IJIRR

International Journal of Information Research and Review
Vol. 04, Issue, 04, pp.4019-4023, April, 2017

Research Article

PRESENTATION STUDY OF DIFFERENT DATA COLLECTION LEVELS, DATA SCHEMES AND MINING TECHNIQUES USED IN VANET

[1,*]Mythili, A. and [2]Dr. Mahendran, S.K.

[1]Bharathiar University, Coimbatore, India
[2]OOty Government Arts College, Nilgiris, India

ARTICLE INFO

Article History:

Received 04th January, 2017
Received in revised form
26th February, 2017
Accepted 11th March, 2017
Published online 30th April, 2017

Keywords:

VANET, Road side Unit (RSU),
Data Collection Schemes (DCS),
Data Mining (DM).

ABSTRACT

The main objective of this article is to find the greatest method for collecting the data for VANET in terms of communication overhead, average latency and packet delivery ratio. This article is used to identify the various levels of techniques for data collection in VANET at single RSU. A VANET is composed of fast-moving mobile nodes (vehicles) that have irregular and short contacts, fixed road-side units (RSUs) that overhear and broadcast to vehicles, and a central server. Vehicles move along roads, collect data and process them, and distribute the data to other vehicles and RSUs. The central server aggregates data collected by vehicles, overviews traffic and road status, and generates keys and certificates when necessary. RSUs overhear the data sent from vehicles, broadcast road-side information to vehicles, and communicate to the central server via backhaul network. Date collected from the different resources is organized using different levels of data. We have presented data mining techniques for mining the data in VANET.

INTRODUCTION

VANET is a promising technology, and it is an explicit type of Mobile Ad-Hoc Networks (MANET). Generally VANET consists of two types of wireless nodes like mobile units and the Road Side Units (RSU). Mobile units in VANET are the extremely dynamic vehicles equipped with a sensing device commonly a Global Positioning System (GPS), and antennas performing as transceivers for transmitting and receiving data or information, collectively known as On Board Unit (OBU) that are used for communication between other vehicles on road or with the fixed units in network. Road side units in VANET are the fixed wireless nodes on the road sides that will offer internet connections to the vehicles through Internet Service Providers (ISP). The VANET architecture is shown in Figure 1, There are two types of communication possible in VANETs Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I). V2V is an infrastructure less approach where communication is among the vehicles moving in the network that can locate each other position and then transmit messages using their On Board Units (OBU). V2I is an infrastructure based approach as it provides communication among vehicles and the fixed units such as Road Side Units or traffic lights (Zeadally *et al.,* 2012).

Corresponding author: Mythili, A.
Bharathiar University, Coimbatore, India.

V2R is called as Vehicle-to-Roadside (V2R) communications. In this protocol, emergency warning messages are simultaneously transmitted via Vehicle-to-Vehicle (V2V) and Vehicle-to-Roadside (V2R) communications in order to achieve multipath diversity routing. In VANET dynamic changes in network will occurs frequently and is also describe the obstacles like tress, buildings, and other vehicles on road. VANET plays a significant role for Safety and Traffic Applications like avoiding road accidents, maintaining road conditions, generating warning alerts, vehicles position tracking and checking traffic violations, vehicle path map. VANET also providing Information like internet access, news, music, videos, parking availability, video conferencing due to its high mobility. VANET topologies continue on changing very fast leading to high communication overhead in exchanging new available topology information. Figure 2 describes the RSU infrastructure communication between the cars for sharing and information from various vehicles. Road Site Unit (RSU) is an important component for helping information dissemination in traffic networks (Bakhouya *et al.,* 2011).

Data collection in Vehicular Adhoc Network

A car may be defined, nowadays, as a set of non limited capabilities sensors that are able to detect and analyze a huge variety of data.

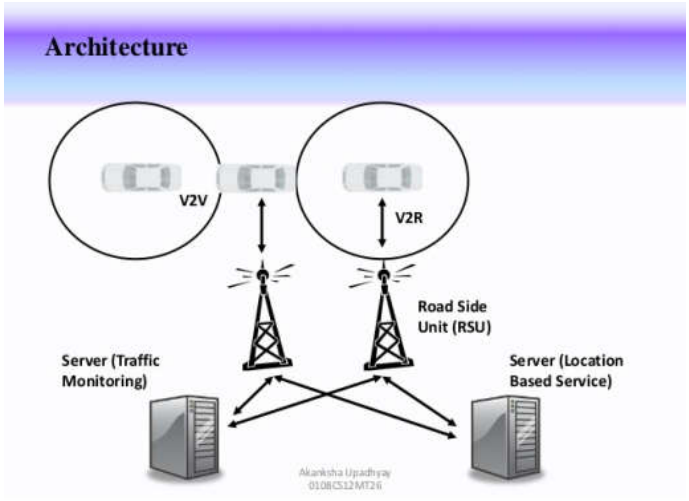

Figure 1. VANET Architecture

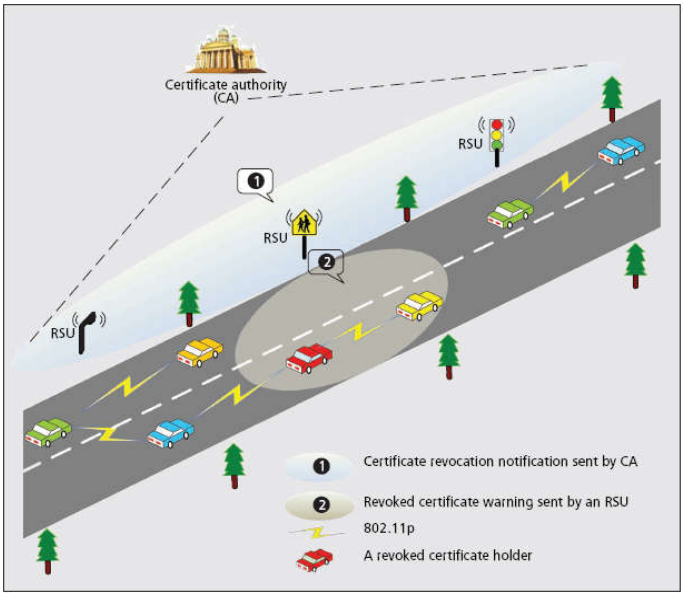

Figure 2. Area consisting of RSUs and moving Vehicles

With the look and emergence of on-board units in newly designed cars, collected data are exchanged between vehicles at each encounter while traveling in the road topology using the newly standardized technique IEEE802.11p or cellular capabilities such as Universal Mobile Telecommunications System (UMTS) offered by the OBU. Several architectures and protocols have been defined to support data collection and exchange between vehicles. Due to their high mobility and unpredictable frequent change of topology, vehicles are usually delegated to themselves and are self-organized with the help of central third parties. So, a network organization could be either decentralized self-organized without any use of external infrastructure, centralized self-organized where a central third party (e.g Road Side Unit (RSU), eNodeB, Base Station (BS), etc.) is in charge of vehicles' organization management or hybrid where the network management is shared between mobile vehicles and central infrastructures (Yang *et al.,* 2010). Data exchange between network members is handled using routing protocols and forwarding strategies based on V2V, V2I and Infrastructure-to-Vehicle (I2V) communications. Various kinds of protocols were planned to handle data routing. The protocols are mainly classified into two categories: (i) topology-based where information about existing links in the network are used to perform packet forwarding and (ii) position-based (or geographic) where neighboring location information are used to forward packets.

Data Collection Levels

Level 0- Preprocessing: At this level, data collected from the different resources are organized and normalized before their treatment.

Level 1- Object Assessment: At this level, a set of procedures are applied to the data given by the preprocessing phase. These procedures include spatio-temporal alignment, correlation, clustering, false positives removal, identity fusion and combination of features extracted from images. Therefore, the inputs are transformed into consistent data structures.

Level 2- Situation Assessment: It consists on a higher level inference based on the given objects descriptions resulted from level 1. In fact, objects are combined and relations are identified to determine their significance into a specific environment. So, the main aim of this phase is the identification of activities and events based on the relations built between objects and their environment.

Level 3- Impact Assessment: This level considers the study of potential risks that could be caused by the previously identified activities. Therefore, it is the level of threat evaluation and counter measures planning.

Level 4- Process refinement: It is the resource management process where task scheduling and priority management are being processed to enhance the efficiency of decisions making and reactions (Yang *et al.,* 2010). A more generalized model describing the hierarchy of the data fusion process that could be applied to all application fields. It is a five-level model that was inspired from the JDL model. It excludes terminologies ambiguity appearing in the first model and gives a clear overview about the fusion degree of gathered data. In the following we describe the different levels of Process refinement model;

- **Data In-Data out (DAI-DAO):** It corresponds to the preprocessing phase of the JDL model where incoming data from different sources are treated and refined separately without any combination. At this level, basic functions and algorithms are applied to the gathered data such as signal amplification and redundancy elimination in text data.
- **Data In-Feature out (DAI-FEO):** At this level, data with the same type of presentation are analyzed and treated together to extract the feature they could present in the environment. This fusion process is similar to the level 1 of the JDL model.
- **Feature In-Feature out (FEI-FEO):** This fusion level consists of the combination and analysis of all features resulted from previous grouped data to give clear and meaningful descriptions of an event or situation in the network.
- **Feature In-Decision out (FEI-DEO):** At this level, the given feature descriptions are introduced in a decision making system to clear the situation and plan for possible actions that could be taken.
- **Decision In-Decision Out (DEI-DEO):** As a final

processing for the data, possible decisions given by the previous fusion level are combined and matched together with other previous stored decisions to give a result.A final action to be taken to face an event or abnormal situation in the environment and react according to that (Samara *et al.,* 2010).

Data Collection Schemes (DCS) in VANET

Various DCSs has been proposed by various authors for collecting the path information for vehicles. The schemes can be categorized into two groups Static and Dynamic. In static schemes path information remain static if due to some reason vehicles change the path that information cannot be provided by the static schemes but dynamic schemes are capable of providing such kind of path information. Global Positioning System (GPS), which was evolved in 1973 by U.S. Department of defense. It consists of twenty four satellites which operate in orbit moving around the earth. Every one out of 24 satellites rotates around the earth twice a day at the height of about 20,200 km. The orbits are placed in a manner such that each region of earth can be monitored by minimum 4 satellites. A GPS receiver constantly receives information from these 4 satellites. Different events are occurring by sensing different activities like temperature, humidity, fire, etc. using GPS. GPS is used to locate the moving nodes of the network so that they can be tracked easily. Time of Arrival (TOA) is used to locate the position of the moving vehicle by using trilateration technique. But it may not be the accurate location information due to various line of sight problems such as obstacles in path like buildings, trees etc. Each vehicle in the VANET should be equipped with GPS, as it is not available all the times and is not even robust enough so it leads to many problems in VANET. A protocol that was used to find the location of a vehicle without the use of GPS due to its above said disadvantages. It makes use of the clustering approach, where cluster heads are appointed and they communicate with the other nodes in the network using a single coordinate system (Kaplan, 1996). Map matching is not a position tracking technique but it can be used with one such technique like GPS.Map matching technique can find the location tracked by GPS on a preloaded map and pinpoint the location of those vehicles on map.

Various match mapping techniques are available for map matching to match the GPS points that consider only short polling time intervals. So a new algorithm was proposed that works for long interval polling (Yang, 2005). Dead reckonings are a method that is used to find location of neighbor nodes. Based on the previous known position of a vehicle and by finding its movement information through speed, distance and time, the latest position of that vehicle can be obtained. The previous known position of a vehicle can be found using GPS or any reference point like river, any crossing, etc. Dead reckoning is used in the situations where GPS fails like under a tunnel a vehicle losses its GPS connection. Next best architecture is cellular architecture. Here area is divided into cells and location of a vehicle is traversed through the signals received from them. During handoff vehicles pass their information to the towers which help in detecting the congestion levels or any cellular activity. Hand off traffic model using mean and variance, and poison method is not valid for longer times. The focus was on to identify the traffic pattern during the hand off process. This will help to identify the traffic pattern in VANET as the nodes are highly mobile. A metric is used to CDT (Cell Dwell Time) that was used to find traffic congestion on the road. CDT is the time till a vehicle is connected to a base station before its hand off to other base station. So more is the value of CDT, more is the traffic congestion on road as it will be connected tone base station for more time. Different update techniques established to find the position of moving vehicles. In this vehicles are treated as clients and central database as a server. Next best scheme is a cluster based communication scheme for VANET. Here the communication is made under the observation of nominated cluster head. The communication based on cluster head is more reliable as failure of cluster will not affect other one (Sakthipriya *et al.,* 2014). Compared the performance of existing routing protocols in VANET by changing the various performance metrics like traffic pattern, speed of vehicles and number of vehicles on the road. It also focused on the amount of memory required for each routing protocol. Dynamic schemes are more efficient in terms of CO, Latency and PDR with respect to the above schemes. Generally it categorized DCS into two groups depending upon their data collection method as RSU initiated or Vehicle initiated. The various DCSs are as follows:

RSU Initiated (RI)

In this scheme, RSU will initiate a beacon message after M seconds to vehicles in its vicinity. Each vehicle in response will generate a packet containing the partial path information to the RSU. A new scheme road side probing where Road Side Units (RSU) commence the probing procedure and enquire each vehicle that is going on the road to collect environmental, traffic, accidents information.

Vehicle Initiated-Broadcast Mode (VIB): In this scheme, vehicle will initiate the packet transmission in broadcast mode to all the RSUs in its range. There are two proposed methods to collect data using broadcast mode, confined two-hop broadcast mechanism and probabilistic confined two-hop broadcast. Their basic need is to detect traffic congestion and improve the detection range of RSU to maximum extent. VIB scheme is further divided into two schemes:

VIB-New Segment (VIB-NS): Vehicle will transmit the packet whenever a new segment is received by it in its path.

VIB-Complete Path (VIB-CP): Vehicle will transmit the packet only when they have collected the complete path information, that is, when it has traversed all the road segments in its path and now it has reached its final destination and stopped moving.

Vehicle Initiated-RSU Find Mode (VIR): Vehicle will initiate the packet transmission as a unicast message for that it will broadcast an RSU to find message and the RSU which is present in its range and is close to it will reply first with a message containing its address. VIR scheme is further divided into two:

VIR-New Segment (VIR-NS): Vehicle will transmit the packet to a particular RSU whenever a new segment is received by it.

VIR-Complete Path (VIR-CP): Vehicle will transmit the packet only when they have collected the complete path information (Gamess and Chachati, 2014).

Data Mining (DM) Techniques in VANET

DM as a technique that is used to obtain unambiguous information from the data collected. Here various DM techniques used in VANETs and the applications they have served when implemented are discussed.

Clustering

Clustering is basically an unofficial learning technique. In this classes are formed and each class is known as a cluster such that intra-cluster association is more than inter-cluster association and each cluster is observed by a Cluster Head (CH). Clustering technique provides more real time information related to traffic jams or accidents on the road. Various clustering techniques for mining data for VANET are;

In this technique where nodes exchange their states using HELLO messages and if it does not receive any message for a time limit it announces itself as the Cluster Head (CH), otherwise it registers itself as the member node to the existing CH. Here nodes know their destination, so they are forwarded to directly. Nodes can act as members for multiple clusters. So they can be referred as Gateways and are will therefore route the packets to their destination. A new distributed scheme known as cluster based multi-channel scheme. This scheme is for improving the QoS in VANET. The main aim is to provide QoS for delivery of the real-time data and thus improving the throughput for the traffic that is non-real-time. Another clustering technique for cluster-head selection which is similar to finding the Minimum Dominating Sets (MDS). This approach is well known as Position based Prioritized Clustering (PPC). A new clustering scheme that categorizes vehicles into groups on the basis of their range of speed. Vehicles which follow same speed are placed in one group or same cluster (Su, ?).

Association Rule based Mining

Associations as the discovery of togetherness or connection of objects. An association rule shows relationship among objects such that existence of one object in a set reveals existence of another object related to it ,that is collecting the past interest behavior of a customer, decision made for an organization can be improved. The association rules are generated according on the user request. The latest scheme called VANET Association Rules Mining (VARM) where each vehicle collects data for each neighbor in its range, and extracts various rules for temporal correlation so as to detect malicious or faulty vehicles, association rule is used to identify the behavior of neighboring vehicles and then have a check for malicious or faulty vehicles. Association rules mining was used to extract control rules for the linguistic information system in the context-aware driving assistance system. That helps to prevent the occurrence of traffic accidents (Provisioning over vehicular ad hoc networks, 2007). There are four ways suggested for improving the efficiency of association rules in the computational cost;

- Reduce the number of elapses over the database,
- Sampling the database,
- Add more constraints on the organization of patterns,
- Use parallelization.

Classification

A model which describes a set of predetermined classes from a set of tuples known as the training set. This technique is supervised, which means that it requires an input from the user, the training set, to build the model and train it accordingly. So many classification techniques used for different applications in VANET, such as Bayesian networks, decision tree induction, case-based reasoning, k-nearest neighbor classifier, fuzzy logic techniques and genetic algorithm. Security engineer¬ing approach for VANETs towards the assessment of the security needed for certain applications. In order to do so, the authors proposed the use of classification DM in order to analyze the large set of VANETs applications, clas sify them according to their security requirements and provide security solution for each class of application (Sujatha *et al.,* 2013).

Sequential Mining

Sequential mining as a technique that is used to discover sequences of events that commonly occur together and frequently. It can be on the basis of time or order of events. An item is called frequent if it occurs more than a predefined threshold value. A new algorithm is designed to find the frequent mining pattern by giving the formal definition of movement patterns. A new protocol DFS_MINE for fast mining of patterns by using depth first approach to find the longest sequential pattern (Lee *et al.,* 2009).

Conclusion

We offered works related most important components in VANET, which are various levels of data and data collections. A secure and efficient collection of data flowing in the vehicular network opens possibilities to develop new applications and enhance existing ones. We have reviewed different level of techniques and methods, protocols developed for data collection which leads to an identification of weaknesses related to real-time traffic consideration. We also discussed various data mining techniques for and analysis to enhance applications efficiency and preserve higher quality of service for the clients. A study that confirmed the data mining techniques to provide QoS for delivery of the real-time data and also improve the throughput. Classification is a Data mining technique provides security solution for each class of application.

Future Work: As a fast emerging area, VANET is imposed on various problems and research oppor¬tunities that are different from traditional Wireless LAN or cellular networks. The areas covered in the above surely will helps the young researchers to identify the various data collection schemes and other future work includes developing a more efficient method for mining the data ,and we also expect the ideas of distributing the DMV's duty to multiple RSUs would help applications other than detection.

REFERENCES

Zeadally, S., Hunt, R., Shyan, Y., Angela Irwin, C., Hassan, A. 2012. Vehicular ad hoc networks (VANETS): Status, results, and challenges, Telecomm nation System.

EURASIP Journal on Wireless Communications and Networking.

Bakhouya, M., J. Gaber, and P. Lorenz, 2011. An adaptive approach for information dissemination in vehicular ad hoc networks. *Journal of Network and Computer Applications*, 34(6):1971–1978.

Yang, Q., Lim, A., Li, S., Fang, J., Agrawal, P. 2010. ACAR: Adaptive Connectivity Aware Routing for vehicular ad hoc networks in city scenarios. *Mobile Network Application.*

Yang, Q., Lim, A., Li, S., Fang, J., Agrawal, P. 2010. ACAR: Adaptive Connectivity Aware Routing for vehicular ad hoc networks in city scenarios. *Mobile Network Application.*

Samara, G., Wafaa, A.H., Sures, R. 2010. Security analysis of Vehicular Ad Hoc Networks. Second International Conference on Network Applications, Protocols and Services.

Kaplan, D., Christopher, J. 1996. Understanding GPS: Principles and applications. Artech House.

Yang, J., Kang, S. 2005. The map matching algorithm of GPS data with relatively long polling time intervals. *Journal of the Eastern Asia Society for Transportation Studies.*

Sakthipriya, N., Sathyanarayanan, P. 2014. A reliable communication scheme for VANET communication environments. *Indian Journal of Science and Technology,* june.

Gamess, E., Chachati, M. 2014. Analyzing routing protocol perfor-mance with NCTUns for vehicular networks. *Indian Journal of Science and Technology,* Sept.

Su, H., Zhang, X. Clustering-based multichannel MAC protocols for QoS.

Provisioning over vehicular ad hoc networks, 2007. IEEE Transactions on Vehicular Technology, Nov.

Sujatha, M., Prabhakar, S., Devi, G. 2013. A Survey of classification techniques in data mining. International Journal of Innovations in Engineering and Technology (IJIET). Aug.

Lee, J., Chen, Y., Chong, W. 2009. Mining frequent trajectory patterns in spatial-temporal databases. *Information Sciences.*

****