Big Data Analytics for Large-scale Wireless Networks

ADITI SHARMA[1], ISHITA MEHRA[2]
[1]Chandigarh College of Engineering and Technology, Chandigarh, India (e-mail: (lco22393@ccet.ac.in)
[2]Chandigarh College of Engineering and Technology, Chandigarh, India (e-mail: (co22334@ccet.ac.in)

ABSTRACT
This study looks at how the huge amount of data flowing through big wireless networks. With more and more people using smartphones and other devices, these networks handle a ton of information. This is called "big data" because it's a lot, comes in different forms like text and images, moves really fast, and can be super useful if it is analyzed in right way. The big data is divided into four stages: first, the raw data is collected (Data Acquisition); then, the data is cleaned up (Data Preprocessing); next, find a place to store it (Data Storage); finally, differ-ent methods are used to get insights from the data (Data Analytics). This paper covers a range of wireless networks like sensor networks, mobile communication networks, and more. It explains why analyzing this data is essential for things like improving user experience, safety, and efficiency in these networks. The goal is to give a clear picture of how this whole process works in the world of big wire-less networks.

KEYWORDS Big Data Analytics, Wireless Networks, Data Preprocessing, Data Acquistion.

I. INTRODUCTION

There's a lot of growth can be seen in wireless communication methods, which are used by many people around the world. These technologies connect various devices, forming large-scale wireless networks. These networks handle a massive amount of data, especially with the increasing use of smartphones, tablets, laptops, sensors, and other smart devices. The amount of mobile data has grown so much that this period is called 'big data era'[1] of artificial intelligence(AI). It means a type of data that has some special characteristics, often called the "4Vs". As shown in Fig 1:

a) Volume: There's a huge amount of data, ranging from a lot to a whole lot (think of it as a really big file).
b) Variety: The data comes in many forms - structured (like in data-bases), semi-structured, unstructured, text, images, videos, and more.
c) Velocity: Data is generated and processed very quickly, often in real-time.
d) Value: The analysis of this big data can provide significant bene-fits, both in business and for society.

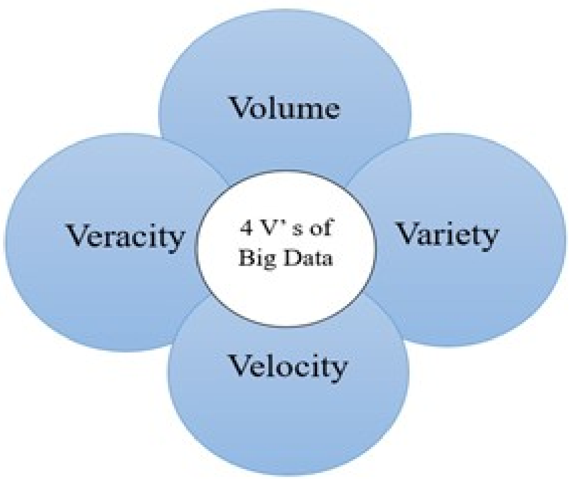

FIGURE 1: 4V's of Big Data

All the information is flowing through wireless networks – messag-es, pictures, videos, and data from various devices. There's so much of it, in different forms, coming at high speeds, and it's valuable for many things like improving network operations, managing networks efficiently, and ensuring security, optimizing traffic systems, manag-ing logistics, and studying social behavior.

This explosion of big data in wireless networks presents challenges, like designing networks that can handle this huge flow of information. But it also brings lots of benefits because analyzing this data can help us understand and improve how these networks function. It's like hav-ing a treasure trove of information that, when properly analyzed, can lead to valuable insights and improvements in various aspects of our connected world. Managing the growing volume and complexity of data poses a difficulty for cybersecurity. Security [2-4] measures need to examine extensive datasets to recognize patterns, anomalies, and potential threats.

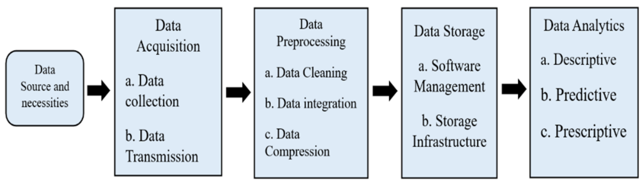

FIGURE 2: Life cycle of big data analytics in large scale wireless net-works

II. LIFE CYCLE OF BIG DATA ANALYTICS

Here the process of Big Data Analytics (BDA) is introduced by a four-stage life cycle as represented in Fig 2. This cycle includes Data Acquisition, Data Preprocessing, Data Storage, and Data Analytics.

a) Data Acquisition: Think of this stage as collecting information. The raw data is gathered from various sources using specialized technologies, like reading RFID tags in the Internet of Things (IoT)[5-6]. This data is then sent to a storage system through wired or wireless networks.
b) Data Preprocessing: Once the raw data is collected, it is needed to be cleaned up before storing it. This involves dealing with the large volume, duplicate entries, and uncertain features in the data. The techniques are used like data cleaning, integration, and com-pression to tidy up the information.
c) Data Storage: Now there is need of a place to keep all this data. Data storage involves managing massive datasets and has two parts: infrastructure (like storage and network devices) and data management software.
d) Data Analytics: The use different schemes to extract valuable in-sights from the massive datasets.

The following diagram represents all these stages.

III. ROLE OF WIRELESS NETWORKS

Throughout these stages, consideration of the types of wireless net-works is important to make sure to adapt the process to the unique fea-tures of larger area wireless networks. In simple words, it's like going through steps to handle a lot of in-formation: first, collecting it, then cleaning it up, finding a good place to store it, and finally, figuring out what useful things could be learnt from it. And all of this is tailored to fit the world of large-scale wire-less networks.
The comparison of the current paper with existing surveys and high-lights the contributions of the current research in the context BDA for wireless networks. There have been many studies on Big Data Analytics, which is like making sense of really large amounts of data. Some studies focused on general computing systems, like computers and data warehouses, but they might not be suitable for large wireless networks. These networks generate a massive amount of data in real-time and have different characteristics.
Some recent studies looked at Big Data Analytics in specific wire-less networks, like sensor networks or mobile communication net-works. However, these studies are often too specific and don't cover the variety of wireless networks that exist. This article covers the big data analytics in large wireless networks in a way that includes different phases and types of networks. The aim is to give readers a clear picture of how Big Data Analytics works in these networks, from collecting data to analyzing it. This is important because wireless networks are different from other systems, and understanding how Big Data Analytics fits into them is crucial.

IV. SOURCES OF DATA

The data in large wireless networks comes from various sources like sensors, cameras, and devices in vehicles, smartphones, and RFID tags. This data is often generated in real-time, like quickly sending a road danger warning to drivers.
Mobile Communication Networks - Mobile networks generate a lot of data related to user activities. This includes data from sensors, user profiles, and the usage of social applications. For example, it can in-clude login information, connections between users, and data from using third-party services like online games.
Vehicular Networks - In vehicular networks, data is gathered from sensors in vehicles for safety and efficiency. This data includes struc-tured information like transportation safety and efficiency, and it's collected in real-time.
Mobile Social Networks - Mobile social networks involve data related to users' social interactions and applications. This includes login data, connections, application data, user profiles, ratings, and interests. Ba-sically, it's about how people connect and interact in the digital world.
Internet of Things (IoT) - IoT connects various devices to the internet, collecting data from the environment. This can be in logistics, envi-ronmental monitoring, smart homes, and more. For example, wireless sensor networks (WSNs)[5] are a part of IoT and are used in things like smart cities and manufacturing. All these data sources generate a huge amount of diverse data, in-cluding structured and unstructured types, and it often needs real-time processing. The table summarizes these data sources.

Necessities of Big Data Analytics for Large Scale Wireless Net-works There is so much data being generated every day, and the need of extracting, classification and processing valuable information from it, is evident which can be performed by machine learning [7-10]. Different networks have different needs:

1) Mobile Communication Networks - BDA helps optimize network costs and energy consumption, improve user experience, and en-hance security.
2) Vehicular Networks - It ensures transportation safety, efficiency, and ride quality by analyzing real-time data.
3) Mobile Social Networks - BDA is essential for understanding user behavior, interests, and improving social welfare.

4) Internet of Things (IoT) - BDA is crucial for managing and making sense of the big amount of data that is generated by IoT devices.

In simpler terms, Big Data Analytics helps us make sense of the mas-sive and diverse data from various wireless networks, leading to im-provements in user experience, safety, efficiency, and overall network optimization. Big data analytics plays a significant role in sustainable development [11-13] by providing valuable insights, facilitating in-formed decision-making, and addressing various challenges associated with environmental, social, and economic sustainability. Parallel computing[14-21] plays a crucial role in data analytics by improving the speed, efficiency, and scalability of data processing tasks. In the context of large-scale data analytics, where datasets can be massive and analytical tasks complex, parallel computing offers several ad-vantages.

V. CONCLUSION

This article explains crucial role of Big Data Analytics (BDA) in navi-gating the challenges posed by large-scale wireless networks. The ex-ponential growth of data in these networks, stemming from the wide-spread use of devices like smartphones and sensors, defines the con-temporary "big data era." Big Data life cycle encompasses all the four stages, serves as a practical guide for handling the vast and diverse data generated by these networks. By exploring the necessities of BDA for specific networks, such as optimizing costs in mobile communica-tion networks or ensuring safety in vehicular networks, this article bridge the gap between theory and practical application. The diverse data sources in large wireless networks, ranging from sensors and cameras to smartphones and IoT devices, present unique challenges. Real-time processing and the time-sensitive nature of this data add complexity to the analysis process. However, the benefits of BDA, including improved user experience, safety, and efficiency, underscore its significance in optimizing these networks. In essence, this study contributes to a understanding of how BDA functions within the dynamic landscape of large-scale wireless net-works. As these networks continue to evolve, our insights aim to guide further advancements, fostering enhanced functionality, security, and efficiency. Ultimately, the effective implementation of BDA in large wireless networks holds the key to unlocking valuable insights and improvements in our increasingly connected world.

REFERENCES

[1] Sharma, A., et al. (2023, January). Fuzzy Based Clustering of Consumers' Big Data in Industrial Applications. In 2023 IEEE International Conference on Consumer Electronics (ICCE) (pp. 01-03). IEEE. http://doi.org/10.1109/ICCE56470.2023.10043451

[2] Dubey, H. A. R. S. H. I. T., Kumar, S. U. D. H. A. K. A. R., & Chhabra, A. N. U. R. E. E. T. (2022). Cyber Security Model to Secure Data Transmission using Cloud Cryptography. Cyber Secur. Insights Mag, 2, 9-12.

[3] Singh, M., Singh, S. K., Kumar, S., Madan, U., & Maan, T. (2021, September). Sustainable Framework for Metaverse Security and Privacy: Opportunities and Challenges. In International Conference on Cyber Security, Privacy and Networking (pp. 329-340). Cham: Springer Inter-national Publishing.

[4] Sharma, A., Singh, S.K., Kumar, S., Chhabra, A., Gupta, S. (2023). Security of Android Banking Mobile Apps: Challenges and Opportunities. In: Nedjah, N., Martínez Pérez, G., Gupta, B.B. (eds) International Conference on Cyber Security, Privacy and Networking (ICSPN 2022). ICSPN 2021. Lecture Notes in Networks and Systems, vol 599. Springer, Cham. https://doi.org/10.1007/978-3-031-22018-0_39.

[5] Singh, R., Singh, S. K., Kumar, S., & Gill, S. S. (2022). SDN-Aided Edge Computing-Enabled AI for IoT and Smart Cities. SDN-Supported Edge-Cloud Interplay for Next Generation Internet of Things, 41-70.

[6] Vats, T., et al. (2023). Explainable context-aware IoT framework using human digital twin for healthcare. Multimedia Tools and Applications, 1-25. Doi: http://dx.doi.org/10.1007/s11042-023-16922-5

[7] Kumar, J., Gupta, D. L., & Umrao, L. S. (2022). Fault-Tolerant Algorithm for Software Preduction Using Machine Learning Techniques. Interna-tional Journal of Software Science and Computational Intelligence (IJSSCI), 14(1), 1-18. http://doi.org/10.4018/IJSSCI.309425Mengi, G., Singh, S. K., Kumar, S., Mahto, D., & Sharma, A. (2021, September). Au-tomated Machine Learning (AutoML): The Future of Computational In-telligence. In International Conference on Cyber Security, Privacy and Networking (pp. 309-317). Cham: Springer International Publishing.

[8] Almomani, A., et al. (2022). Phishing Website Detection With Semantic Features Based on Machine Learning Classifiers: A Comparative Study. International Journal on Se-mantic Web and Information Systems (IJSWIS), 18(1), 1-24. http://doi.org/10.4018/IJSWIS.297032

[9] Gupta, S., Agrawal, S., Singh, S. K., & Kumar, S. (2023). A Novel Transfer Learning-Based Model for Ultrasound Breast Cancer Image Classification. In Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022 (pp. 511-523). Singapore: Springer Nature Singapore.

[10] Sharma, A., Singh, S. K., Chhabra, A., Kumar, S., Arya, V., & Moslehpour, M. (2023). A Novel Deep Federated Learning-Based Model to Enhance Privacy in Critical Infrastructure Systems. International Jour-nal of Software Science and Computational Intelligence (IJSSCI), 15(1), 1-23. http://doi.org/10.4018/IJSSCI.334711.

[11] Chhabra, A., et al. (2024). Sustainable and Intelligent Time-Series Models for Epidemic Disease Forecasting and Analysis. Sustainable Technology and Entrepreneurship, 100064.

[12] S. Kumar, S. K. Singh and N. Aggarwal, "Sustainable Data Dependency Resolution Architectural Framework to Achieve Energy Efficiency Us-ing Speculative Parallelization," 2023, IEEE 3rd International Confer-ence on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2023, pp. 1-6, http://doi.org/10.1109/CISCT57197.2023.10351343.

[13] Peñalvo, F. J. G., Maan, T., Singh, S. K., Kumar, S., Arya, V., Chui, K. T., & Singh, G. P. (2022). Sustainable Stock Market Prediction Framework Using Machine Learning Models. International Journal of Software Science and Computational Intelligence (IJSSCI), 14(1), 1-15.

[14] Kumar, S., Singh, S. K., & Aggarwal, N. (2023). Speculative Parallelism on Multicore Chip Architecture Strengthen Green Computing Concept: A Survey. In Advanced Computer Science Applications (pp. 3-16). Ap-ple Academic Press.

[15] Kumar, S., Singh, et al. (2022). An efficient hardware supported and parallelization archi-tecture for intelligent systems to overcome speculative overheads. In-ternational Journal of Intelligent Systems, 37(12), 11764-11790.

[16] Kumar, S. S., Singh, S. K., Aggarwal, N., & Aggarwal, K. (2021). Efficient speculative parallelization architecture for overcoming speculation overheads. In International Conference on Smart Systems and Advanced Computing (Syscom-2021) (Vol. 3080, pp. 132-138).

[17] Rastogi, A., Sharma, A., Singh, S., & Kumar, S. (2017). Capacity and Inclination of High Performance Computing in Next Generation Computing. Proceedings of the 11th INDIACom. IEEE.

[18] Al Abass, A., Bani-Mohammad, S., & Ababneh, I. (2022). Performance Evaluation of Contiguous and Noncontiguous Processor Allocation Based on Common Communication Patterns for 2D Mesh Interconnec-tion Network. International Journal of Cloud Applications and Compu-ting (IJCAC), 12(1), 1-21. http://doi.org/10.4018/IJCAC.295239

[19] Sissodia, R., Rauthan, M. M., & Naithani, K. (2022). A Survey on Multi-Objective Tasks and Workflow Scheduling Algorithms in Cloud Computing. International Journal of Cloud Applications and Computing (IJCAC), 12(1), 1-16. http://doi.org/10.4018/IJCAC.297100

[20] Chander, S., Vijaya, P., & Dhyani, P. (2022). A Parallel Fractional Lion Algorithm for Data Clustering Based on MapReduce Cluster Framework. International Journal on Semantic Web and Information Systems (IJSWIS), 18(1), 1-25. http://doi.org/10.4018/IJSWIS.297034

[21] Ling, Z. & Hao, Z. J. (2022). Intrusion Detection Using Normalized Mutual Information Feature Selection and Parallel Quantum Genetic Algorithm. International Journal on Semantic Web and Information Systems (IJSWIS), 18(1), 1-24. http://doi.org/10.4018/IJSWIS.307324