Big Data Challenges and Data Aggregation in Wireless Sensor Networks.

Zubair Ahmad Mir, Dr Surendra Yadav and Tanveer Ahmad Dar Department of Computer Science and Engineering, Vivekananda Global University, Jaipur, Rajasthan

mirzubair026@gmail.com[1], syadav66@gmail.com[2] Tanveerabdullah786@gmail.com[3]

ABSTRACT

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. This characteristic refers to event based detection of data collection. In continuous sampling data, the sensors regularly samples and reports the data to the sink where the snapshot of area for the collected data is generated. One of the most widespread and important applications in wireless sensor networks (WSNs) is the continuous data collection, such as monitoring the variety of ambient temperature and humidity. Due to the sensor nodes with a limited energy supply, the reduction of energy consumed in the continuous observation of physical phenomenon plays a significant role in extending the lifetime of WSNs.

Keywords: wireless sensor networks; big data; infrastructure; data collection; data processing

INTRODUCTION

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). However, accompanying those enticing opportunities and prospective rewards, there are significant challenges and substantial risks associated with big data. One of the biggest challenges for big data is increased security risk (N. Kato 2014). Security for big data is magnified by the volume, variety, and velocity of big data. With the proliferation of the Internet and the Web, pervasive computing, mobile commerce, and large scale cloud infrastructures, today's data are coming from diverse sources and formats, at a dynamic speed, and in high volume (Sotto, 2013). Traditional security measures are developed for clean, structured, static, and relatively low volume of data. Undeniably, big data presents huge challenges in maintaining the integrity, confidentiality, and availability of essential data and information (ADMA, 2013).

BIG DATA SECURITY MANAGEMENT

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). The NIST defines information security as "The protection of information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide

confidentiality, integrity, and availability." Although security and privacy principles for traditional data can be applied to big data, hence many of the existing security technologies and best practices can be extended to the big data ecosystem, the different characteristics of big data require modified approaches to meet the new challenges to effective big data management. There are several areas in which big data faces different or higher risk than traditional data (Morillo, P.; 2018). Higher volume translates into higher risk of exposure when security breach occurs. More variety means new types of data and more complex security measure. Increased data velocity implies added pressure for security measures to keep up with the dynamics and faster response/recovery time. Most organizations are just starting to embrace big data, thus the big data governance and security (Kim, B.; Park, H.; Kim 2017).

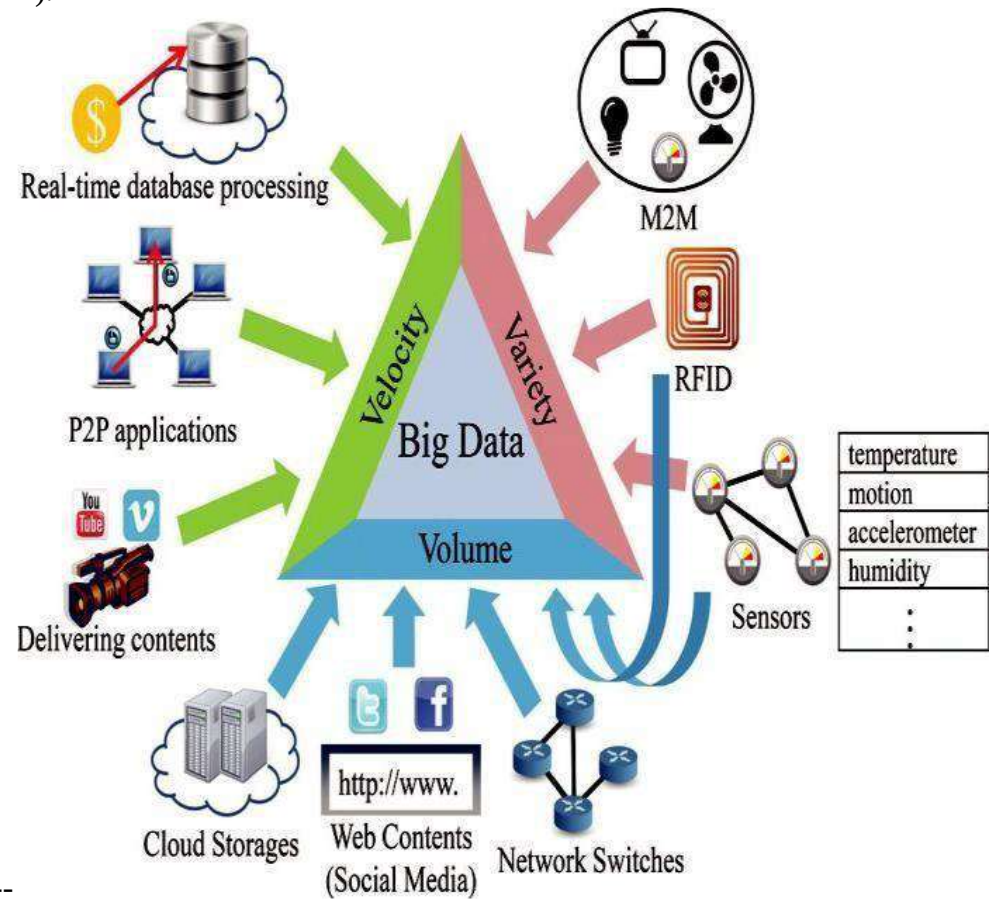

Figure : Trends of Big Data Collection

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).

BIG DATA AGGREGATION PROPOSED STRATEGIES IN WIRELESS SENSOR NETWORKS

Data aggregation is one of the key challenges in big data wireless sensor networks (W.-C. Lan, 2013). Data aggregation allows combining data from different sources to eliminate the redundancy, and reduce consequently the consumption of resources available in the network. Data aggregation is a subset of data fusion that involves the use of techniques that combine and gather data from multiple sources to make more effective and potentially more accurate inferences, reparations and

associations (L. Karim, 2016) , (M. S. Al-Kahtani, 2015). Strategies are proposed to deal with this challenge. They are mainly based on the correlation between the data aggregation, the clustering and the energy consumption challenges of big sensor data. In the following, we survey the big data aggregation strategies proposed for wireless sensor networks:

Big Data Security Challenges

Security has become an increasing concern as more and more people are connected to the global network economy. According to the Internet Live Stats (IBM, 2014) website, the number of worldwide Internet users is approaching three billions (Internet Users). There has never been a lack of security attacks, from the early days of telephone phracker to modern time cyber criminals (Rowe, 2012). As the big data increases in volume, speed, and variety, the likelihood that data breach is expected to increase significantly. Recently, it is estimated that the average security breaches results in a loss of $40 million for American companies (Khalid, Z.; Fisal 2014) . Symantec, an Internet security firm, estimated that the worldwide cybercrime cost is over $100 billion more than the cost of illegal drug market (R. J., & Kwon 2014). A general big data security management (BDSM) framework we adopted from the Plan-Do-Check-Act Deming cycle. The Design/Planning phase defines, identifies, and evaluates security risk, and develop appropriate policies, procedures, control, and measure according to desired security level. This phase also includes a security modeling that builds a threat model and lays out strategies for most security/privacy breach scenarios (Demchenko, Y., Grosso, 2013). The design blueprint will be turned into operational systems at the implementation phase. The operation phase delivers the desired functions and services with security and privacy protection and compliance to laws and regulations (Zhu, C.; Shu, 2017), (Meng, W.; 2018).

Top Ten Big Data Security and Privacy Challenges

Big Data Challenge	Use Case Example
Secure computations in distributed programming Frameworks	Prevent untrusted mappers from snooping on requests or alter scripts/results.
Security best practices for non-relational data Stores	Migrate from traditional relational databases to NoSQL databases that are more suitable to process large volume dynamic data.
Secure data storage and transactions logs	Adopt an auto-tier storage system which prioritizes data storage by putting less utilized data to a lower tier.
End-point input validation/filtering	Develop effective algorithms to validate input data from diverse and dynamic data sources
Real-time security monitoring	Implement real-time security analytics that support real-time monitoring, querying, and decision making.
Scalable and composable privacy-preserving data mining and analytics	Improve scalable privacy-preserving data mining algorithms and prevent untrusted user from accessing sensitive data.
Cryptographically enforced data centric security	Design innovative techniques to index, analyze, and process encrypted data with diverse sources. Maintain data confidential and integrity
Granular access control	Establish access control at a granular scale so that individual user roles and responsibilities can be set to access only authorized data, functions, or services
Granular audits	Update auditing with extended scope and granularity for real-time security and meeting compliance requirements

Data provenance	Keep track of metadata of data creation to support effective digital forensic and compliance auditing

CONCLUSION

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. As big sensor data aggregation represents our principle point of interest, we survived its proposed strategies in detail. In the future, we aim to propose novel strategies for big data challenges and issues in heterogeneous wireless sensor networks.

References

1. Sotto, L. J. (2013). *Privacy and Data Security Law Deskbook*. Frederick, MD: Aspen Publishers.
2. Rowe, N. (2012). *The Big Data Imperative: Why Information Governance Must be Addressed*. Aberdeen Group Research Report.
3. ADMA. (2013). *Best Practice Guideline: Big Data. A guide to maximising customer engagement opportunities through the development of responsible Big Data strategies*. Retrieved from: http://www. adma.com.au.
4. D.Takaishi, H. Nishiyama, N. Kato and R. Miura, “Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks”, IEEE Transactions on Emerging Topics in Cloud Computing, 2014.
5. M. Chen, S. Mao, and Y. Liu, ``Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171_209, 2014.
6. A. A. Tole, ``Big data challenges," *Database Syst. J.*, vol. 4, no. 3, pp. 31_40, 2013.
7. 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.
8. 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.
9. 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.
10. T.-Y. Tsai, W.-C. Lan, C. Liu, and M.-T. Sun, ``Distributed compressive data aggregation in large-scale wireless sensor networks," *J. Adv. Comput.Netw.*, vol. 1, no. 4, pp. 295_300, Dec. 2013.
11. L. Karim and M. S. Al-kahtani, ``Sensor data aggregation in a multi-layer big data framework," in *Proc. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2016, pp. 1_7.
12. M. S. Al-Kahtani, ``Ef_cient cluster-based sleep scheduling for M2M communication network," *Arabic J. Sci. Eng.*, vol. 40, no. 8, pp. 2361_2373, 2015.
13. IBM. (2014). *IBM X-Force Threat Intelligence Quarterly, IBM Security Systems*. First Quarter 2014. Retrieved from: http://www-03.ibm.com/security/xforce/

14. Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*, 6780. doi:10.1016/j.dss.2013.08.008
15. Demchenko, Y., Grosso, P., de Laat, C., & Membrey, P. (2013). Addressing big data issues in Scientific Data Infrastructure. In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*. IEEE. doi:10.1109/CTS.2013.6567203
16. Khalid, Z.; Fisal, N.; Rozaini, M. A Survey of Middleware for Sensor and Network Virtualization. Sensors **2014**, 14, 24046–24097.
17. 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**, 55, 24–30.
18. 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.
19. 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.
20. Morillo, P.; Orduña, J.; Fernández, M.; García-Pereira, I. Comparison of WSN and IoT Approaches for A Real-time Monitoring System of Meal Distribution Trolleys: A Case Study. Future Gener. Comput. Syst. **2018**, 87, 242–250.
21. Kim, B.; Park, H.; Kim, K.; Godfrey, D.; Kim, K. A Survey on Real-Time Communications in Wireless Sensor Networks. Wirel. Commun. Mob. Comput. **2017**, 2017, 1864847.