Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 1, January 2024, pp. 314~323
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i1.pp314-323

Big data vehicle density management in vehicular ad-hoc network

Mouad Tantaoui[1], Mehdi Moukhafi[1], Idriss Chana[2]
[1]Department of Sciences, IEVIA Team, Image Laboratory, ENS, Moulay Ismail University of Meknes, Meknes, Morocco
[2]SCIAM Team, Image Laboratory, ESTM, Moulay Ismail University of Meknes, Meknes, Morocco

Article Info

Article history:

Received Sep 19, 2023
Revised Oct 31, 2023
Accepted Nov 2, 2023

Keywords:

Big data
Intelligent transport system
Machine learning
Traffic congestions prediction
Traffic management
Vehicular ad-hoc network

ABSTRACT

Smart city project is today a domain of interest to community research which play well-known role in road traffic management. Data exchange became complicated in terms of capacity in the intelligent transport system (ITS), and without the raise of big data, the treatment is very difficult to manage. vehicular ad-hoc network (VANETs) faces many challenges mainly the voluminous data generated by different actors of VANET environment. We propose a real time anomalies detection system in an instantaneous way with parallel data treatment. The system method intends to compute precisely vehicle density at each section on each road, which help to handle the traffic and forward to vehicles information about the road and the best safe path to reach their destination. Also, we build anomalies prediction system based on machine learning framework, it is a good solution for avoiding traffic congestion and limiting the risk of accidents. The simulation results demonstrate that the proposed system method reduces congestion greatly by taking into account the load balancing and therefore avoids saturation and reduces accidents. It should also be noted that the results obtained show that the system is characterized by low latency and high accuracy.

Corresponding Author:

Mouad Tantaoui
Department of Sciences, IEVIA Team, Image laboratory, ENS, Moulay Ismail University of Meknes
Meknes, Morocco
Email: tantaoui.mouad@gmail.com

1. INTRODUCTION

Intelligent transport systems (ITS) play well-known role in road traffic management through optimization and information communication. Though, data exchange, moreover, begin to be questionable in terms of capacity, with the raise of big data and more player get involved in the ITS, road and highway jams, different vehicles type with more and more vehicles in the road. However, what make the communication (vehicle-vehicle and vehicle-infrastructure communication [1]) between those players more difficult is the variety and amount of data, beside this, the real-time system is what make the ITS more efficient. Those issues make another related context more fragile for jams, junction bad design, and transport incidents. Overall, those issues impact the proper functioning of an infrastructure, through the congestion they generate will rise more issues. Thus, the management of traffic is a key for several problems. Several research projects and applications made vehicular ad-hoc network (VANETs) more suitable for ITS implementation. One of them is the raise of on-board unit (OBU), which make communication between mobile nodes more efficient. Also, there is roadside units via dedicated short-range communications (DSRC) that is suitable for fixed and non-fixed node communication with the low level and high-level communication gateway. The centralization of those applications is more demanded for making management and the deployment more efficient. Fazio *et al.* [2], linked the problem of finding the shortest path features to big data characteristics; they evaluated the performance of Dijkstra algorithm utilized for routing on hadoop map reduce environment to improve the computation time. Their goal is to reduce the routing path computation time in a network environment containing several points linked to each other. Rahim *et al.* [3], the scholars effectively address the quandary of surpassing a larger automobile by fabricating an impromptu linkage founded on 5G technology with the target vehicle that necessitates overtaking.

In the study conducted by Cheng *et al.* [4], an extensive volume of speed-related data was subjected to big data analytics to uncover intricate patterns and interdependencies among various factors and vehicle speed. The analysis leveraged the combination of big data analytics and the adaptive neuro-fuzzy inference system (ANFIS) model, which served as the foundation for the proposed algorithm. Al-Najada and Mahgoub [5], designed a real time system based on Lambda architecture (LA) proposed in [6]. It performs online streamed data emanating from vehicles within road and data of real time average speed coming from vehicles sensors to range of different road side units (RSU) to firstly compute precise estimated time of arrival (ETA) utilizing a linear regression (LR) model. It also predicts traffic anomalies utilizing two main classifiers of machine learning techniques and updates ETA in case of having traffic anomaly detected. Mahajan and Kaur [7], the principal goal of this work is to simulate ITS scenario and compute the traffic density for various cities on different roads of the city and establish a comparative evaluation of traffic densities in various cities relying on scenarios and to ameliorate the precision of the density prediction mechanism in ITS environment. Lin and Wang [8], to counter selfish and malicious vehicles, the authors designed a system whose goal is to send reports in dynamic manner. The designed system relies on an encryption technique to send messages.

Tantaoui *et al.* [9], proposed an architecture system for near real time massive data computing in vehicular ad-hoc networks, which utilize latest trends and evolvement with respecting emerging big-data patterns. The proposed architecture includes centralized data storage technique for batch treatment and distributed data storage technique for streaming computing in real time processing. Tantaoui *et al.* [10], establish a prediction technique in real time utilizing big data technologies to ameliorate the VANET management. They used traffic density and average speed to predict the risk of vehicle accident instantly with parallel data processing. Zhao *et al.* [11], the trichotomy AdaBoost algorithm was utilized to train many classifiers from the experimental dataset and after that harmonizing them into a strong classifier to construct the prediction model. Malik *et al.* [12], they undertake a comprehensive analysis of resource allocation and mode selection mechanisms in the context of device-to-device (D2D) and cooperative communication techniques. The proposed technique aims to improve the efficiency and performance of D2D and cooperative communications in a cellular network environment. Bukhari *et al.* [13], introduce a method that leverages the apache hadoop platform to address the challenges associated with managing the escalating volume of demographic data. The system is designed to effectively handle the complexities and scale of demographic data management. In the research study referenced as article [14], the assessment of five classifiers was conducted on two extensive sets of workshop data using the H2O and WEKA extraction tools. Notably, the experimental findings revealed that Naïve Bayes (NB) yielded optimal outcomes, exhibiting the shortest computational time along with commendable area under the curve (AUC) and accuracy (ACC) metrics. Furthermore, Alinani [15], internet architecture called named data networking (NDN) has emerged as a promising solution to overcome the limitations prevalent in current VANET networks. NDN offers a robust framework for various applications, including object tracking and real-time vehicle monitoring. The research project described in article [16] presents a novel approach to enhance the efficiency and security of data collection in vehicular networks. Central to this approach is the utilization of asymmetric encryption, which facilitates secure communication between the vehicles and RSUs. Prior to the collection of vehicle data by the RSU, a robust and secure authentication process is established between the vehicle and the RSU.

Cheng *et al.* [17], take a first-hand look at VANET technologies for transmitting big data with good performance. Then, they established a case study which contains machine learning diagrams in order to accurately detect poor communication state. Contreras-Castillo *et al.* [18], analyze and talk about big data alternatives that can be proposed to address some of the emerging challenges of VANETs. In the scholarly paper cited as [19], the authors delve into the comprehensive analysis of channel estimation techniques designed specifically for massive multiple-input multiple-output (MIMO) systems. The study involves a meticulous comparison among various channel estimation methodologies, shedding light on their respective strengths and limitations. Furthermore, Du *et al.* [20] takes a historical perspective on the evolutionary trajectory of data handling systems. The article also delves into the contemporary landscape of big data handling systems, particularly focusing on aspects such as data storage, model frameworks, and query engines. Lastly, Shahin *et al.* [21], elaborated an optimization strategy which manage the configuration and localization of fog. The main idea proposed is to construct a system that allocate the processing capacity of each node included in fog relying on the quantity of processing demand. Tantaoui *et al.* [22], proposed a method that aim to forward to vehicles the accurate estimated time of arrival all along their path. With the support of big data technologies, they succeeded in ameliorating traffic management. Raj and D'Souza [23], analyze the hadoop architecture and hadoop eco system. Mouad *et al.* [24], proposed a big data-based technique to balance different roads of the city by directing vehicles to their destination relying on the estimated time of arrival. The research paper identified as [25], introduces a method that proposes a flexible transfer especially for VANET by including fog computing and software defined networking (SDN) technologies. The experiment evaluation proves a good transfer performance amelioration of the proposed solution. On the other hand, Khudhair *et al.* [26], they establish a clustering strategy to ameliorate the performance of wireless networks and having the clusters more stable.

These methods mentioned above have good impacts on traffic management in various aspects, however, they have some limitations as the VANET network is distinguished by its high mobility, it is essential that there will be a way to collect and process the information very quickly and have them updated as frequently as possible. In the method of the implementation of dijkstra's algorithm on hadoop mapreduce, it makes it possible to find the shortest path to reach the destination, but this metric does not allow to have a faster path to the desired destination in terms of time, it is obligatory to add other metrics to get good results. Speed prediction allows you to know the speed of vehicles on different roads in the future, but speed alone cannot detect risky situations which will depend on other factors. The method of calculating the estimated time of arrival of vehicles in real-time is a good method of knowing exactly the time of arrival of vehicles, but it lacks a way to select the optimal paths and direct vehicles to an optimal path. Estimating the traffic density for different cities on different roads is not sufficient to have good control for traffic management. For the problem of trajectory planning for fully autonomous vehicles, this work presents a method of optimal and safe trajectory selection in autonomous vehicles, the weak point of this method is in the criterion of trajectory selection, the estimated time of arrival is calculated for each segment by dividing the length of each segment by the speed limit assigned to this segment, it does not take into account the variations in the state of the traffic at time t. Our method takes into account frequent changes in traffic conditions in real-time, it makes it possible to control the traffic, detect risky situations on the different roads and then direct vehicles to optimal paths in terms of speed and safety, and it can also predict accidents and warn vehicles of the risk. In our previous method [24], its priority was speed more than safety; its main metric was the travel time on different sections of road. On the other hand, in this paper, we built a method whose main metric is the density of the vehicle in the different road sections, which will result in a better balance of the load, and therefore this last technique is better in terms of road safety. Table 1 presents a summary of study comparison of relevant methods and their limitations that we have tried to overcome and correct in our method explicated in results and discussion section.

Table 1. Study comparison of relevant literature

The method	Metric	Limitations
Dijkstra's algorithm	Shortest path	Does not allow to have a faster path to the desired destination in terms of time
Speed prediction method	Speed	Cannot detect risky situations which will depend on other factors
Estimated time of arrival	Time	It lacks a way to select the optimal paths and direct vehicles to an optimal path
Estimating the traffic density	Density	It is not sufficient to have good control for traffic management
Trajectory planning for fully autonomous vehicles	Static trajectory planning	It does not take into account the variations in the state of the traffic at time t
Travel time on different sections of road	Travel time	Its main objectif is to reach destination quickly, but it lacks security
Prediction technique in real time	Density and average speed	It does not manage traffic and does not direct vehicles to optimal paths

The rest of paper is organized as follows: section 2 presents the proposed method where we explain our approach which begins by building the database containing the density of each section in real time. Then the routing of vehicles then finally the detection of anomalies. In section 3, results and analysis, we present the outcomes of our experimentation and discuss the performance metrics achieved by our model compared with other methods mentioned above.

2. METHOD

2.1. Traffic management

Saturation usually arises in areas where supply is not uniform or where demand is not uniform (arrival of a cross-section). We can distinguish the loaded situation in which the level of service deteriorates continuously and the appearance of saturation translated by a sudden break in the average level of service. The preventive measures against saturation consist in increasing the supply of the network and the regulation so that the supply (roads) remains higher than the demand (vehicles), in our cases more road path. The curative measures consist in controlling and improving the level of service when the instantaneous demand exceeds the supply. In this case, it is a matter of transferring the excess demand to areas where saturation is artificially created and controlled: where supply is least dependent on demand.

The detailed processing below needs high-performance technologies so that they can be carried out quickly, we used the Lambda architecture, including apache storm for real-time analysis and distributed computation to achieve faster flow processing. Additionally, we integrated Hadoop MapReduce for handling large-scale data processing, which constructs the database utilized by the speed layer. The experience is carried out on the map of Casablanca, Morocco which constitutes as input value of the sumo simulator which gives us a trace file of the generated traffic containing the different roads divided into small sections, the detailed processing below was done on this trace file.

2.1.1. Data collection

To make a more suitable traffic management and avoid saturation, we proposed a solution consist of building a real-time database that contains informations about density of roads indexed for each city at any moments, then, the database in question will propose a better routing for each vehicle to assure a better load balancing by cities (urban traffic) and highway. The routing mechanism will help the different player in the system to avoid saturated road and propose better ways with a multiselection. Every vehicle within the system records its entry time as well as its exit time when traversing a particular road section. Subsequently, it transmits this information along with its unique identifier (ID) to RSU. The purpose of this data transmission is to enable the RSU to calculate the vehicle density of the corresponding road section. By accumulating these recorded traces from all vehicles passing through each section, it becomes feasible to construct a real-time database containing the count of vehicles present on all road sections at any given moment. This database provides valuable insights into the current traffic conditions and aids in monitoring and managing the overall flow of vehicles within the system. The database will be an RSU layer that catch information sent by vehicles in each step (road entrance, speed, and direction) which can be identified by an ID. This will create a set of information about each vehicle that crossing the road in real-time. The Figure 1 show the simulation setup of our generated traffic on the map of Casablanca, we generated cars, trucks, bus, and we chose the value 5 for through traffic factor parameter and the value 12 for the count parameter which means the number of vehicles generated per hour and lane-kilometer, Figure 2 show an extract from the traffic simulation on sumo. So, sumo generates the traffic which will be used by our system to build the database described above and resumed in Figure 3. The simulator randomly chooses a departure and arrival edge for the vehicle, and our system chooses the optimal route detailed Figures 1 and 2 which will be sent to the simulator.

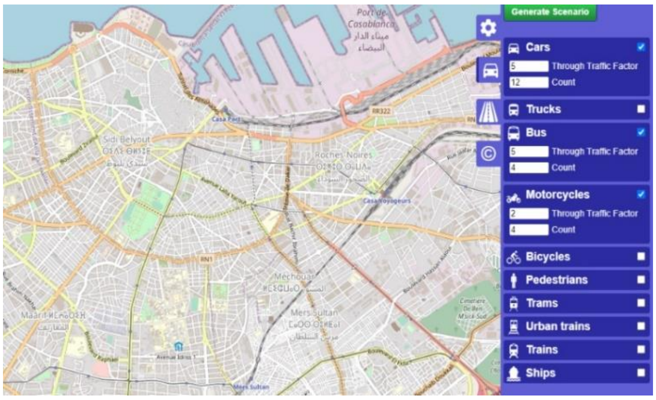


Figure 1. Simulation setup of the generated traffic

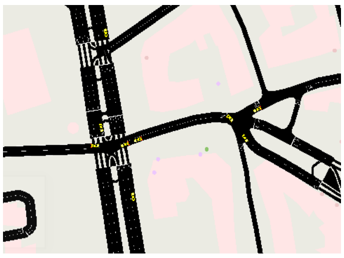


Figure 2. Extract from the traffic simulation on sumo

2.1.2. Routing

The database (which has already been getting information from vehicles), will be the source of inputs to build a route for specific car by demand. The demand will be sent by the vehicle with specific information (destination, weight, and type of vehicle), with the respect of road density and avoiding saturation (scheduling) by taking into account the load balancing, the vehicle will receive a response that includes the whole routing side, with the estimation time, and information about the road (stop sign and traffic colors) see Figures 3 and 4. The system chooses an optimal route with minimal density using vehicle density base which will be sent to the vehicle.

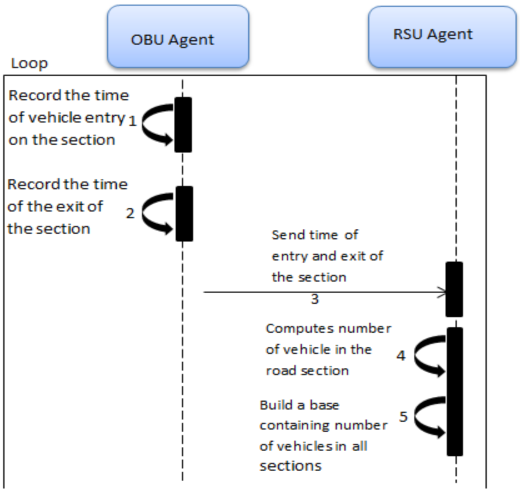


Figure 3. Sequence diagram of constructing the database

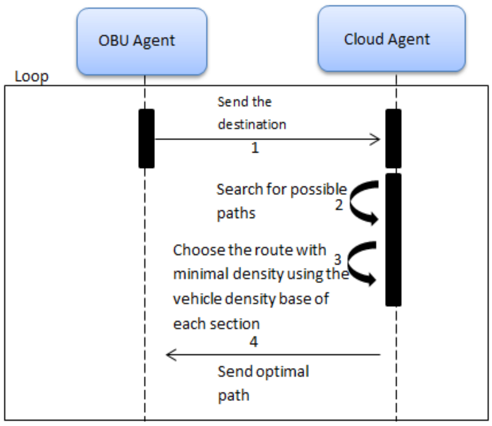


Figure 4. Sequence diagram of routing

2.1.3. Anomaly detection

Upon receiving the latest vehicle density information, the system proceeds to compare these values with the corresponding data stored in the database. If the disparity between the received values and the stored values exceeds a predefined threshold, it indicates the presence of an anomaly or event in the road section, characterized by unusually high vehicle density. Upon detecting such a change in vehicle density, the system promptly updates the database entries associated with the affected section to reflect the updated information. This ensures that the database remains up-to-date and accurately represents the current state of vehicle density for each section. When a vehicle commences its journey, it determines its desired destination. Subsequently, the system leverages the stored vehicle density data to select the optimal path for the vehicle to reach its destination. This selection process involves assessing the vehicle density of different sections along potential routes. By summing the number of vehicles in each section of a candidate path, the system identifies the path with the lowest vehicle density. The change that will be leveled-up to the system after comparison, will be stored to the database as first and after, will be sent to each vehicle that has been routed to the specific section, to be informed about the anomaly as an alert first, but also to be sure about which new decision need to be taken (change the road or stand still). And also, if a section is under a normal density, we can select it for new path or to remove the alert about the current anomaly.

This mechanism serves to assist vehicles in avoiding road sections that exhibit anomalies or congestion, thereby mitigating the likelihood of accidents. At the outset of a vehicle trip, after establishing the optimal route, the vehicle periodically sends notifications to the system, detailing its entry into a specific section and the designated route. The purpose of these notifications is to enable the system to monitor for any anomalies or changes in vehicle density in the subsequent sections of the established route. If such anomalies or changes are detected, the system initiates a search for an alternative, more optimal route and communicates the new path to the vehicle. However, if no anomalies are identified, the vehicle proceeds along the originally established route. These checks regarding the state of the next section are performed continuously throughout the vehicle's journey, triggered at each section entry point, the proposed method can be summarized in Figure 5. The frequency of these checks contributes to the precision of ETA for vehicles.

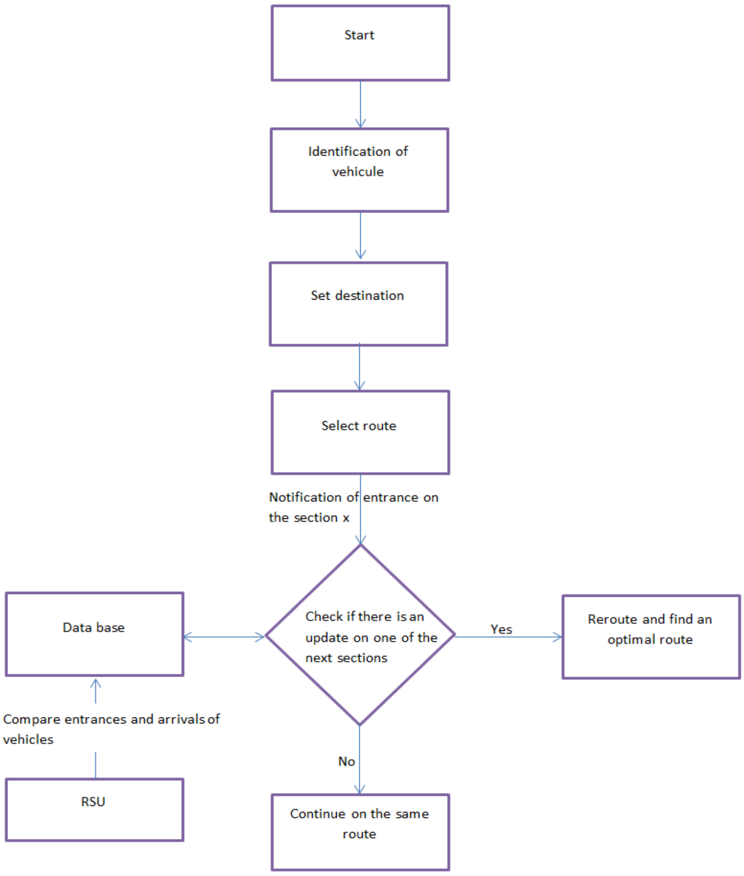


Figure 5. summary of all parts of the methodology

2.2. Prediction of anomalies

To be prepared as soon as possible to prevent a higher number of human losses is a considerable aspect for many companies and government. Yet, there is no system accurate enough that detects anomalies in both type in cities road and highway in real-time. That's why a prediction for anomalies pattern can be a new jump into road safety. Training data will be way more difficult due to the lake of information, however, our system-and with help of the database in question-can provide a way more clear data set with labeled inputs for a historic analyzes by the machine learning framework.

In the realm of machine learning, there are various approaches for making predictions. Some studies utilize the NB classifier, while others employ the discriminant random forest (DRF) algorithm. However, it is essential to conduct a comparative analysis between these two methods to determine their suitability for specific needs. The selection of one method over the other is not solely based on the model's capacity but rather on the specific requirements of the application. The NB classifier belongs to the family of linear classifiers. It makes predictions by assuming that the features are conditionally independent given the class. On the other hand, the DRF algorithm is a machine learning technique that combines the concepts of random subspaces and bagging. In this context, the class attribute of the model is the congestion degree, indicating the level of traffic congestion. It is important to note that the choice between NB and DRF depends on the specific requirements and characteristics of the problem at hand. Each method has its strengths and limitations, and a thorough comparison study can help determine the most appropriate approach for a given scenario. The class attribute of this model pertains to the degree of congestion.

The NB and DRF methods both exhibit exceptional accuracy in predicting the congestion degree. Table 2 provides a concise overview of the classification results obtained from these two classifiers. The dataset used in this analysis categorizes congestion into three levels: minor, intermediate, and major. For example, utilizing the NB method, as shown in Table 2, the accuracy (ACC) can reach approximately 84.2%. Accuracy measures the number of correct predictions made by the model across all predictions. Additionally, the area AUC is utilized as a performance metric for classification problems, indicating the model’s ability to distinguish between different classes. Comparatively, the DRF method achieves accuracy values of around 89.6%. Notably, the NB classifier exhibits the fastest computation time, taking approximately 0.07 seconds. While DRF yields better classification results than NB, it requires more time for computation (around 14 seconds). In real-time scenarios where prompt decision-making is crucial, utilizing NB with a reduced feature set can provide quick and reasonably accurate decisions. Subsequently, alerts can be sent to participating vehicles and drivers, enabling them to make more informed decisions in response to the prevailing traffic conditions.

Table 2. DRF and NB classification results

Classifier	Time of computation	ACC	AUC
Naïve bayes	0.07	84.2	67
Random forest	14	89.6	62

3. RESULTS AND DISCUSSION

The suggested strategy involves constructing a dynamic database that stores real-time information regarding the density of roads categorized by individual cities. This comprehensive database enables the provision of optimized routing solutions for vehicles, ensuring a balanced distribution of traffic across urban and highway areas. To illustrate the effectiveness of continuously updating the database, we present a simulation conducted on the road network of Casablanca, Morocco. The roads are divided into sections. The simulation employs the sumo simulator to generate traffic, vehicle’s intended destination. To highlight the practical value of real-time updates and the importance of maintaining an up-to-date database, we selectively present the paths assigned to four vehicles in Table 3. These vehicles share the same source and destination. The assigned paths are determined by our proposed method, which leverages the vehicle density data stored in the database to select the optimal route for each vehicle. For instance, at 2:00 PM, vehicle ID one starts its journey from road ID one, section 1, aiming to reach road number nine, section 3. The assigned path, indicated in the ‘path’ field of Table 3, consists of a sequence of road and section identifiers (x, y), where x represents the road ID and y represents the corresponding section ID. Following vehicle ID one, vehicle ID two commences its trip at 2:02 PM, also heading towards the same destination as vehicle ID one. While a navigation map module utilizes the database to determine the most efficient path with minimal density for each remarkably, the system assigns the same path to vehicle ID two, as there have been no modifications in the database, preserving the optimality of the initial path. In contrast, vehicle ID three, starting at 2:18 PM, shares the same destination as vehicles ID one and two. However, the system assigns a different path for vehicle ID three due to increased vehicle density in section 5 of route three. This change in path selection is a result of the detected anomaly, which could be attributed to accidents, congestion, or other unforeseen events. The updated database reflects the escalating vehicle density at coordinate (3,5), prompting the system to choose a new, more optimal path compared to the previous routes.

Table 3. Vehicle itinerary assigned by our methodology

Time	Vehicle ID	Source	Path	Duration
14:00	1	(1,3)	(1,3) -> (1,4) -> (3,4) -> (3,5) -> (5,6) -> (5,7) -> (8,1) ->(9,5)	50 min
14:02	2	(1,3)	(1,3) -> (1,4) -> (3,4) -> (3,5) -> (5,6) -> (5,7) -> (8,1) -> (9,5)	52 min
14:18	3	(1,3)	(1,3) -> (1,4) -> (3,7) -> (5,7) -> (6,8) -> (6,9) -> (8,1) -> (9,5)	56 min
14:21	4	(1,3)	(1,3) -> (1,4) -> (3,7) -> (5,7) -> (6,8) -> (6,9) -> (8,1) -> (9,5)	57 min

As we can remark in Table 3, the new path is third path in Table 2 that has time duration of fifty-six minutes; this third path is much better than the old path assigned to vehicle of id one and two because after the anomaly occurred, its time duration has become equal seventy-one minutes. Because of several variations in the traffic condition in the VANET network, there are several real-time modifications on the database; the Table 4 exemplifies an extract of this database. The proposed mechanism helps to avoid congestion and to reduce the seriousness that there will be an accident, the methods mentioned in the related work make predictions of having an accident/congestion or not, and as explicated in the Table 1, each method mentioned has its limitations that we have tried to overcome and correct in our approach, among the strengths of our method is that it detects in real-time what is happening in different road sections, it is not just a prediction like those methods did, but it illustrate what is actually occurring in the roads relying on vehicle density and time spent by the vehicles in real-time, in addition to that, when we remark that there is a high density on some sections so we will know that it is congested (high density) or there is an anomaly in that section so our system will redirect next vehicles to another path. It avoids saturation (scheduling) by taking into account the load balancing in terms that our mechanism makes fill roads by the vehicles, and once they are overwhelmed and busy, the method reroutes the following vehicles to a road which is less loaded than the others, our method develops this effect load balancing in an automatic way. This mechanism is demonstrated in Table 4, road section (1,3) is affected to many paths forwarded to lot of vehicles, so after a while (15 min), it gets dense (medium density), and unlike (1,5), the system method doesn't choose it in the path, so after a few times it will be low congested. For (3,5), its vehicle density moved from 13 to 35 because there is an anomaly occurred so we stop to utilize this section in paths until it will be normal. All these updates on the database let our method more precise. The architecture proposed is composed of centralized batch data storage, processing technique and a distributed data storage mechanism for real-time processing. For technologies adopted on the experiment, we took advantages of Lambda architecture, we incorporate apache storm for the real-time analysis to compute distributed and having faster flow processing. And we incorporate hadoop mapreduce concerning processing the data mass that constructs the database that will be an input of speed layer which is a fast-processing layer in real-time. Our proposed method addresses the issue of congestion by prioritizing load balancing. It ensures that roads are filled with vehicles up to a manageable capacity. Once a road becomes overwhelmed and congested, our method automatically reroutes incoming vehicles to less congested roads, thus achieving a balanced load distribution. This load balancing effect is exemplified in Table 5, where road section (1,3) receives a high number of vehicles, leading to medium density after a certain period of time. On the other hand, road section (1,5) is not chosen as a preferred path, resulting in lower congestion levels over time.

Table 4. An extract of the database containing all road sections density

(road, section)	Density at 14:00	Density at 14:10	Density at 14:20
(1,3)	13	20	27
(1,5)	17	16	20
(2,2)	9	8	10
(4,1)	12	14	15
(3,5)	13	35	30
(7,3)	24	23	21

Table 5. Sections status at T and T+15 min

Time	Road section	Density at T	Density at T+15 min
14:00	(1,3)	Low	Medium
14:00	(3,5)	Medium	Very high
14:10	(1,5)	Medium	Low
15:00	(x, y)	Low	Low

Furthermore, in the case of road section (3,5), its vehicle density increases significantly from 13 to 35 due to an anomaly. As a result, the system suspends the use of this section in the designated paths until the density returns to normal. These continuous updates in the database enhance the precision of our method in estimating vehicle density and predicting the time of arrival, enabling us to guide vehicles through different paths. In conclusion, our method has total control over traffic in terms of management to ultimately have less congestion and fewer accidents, which is not done by other methods. The experimental results and analysis demonstrate that our proposed system method, which incorporates big data technologies, provides an optimal solution for transferring data and enables near real-time processing suitable for an intelligent transportation system within a vehicular ad-hoc environment.

4. CONCLUSION

Dealing with the vast amount of data generated in VANET environments poses a significant challenge, necessitating the use of big data technologies to extract valuable insights from this voluminous data. Traffic management has long been a crucial area of study to ensure the safety of vehicles and pedestrians. In this context, our research focuses on leveraging big data tools to enhance traffic management and specifically on predicting the risk of vehicle accidents. To achieve this, we have developed a real-time anomaly detection system that enables instant detection of anomalies through parallel data processing, thereby facilitating faster execution times. Our system aims to accurately compute the vehicle density at each section of every road, providing valuable information to effectively manage traffic and communicate the best and safest routes to vehicles for reaching their destinations. By promptly addressing potential risks, we aim to mitigate human losses and minimize the occurrence of traffic congestion and accidents.

The foundation of our anomaly prediction system lies in a machine learning framework, which proves to be a viable solution for avoiding traffic congestion and reducing the likelihood of accidents. Through simulations, we have demonstrated that our system significantly mitigates congestion and effectively improves traffic flow, leading to a reduction in accidents. Furthermore, our system exhibits low latency and achieves high precision in its predictions. In our future research endeavors, we plan to capitalize on the database of vehicle density per section that we have established in this system. By integrating this database with machine learning techniques, we aim to further enhance our results and generate even more effective traffic management strategies. Through the synergy of data-driven insights and advanced machine learning algorithms, we aspire to continue making valuable contributions to the field of traffic management and safety.

REFERENCES

[1] M. M. Hamdi, L. Audah, S. A. Rashid, A. H. Mohammed, S. Alani, and A. S. Mustafa, "A review of applications, characteristics and challenges in vehicular Ad-hoc networks (VANETs)," in *HORA 2020 - 2nd International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings*, Jun. 2020, pp. 1–7, doi: 10.1109/HORA49412.2020.9152928.

[2] M. Fazio *et al.*, "A map-reduce approach for the dijkstra algorithm in SDN over osmotic computing systems," *International Journal of Parallel Programming*, vol. 49, no. 3, pp. 347–375, Jun. 2021, doi: 10.1007/s10766-021-00693-3.

[3] A. Rahim, D. P. K. Mallik, and D. V. A. Sankar, "Fractal antenna design for overtaking on highways in 5G vehicular communication Ad-hoc networks environment," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1s6, pp. 157–160, Dec. 2019, doi: 10.35940/ijeat.a1031.1291s619.

[4] Z. Cheng, M. Y. Chow, D. Jung, and J. Jeon, "A big data based deep learning approach for vehicle speed prediction," in *IEEE International Symposium on Industrial Electronics*, Jun. 2017, pp. 389–394, doi: 10.1109/ISIE.2017.8001278.

[5] H. Al-Najada and I. Mahgoub, "Anticipation and alert system of congestion and accidents in VANET using big data analysis for Intelligent Transportation Systems," in *2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016*, Dec. 2017, pp. 1–8, doi: 10.1109/SSCI.2016.7850097.

[6] N. Marz and J. Warren, "Big data: principles and best practices of scalable realtime data systems," in Shelter Island, Ny: Manning, 2015.

[7] A. Mahajan and A. Kaur, "Educational function of museums, methods of teaching 'the medieval history of Kazakhstan', museum pedagogy, open-air museum," *Indian Journal of Science and Technology*, vol. 9, no. 42, Nov. 2016, doi: 10.17485/ijst/2016/v9i42/98970.

[8] B.-R. Lin and T.-P. Wang, "Dynamic reporting mechanisms for trust management in vehicular ad-hoc networks," in *in the 10nd International Conference on Advanced Information Technologies (AIT)*, 2016, no. 140, pp. 813–819, doi: 10.1129/ait.2016.764897.

[9] M. Tantaoui, M. D. Laanaoui, and M. Kabil, "Big data accident prediction system in green networks and intelligent transportation systems," in *Advances in Science, Technology and Innovation*, 2021, pp. 121–127.

[10] M. Tantaoui, M. D. Laanaoui, and M. Kabil, "Real-time prediction of accident using big data system," in *ACM International Conference Proceeding Series*, Mar. 2020, pp. 1–5, doi: 10.1145/3386723.3387886.

[11] H. Zhao, H. Yu, D. Li, T. Mao, and H. Zhu, "Vehicle accident risk prediction based on AdaBoost-SO in VANETs," *IEEE Access*, vol. 7, pp. 14549–14557, 2019, doi: 10.1109/ACCESS.2019.2894176.

[12] P. K. Malik, D. S. Wadhwa, and J. S. Khinda, "A survey of device to device and cooperative communication for the future cellular networks," *International Journal of Wireless Information Networks*, vol. 27, no. 3, pp. 411–432, Sep. 2020, doi: 10.1007/s10776-020-00482-8.

[13] S. S. Bukhari, J. Park, and D. R. Shin, "Hadoop based demography big data management system," in *Proceedings - 2018 IEEE/ACIS 19th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2018*, Jun. 2018, pp. 93–98, doi: 10.1109/SNPD.2018.8441032.

[14] H. Al Najada and I. Mahgoub, "Big vehicular traffic data mining: towards accident and congestion prevention," in *2016 International Wireless Communications and Mobile Computing Conference, IWCMC 2016*, Sep. 2016, pp. 256–261, doi: 10.1109/IWCMC.2016.7577067.

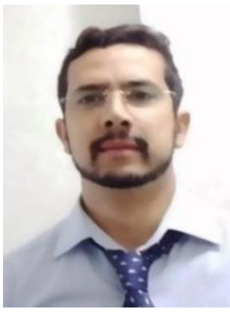
[15] A. Alinani and K. Alinani, "Real-time push-based data forwarding for target tracking in vehicular named data networking," *in 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, Oct. 2018, pp. 1587-1592, doi: https://doi.org/10.1109/smartworld.2018.00272.

[16] A. Malik and B. Pandey, "Asymmetric encryption based secure and efficient data gathering technique in VANET," in *Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering*, Jan. 2017, pp. 369–372, doi: 10.1109/CONFLUENCE.2017.7943177.

[17] N. Cheng *et al.*, "Big data driven vehicular networks," *IEEE Network*, vol. 32, no. 6, pp. 160–167, Nov. 2018, doi: 10.1109/MNET.2018.1700460.

[18] J. Contreras-Castillo, S. Zeadally, and J. A. G. Ibañez, "Solving vehicular ad-hoc network challenges with big data solutions," *IET Networks*, vol. 5, no. 4, pp. 81–84, Jul. 2016, doi: 10.1049/iet-net.2016.0001.
[19] N. Shaik and P. K. Malik, "A retrospection of channel estimation techniques for 5G wireless communications: opportunities and challenges," *International Journal of Advanced Science and Technology*, vol. 29, no. 05, pp. 8469–8479, 2020, doi: 10.1178/ijast.2020.654097.
[20] X. Y. Du, W. Lu, and F. Zhang, "History, present, and future of big data management systems," *Ruan Jian Xue Bao/Journal of Software*, vol. 30, no. 1, pp. 127–141, 2019, doi: 10.13328/j.cnki.jos.005644.
[21] R. Shahin, S. M. Saif, A. A. El-Moursy, H. M. Abbas, and S. M. Nassar, "Fog-ROCL: a fog based RSU optimum configuration and localization in VANETs," *Pervasive and Mobile Computing*, vol. 94, p. 101807, Aug. 2023, doi: 10.1016/j.pmcj.2023.101807.
[22] M. Tantaoui, M. D. Laanaoui, and M. Kabil, "Towards an efficient vehicle traffic management using big data," in *ACM International Conference Proceeding Series*, Oct. 2019, pp. 1–6, doi: 10.1145/3368756.3369080.
[23] A. Raj and R. D'Souza, "A review on hadoop eco system for big data," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 343–348, Feb. 2019, doi: 10.32628/cseit195172.
[24] T. Mouad, L. M. Driss, and K. Mustapha, "Big data traffic management in vehicular ad-hoc network," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4, pp. 3483–3491, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3483-3491.
[25] M. K. Murtadha and B. M. Mushgil, "Flexible handover solution for vehicular ad-hoc networks based on software defined networking and fog computing," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 1570–1579, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1570-1579.
[26] H. A. Khudhair, A. T. Albu-Salih, M. Q. Alsudani, and H. F. Fakhruldeen, "A clustering approach to improve VANETs performance," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 5, pp. 2978–2985, Oct. 2023, doi: 10.11591/eei.v12i5.5086.

BIOGRAPHIES OF AUTHORS



Mouad Tantaoui is a computer science engineer from National School of Computer Science and Systems Analysis (ENSIAS) in 2013, and obtained a doctoral degree in Computer Science from Faculty of Science and Technology, Hassan 2 University, Mohammedia, now he holds the position of a Professor at Moulay Ismail University, Meknes. He is part of the research team "Electrical Engineering, Vision, and Artificial Intelligence (IEVIA)" of the "Computer Science, Applied Mathematics, and Electrical Engineering (IMAGE)" Laboratory. His research endeavors are concentrated in the realm of computer security, approached through an artificial intelligence-based perspective. He can be contacted at email: tantaoui.mouad@gmail.com.

Mehdi Moukhafi with a master's degree in quality from the Faculty of Sciences Dhar El Mehraz, Fez in 2014, He further advanced his education by obtaining a doctoral degree in Computer Science from Moulay Ismail University, Meknes. Presently, they hold the position of a professor at the same university. His expertise flourishes within the research team "electrical engineering, vision, and artificial intelligence (IEVIA)" of the "computer science, applied mathematics, and electrical engineering (IMAGE)" laboratory. His research endeavors are concentrated in the realm of computer security, approached through an artificial intelligence-based perspective. He can be contacted at email: mehdi.moukhafi@gmail.com.

Idriss Chana received the Ph.D. degree from Mohamed V University of Rabat, Morocco in 2013. He is currently an Associate Professor at Moulay Ismail University of Meknès, Morocco. His research interests include information and communication technologies and artificial intelligence. A large part of his research projects is related to error correcting codes and turbo codes. Idriss Chana has published more than 40 papers in major journals and conferences in information theory and articial intelligence. He can be contacted at email: idrisschana@gmail.com.