Analysis on Big Data: specially its Security Challenges and Tools

Zubair Ahmad Mir[1] and Dr Surendra Yadav[2]
mirzubair026@gmail.com[1] and syadav66@gmail.com[2]
Department of Computer Science and Engineering, Vivekananda Global University jaipur Rajasthan

Abstract:
Before discovering meaningful knowledge from big data systems, it is first necessary to build a data-gathering infrastructure. Big data is a buzzword today, and security of big data is a big concern. Traditional security standards and technologies cannot scale up to deliver reliable and effective security solutions in the big data environment. The data is expected to be collected by the sensors and report to the sink only when an event of interest is occurred, thus leading to error sensitivity and intolerance delay in WSN. Among many feasible data sources, wireless sensor networks (WSNs) are rich big data sources: a large amount of data is generated by various sensor nodes in large-scale networks. Analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge for decision making. Therefore, big data analysis is a current area of research and development. The basic objective of this paper is to explore the potential impact of big data challenges,

Keywords:
Big data; wireless sensor networks; infrastructure; data processing, Massive data; Structured data; Unstructured Data; data collection.

Introduction
When constructing a big data system, data collection, storage, processing, analysis, and visualization are steps that need to be followed in the said order. Big Data has become an entrenched part of discussions of new development in information technology, businesses, governments, markets, and societies in recent years. It has inspired noteworthy excitement about the potential opportunities that may come from the study, research, analysis, and application of big data (L. Singh, 2017). In ongoing research on big data systems, the research communities focus on fundamental aspects of dealing with big data: specific platforms, technology, beneficial applications, standards, and best practices (for applications in social web, financial issues, and so on). Moreover, there are many platforms and tools that can implement these functions in the real world. Thus, data-intensive applications are now being developed to benefit from them. (Chen et al.1) present state-of-the-art technologies to deal with the problems of big data systems (specifically, computing infrastructure for data deluge that includes granular computing, cloud computing, bio-inspired computing, and quantum computing). Moreover, seven principles to be applied in big data systems are defined and explained. In addition, (Chen et al. 2) provided a general overview of big data systems, focusing on the four phases of value chain of big data.

CHALLENGES IN BIG DATA ANALYTICS
Recent years big data has been accumulated in several domains like health care, public administration, retail, biochemistry, and other interdisciplinary scientific researches. Web-based applications encounter big data frequently, such as social computing, internet text and documents, and internet search indexing. Social computing includes social network analysis, online communities, recommender systems, reputation systems, and prediction markets where as internet search indexing includes ISI, IEEE

Xplorer, Scopus, Thomson A standard process to this end is to transform the semi structured or unstructured data into structured data, and then apply data mining algorithms to extract knowledge. A framework to analyze data was discussed by Das and Kumar (T. K. Das). Similarly detail explanation of data analysis for public tweets was also discussed by Das et al in their paper (D. P. Acharjya).

Data Management

There is a book chapter on big data management in WSN edited by (Hung and Hsieh), where an overview of big data management on WSN is presented. Moreover, big data tools and frameworks are introduced to evaluate performance of query processing and data collection. As the major challenge in data management in WSNs, the authors focus on energy preservation. In the aspects of energy efficiency, the authors emphasize decentralization, which is one of the promising ways to achieve energy preservation by distributing the computation tasks among sensor nodes. Moreover, (Medlej) proposes novel big data management techniques for periodic sensor networks to overcome the limitations imposed by WSN as well as adapt key features of sensor data. The main contribution of this thesis is to propose an adaptive sampling approach for periodic data collection by allowing each sensor node to adjust its sampling rates in parallel with the physical changing dynamics.

Applications of WSN-Based Big Data Systems

Before starting a detailed analysis of related work, it is highly desirable to understand which big data applications can be implemented and deployed through WSNs. Since a WSN is usually built to meet application-specific requirements, it is reasonable to review big data applications prior to addressing their technical issues. The following monitoring applications can benefit from using WSNs: smart grids, monitoring human body, and monitoring the environment. In case of smart grids (Jaradat, M.; Jarrah, M), smart sensor networks are introduced in big data systems for energy management. These systems run smart grid applications that include power monitoring, demand-side energy management, coordination of distributed storage, and integration of renewable energy generators. Additionally, techniques used to manage big data generated by sensors and meters are proposed. Moreover, feasible recommendations and practices for smart grid are discussed. The authors in (Jangili, S.; Bikshalu) focus on managing data efficiently and extracting required information from the big data system. Because reliability and low latency are two objectives in a smart grid, streaming processing on fog computing architecture is studied for real-time applications over a well-designed platform.

Data Storage and Analysis

In recent years the size of data has grown exponentially by various means such as mobile devices, aerial sensory technologies, remote sensing, radio frequency identification readers etc. These data are stored on spending much cost whereas they ignored or deleted finally because there is no enough space to store them. Therefore, the first challenge for big data analysis is storage mediums and higher input/output speed. In such cases, the data accessibility must be on the top priority for the knowledge discovery and representation. The prime reason is being that, it must be accessed easily and promptly for further analysis. In past decades, analyst use hard disk drives to store data but, it slower random input/output performance than sequential input/output(O. Y. Al-Jarrah). To overcome this limitation, the concept of solid state drive (SSD) and phrase change memory (PCM) was introduced. Analysis of large dataset requires more computational complexities. The major issue is to handle inconsistencies and uncertainty present in the datasets. In general, systematic modeling of the computational complexity is used. It may be difficult to establish a comprehensive mathematical system that is broadly applicable to Big Data. But a domain specific data analytics can be done easily by understanding the particular complexities. A

series of such development could simulate big data analytics for different areas. Much research and survey has been carried out in this direction using machine learning techniques with the least memory requirements. The basic objective in these research is to minimize computational cost processing and complexities (Changwon. Y, Luis.), (B. Suri).

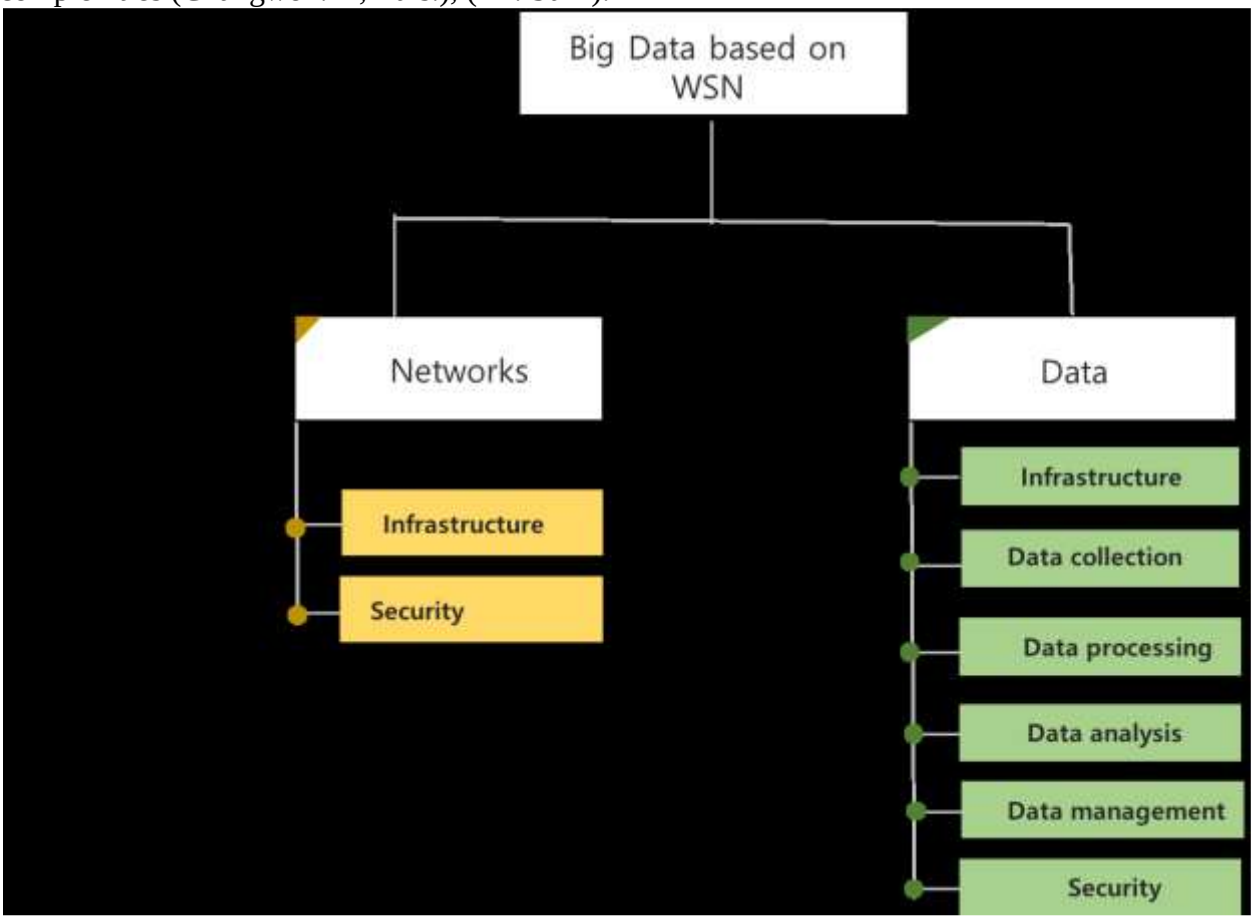

Figure . Classification of related work.

BIG DATA CHALLENGES IN WIRELESS SENSOR NETWORKS

The notable growth and emergence of different network technologies and the explosion in their utilization led to an impressive augmentation in big data generation and handling (M. Chen, 2014), (A. A. Tole, 2013). Thus, and due to this increasing and the huge volume of big data, the development of big data application meets obstacles and challenges that must be overcome to efficiently manipulate the impressive volume of data deployed (J. Cao 2016).

Security

To detect intrusion efficiently and correctly, a WSN is usually a good solution: it can defend against the insider attacks using a relevant trust-based mechanism. However, due to excessive data, effectiveness of trust computation can be degraded significantly. The complexity resulting from the five dimensions (5 V's) of big data makes the management of big data more challenging than traditional database or knowledge base management (B. Saneja, 2017). Major challenges in handling big data include storage challenge, which must deal with increased size, cost, and scalability requirements; network challenge, which involves the accessibility, reliability, and security of obtaining and sharing data with both internal constituents as well as external customers, suppliers, and other partners; data integrity challenge, which

necessitates data authentication, validation, consistency checking, and backup; and metadata challenges, which requires the establishment of data ontology and data governance (Kucukkecesi, C.; Yazici, 2018). To prevent this degradation, (Meng et al.)proposed Bayesian-based trust management with traffic sampling under a hierarchical structure. Their experimental results prove that the proposed approach improved trust management in a high-traffic case by detecting malicious nodes quickly. Security in multimedia big data applications for smart city in trust-assist sensor cloud (TASC) is addressed in (Zhu, C.; Shu,). Two types of TASC: TASC-S (TASC with a single trust value threshold), and TASC-M (TASC with multiple trust value thresholds), are proposed and evaluated through extensive simulation results in the aspect of throughput. Further,(Wu et al.) proposed a dynamic trust relationships aware data privacy protection (DTRPP) mechanism by combining key distribution with trust management. The major contribution of this work is to evaluate the trust value of a public key according to both the number of supporters and the trust degree of the public key.

Conclusions

In this paper, we presented technical research challenges for big data systems based on WSN in cases when WSN was regarded as one of the major data sources. Prior to looking into major contributions of each work, we presented the opportunities of big data systems in WSN. Big sensor data continue to increase every day. Their variety, volume and velocity are also expanding. The big data paradigm in wireless sensor networks requires energy efficient clustering, processing, and securing. These requirements represent the main big data challenges in wireless sensor networks. The data aggregation is one of the principle big sensor data processing challenges. In this paper, we introduced big data in wireless sensor networks. We presented a view of big data concepts and analytic tools and survived the works proposed for integrating them in wireless sensor networks. We also proposed a classification for big sensor data challenges and reviewed the proposed solutions for these challenges. Subsequently, previous research was categorized according to the research area and objective. Specifically, we focus on data collection and in-network processing while considering the unique characteristics of network properties. . In the future, we aim to propose novel strategies for big data challenges and issues in heterogeneous wireless sensor networks.

References

1. Chen, C.L.P.; Zhang, C. Data-intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. Inf. Sci. 2014,
2. Chen, M.; Mao, S.; Liu, Y. Big Data: A Survey. Mob. Netw. Appl. 2014,
3. Meng, W.; Li, W.; Su, C.; Zhou, J.; Lu, R. Enhancing Trust Management for Wireless Intrusion Detection via Traffic Sampling in the Era of Big Data. IEEE Access 2018, 6, 7234–7243. [CrossRef]
4. Zhu, C.; Shu, L.; Leung, V.; Guo, S.; Zhang, Y.; Yang, L. Secure Multimedia Big Data in Trust-Assisted Sensor-Cloud for Smart City. IEEE Commun. Mag. 2017,
5. Wu, D.; Si, S.; Wu, S.; Wang, R. Dynamic Trust Relationships Aware Data Privacy Protection in Mobile Crowd-Sensing. IEEE Internet Things J. 2017, 5, 2958–2970. [CrossRef].
6. Hung, C.; Hsieh, C. Chapter 5: Big Data Management onWireless Sensor Networks. In Big Data Analytics for Sensor-Network Collected Intelligence; Morgan Kaufmann: Burlington, MA, USA, 2017.
7. Medlej, M. Big Data Management for Periodic Wireless Sensor Networks. Ph.D. Thesis, Universite de Franche-Comte, Besançon, France, 2014.

8. Jaradat, M.; Jarrah, M.; Bousselham, A.; Jararweh, Y.; Al-Ayyouba, M. The Internet of Energy: Smart Sensor Networks and Big Data Management for Smart Grid. Procedia Comput. Sci. 2015,
9. Jangili, S.; Bikshalu, K. Smart Grid Administration using Big Data andWireless Sensor Networks. Int. J. Adv.Res. Sci. Eng. 2017, 6, 629–636.
10. T. K. Das and P. M. Kumar, Big data analytics: A framework for unstructured data analysis, International Journal of Engineering and Technology, 5(1) (2013), pp.153-156.
11. T. K. Das, D. P. Acharjya and M. R. Patra, Opinion mining about a product by analyzing public tweets in twitter, International Conference on Computer Communication and Informatics, 2014.
12. O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis and K. Taha, Efficient machine learning for big data: A review, Big Data Research, 2(3) (2015), pp.87-93.
13. Changwon. Y, Luis. Ramirez and Juan. Liuzzi, Big data analysis using modern statistical and machine learning methods in medicine, International Neurourology Journal, 18 (2014), pp.50-57.
14. P. Singh and B. Suri, Quality assessment of data using statistical and machine learning methods. L. C.Jain, H. S.Behera, J. K.Mandal and D. P.Mohapatra (eds.), Computational Intelligence in Data Mining, 2 (2014), pp. 89-97.
15. L. Singh and D. Kumar, ``A big data analysis for risk identi_cation on wireless sensor network,"WorldWide J. Multidiscipl. Res. Develop., vol. 3,no. 8, pp. 337_343, 2017.
16. B. Saneja and R. Rani, ``An ef_cient approach for outlier detection in big sensor data of health care," Int. J. Commun. Syst., vol. 30, no. 17, p. e3352,Nov. 2017.
17. Kucukkecesi, C.; Yazici, A. Big Data Model Simulation on a Graph Database for Surveillance in Wireless Multimedia Sensor Networks. Big Data Res. 2018, 11, 33–43.
18. M. Chen, S. Mao, and Y. Liu, ``Big data: A survey," Mobile Netw. Appl., vol. 19, no. 2, pp. 171_209, 2014.
19. A. A. Tole, ``Big data challenges," Database Syst. J., vol. 4, no. 3, pp. 31_40, 2013.
20. B. Liu, J. Cao, J. Yin, W. Yu, B. Liu, and X. Fu, ``Disjoint multi mobile agent itinerary planning for big data analytics," EURASIP J. Wireless Commun. Netw., vol. 1, p. 99, Dec. 2016.