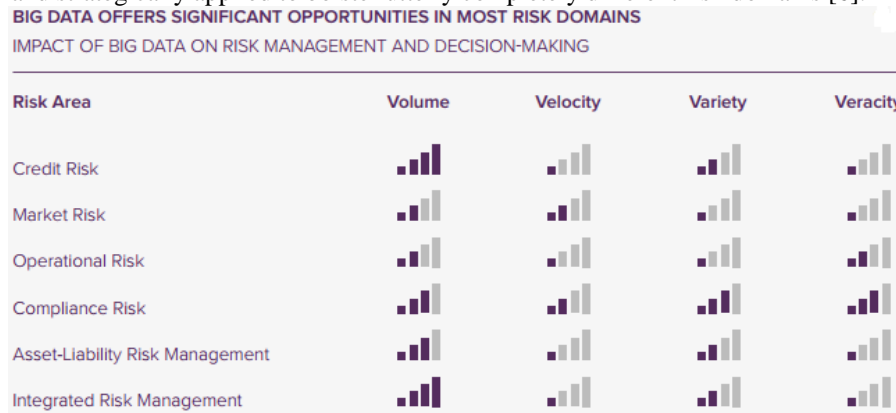


Figure 1 Big data with four V's[2]

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

Risk management faces new demands and challenges. In response to the crisis, regulators unit of measurement requiring plenty of careful data and additional and additional refined reports. Banks unit of measurement expected to

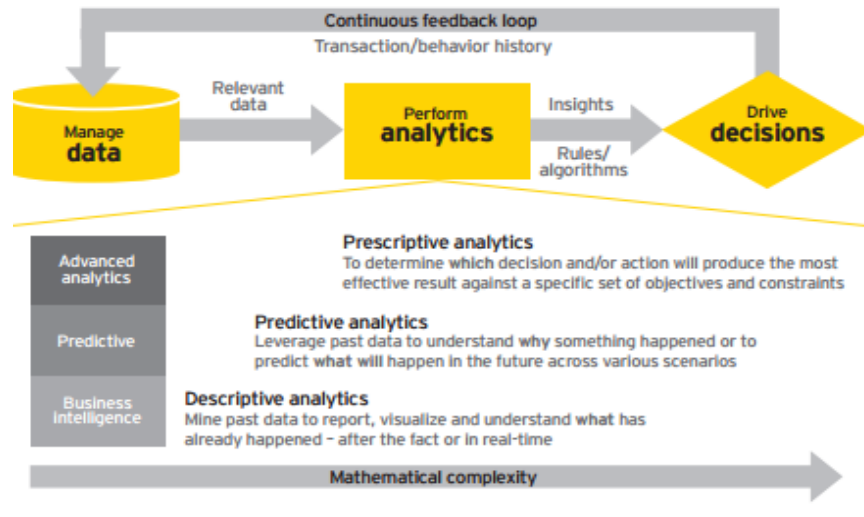
conduct regular and comprehensive bottom-up stress tests for kind of things across all quality classes. Recent, very exposed ‘rascal trader’ and concealing scandals have prompted extra business needs improved risk looking at and modeling [6]. Immense data technologies gift modern opportunities to influence these challenges. Vast, comprehensive and shut to amount of your time data has the potential to spice up looking at of risk (while reducing noise-to-signal ratios), risk coverage, and so the soundness and oracular power of risk models. In a very kind of key domains - notably operational and compliance risk - immense data technologies will alter the event of models that will support everyday Risk Officer Decision-making. Ready to technique immense amounts of data in fast timeframes, the technologies can also accommodate new wants for state of affairs stress tests at the trade, counterparty and portfolio levels. The majority of benefits - and challenges - offered by immense data stem from its giant volume and choice (figure.1.2). However, utterly completely different risk domains stand to take advantage of immense data technologies in varied ways in which. Immense data area unit typically targeted to Associate in Nursing organization’s specific wishes - whether or not or not they unit of measurement for larger volume, variety, rate or honesty - and strategically applied to bolster utterly completely different risk domains [6].



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

### III.BIG DATA AND ANALYTICS

Big information poses every opportunities and challenges for businesses. Thus on extract value from Brobdingnagian information, it ought to be processed and analyzed terribly very timely manner, and thus the results ought to be accessible in such the best method on be ready to result positive modification or influence business selections. The effectiveness to boot depends on a company having the proper combination of people, technique and technology. By pure definition, analytics is that the invention and communication of meaty patterns in information - aside from business, analytics have to be compelled to be viewed as a result of the intensive use of data, maths and analysis, victimization instructive and vertical models to drive fact-based business management selections and actions. Analytics helps to optimize key processes, functions and roles [7]. It’s leveraged to combination every internal and external information. It permits organizations to satisfy neutral news demands, manage massive information volumes, turn out market edges, manage risk, improve controls and, ultimately, enhance structure performance by turning information into intelligence. Analytics can establish innovative opportunities in key processes, functions and roles. It creates a catalyst for innovation and alter -and by tough the institution, it’ll facilitate to form new potentialities for the business and its customers. refined techniques can alter companies to induce root causes, analyze little segments of their markets, work on processes and make correct predictions concerning future events or customers’ propensity to buy for, churn or engage[1][3][7].



**Figure 3: Big data and analytics [3]**

It is not enough for companies to simply understand current technique or operations with a browse on up what already exists, once there is presently the aptitude to question if a technique has connection to the business, or whether or not or not there is a greenhorn manner of resolution a particular issue. The key driver for innovation within organizations is to constantly challenge existing practices rather than consistently accept constant. Most organizations have difficult and fragmented style landscapes that build the cohesive collation and dissemination of data difficult. New Associate in Nursingalytic solutions are participating during a crucial role in sanctioning an economical Intelligent Enterprise (IE). Associate i.e. helps to form one browse across your organization by utilizing a combination of traditional news and data mental representation [6] :

- Info from multiple offer systems is cleansed, normalized and collated
- External feeds are going to be gathered from the most recent analysis, best follow pointers, benchmarks and completely different on-line repositories
- Use of redoubled mental representation techniques, benchmarking indexes and dashboards can inform management and shoppers via smartphones, laptops, tablets, etc., in-house or remotely all companies got to begin puzzling over assortment and pattern relevant large info.

Data-driven selections can cut back inability between the business, legal and IT, optimize existing information assets and address disconnects between fully completely different functions of an organization. However, it's worth noting that the foremost effective info and conjointly the foremost advanced analytical tools and techniques mean nothing if they're not being leveraged by those who are asking the proper queries. Big data, rising storage technology platforms and conjointly the newest analytical algorithms are enablers to business success — not a guarantee of it [8].

#### **IV.DATA MANAGEMENT ON WSNs**

The purpose of data management in device networks is to separate the logical scan (name, access, operation) from the physical scan of the information. Users and applications needn't worry concerning the little print of device networks, but concerning the logical structures of queries. From an information management purpose of scan, the data the data the information management system of a device network are seen as a distributed info system, but it's very totally different from ancient ones. Data management system of a device network organizes and manages perceptible data from the inspected area and answers queries from users or applications. This chapter discusses the ways and techniques of knowledge management in device networks, further because the distinction between data management systems in device networks and in ancient distributed data systems, the look of a information management system throughout a device network, the information model and additionally the search language, the storing and categorization techniques of device data, the operative algorithms, the question method techniques, and a couple of samples of knowledge management systems in device networks: TinyDB and Felis concolor.

#### **V.RELATED WORK**

Ms.R.S.Mahathi et.al.[2016] have studied the device nodes energy plays a serious role in Wireless device networks. In most of the cases the failure of a device node happens as a result of lack of energy so the period of nodes are

restricted. Hence, the energy consumption ought to be managed well to maximize the network period. There's varied numbers of protocols and algorithms were planned to decrease the consumption of energy. Here, a time period huge information gathering formula (RTBDG) is employed for gathering the info in time period. In RTBDG formula, device nodes will screen collected information from the sensors and so clump information transmission structure is established supported the received signal strength indicator (RSSI) and residual energy data [12] Panneer Selvam G. et.al.[2016] have planned the reduction of battery exhausting within the wireless device network victimization huge information technique, and avoiding the failure of nodes in wireless device network, conjointly to avoid perennial transmission and energy wastage in network, information collision through the perennial transmission. It'll monitor price the worth}s of device and create the choice whether or not the value is risk or not, if the worth is found as risk the info are going to be pre-request and transmitted to the server.[15] Jie Cheng et.al.[2016] have planned information gathering in device networks is needed to be economical, all-mains and strong. Recently, compressive sensing (CS) based mostly information gathering shows promise in meeting these needs. Existing CS-based information gathering solutions need that a remodel that best sparsifies the device readings ought to be employed in order to scale back the number of information traffic within the network the maximum amount as doable. As a result, it's terribly possible that completely different transforms need to be determined for various device networks that seriously affect the ability of CS-based schemes. Additionally, the present schemes lead to important errors once the rate of device information is low (equivalent to the case of high packet loss rate) as a result of caesium inherently needs that the quantity of measurements ought to exceed a particular threshold. This paper presents STCDG, associate economical information gathering theme supported matrix completion. [9]

Xuejun ding et.al.[2016] have planned typical residence, office, and producing environments were chosen. The signal transmission characteristics of an internal WSN were obtained by analyzing the take a look at information. in line with these characteristics, a time period huge information gathering (RTBDG) formula supported an internal WSN is planned for the chance analysis of commercial operations. During this formula, device nodes will screen the info collected from the surroundings and instrumentation in line with the wants of risk analysis. Clump information transmission structure is then established on the premise of the received signal strength indicator (RSSI) and residual energy data. Experimental results show that RTBDG not solely with efficiency uses the restricted energy of network nodes however conjointly balances the energy consumption of all nodes.[24] S. Arivoli et.al.[2016] have planned Novel energy-aware routing algorithms to be planned for wireless device Networks, known as reliable minimum Hybrid Dynamic Energy Routing Protocol (HDERP). HDERP is employed attributable to its responsibility, higher life amount and effective utilization of energy. By effective utility of battery power in sensors nodes the life spam of the network is accumulated. Responsibility is ensured by the routing formula that determines the trail of transmission of information .Hence by victimization HDERP responsibility is ensured. Hybrid Cryptography provides the hybrid cryptography technique. The most objective of this paper is to scale back the end-to-end delay and energy consumption when put next to the present technique. Just in case of protective the info from attackers on the net, secret writing ways like Advanced secret writing customary and Elliptic curve cryptography are used.[18] T. Sujithra et.al [2016] has studied massive knowledge Gathering victimization native knowledge Collector (EEBDG-LC) and Energy economical massive knowledge Gathering victimization native knowledge Collector with Threshold (EEBDG - LCWT). Initial approach concentrates on putting area knowledge collect or in each center of mass of the region. It will increase the speed of mobile part knowledge gathering. The most goal of the second approach is to scale back the traffic within the native sensing region of EEBDG-LC supported the edge worth. During which node reaches the edge worth area unit solely allowed to transmit knowledge to the native knowledge collector. Others head to the sleep mode directly. Thus, will increase the period of the device network, and packet delivery quantitative relation. Numerous knowledge gathering mechanisms like mobile part knowledge gathering and knowledge gathering victimization UAV are used and comparison between these 2 has been done. The effectiveness of our approach is valid through intensive simulations.[21] Abdullah I. Alhasanat et.al.[2015] have projected Energy consumption is an important concern to Wireless device Networks (WSNs).The major reason for the energy consumption in WSNs is thanks to the information aggregation. Efficient thanks to perform such a task are accomplished by victimization clump. In clump, nodes area unit sorted into clusters wherever variety of nodes, known as cluster heads, area unit chargeable for gathering knowledge from alternative nodes, mixture them and transmit them to the bottom Station (BS). During this paper they created a replacement algorithmic rule that targeted on reducing the transmission bathtub between device nodes and cluster heads. a correct utilization and reserving of the out there power resources is achieved with this system compared to the well-known LEACH\_C algorithmic rule.[4] Vijayalaxmi et.al. [2015] have projected the massive knowledge has become a preferred topic attributable to huge growth within the field of data technology. The distributed device network is that the production of huge knowledge. A single device in a very network might not provide a big knowledge, however the knowledge detected

by millions of sensors manufacture a massive knowledge. massive knowledge gathering in densely distributed device network with energy expeditiously is a difficult task. One of the most economical solutions for this challenge is the use of sink node. This methodology will confirm the sink node's mechanical phenomenon and knowledge gathering methodology through clump. [22] Kan Yu et.al. [2014] have projected a Wireless technologies are progressively applied in industrial automation systems thanks to versatile installation, mobility, and value reduction. not like ancient wireless device networks (WSNs), industrial wireless device networks (IWSNs), once increasing from wireless observation to wireless management, have a lot of demanding necessities on responsibility, period of time performance, and lustiness in a very variety of commercial applications. During this paper, they explained the first challenges of planning applicable routing protocols and gift a reliable period of time flooding-based routing protocol for IWSNs (REALFLOW). Rather than ancient routing tables, connected node lists area unit generated in a very straightforward distributed manner, serving for packet forwarding. Performance evaluations via simulations verify that important enhancements of responsibility, period of time performance, and network recovery time are often achieved by REALFLOW, compared with ancient routing protocols.[10]

## VI.CONCLUSION

The term big data sets so large or complex data sets where the traditional data processing applications are inadequate. The main source of big data is wireless sensor networks. The wireless sensor networks are nothing but the collection of many hundreds and thousands of sensor nodes. Because of sensor nodes small size, flexibility, low cost and many other characteristics the Wireless sensor networks are used for monitoring the environment, industries, military and some other technologies. Wireless Sensor Networks has some of the limitations and several types of challenges. In future this can be implemented in various applications like industries, manufacturing plants etc. As in future work, when the network is designed then the intermediate nodes in each cluster can serve as cache points in WSN with mobile elements, in an attempt to reduce the contact time of mobile sink, and then reduce the latency of the network. Also, it would be useful if the residual energy is taken into account for cluster heads selection in each transmission round.

## REFERENCES

- [1] A. A. Kumar S. et.al. "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
- [2] A. Maskooki et.al. "Adaptive routing for dynamic on-body wireless sensor networks," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 2, pp. 549-558, 2015.
- [3] A. Rajini, et.al. "A hybrid metaheuristic algorithm for classification using microarray data", Int. J. Scientific & Engineering Research, 3(2), 1-9 (2012).
- [4] 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.
- [5] Antu Raj S et.al. "A New Approach for Big Data Gathering in Dynamic Wireless Sensor Networks" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013).
- [6] Arivoli, Subramanian et.al. "Larvicidal activity of *Citrullus colocynthis* (L.) Schrad (Cucurbitaceae) isolated fractions against *Aedes aegypti* (L.), *Anopheles stephensi* Liston and *Culex quinquefasciatus* Say (Diptera: Culicidae)." Indian Journal of Applied Research 5, no. 8 (2016).
- [7] B. C. Villaverde et.al. "InRout-A QoS aware route selection algorithm for industrial wireless sensor networks," Ad Hoc Networks, vol. 10, no. 3, pp. 458-478, 2012.
- [8] D. Boyd et.al. "Six provocations for big data," A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, 2011.
- [9] Jie Cheng et.al. "STCDG: An Efficient Data Gathering Algorithm Based on Matrix Completion for Wireless Sensor Networks" IEEE TRANSACTION ON WIRELESS COMMUNICATIONS-2016.
- [10] K. Yu et.al. "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.
- [11] 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.
- [12] Ms.R.S.Mahathi et.al. "A Real-time Big Data Gathering Algorithm in Wireless Sensor Networks" IJEST ISSN (2455-8494) Special Issue Vol.01, No. 01, May 2016.
- [13] N. Marchenko et.al. "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.

- [14] Neethu M. Nair et.al. "Survey on Data Collection Methods in Wireless Sensor Networks" International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 12, December – 2013.
- [15] 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
- [16] R. Abrishambaf et.al. "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.
- [17] R. Elankavi et.al. "Data Collection in Wireless Sensor Networks - A Literature Survey" International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 5, July 2013.
- [18] S. Palit et.al. "A cryptanalytic attack on the knapsack cryptosystem using binary Firefly algorithm", in: 2nd Int. Conference on Computer and Communication Technology (ICCT), 15-17 Sept 2011, India, pp. 428- 432 (2011).
- [19] Sina K. Azad et.al. "Optimum Design of Structures Using an Improved Firefly Algorithm" International Journal of Optimisation in Civil Engineering, 1(2), 327-340(2011).
- [20] Sujithra, T et.al. "Swift and Energy Efficient Big Data Gathering Approaches in Wireless Sensor Networks." International Journal of Computer Applications 138, no. 10 (2016): 14-19.
- [21] T.M. Choi, et.al. "Coordination and risk analysis of VMI supply chains with RFID technology," IEEE Transactions on Industrial Informatics, vol. 7, no.3, pp. 497-504, 2011.
- [22] V. Shrinithi et.al. "A Survey on Data Collection in Wireless Sensor Network with Mobile Elements" International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 3 Issue 2 November 2013.
- [23] W. Shen et.al. "Priority MAC: 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.
- [24] 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.
- [25] Z. Kremljak et.al. "Types of risk in a system engineering environment and software tools for risk analysis," Procedia Engineering, vol. 69, pp. 177-183, 2014