The Intelligent Mechanism for Data Collection and Data Mining in the Vehicular Ad-Hoc Networks (VANETs) Based on Big-Data-Driven

Mahshad Mahmoudian
Department of Computer Engineering,
Najafabad Branch, Islamic Azad University,
Najafabad, Iran.
mah.mahmoudian@ieee.org

S. Mohammadali Zanjani*
Smart Microgrid Research Center,
Najafabad Branch, Islamic Azad University,
Najafabad, Iran.
sma_zanjani@pel.iaun.ac.ir

Hossein Shahinzadeh
Department of Electrical Engineering
Amirkabir University of Technology
Tehran, Iran
h.s.shahinzadeh@ieee.org

Yasin Kabalci
Department of Electrical Engineering
Nigde Ömer Halisdemir University
Nigde, Turkey
yasinkabalci@ohu.edu.tr

Ersan Kabalci
Department of Electrical Engineering
Nevsehir Haci Bektas Veli University
Nevsehir, Turkey
kabalci@nevsehir.edu.tr

Farshad Ebrahimi
Department of Electrical and Computer Engineering
University of Houston
Houston, TX 77004, USA
febrahimi@uh.edu

***Abstract*— Big data technology has attracted the main attention of researchers in almost all sciences. The Vehicular Ad-Hoc Network (VANET) enables information exchange between vehicles, other devices, and public networks, playing a key role in road safety and intelligent transportation systems. With the proliferation of connected vehicles and the development of novel mobile apps and technologies, VANETs will generate vast quantities of data that need to be transmitted quickly and reliably. Furthermore, analyzing a wide range of data types can enhance VANET's performance. By utilizing big data technologies, the Ad-Hoc Vehicular Network can extract valuable insights from a large amount of operational data, thus improving traffic management processes, including planning, engineering, and operations. VANETs have access to big data during real-time operations. This paper presents VANET features as big data features in the literature, followed by a discussion of methods for utilizing big data to study VANET features. Combining automotive ad networks and big data facilitates the easy management of a large number of driving factors, as the data mining process in big data enables quick decision-making based on statistical analysis or graphical representations of data.**

Keywords*— Vehicular ad-hoc networks, Vehicles, Big Data, Hadoop, Map Reduce, Security, Sensors, Traffic Management, Smart Cities, Internet of Things (IoT), Mobile ad-hoc networks.

I. Introduction

Recent years have witnessed extensive research in the field of big data technology due to the exponential growth in data capacity and complexity. The term "big data" is still evolving and developing. With the increasing amount of information available on the internet, the web is expanding rapidly, but only a few truly comprehend its significance on a "big" scale [1]. In modern computing, big data has become the most commonly used buzzword, and big data technologies represent a new generation of technologies and architectures designed to extract value from vast and diverse datasets, facilitating recording, discovery, and analysis. In the future, big data can enable Vehicular Ad-Hoc Networks (VANETs) to support a wide range of potential services and applications, including those for smart cities and intelligent transportation [2-3]. This will fundamentally transform various aspects of social life, such as government, communications, transportation systems, and even human lifestyles [4]. Figure 1 illustrates the role of big data in VANET and its applications. For instance, cars and roadside equipment can gather information on road traffic and transmit it to a cloud server for the Intelligent Transportation System (ITS) [5]. Based on comprehensive traffic data, real-time traffic prediction and management can be performed to detect road anomalies, alleviate traffic congestion, and reduce pollution. Therefore, there is hope that VANETs will be employed in the near future to fulfill the demands of big data customers and mobile phone users, enabling a multitude of services and programs [6-7]. Big data trends can offer both opportunities and challenges for VANETs. On one hand, the utilization of big data in VANETs necessitates significant heterogeneous resources and entails diverse requirements. VANETs need to effectively support big data by providing high data transfer rates, ample network capacity, integration of different networks, and Quality of Service guarantees. Moreover, VANETs will play a vital role in the future for data gathering, storage, and processing [8]. On the other hand, big data in VANETs, encompassing GPS, vehicle tracking, road traffic statistics, and network measurements, is a valuable source of insights. When appropriately employed, big data can reveal various network characteristics, evaluate network performance, and employ state-of-the-art methods such as big data analytics and machine learning techniques for network management [9-11]. The objective of this article is to examine the impact of big data on VANETs, present new challenges and opportunities, and discuss relevant approaches to address these issues. Hence, this article concentrates on two interconnected subjects: the effectiveness of supporting big data in VANETs and leveraging big data to enhance the understanding and advancement of VANETs [12].

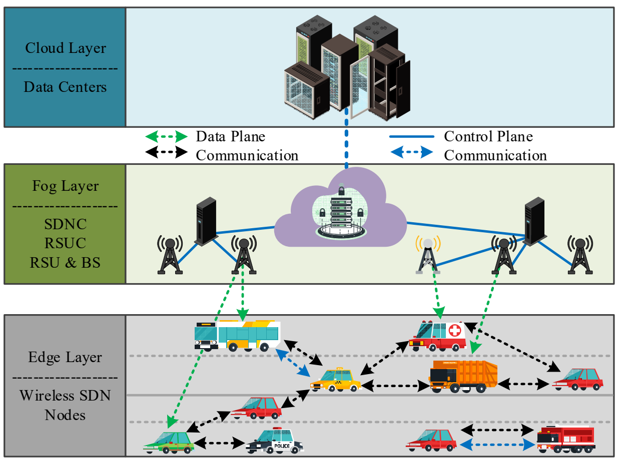

Fig. 1. Big Data in VANETs and applications

II. Characteristics of Big Data

Big data is characterized by four primary features, referred to as the "4V" (Figure 2 shows these characteristics) [13-14]:

* Assistant Professor, Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran.

979-8-3503-0198-4/23/$31.00 ©2023 IEEE

• Volume: A substantial amount of data is collected and analyzed, necessitating the use of distributed cloud storage. Security challenges arising from the large volume of data require the use of encryption and decryption rules, which can reduce performance speed.

• Velocity: Data is gathered and processed in real-time, requiring immediate data extraction. Achieving short response times while ensuring security and privacy is extremely challenging.

• Variety: Data can come from a wide range of sources, such as sound, images, text, web analytics data, and related data. Ensuring the security and integrity of data can be very challenging due to these data types.

• Veracity: Data may contain noise, so it must be detected and removed before analyzing it to obtain an accurate dataset.

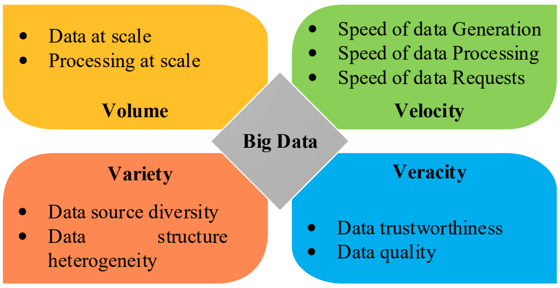

Fig. 2. 4V characteristics of big data

III. Big Data in Vehicular Networks

Lately, all departments have begun to work with digital data that contains a vast amount of information. Departments dealing with big data include networking, engineering, healthcare, transportation, retail, telecommunications, government services, media, entertainment, and video surveillance [15]. This section centers on the concept of VANETs in the transportation system, which has real-time access to big data for safe and efficient traffic flow planning. Big data in VANETs is obtained from multiple heterogeneous sources that provide various features, such as volume, structure, and value, causing processing delays. Big data in VANETs is classified based on the data resources as follows:

• *Vehicle sensor data*: In order to gather data about the vehicle and its surroundings, modern automobiles are fitted with a variety of sensors (such as tire pressure sensors, speedometers, etc.). Numerous applications, including online car diagnostics, improved traffic safety, intelligent charge planning, and accident detection, may be made possible by the data from these sensors [16].

• *Global Positioning System (GPS) data*: Vehicles' accurate locations and structured data, such as longitude, latitude, altitude, and speed, may be provided via GPS devices. Numerous uses for GPS data include navigation, traffic control, route optimization, content storage, and vehicle sharing. Moreover, the collection of vehicle data obtained through long-term GPS tracking in a geographical area can be used for analyzing VANET features (like network connection) and designing effective procedures (like routing protocols for transportation networks that can tolerate delays and expanding radio access networks) [17].

• *Self-driving vehicle data*: Self-driving vehicles produce vast amounts of data. Precise awareness of the surroundings is crucial for autonomous technology to make informed decisions regarding vehicle control. These vehicles are equipped with advanced devices such as cameras and Light Detection and Ranging Sensors (LIDARs), which generate substantial amounts of data. Conventional sensors are inadequate in providing essential data, such as real-time road views and precise distances on 3D maps. Therefore, these advanced devices consistently gather large amounts of data, including high-quality videos [18].

• *Mobile services data*: Mobile apps, such as video or audio streaming, online gaming, social networking, and user-generated content, need to create large amounts of data to provide entertainment and work in the car [19].

• *Real-time data*: VANETs data needs to be automatically updated and stored in databases, necessitating the utilization of large tables for decision-making in routing. These maps are a component of big data, which encompasses the extensive amount of data collected from various sources [20].

• *Variable density network*: This section encompasses a diverse range of structured and unstructured data from various sources, which is generated by different active GPS sensors and exhibits highly variable data densities [21].

• *Dynamic topology and motion modeling*: In VANETs, nodes are connected to vehicles. Due to the highly dynamic nature of these connections, the topology rapidly changes, leading to frequent network segmentation based on effective diameter and density. This characteristic of VANETs supports the need for high-speed transmission of large amounts of data. This encompasses not only the volume of information generated but also the processing time required for input data and the frequency of data delivery [22].

• *Large-scale network and high computational capability*: In a network with a substantial number of active GPS devices, having accurate data is crucial for predicting node behavior and making optimal routing decisions [23].

• *Anonymous users and potential infrastructure support*: Identifying vehicles in a specific area and identifying neighboring nodes with the support of existing infrastructure are fundamental requirements in most applications in vehicular networks. This necessitates accurate data collection methods, processing techniques, infrastructure, and secure data exchange to ensure the data's integrity and protection against unauthorized access or alterations throughout its existence [24].

IV. Big Data Analysis in VANETs

To enhance transportation systems and supply chain management, a multitude of factors related to the road, cars, and drivers can be measured [25]. Traffic flow may vary based on local weather conditions, car performance, and driving behavior, which can be derived from sensor data obtained from roads and vehicles. The collection of data from millions of kilometers of roads, vehicles, and drivers has the potential to yield remarkable results [26]. Currently, with the aid of modern "Big Data" technology, this data can be analyzed within a reasonable timeframe to derive meaningful insights. For processing large datasets in the petabyte range, the Hadoop framework is widely used as a distributed computing environment specifically designed for big data [27]. Hadoop is a distributed computing platform designed to facilitate big data processing. It offers distributed computing and storage capabilities, enabling programs to operate with millions of nodes and large datasets [28]. Hadoop is extensively utilized across various commercial operating systems. The primary components of Hadoop include the Hadoop Distributed File System (HDFS) and MapReduce. Hadoop serves as a framework comprising software and libraries that enable the processing of vast amounts of distributed data [29]. Moreover, it can be implemented using affordable, standard computers.

A. Hadoop Distributed File System:

HDFS (Hadoop Distributed File System) is specifically designed to store massive volumes of data, ranging from petabytes to terabytes, across a cluster of machines. This distributed system ensures reliable data storage and is typically implemented using commodity hardware [30]. HDFS utilizes blocks to store files or segments of files, following a write-once, read-many-times approach for data access. Due to its intricate nature, utilizing HDFS requires the expertise of experienced programmers to develop complex read or write programs [31]. Additionally, as a logical data structure for data implementation, HDFS may not be readily accessible. To enhance its usability, it is advisable to leverage other well-known distributed systems in the big data domain, such as key-value pairs, document-oriented, graph, and columnar systems [32].

B. *MapReduce*

MapReduce is a scalable programming technique initially developed by Google, which allows for the concurrent processing of extensive data volumes on supercomputers. While MapReduce follows a simple principle, it can become complex, and implementing it using this model may be challenging due to structural intricacies across different algorithms. However, MapReduce provides a set of library functions that abstract away the underlying complexities, simplifying the programmer's perspective. Notable features of MapReduce include automatic task parallelization, load balancing, optimization of data transfer between disk and network, as well as handling machine failures for fault management [33-35].

Any enhancements made to a library are propagated to all the components that utilize it. This methodology comprises two primary steps known as "Map" and "Reduce." In the Map step, the main node receives inputs and divides them into smaller sub-problems. Subsequently, the main node assigns these sub-problems to individual nodes responsible for executing the tasks. This process may be recursively repeated, resulting in a multi-level structure. The solutions to these sub-problems are then processed and sent back to the main node. In the Reduce step, the main node collects the results and combines them to generate the final output. During this phase, operations like filtering, summarizing, or transforming can be performed on the results [36].

In Figure 3, these two steps are depicted along with the corresponding explanations. The Map function operates on a pair of orders (key, value), transforming it into a list of ordered pairs.

$(Map\ (k_1, v_1)) \geq list\ (k_2, v_2)$

Once all the pairings with the same key have been collected from the lists, the MapReduce framework groups them together. Subsequently, a group is created for each unique key generated. Each group then undergoes the Reduce function for further processing.

$(Reduce\ (k_2, list\ (v_2)) \geq list\ (v_3)$

Finally, the MapReduce framework transforms a list of (keys, values) into a list of values.

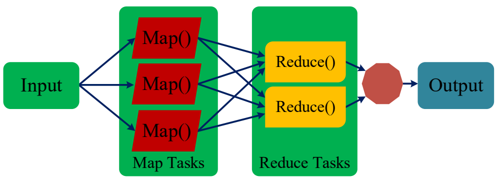

Fig. 3. An example of the MapReduce model

V. Working with Big Data in Vehicular Networks

The vast amount of intelligent data in VANETs can be leveraged to characterize and evaluate their performance, as well as to develop new protocols [37]. Figure 4 illustrates the utilization of two prevalent datasets in VANETs: vehicle mobility tracking data and VANET measurement data. These datasets provide valuable insights into operational models, mobility patterns, and vehicle motion prediction. Armed with this knowledge, intelligent protocol design and enhanced VANET features can be achieved [38].

A. *Moving Vehicle Data Tracking*

It goes without saying that VANETs face challenges due to the increased mobility of vehicles. However, network mobility also presents advantages, such as mobility-aware protocols and the ability to disseminate information with delay tolerance. By analyzing vehicle mobility data, valuable information, such as practical mobility models, spatial and temporal density distributions, and network connections, can be obtained [39]. Currently, large-scale taxi monitoring information from several cities, including San Francisco, Shanghai, and Shenzhen, is stored in numerous databases. The primary contents of these databases include the time of movement, vehicle speed, driving route, and vehicle position, and they can be utilized for future VANET research [40].

In the design of VANET protocols, location-based mobility models and performance evaluation are widely utilized [41]. However, the tracking data of vehicles are prone to errors due to the time intervals at which they are reported. Moreover, there exists a spatial gap between two consecutive rows of data, necessitating the need for data preprocessing mechanisms. For instance, leveraging the predictable nature of vehicle movement, the distance between two vehicles can be estimated by analyzing maps, road signs, and tracking data of passing vehicles [42]. By incorporating corrected tracking information, a more reliable mobility model can be generated.

B. *VANETs Measurement Data*

Communication measurement plays a crucial role in understanding the characteristics of VANETs, considering various factors that impact the modeling process, such as cellphone networks, pedestrians, the environment, and obstacles. For conducting experiments and gathering accurate measurement data, communication devices are installed in vehicles and roadside locations, utilizing the IEEE 802.11p protocol [43]. The collected data from these measurements are diverse and based on specific features of interest. These tests are conducted across a range of environments, including urban, suburban, rural, and highway settings [44].

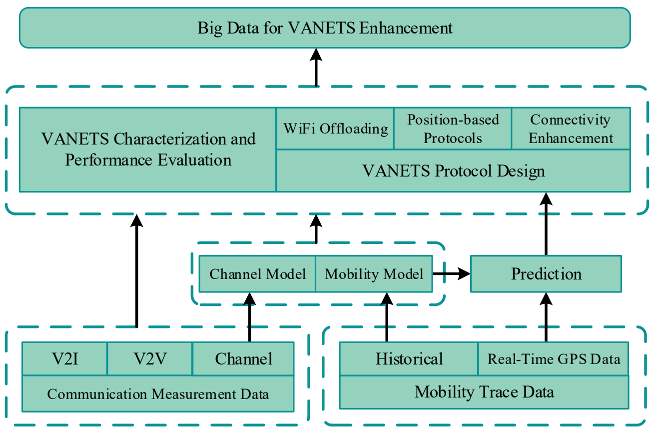

Fig. 4. Deployment of big data in vehicular networks

VI. Support for Big Data in VANET

To ensure the efficient operation of a big data system, four crucial components must be well-supported: information gathering, storage, transfer, and computation [45]. In VANETs, vehicle sensors are utilized to collect and store raw data. Data processing is then performed to extract valuable information from the large volume of raw data. Once the raw or processed data is collected, there is a requirement to transfer the data to data storage systems for further analysis and processing. Hence, VANETs must possess the capability to support these essential functions of big data [46].

Traditional VANETs rely on Dedicated Short-Range Communications (DSRC) and IEEE 802.11p technologies for data transmission. However, the data transfer primarily depends on multi-purpose Media Access Control (MAC) protocols and routing protocols [47]. Nonetheless, decentralized protocols and bandwidth constraints impose limitations on these traditional VANET technologies. These limitations manifest as network resource depletion and a lack of flexibility in supporting large data with diverse Quality of Service (QoS) requirements [48]. Consequently, these technologies struggle to meet the hardware requirements of large data applications. Furthermore, issues related to energy efficiency, storage, and computing capacities in current VANETs have not been adequately addressed [49].

In this section, various technologies for enhanced support of large data, including 5G technology, will be discussed. Figure 5 illustrates how 5G macro cells offer comprehensive communication support, while small 5G communications, local wireless networks, radio networks, and device-to-device communications are also recommended. Furthermore, data pipelines are proposed to facilitate the cost-effective transmission of large VANET data [50].

Fig. 5. Large data support in the global vehicular network

A. 5G Technology

5G technology serves as a supportive platform for handling large-scale data in Vehicular Ad-Hoc Networks (VANETs), a wide mobile phone network. Similar to the fourth-generation Long-Term Evolution (LTE) network, which aims to cater to the increasing volume of information and diverse mobile phone services with various Quality of Service (QoS) requirements, next-generation networks like 5G offer solutions to address these challenges [51]. Leveraging Software Defined Networking technologies, 5G networks are designed as a service delivery platform. 5G networks exhibit impressive performance metrics, including data transmission rates of up to 10 gigabytes per second and latency of less than 1 millisecond. Moreover, they effectively support low-power and highly reliable communications for emerging internet applications [52]. 5G has identified the requirements for common Vehicle-to-everything (V2X) scenarios, such as transportation, advanced driving, advanced sensors, and remote driving (shown in Figure 6). To accurately characterize and support diverse services, 5G defines three application categories with their respective key performance metrics: Improved Mobile Broadband, Ultra-Reliable and Low Latency Communication, and widespread Machine Type Communication. Alongside the aforementioned core technologies, these categories ensure optimal performance for data collection and the transmission of large VANET datasets [53-54].

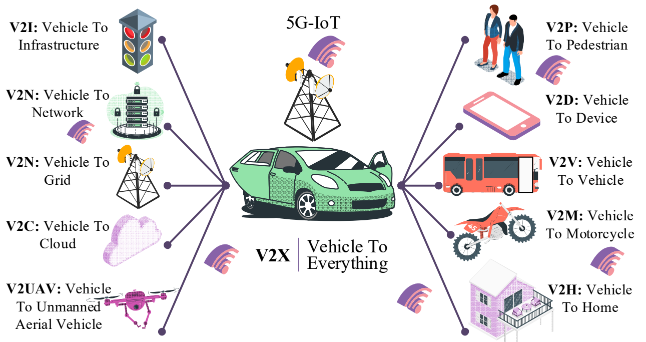

Fig. 6. View of the application of 5G-IoT technology in Vehicle-to-Everything (V2X) communication

With advancements such as advanced channel plans, millimeter waves, and small and large cell networks, Enhanced Mobile Broadband (eMBB) has the capability to deliver data at a maximum speed of 10 gigabits per second with a mobile data volume of 10 $Tb/s/km^2$ [55]. Consequently, utilizing a 5G network enables the support and realization of innovative applications that involve large volumes of vehicle data. Transmission of safety messages, for example, requires extremely low latency and very high reliability from information services in VANETs. 5G can meet the requirements of Ultra-Reliable and Low Latency Communications (URLLC), providing latency of fewer than 5 milliseconds and reliability exceeding 99.999% [56-57].

Considering emerging technologies like machine-to-machine communications and the Internet of Things, which often have limited bandwidth, massive Machine-Type Communications (mMTC) aims to support widespread device connectivity with low power consumption and low latency [58]. In VANETs, a significant amount of data is generated by lightweight devices, such as sensors installed in vehicles or along roads. 5G technologies can support these massive connections simultaneously, ensuring reliable data transmission and extended device battery life, thus facilitating efficient data collection services. Figure 7 shows a perspective of various ways of communication between the V2X with 5G-NB-IoT technology in the smart city [59-61].

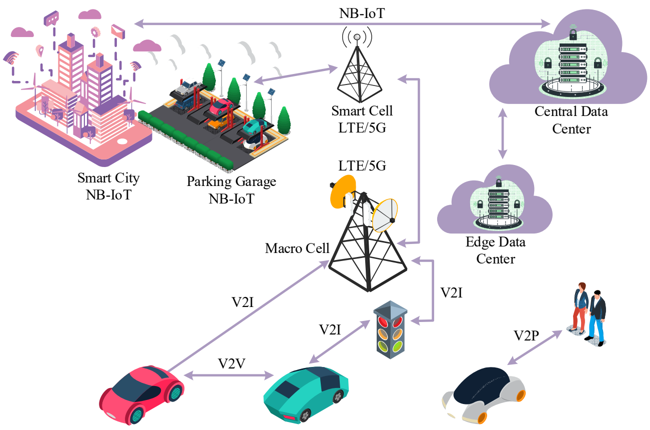

Fig. 7. Types of communication between V2X and 5G-NB-IoT technology in smart city

B. Data Pipes

While 5G networks have the potential to significantly increase network capacity, the addition of more data introduces a substantial load to the network, increasing the risk of congestion [62]. Moreover, the deployment and commercialization of 5G networks began in 2020 and will require additional time to become widely available. Consequently, 4G LTE networks will soon face challenges in efficiently handling large data volumes [63]. Additionally, transmitting a substantial amount of information over mobile networks can result in significant costs. Therefore, data pipelines are necessary to support the transfer of large data, and cost-effective alternatives such as device-to-device (D2D) communications, cognitive radio networks (CRNs), and wireless local area networks (WLANs) can be utilized to offload large data from cellular networks to VANET [64].

VII. Conclusion

Many researchers are currently interested in big data due to the generation of a large amount of data in daily activities. Previous studies have identified volume, variety, velocity, accuracy, and value as significant characteristics of big data. This paper focuses on analyzing big data in VANET and utilizes a hybrid approach using the Hadoop framework. However, it is important to note that Hadoop and big data are distinct concepts that cannot be directly compared. Hadoop can be seen as a potential solution for managing big data challenges. The integration of Hadoop and big data in VANET leads to the development of diverse services with significant potential. While numerous studies have been conducted by researchers, a definitive solution has yet to be found. Nonetheless, the existing background research will undoubtedly contribute to identifying new solutions to address current weaknesses.

References

[1] Shahinzadeh, H., Zanjani, S. M., Moradi, J., Fayaz-dastgerdi, M. H., Yaïci, W., & Benbouzid, M. (2022, October). The Transition Toward Merging Big Data Analytics, IoT, and Artificial Intelligence with Blockchain in Transactive Energy Markets. In *2022 Global Energy Conference (GEC)* (pp. 241-246). IEEE.

[2] Mchergui, A., Moulahi, T., & Zeadally, S. (2022). Survey on artificial intelligence (AI) techniques for vehicular ad-hoc networks (VANETs). *Vehicular Communications*, *34*, 100403.

[3] Rashid, K., Saeed, Y., Ali, A., Jamil, F., Alkanhel, R., & Muthanna, A. (2023). An Adaptive Real-Time Malicious Node Detection Framework Using Machine Learning in Vehicular Ad-Hoc Networks (VANETs). *Sensors*, *23*(5), 2594.

[4] Kabalcı, Y., & Ali, M. (2019, June). Energy Internet: A Novel Vision for Next-Generation Smart Grid Communications. In *2019 1st Global Power, Energy and Communication Conference (GPECOM)* (pp. 96-100). IEEE.

[5] Moazzami, M., Sheini-Shahvand, N., Kabalci, E., Shahinzadeh, H., Kabalci, Y., & Gharehpetian, G. B. (2021, October). Internet of things architecture for intelligent transportation systems in a smart city. In *2021 3rd Global Power, Energy and Communication Conference (GPECOM)* (pp. 285-290). IEEE.

[6] Shahwani, H., Shah, S. A., Ashraf, M., Akram, M., Jeong, J. P., & Shin, J. (2022). A comprehensive survey on data dissemination in Vehicular Ad Hoc Networks. *Vehicular Communications*, *34*, 100420.

[7] Gaouar, N., & Lehsaini, M. (2021). Toward vehicular cloud/fog communication: A survey on data dissemination in vehicular ad hoc networks using vehicular cloud/fog computing. *International Journal of Communication Systems*, *34*(13), e4906.

[8] Ameur, A. I., Lakas, A., Yagoubi, M. B., & Oubbati, O. S. (2022). Peer-to-peer overlay techniques for vehicular ad hoc networks: Survey and challenges. *Vehicular Communications*, *34*, 100455.

[9] Singh, D., Maurya, A. K., & Yadav, R. S. (2022). A trust-based clustering approach to form stable clusters in vehicular ad hoc networks. *Journal of Ambient Intelligence and Humanized Computing*, 1-20.

[10] Gaba, P., & Raw, R. S. (2022). B-VANET: A blockchain-based vehicular ad-hoc network for data validation. *Peer-to-Peer Networking and Applications*, *15*(6), 2650-2669.

[11] Hussein, N. H., Yaw, C. T., Koh, S. P., Tiong, S. K., & Chong, K. H. (2022). A Comprehensive Survey on Vehicular Networking: Communications, Applications, Challenges, and Upcoming Research Directions. *IEEE Access*, *10*, 86127-86180.

[12] Rajput, N. S., Singh, U., Dua, A., Kumar, N., Rodrigues, J. J., Sisodia, S., ... & Lakys, Y. (2023). Amalgamating Vehicular Networks With Vehicular Clouds, AI, and Big Data for Next-Generation ITS Services. *IEEE Transactions on Intelligent Transportation Systems*.

[13] Zanjani, S. M., Zanjani, S. H., Shahinzadeh, H., Rezaei, Z., Kaviani-Baghbaderani, B., & Moradi, J. (2022, November). Big Data Analytics in IoT with the Approach of Storage and Processing in Blockchain. In *2022 6th Iranian Conference on Advances in Enterprise Architecture (ICAEA)* (pp. 1-6). IEEE.

[14] Shi, Y. (2022). Advances in big data analytics. *Adv Big Data Anal*.

[15] Mahi, M. J. N., Chaki, S., Ahmed, S., Biswas, M., Kaiser, S., Islam, M. S., ... & Whaiduzzaman, M. (2022). A review on VANET research: Perspective of recent emerging technologies. *IEEE Access*.

[16] Xu, X., Wang, Y., & Wang, P. (2022). Comprehensive review on misbehavior detection for vehicular ad hoc networks. *Journal of Advanced Transportation*, *2022*.

[17] Ahmad, W., Husnain, G., Ahmed, S., Aadil, F., & Lim, S. (2023). Received Signal Strength-Based Localization for Vehicle Distance Estimation in Vehicular Ad Hoc Networks (VANETs). *Journal of Sensors*, *2023*.

[18] Younas, S., Rehman, F., Maqsood, T., Mustafa, S., Akhunzada, A., & Gani, A. (2022). Collaborative Detection of Black Hole and Gray Hole Attacks for Secure Data Communication in VANETs. *Applied Sciences*, *12*(23), 12448.

[19] Rashid, S. A., Hamdi, M. M., & Alani, S. (2020, June). An overview on quality of service and data dissemination in VANETs. In *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1-5). IEEE.

[20] Shen, J., Liu, D., Chen, X., Li, J., Kumar, N., & Vijayakumar, P. (2019). Secure real-time traffic data aggregation with batch verification for vehicular cloud in VANETs. *IEEE Transactions on Vehicular Technology*, *69*(1), 807-817.

[21] Sindhwani, M., Sachdeva, S., Arora, K., Yoon, T., Yoo, D., Joshi, G. P., & Cho, W. (2022). Soft Computing Techniques Aware Clustering-Based Routing Protocols in Vehicular Ad Hoc Networks: A Review. *Applied Sciences*, *12*(15), 7922.

[22] Han, Z., Xu, C., Ma, S., Hu, Y., Zhao, G., & Yu, S. (2023). DTE-RR: Dynamic Topology Evolution-Based Reliable Routing in VANET. *IEEE Wireless Communications Letters*.

[23] Khattak, H. A., Islam, S. U., Din, I. U., & Guizani, M. (2019). Integrating fog computing with VANETs: A consumer perspective. *IEEE Communications Standards Magazine*, *3*(1), 19-25.

[24] Zhou, X., He, D., Khan, M. K., Wu, W., & Choo, K. K. R. (2022). An Efficient Blockchain-based Conditional Privacy-preserving Authentication Protocol for Vanets. *IEEE Transactions on Vehicular Technology*.

[25] Zaraket, K., Bennis, I., Harb, H., Jaber, A., & Abouaissa, A. (2022). A Comparative Study of Recent Advances in Big Data Analytics in Vehicular Ad Hoc Networks. *2022 International Wireless Communications and Mobile Computing (IWCMC)*, 841-846.

[26] Gao, J., Manogaran, G., Nguyen, T. N., Kadry, S., Hsu, C. H., & Kumar, P. M. (2022). A vehicle-consensus information exchange scheme for traffic management in vehicular ad-hoc networks. *IEEE Transactions on Intelligent Transportation Systems*, *23*(10), 19602-19612.

[27] Zang, M., & Yan, Y. (2021, April). Machine learning-based intrusion detection system for big data analytics in VANET. In *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)* (pp. 1-5). IEEE.

[28] Aljehane, N. O., & Mansour, R. F. (2022). Big data analytics with oppositional moth flame optimization based vehicular routing protocol for future smart cities. *Expert Systems*, *39*(5), e12718.

[29] Mouad, T., Driss, L. M., & Mustapha, K. (2021). Big data traffic management in vehicular ad-hoc network. *International Journal of Electrical and Computer Engineering*, *11*(4), 3483.

[30] Ikegwu, A. C., Nweke, H. F., Anikwe, C. V., Alo, U. R., & Okonkwo, O. R. (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing*, *25*(5), 3343-3387.

[31] Beena, T. L. A., Sumathi, D., Krishnamurthi, R., & Selvaraj, P. (2020). Big Data Management Solutions for IoT: Case Study–Connected Car. In *The Internet of Things and Big Data Analytics* (pp. 301-318). Auerbach Publications.

[32] Govindasamy, A., & Thiruvengadam, S. J. (2020). Data Science in Vehicular Ad-Hoc Networks. In *IoT and Cloud Computing Advancements in Vehicular Ad-Hoc Networks* (pp. 191-202). IGI Global.

[33] Tantaoui, M., Laanaoui, M. D., & Kabil, M. (2020). Vehicle Traffic Management with the Help of Big Data Technologies. In *Innovations in Smart Cities Applications Edition 3: The Proceedings of the 4th*

International Conference on Smart City Applications 4 (pp. 894-905). Springer International Publishing.

[34] Ikegwu, A. C., Nweke, H. F., Anikwe, C. V., Alo, U. R., & Okonkwo, O. R. (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing*, *25*(5), 3343-3387.

[35] Hamdaoui, I., El Fissaoui, M., El Makkaoui, K., & El Allali, Z. (2022, March). An intelligent traffic monitoring approach based on Hadoop ecosystem. In *2022 5th International Conference on Networking, Information Systems and Security: Envisage Intelligent Systems in 5g//6G-based Interconnected Digital Worlds (NISS)* (pp. 1-6). IEEE.

[36] Beena, T. L. A., Sumathi, D., Krishnamurthi, R., & Selvaraj, P. (2020). Big Data Management Solutions for IoT: Case Study–Connected Car. In *The Internet of Things and Big Data Analytics* (pp. 301-318). Auerbach Publications.

[37] Gupta, N., Manaswini, R., Saikrishna, B., Silva, F., & Teles, A. (2020). Authentication-based secure data dissemination protocol and framework for 5G-enabled VANET. *Future Internet*, *12*(4), 63.

[38] Sbayti, O., & Housni, K. (2023, March). Evaluations of Some Routing Protocols Metrics in VANET. In *Proceedings of the 6th International Conference on Big Data and Internet of Things* (pp. 524-536). Cham: Springer International Publishing.

[39] Garg, S., Singh, A., Kaur, K., Aujla, G. S., Batra, S., Kumar, N., & Obaidat, M. S. (2019). Edge computing-based security framework for big data analytics in VANETs. *IEEE Network*, *33*(2), 72-81.

[40] Rao, R. S., Singh, N., Kaiwartya, O., & Das, S. (Eds.). (2020). *Cloud-Based Big Data Analytics in Vehicular Ad-Hoc Networks*. IGI Global.

[41] Tantaoui, M., Laanaoui, M. D., & Kabil, M. (2020). Vehicle Traffic Management with the Help of Big Data Technologies. In *Innovations in Smart Cities Applications Edition 3: The Proceedings of the 4th International Conference on Smart City Applications 4* (pp. 894-905). Springer International Publishing.

[42] Khatri, S., Vachhani, H., Shah, S., Bhatia, J., Chaturvedi, M., Tanwar, S., & Kumar, N. (2021). Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges. *Peer-to-Peer Networking and Applications*, *14*, 1778-1805.

[43] Shah, A. S., Ilhan, H., & Tureli, U. (2019, July). Modeling and performance analysis of the IEEE 802.11 P MAC for VANETs. In *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)* (pp. 393-396). IEEE.

[44] Khan, S., Sharma, I., Aslam, M., Khan, M. Z., & Khan, S. (2021). Security challenges of location privacy in VANETs and state-of-the-art solutions: A survey. *Future Internet*, *13*(4), 96.

[45] Moradi, J., Shahinzadeh, H., Nafisi, H., Marzband, M., & Gharehpetian, G. B. (2019, December). Attributes of big data analytics for data-driven decision making in cyber-physical power systems. In *2020 14th international conference on protection and automation of power systems (IPAPS)* (pp. 83-92). IEEE.

[46] Bhuvana, J., HASHMI, H., Adhvaryu, R., Kashyap, S., Kumari, S., & Wadhwa, D. (2023). Intelligent analytics algorithms in breach detection systems for securing VANETs and data for smart transportation management.

[47] Malik, S., & Sahu, P. K. (2019). A comparative study on routing protocols for VANETs. *Heliyon*, *5*(8), e02340.

[48] Ayaz, M., Iqbal, J., Adnan, M., Hussain, S. S., Alabrah, A., Amin, N. U., ... & Md, S. (2022). A Content Dissemination Technique Based on Priority to Improve Quality of Service of Vehicular Ad Hoc Networks. *Journal of Advanced Transportation*, *2022*.

[49] Pereira, J., Ricardo, L., Luís, M., Senna, C., & Sargento, S. (2019). Assessing the reliability of fog computing for smart mobility applications in VANETs. *Future Generation Computer Systems*, *94*, 317-332.

[50] Hussain, R., Hussain, F., & Zeadally, S. (2019). Integration of VANET and 5G Security: A review of design and implementation issues. *Future Generation Computer Systems*, *101*, 843-864.

[51] Gupta, S., & Khaitan, V. (2021). Reliability and survivability analysis of long-term evolution vehicular ad-hoc networks: an analytical approach. *Journal of Network and Systems Management*, *29*, 1-34.

[52] Storck, C. R., & Duarte-Figueiredo, F. (2020). A survey of 5G technology evolution, standards, and infrastructure associated with vehicle-to-everything communications by internet of vehicles. *IEEE access*, *8*, 117593-117614.

[53] Ashraf, S. A., Blasco, R., Do, H., Fodor, G., Zhang, C., & Sun, W. (2020). Supporting vehicle-to-everything services by 5G new radio release-16 systems. *IEEE Communications Standards Magazine*, *4*(1), 26-32.

[54] Thaker, J., Jadav, N. K., Tanwar, S., Bhattacharya, P., & Shahinzadeh, H. (2022, September). Ensemble Learning-based Intrusion Detection System for Autonomous Vehicle. In *2022 Sixth International Conference on Smart Cities, Internet of Things and Applications (SCIoT)* (pp. 1-6). IEEE.

[55] Gupta, R., Jadav, N. K., Nair, A., Tanwar, S., & Shahinzadeh, H. (2022, September). Blockchain and AI-based Secure Onion Routing Framework for Data Dissemination in IoT Environment Underlying 6G Networks. In *2022 Sixth International Conference on Smart Cities, Internet of Things and Applications (SCIoT)* (pp. 1-6). IEEE.

[56] Han, M., Lee, J., Rim, M., & Kang, C. G. (2021). Dynamic bandwidth part allocation in 5G ultra reliable low latency communication for unmanned aerial vehicles with high data rate traffic. *Sensors*, *21*(4), 1308.

[57] González, E. E., Morales, F. D., Coral, R., & Toasa, R. M. (2021). Fifth-generation networks and vehicle-to-everything communications. In *Information Technology and Systems: ICITS 2021, Volume 1* (pp. 350-360). Springer International Publishing.

[58] Terzija, V., Cetenovic, D., Milovanovic, D. A., & Bojkovic, Z. S. (2020). Integration of Multienergy Systems into Smart Cities: Opportunities and Challenges for Implementation on 5G-Based Infrastructure. *5G Multimedia Communication*, 211-233.

[59] Xie, Q., Ding, Z., & Zheng, P. (2023). Provably Secure and Anonymous V2I and V2V Authentication Protocol for VANETs. *IEEE Transactions on Intelligent Transportation Systems*.

[60] Mlongeni, M., Abu-Mahfouz, A. M., & Hancke, G. P. (2022, October). Towards Building a Secure NB-IoT Environment on 5G Networks: A User and Device Access Control System Review. In *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society* (pp. 1-6). IEEE.

[61] Tahir, M. N., Leviäkangas, P., & Katz, M. (2022). Connected vehicles: V2V and V2I road weather and traffic communication using cellular technologies. *Sensors*, *22*(3), 1142.

[62] Prasad, V. K., Tanwar, S., & Bhavsar, M. D. (2021). Advance cloud data analytics for 5G enabled IoT. *Blockchain for 5G-Enabled IoT: The new wave for Industrial Automation*, 159-180.

[63] Berto, F., Ardagna, C., Torrente, M., Manenti, D., Ferrari, E., Calcante, A., ... & Ciani, L. (2022, December). A 5G-IoT enabled Big Data infrastructure for data-driven agronomy. In *2022 IEEE Globecom Workshops (GC Wkshps)* (pp. 588-594). IEEE.

[64] Alalewi, A., Dayoub, I., & Cherkaoui, S. (2021). On 5G-V2X use cases and enabling technologies: A comprehensive survey. *Ieee Access*, *9*, 107710-107737.