

## Big data traffic management in vehicular ad-hoc network

Tantaoui Mouad<sup>1</sup>, Laanaoui My Driss<sup>2</sup>, Kabil Mustapha<sup>3</sup>

<sup>1,3</sup>Hassan II University, Mohammedia, Morocco

<sup>2</sup>Cadi Ayyad University, Marrakech, Morocco

### Article Info

#### Article history:

Received Aug 5, 2020

Revised Dec 19, 2020

Accepted Jan 19, 2021

#### Keywords:

Big data

Intelligent transportation systems

Machine learning

Traffic congestions prediction

Traffic management

Vehicular ad hoc network

### ABSTRACT

Today, the world has experienced a new trend with regard to data system management, traditional database management tools have become outdated and they will no longer be able to process the mass of data generated by different systems, that's why big data is there to process this mass of data to bring out crucial information hidden in this data, and without big data technologies the treatment is very difficult to manage; among the domains that uses big data technologies is vehicular ad-hoc network to manage their voluminous data. In this article, we establish in the first step a method that allow to detect anomalies or accidents within the road and compute the time spent in each road section in real time, which permit us to obtain a database having the estimated time spent in all sections in real time, this will serve us to send to the vehicles the right estimated time of arrival all along their journey and the optimal route to attain their destination. This database is useful to utilize it like inputs for machine learning to predict the places and times where the probability of accidents is higher. The experimental results prove that our method permits us to avoid congestions and apportion the load of vehicles in all roads effectively, also it contributes to road safety.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Tantaoui Mouad

Department of mathematics and applications

Hassan II University, FST, Mohammedia, Morocco

Faculté des Sciences et Techniques de Mohammedia BP 146 Mohammedia 28806 Morocco

Email: mouad.tantaoui-etu@etu.univh2c.ma

## 1. INTRODUCTION

Today cities are developing more and more until it becomes what we call smart cities, our goal is to ameliorate traffic management in intelligent transport system (ITS), this is a research area that catch the attention of research community mainly in the field of computer science because of social and economic challenges. The diversity of data sources produces voluminous data described by its different type and velocity, those data have become difficult to handle, in this context, big data technologies is the solution to handle this massive data in real time, one of the research areas that rely on big data technologies is vehicular ad-hoc network for managing their voluminous data. Our cities have experienced a growing trend of non-stop vehicles. The bulky increase in vehicles number has engendered lot of social and environmental problems in everyday life within cities like many traffic jams, road accidents and breathtaking air pollution, in this direction, ITS is research domain of interest to the research community principally in computer science domain due to challenges raised. Our goal is to find out a way for predicting traffic density on all roads and various cities which would permit us afterward to predict places where the probability of accidents is high. To attain this goal, we have to establish a road density prediction method utilizing voluminous data gathered during vehicle-vehicle and vehicle-infrastructure communication [1]. Vehicular ad hoc network (VANETs) has made since their birth, an increasingly advancement. Many standards, applications and methods of data

processing have been established to remedy the characteristics of this new gender of networks. The main challenges increase especially from the high mobility of vehicles and the spatio-temporal diversity of traffic density. Among the main components of the architecture of VANET are on-board unit (OBU) which are on mobile node for facilitating communication with other mobile nodes such as vehicles and fixed stations like roadside units via dedicated short-range communications (DSRC) and the ability to communicate utilizing cellular radio networks like GSM, 4G, Wi-Fi and WiMAX. Also, there are road-side units (RSUs) which are the base stations that handle VANET applications and manage actions to share and process information and also disseminate data, provides traffic directories, behave as location servers, and connect to the Internet and external centralized or distributed servers. And centralized cloud which is a gender of computer architecture where all or most of the treatment or computation is processed on such a central server. Centralized computing enables the deployment of all IT resources, administration and management of the central server.

## 2. RELATED WORKS

Actually, VANET environment is becoming as a big data problem. Therefore, various big data tools can be used to manage data from VANETs for improving traffic management. One of the important works that interest us was about implementing VANET Dijkstra algorithm [2], they implemented the algorithm in Hadoop MapReduce environment and the compared it to a dijkstra implemented in a simple NetBeans. Another work which is very important is detecting within the VANET network the vehicles that can proceed as information hubs whose role is to gather information from the network and share it. Ranking algorithm is developed in this context like for example InfoRank [3]. Speed prediction algorithms are essential for managing traffic in the field of intelligent transport system; Big data based deep learning speed prediction (BDDL-SP) [4] is a speed prediction algorithm that can predict the speed of a vehicle in highways and urban areas road networks. There are new systems to manage voluminous Data generated in real-time by different agent such as vehicles and roadside units (RSUs) ensuring a rapid treatment of data in the purpose to make road management decisions. Those new systems can be utilized for example to compute the estimated arrival time of vehicles or predict accidents and congestions with the help of naive bayes and distributed random forest (DRF) [5]. One of these systems is the Lambda architecture (LA) [6] which is data treatment architecture that allows handling massive data in batch and in real time. In article [7], an experiment on ITS environment was implemented to assess the traffic density estimation about different cities on different roads and carry out a comparative evaluation relying on that parameter. In article [8], the authors suggest a system to send with dynamic manner reports against selfish and malicious vehicles. The proposed system utilizes an encryption mechanism to exchange messages. In article [9], in the first hand, authors proposed architecture for large on-vehicle datasets which administer centralized access to massive data. The proposed system integrates centralized data storage and processing mechanism, and a distributed data storage mechanism for real-time processing and analysis. In article [10], a routing protocol was proposed which is based on road vehicle density in real time. The computation of road density is based on each vehicle to which it belongs by utilizing tag messages and the road information table.

In article [11], the synthetic minority oversampling technique is used for reconstructing the experimental dataset, the minority samples in the study dataset were oversampled and new samples were synthesized for completing the missing data. In article [12], in the first hand, they reviewed VANET technologies for efficient and reliable data transmission. And then, they presented the methods used by Big Data for studying the characteristics of VANETs and improving their performance. In article [13], a new routing protocol is proposed which uses (link guarantee) and (forwarding movement distance) a node to select the next hop node. They used the weighted function by normalizing all quality-of-service metrics. In article [14], the H2O and WEKA extraction tools are used for evaluating five classifiers on two large sets of workshop data. The classifiers utilized are: AdaboostM1, C4.5, random forest (RF), naive bayes, (with the C4.5 basic classifier) and Bagging. The selection of attributes is applied and also the problem of class imbalance is tackled. Their experiences showed that naive bayes (NB) gave the optimal results, with the shortest calculation time and a practical area under the curve (AUC) and accuracy (ACC). In paper [15], named data networking (NDN) is new Internet architecture has been established to settle the VANET networks weaknesses and to manage countless applications such as object tracking, tracking a mobile vehicle and handling an effective communication channel in the VANET. In article [16], this research project proposes an efficient and secure data collection technique that ensures the security and confidentiality of data exchanged between vehicles and RSUs. It is based on asymmetric encryption that ensures secure communication between vehicles and RSUs. In this technique, secure authentication is established between the vehicle and the RSU before the RSU begins to collect the vehicle data. In article [17], by significantly expanding the scale of the network and performing real-time and long-term information processing, vehicular VANETs are moving towards the Internet of vehicles (IoV), which promises effective and intelligent

prospects for a future transport system. On the other hand, vehicles are not just consumers; they also generate huge amounts and types of data, called big data. In this article, they first examined the relationship between IoV and big data in the vehicular environment, mainly on how IoV supports transmission, storage, computing using big data and how IoV pulls benefit of big data for characterization, performance evaluation and big data support for a communication protocol design. In article [18], autonomous vehicle (AV) technology leads to many economic and social benefits and impacts. The trajectory planning is one of the essential and critical tasks of driving the autonomous vehicle. In this article, they tackled the problem of trajectory planning for fully autonomous vehicles. The constructed methods are intended for autonomous vehicles in a cloud based smart vehicle environment.

This article presents an optimal and safe trajectory selection method in autonomous vehicles. The selection of the safety trajectory in this work is mainly based on the exploitation of Big Data and the analysis of real-life accident data and real-time connected vehicle data. In the paper [19], authors analyze channel estimation techniques for Massive multiple-input multiple-output systems. They did a comparison among the different channel estimation techniques. For the paper [20], they talked about the history of the evolution of data handling systems, and they discuss the existing state of big data handling systems in the context of data storage, model, and query engines of big data handling systems. In the article [21], Shaofu Lin *et al.* used Guizhou as an example to produce a spatio-temporal big data handling system leaned on GIS bussing technique. In the article [22], authors produced and created a new distributed big data management system (DBDMS) and it offers big data real-time gathering, search and perpetual storage. Hadoop eco system seem to be very important, in the article [23], authors analyze the Hadoop architecture and Hadoop eco system. In article [24], Brunswicker *et al.* developed on research issue which is related to open digital collaboration and established the data analytical challenges which need to be revealed to answer these important research issues. In the paper [25], Bukhari *et al.* proposed a method based on big data demography managing system using apache Hadoop platform to settle troubles of demography high rising data managing. In the paper [26] Malik *et al.* have analyzed the various mechanisms of resource allocation, mode selection for underlying communications in the sense of device to device and cooperative communication techniques and they establish a new technique LTE-Advanced Pro. In the paper [27], authors settle the issue of overtaking larger vehicle by constructing ad-hoc connection based on 5G technology with the vehicle to be overtaken.

### 3. RESEARCH METHOD

#### 3.1. Anomaly detection

In this paper, it's about traffic management in real-time, we propose a system construct that firstly build a base that contain the estimated spending time of each section of all roads in the city at the present. Secondly, the method will be able to detect anomalies in the road so that, with the help of the base, we can compute the time required to reach any destination from any source in real time. More precisely, for the construction of the base, each vehicle logs the time of entry at the beginning and the end of a section, after that it transmits this information and its identifier (ID) to the RSU in order to compute the spending time of the road section. So, we have the traces of time spent to cross each section of all vehicles which will allow us to have a base containing the estimated spending time of all the road sections in real time. This approach is presented in Figure 1. After that, when a vehicle asks for a route to its destination, our system will be able to transmit the route to reach its destination to the vehicle and the estimated arrival time along its way. This approach is presented in Figure 2. Another utility of this base, we can locate an accident or an anomaly in a section when we receive a recent spending time of vehicles very bigger compared to what there is in the base. This approach is presented in Figure 3.

In other word, the system will receive spending time of all vehicles all along their ways, then we will compare these values which we have just received with the corresponding values which are in the base, if the difference crosses a threshold, then we will deduce that there is an anomaly or an event since the majority of vehicles spend too much time in a section x. After noticing that there is a change in terms of time spent on a particular section, our system updates immediately the base spending time of the section affected by the change. And the opposite is true, if the time spent in a section is reduced, so we deduce that there is no more traffic jam for example in that section, so we update the base as well. Therefore, our database is always up to date in real time traffic status and changes of different sections. From the start of a vehicle, it establishes the desired destination; the system will choose the best path to reach its destination relying on the base of the spending time by choosing the path with the minimum expenditure time by summing the spending time of different sections of the path drawn. We can summarize all our use cases in Figure 4.

This mechanism will help vehicles to avoid sections of road where there is an anomaly or congestion; not only that, but it reduces automatically the gravity of having an accident. After having established the optimal route at the beginning of the vehicle trip, at each entry in a section, the vehicle sends

a notification of its entry and its route to the system to check if there is an anomaly or a modification in terms of estimated time for crossing following sections of the route established, if yes, the system will look for another more optimal route and it will send him the new path, otherwise the vehicle continues in his route established beforehand. These checks of the state of the next section are carried out all along its path and at each entry of a section; this is due to the frequent change of the state of the traffic within the VANET network. This characteristic of these frequent checks gives our method more precision regarding the estimated time of arrival of the vehicles.

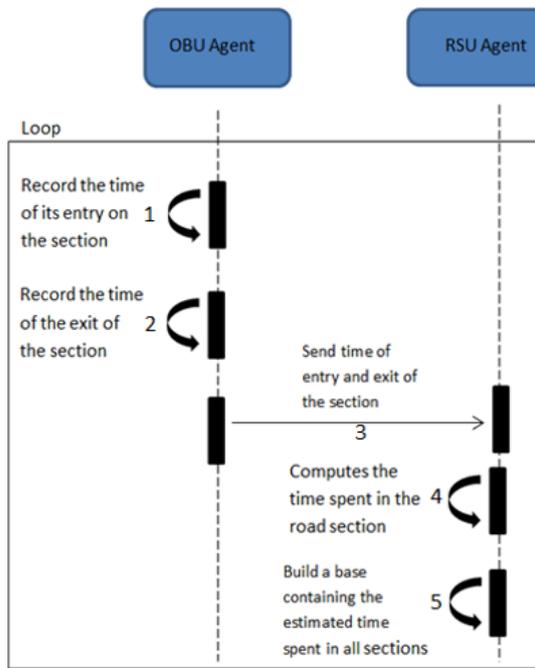


Figure 1. Sequence diagram of building a base containing the estimated spent time of all the sections

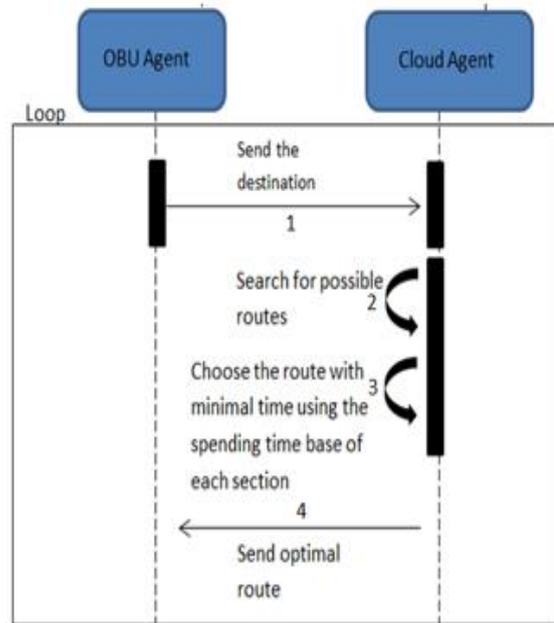


Figure 2. Sequence diagram of sending to the vehicle the estimated time of arrival all along its way

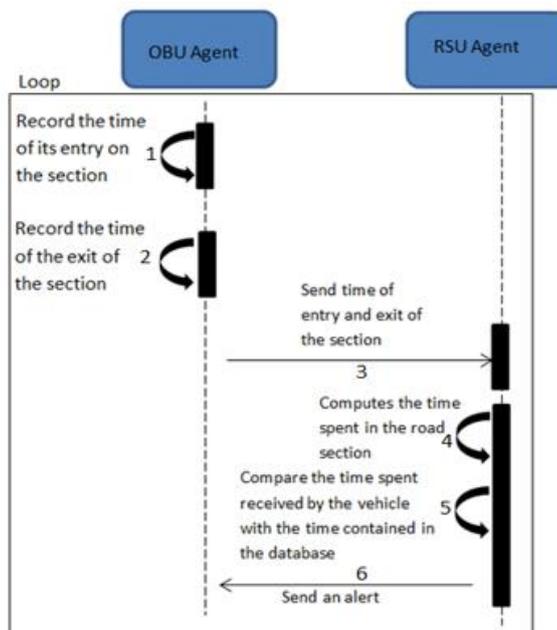


Figure 3. Sequence diagram of detecting an accident or an anomaly in a section

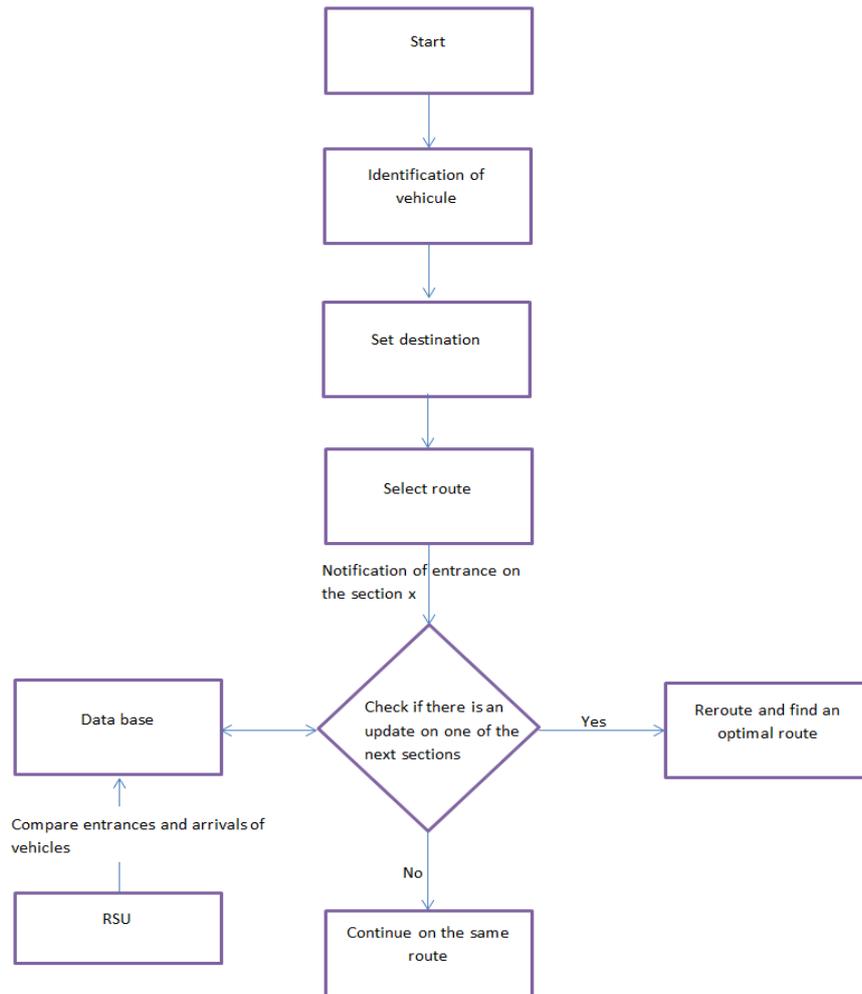


Figure 4. Summary of all parts of the methodology

### 3.2. Prediction of anomalies

Our goal in this part is to predict the areas and the corresponding timestamp where the risk of congestions or accidents is high, we utilized the database constructed above comprising the estimated spending time of different road sections, and we add to this database the average speed of each road section, density traffic of it, and the timestamp, and the results of anomalies detection part done above, we utilized these attributes like our dataset inputs for machine learning. Two different classifiers are utilized and compared: naive bayes (NB) and discriminant random forest (DRF) to discern areas where there will happen an anomaly. The naive Bayesian classification is a type of simple probabilistic Bayesian classification based on Bayes' theorem with a strong (called naive) independence of the assumptions. It implements a naive Bayesian classifier, or naive Bayes classifier, belonging to the family of linear classifiers. While for DRF, it is part of machine learning techniques. This algorithm [28] combines the concepts of random subspaces and bagging. Discriminant random forest algorithm trains on multiple decision trees trained on slightly different data subsets. The basis of the calculation is based on decision tree learning. Breiman's [29] proposal aims to correct several known drawbacks of the initial method, such as the sensitivity of single trees to the order of predictors, by calculating a set of  $X$  partially independent trees. The class feature of our classification is the congestion degree. First, both the NB and DRF methods can achieve superb accuracy values. The classification result has three different values: minor, intermediate and major. For the naïve bayes method, in Table 1, accuracy (ACC) attains about 83.5%, (in classification methods, Accuracy is the number of correct predictions made by the model over all different predictions done, and area under the curve (AUC) is a kind of performance for classification method that express how much model is able to distinguish between classes.), and for DRF, the values attain about 88.3%. DRF classifier had the biggest time of computation (nearly 15 s). Table 1 and Figure 5 show the summarized results of classification for the two classifiers. For DRF, despite of the classification results is better that the naïve bayes results, it took more time (15 s). For

real-time mode, with fewer features, using naïve bayes would be good in providing most likely the right decision quickly, and then the alert will be sent to the participating vehicle and the driver to have better decision.

**Table 1. DRF and NB classification results**

Classifier	Time of computation	ACC	AUC
Naïve Bayes	0.04	83.5	63
Random Forest	15	88.3	61

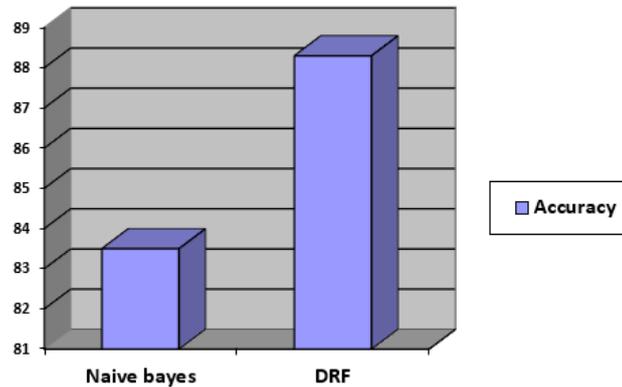


Figure 5. DRF and NB accuracy results

#### 4. RESULTS AND DISCUSSION

The proposed method consists at establishing the optimal path to the destination by choosing the sections of road that have the minimum spending time relying on utilizing big data technologies to assure a fast treatment in real time. In the simulation below, we took the map of Casablanca city, Morocco; and then we cut the roads into sections as is shown in Figure 6. The simulation was done by simulator sumo that generates traffic, and we used a navigation map module that utilizes our database containing time spent on each road section and find the best route with minimal time. To show the usefulness of immediate database changes and always keeping the database up to date, we mentioned in the table only four vehicles which have the same destination and source.

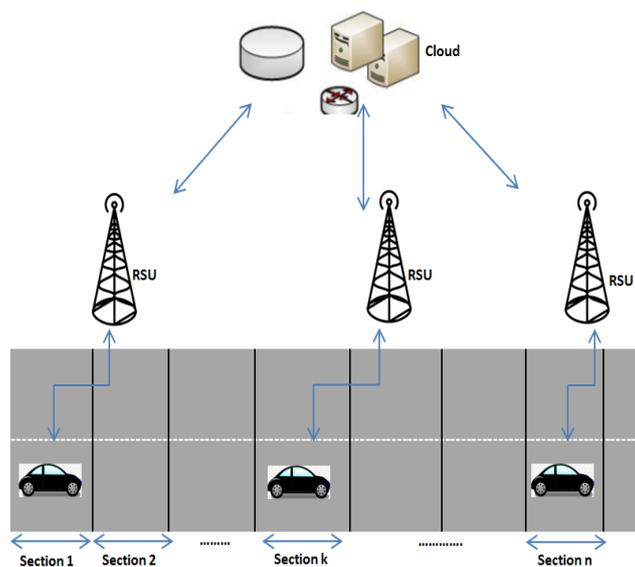


Figure 6. VANET components and cutting roads into sections

The Table 2 shows the itinerary of the vehicles assigned by our proposed method according to the destination established beforehand by the vehicle, the system is based on the database of the spending time of each section of road to assign the vehicles an optimal route. In the example shown below, at twelve o'clock noon, the vehicle number one start its trip from road number one with section number one and wishes to reach road number nine with section number three as its desired destination, our system affected to the vehicle a route indicated in the 'route' field of the Table 2, it is a sequence of couples (x, y) that the vehicle will follow throughout its trip, x being the identifier of the road and y being the section of the road x. For the vehicle number two, it arrived just after vehicle number one at 12:02 and wants to go to the same destination as vehicle number one, as we can see, the system has affected to it the same route because the base is the same and so the optimal route stays the same as the first one. For the vehicle number three that arrived at 12:17 and has the same destination as vehicle number one and two, we noticed that the system has assigned a different route than those of the two previous vehicles, this is because the spending time of section 5 of route 2 has become bigger because of a particular anomaly that can be either an accident, congestion, and an event (Anomaly detected in (2, 5), its spending time is becoming bigger and is updated in the base) and this will push the system to look for a new route which is more optimal than the old one.

As we can see in Table 2, the new route is third route in Table 2 that has duration of fifty minutes; this third route is more optimal than the old route affected to vehicle number one and two because after the anomaly detected, its duration has become equal to sixty minutes. Because of the frequent changes in the traffic state in the VANET network, there are many real-time updates on the database; the Table 3 illustrates an extract of this database. The proposed method gave us a way to avoid congestions and to reduce the gravity to have an accident, It is another way to give to user a safe itinerary to reach their destination more accurately, the other methods in the related method mentioned in the literature predict if there is an accident or congestion in what will happen, the advantage of our method is that it knows in real time what is happening in the different road sections, it is not a prediction, but it illustrate what is actually happening in the roads based on the time spent by the vehicles live, also when we see that vehicles spend too much time 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 route. It also helps to have as a kind of balance of loads in terms that our method makes fill roads by the vehicles, and as soon as that starts to be busy, it reroutes the following vehicles to a road that is less loaded than the others, our method develops this effect of balance of load in an automatic manner.

Table 2. Vehicle itinerary assigned by our method

Time	Vehicle ID	Source	Route	Duration
12:00	1	(1,1)	(1,1)->(1,2)->(2,4)->(2,5)->(4,6)->(4,7)->(6,2)->(9,3)	45 min
12:02	2	(1,1)	(1,1)->(1,2)->(2,4)->(2,5)->(4,6)->(4,7)->(6,2)->(9,3)	47 min
12:17	3	(1,1)	(1,1)->(1,2)->(3,1)->(5,5)->(4,8)->(4,9)->(6,2)->(9,3)	51 min
12:19	4	(1,1)	(1,1)->(1,2)->(3,1)->(5,5)->(4,8)->(4,9)->(6,2)->(9,3)	53 min

Table 3. An extract of the database containing all road sections duration time

(road, section)	Estimated duration time at 12:00	Estimated duration time at 12:10	Estimated duration time at 12:20
(1,1)	6 min 30 sec	7 min 20 sec	8 min 52 sec
(1,2)	7 min 13 sec	7 min 13 sec	7 min 13 sec
(1,3)	8 min 19 sec	8 min 43 sec	8 min 32 sec
(2,1)	6 min 41 sec	6 min 23 sec	6 min 55 sec
(2,3)	5 min 22 sec	5 min 29 sec	5 min 28 sec
(3,4)	7 min 50 sec	7 min 45 sec	7 min 30 sec
(4,1)	8 min 21 sec	8 min 32 sec	8 min 40 sec
(2,5)	7 min 54 sec	13 min 41 sec	12 min 36 sec
(7,3)	9 min 22 sec	9 min 45 sec	9 min 10 sec

This approach is illustrated in Table 4, road section (1, 1) is affected to many routes sent too many vehicles, so after a while (20 min), it becomes a little congested (medium density), and unlike (1, 3), the system doesn't used it in the routes, so after a while it becomes low congested. For (2, 5), its estimated duration time moved from 7min to 13min because of an anomaly so we stop to use this section in routes until it become normal. Those many updates at the level of the database let our method more accurate concerning the estimated time of arrival of vehicles and to route vehicles in different paths. Our architecture is composed of centralized batch data storage and processing mechanism and a distributed data storage mechanism for real-time treatment and analysis to manage and process the flow from the vehicle in a particular area. For the

technologies used on the simulation, we benefited from LA advantages, we used apache storm for the real-time analysis to process distributed and faster flow treatment. While for processing the data mass, we utilized Hadoop MapReduce that constructs the database that will be utilized by speed layer which is a fast-processing layer in real time. Thus, the experimental result and analysis proved that the proposed architecture model is an optimal solution that utilizes big data technologies, destined to transfer data in a near real-time treatment destined for intelligent transportation system in a vehicular ad-hoc environment.

Table 4. Status of different road section at T and T+ 20min

Time	Road section	Density at T	Density at T+20min
12:00	(1,1)	low	Medium
12:00	(2,5)	Medium	Very high
12:10	(1,3)	Medium	Low
13:00	(x,y)	low	Low

## 5. CONCLUSION

VANET engender voluminous data which is complicated to handle; therefore, big data technologies are important to mine meaningful data from these big data. Road accidents are threatening the lives of human beings, and in the purpose of finding out an alternative for the issue of predicting the high probability of vehicle accidents, this paper utilizes the benefits of big data to enhance traffic management, we established real-time anomalies detection system in real time employing parallel data treatment, this gave to our method fast execution time. Our method aims to know exactly the time spent to cross each section on each road, which will give us a control to well manage the traffic and send to vehicles an accurate estimated time of arrivals and the best safe route to reach their destination. And also, anomalies prediction system was developed with the help of machine learning techniques. That will help us to avoid traffic congestion and reduce the risk of accidents. The results of our experiment show that our method decreases congestion significantly and so decrease accidents and our results have low latency and high accuracy. Our future research is to more exploit the base of spending time per section established on this method and merging it with machine learning techniques to have a high control on managing traffic and having better results.

## REFERENCES

- [1] Da Cunha, F. D., Boukerche, A., Villas, L., Viana, A. C., and Loureiro, A. A., "Data communication in VANETs: a survey, challenges and applications," Diss. INRIA Saclay; INRIA, 2014.
- [2] B. Punam, and V. Jindal, "Use of big data technology in vehicular ad-hoc networks," *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2014, pp. 1677-1683.
- [3] Khan, Junaid, Yacine Ghamri-Doudane, and Ali El Masri, "Vers une approche centrée information (ICN) pour identifier les véhicules importants dans les VANETs," *Proceedings of the conference CFIP NOTERE 2015*, 2015.
- [4] Cheng, Z., Chow, M. Y., Jung, D., and Jeon, J., "A big data based deep learning approach for vehicle speed prediction," *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, 2017, pp. 389-394.
- [5] Al Najada, Hamzah, and Imad Mahgoub, "Anticipation and alert system of congestion and accidents in VANET using Big Data analysis for Intelligent Transportation Systems," *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1-8.
- [6] Marz, Nathan, and James Warren., "Big Data: Principles and best practices of scalable real-time data systems," New York; Manning Publications Co., 2015.
- [7] Mahajan, Anshul, and Arvinder Kaur, "Predictive Urban Traffic Flow Model using Vehicular Big Data," *Indian Journal of Science and Technology*, vol. 9, no. 42, pp. 1-8, 2016.
- [8] Bo-Rang Lin, Tsan-Pin Wang, "Dynamic Reporting Mechanisms for Trust Management in Vehicular Ad-hoc Networks," *AIT*, 2016
- [9] Daniel, Alfred, Anand Paul, and Awais Ahmad., "Near real-time big data analysis on vehicular networks," *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*, 2015, pp. 1-7.
- [10] Yu, Hyun, Joon Yoo, and Sanghyun Ahn, "A VANET routing based on the real-time road vehicle density in the city environment," *2013 Fifth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 333-337, 2013, doi: 10.1109/ICUFN.2013.6614836.
- [11] Zhao, H., Yu, H., Mao, T., Zhang, M., and Zhu, H., "Vehicle Accident Risk Prediction Over AdaBoost from VANETs," *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2018, pp. 39-43, doi: 10.1109/IHMSC.2018.10115.
- [12] Cheng, N., Lyu, F., Chen, J., Xu, W., Zhou, H., Zhang, S., and Shen, X., "Big data driven vehicular networks," *IEEE Network*, vol. 32, no. 6, pp. 160-167, 2018.
- [13] Gawas, Mahadev A., Manas Mulay, and V. Bhatia, "Cross Layer Approach for Neighbor Node Selection in VANET Routing," *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, 2018, pp. 1-6, doi: 10.1109/CSNDSP.2018.8471840.

- [14] Al Najada, Hamzah, and Imad Mahgoub., "Big vehicular traffic data mining: Towards accident and congestion prevention," *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2016, pp. 256-261, doi: 10.1109/IWCMC.2016.7577067
- [15] Alinani, Annadil and Alinani, Karim., "Real-Time Push-Based Data Forwarding for Target Tracking in Vehicular Named Data Networking," *2018 IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)*, 2018, pp. 1587-1592, doi: 10.1109/SmartWorld.2018.00272.
- [16] M. Arun, and B. Pandey, "Asymmetric encryption based secure and efficient data gathering technique in VANET," *2017 7th International Conference on Cloud Computing, Data Science and Engineering-Confluence*, 2017, pp. 369-372, doi: 10.1109/CONFLUENCE.2017.7943177.
- [17] Xu, W., Zhou, H., Cheng, N., Lyu, F., Shi, W., Chen, J., and Shen, X., "Internet of vehicles in big data era," *IEEE/CAA Journal of Automatica Sinica*, 2017, vol. 5, no. 1, pp. 19-35, doi: 10.1109/JAS.2017.7510736.
- [18] Al Najada, Hamzah, and Imad Mahgoub, "Autonomous vehicles safe-optimal trajectory selection based on big data analysis and predefined user preferences," *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2016, pp. 1-6, doi: 10.1109/UEMCON.2016.7777922.
- [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. 5, pp. 8469-8479, 2020.
- [20] Xiaoyong, Lu Wei, and Zhang Feng, "History, current status and future of big data management systems," *Journal of Software*, vol. 30, no 1, pp. 127-141, 2019, doi: 10.13328/j.cnki.jos.005644.
- [21] Lin, S., Luo, B., Luo, J., Wang, Z., and Wei, Y., "Guizhou Big Data Management System and Key Technology," *2018 26th Int. Conference on Geoinformatics*, pp. 1-7, 2018, doi: 10.1109/GEOINFORMATICS.2018.8557109.
- [22] Chen, H. Y., "Design and Realization of a Distributed Big Data Management System," *Advanced Materials Research*, vol. 1030, pp. 1900-1904, 2014, doi:10.4028/www.scientific.net/amr.1030-1032.1900.
- [23] Raj, A., and D'Souza, R., "A Review on Hadoop Eco System for Big Data," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 2019, doi: 10.32628/CSEIT195172.
- [24] Brunswicker, S., Bertino, E., and Matei, S., "Big data for open digital innovation—a research roadmap," *Big Data Research*, vol. 2, no. 2, pp. 53-58, doi: 10.1016/j.bdr.2015.01.008.
- [25] Bukhari, S. S., Park, J., and Shin, D. R., "Hadoop based Demography Big Data Management System," *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2018, pp. 93-98, doi:10.1109/snpd.2018.8441032.
- [26] Malik, P. K., Wadhwa, D. S. and Khinda, J. S., "A Survey of Device to Device and Cooperative Communication for the Future Cellular Networks," *Int J Wireless Inf Networks*, pp. 1-22, 2020, doi: 10.1007/s10776-020-00482-8.
- [27] A. Rahim, P. K. Mallik, V. A. Sankar Ponnappalli., "Fractal Antenna Design for Overtaking on Highways in 5G Vehicular Communication Ad-hoc Networks Environment," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 1S6, 2019, pp 157-160, doi: 10.35940/ijeat.A1031.1291S619.
- [28] Ho, Tin Kam, "Random Decision Forests," *Proceedings of 3rd international conference on document analysis and recognition*, 1995, pp. 278-282.
- [29] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.