Communications Software Technical Committee – CommSoft

IEEE COMMUNICATIONS SOCIETY

Communications Software Technical Committee - CommSoft

http://committees.comsoc.org/commsoft/policies.html

IEEE CommSoft E-Letters

Vol. 6, No. 1, May 2017

CONTENTS

MESSAGE FROM COMMSOFT TC CHAIR

The Technical Committee on Communications Software (TC-COMMSOFT) examines methods and techniques devoted to advancement of the formal methods and tools, use of system analysis and design, methodology for development of communications protocols as well as application of general Software Engineering approaches for the purpose of development of communications applications. The issues addressed by the TC-COMMSOFT include domain-specific languages and practices of using them. Developing of "middleware" between networks and applications and the usefulness and usability of it is also a topic.
In this volume, we have selected 2 papers on three various subjects: Big data for VANETs and an overview of multipath routing protocol .

How to use Big Data with VANET

Emilien Bourdy
CReSTIC,
University of Reims Champagne-Ardenne, France
Email: Emilien.bourdy@univ-reims.fr

Abstract: In this article, we will discuss if it's possible and how to use Big Data mechanism with vehicular ad-hoc networking (VANET). Actually, connected vehicles are being manufactured and will become an important part of vehicle market. Topology in this type of network is in constant evolution accompanied with massive data coming from increasing volume of connected vehicles in the network.

Keywords: ITS, VANET, Big Data

1. Introduction

Over the past few years, the volume of data has increased tremendously and in a large scale within various domains, particularly, in vehicles with cooperative Intelligent Transport Systems (ITS). We can observe the emergence of connected vehicles, which will generate high amount of data in the near future. In Vehicular Ad-hoc Network (VANET), the data are generated either from the infrastructure by the Road Side Unit (RSU), or from the vehicles themselves by the On-Board Unit (OBU). RSU and OBU communicate with each other using IEEE 802.11p (vehicular Wi-Fi) [1] standard, which is called ITS-G5 in Europe. With a prediction of 35 % of market share among vehicles that will be marketed in 2022 along with revenues of around 113 billion euros [2, 3], ITS data volume is becoming massive in such a way that they can generally be considered as Big Data and they will have the same problematic.

In existing papers and projects, Big Data is usually presented in a "macro" view [4]: they use Big Data and ITS together as infrastructure or vehicle in order to interact with other entities (i.e., operators). For example, it will be used to ensure logistics between train and plane. And when it is used in ITS, it is in the broad sense, by exploiting Big Data with vehicles and infrastructure [5, 6, 7]. In this article, we focus on the use of Big Data within the network and data are used and produced by vehicle in a "micro" view: data are collected from the network (vehicles) and used for traffic management or for driver behavior modeling. The rest of this paper is organized as follow. When we use Big Data mechanism, some steps are needed: data generation (section 2), pre-processing (section 3), communication to the data center (section 4), data storage and analyses (section 5). Finally, section 6 provides conclusions.

2. Data Generation

2.1 Types of data

In general, there are two types of data in VANET: message data and sensor data. Message data are deployed for cooperative driving and sensor data for helping driver to make decision or to analyze driver's behavior within a specific situation. A simple example is a truck carrying dangerous goods that will be displayed in Cooperative Awareness Messages (CAM) [8]. This is an example among so many others of VANET's usage. Automatic Breaking System (ABS) and cruise control are examples of driving assistance systems that can be activated in this situation. As for sensor data, if a vehicle knows that the road is slippery by noticing that the braking is not efficient or by observing high level of humidity, this vehicle can send a message to others vehicles that do not have this information, and hence all the neighborhood can be aware of the situation.

In Europe, VANET uses the standards from ETSI (European Telecommunications Standards Institute) who defines two main types of messages called Cooperative Awareness Message (CAM [8]) and Decentralized Environmental Notification Message (DENM [9]). The first one is used to indicate the position of the ITS-Station (ITS-S) to its neighborhood and the second one is used to warn them of events. More precisely, the CAM is used to collaborate with the neighborhood by indicating positions, speeds, directions, etc. It needs GPS and potentially Control Area Network (CAN) bus, which is a vehicle integrated bus in which all sensors data are provided. Connected vehicle will use this CAN bus to provide more details and better collaborate. CAM will be used in majority by car manufacturers that have access to the CAN bus. Instead of CAM, which are sent automatically and periodically, DENM are used for events and so they are either using CAN bus to produce automatic event (like emergency braking) or manual event with the use of Human Machine Interface (HMI).

Furthermore, ETSI also defines five other types of messages: POI (Point of Interest) for warning driver of interesting locations like gas station; SPAT (Signal Phase and Timing) for traffic lights; MAP for road topology; IVI (In Vehicle Information) for communicating variable-message sign data inside vehicles; EV-RSR (Electric Vehicle Recharging Spot Reservation) for electric vehicles. These messages are essentially used in one-way direction: from infrastructure to vehicle, because they are service-related and are "static" messages (no need of the CAN bus), hence, we will not go into details for these types of messages.

2.2 Volume, frequency and variety

In terms of frequency, CAMs are sent from 1 to 10 messages per second in functions of the situation of the ITS-S. For example, Road-Side Unit (RSU) or vehicle that moves slowly will send one CAM per second, and vehicle that moves fast will send 10 CAM per second. A vehicle is considered as a fast-moving vehicle if one of the following conditions is validated:

- absolute difference between two headings exceeds 4°,
- distance between two positions exceeds 4 m,
- absolute difference between two speeds exceeds 0.5 m/s.

On the other hand, DENM are sent only one time unless a transmission interval is defined. CAM are sent to ITS-S' neighborhood, and DENM are sent to a geographic area and so multi-hop transmission is needed.

In terms of volume, CAM size transmitting from vehicle ranges from 41 bytes to 430 bytes and from 26 bytes to 160 bytes those transmitting from RSU. However, in practice, they alternate four "light" CAM and one "heavy" CAM per 500 ms. This means that the higher the number of connected vehicles, the higher amount of data per second increases drastically only with CAM.

In terms of variety, even if CAM and DENM are normalized, some fields are optional and are implementation-dependent because of which, messages can vary from an ITS-S to another. This will be even more complicated if we take into consideration the sensor data. For example, in the European Scoop@f project [10], there are 3,000 connected vehicles within 2,000 kilometers of road on five sites with different types of road: "Ile-de-France" (Paris and suburb areas), "East Corridor" between Paris and Strasbourg, Brittany, Bordeaux, and Isère; where each of them has its own hardware and software providers. In this project, there are two RSU vendors and six OBU vendors. For OBU, two types are defined: car manufacturers and road manager (RM). For RM, they may have two types of utilization: as simple users with same use cases than car manufacturer (emergency braking, stationary vehicle) or as mobile RSU with road operator use-cases like roadwork warning.

2.3 The 4 V

Big Data comes generally with the 4 V: Variety, Velocity, Veracity, and Volume [4]. Variety means that generated data changes from a station to another. Velocity means that data from a station change quickly. Veracity means that data is coherent with the situation (for example, a station located in Paris will not indicate a position at New York). Finally, Volume means that there is a huge amount of data so that they cannot be analyzed by classic computers. In VANET, we can also find the 4 V:

1. Variety: Different ITS manufacturers implement standard in different ways (e.g. some manufacturers have access to CAN bus and some others not).

3. Velocity: With a frequency of 1 to 10 messages per second for each moving vehicle, and for each message, some data will change quickly from one message to another like position, timestamp, speed, etc.

4. Veracity: All data need to be coherent with the situation to have an efficient cooperative ITS.

5. Volume: With all data generating from messages and sensors and the number of vehicle, data volume increases fast. In [2, 3], they predict that connected vehicle market will grow from 5 million units to 35 % of the vehicle marketed in 2022.

For the moment, Veracity and Volume are not a big concern because there are not yet enough connected vehicles. On the contrary, Variety and Velocity already raise new issues. In the near future, when more vehicles will be connected, then the huge volume of data will require Big Data algorithm (e.g., 60 bytes per vehicle per 100 ms). Veracity will be essential as well, because we cannot have an efficient cooperative ITS with incorrect data. Of course, if the car indicates a location of Berlin instead of Paris, the data will be absurd, and will be detected easily. However, when the car is near an access road, it is more difficult to know if it is on the access road or still on the highway. Deep-learning approaches will be useful to determine situations and provide recommendations. We can determine two types of data processing: off-line and on-line. Deep-learning are more often off-line processing that need an important volume of data to be processed *a priori*. Use cases definition will be trained and validated off-line, and then we will use prediction or recommendation in an on-line manner. Moreover, we also need traceability, because if there is a problem, we can go back to look at the traces. We know that this situation is going to happen soon, so we have to consider it and begin to prepare solutions.

3. Pre-processing

3.1 Message pre-processing

For a better understanding of how pre-processing is conducted, we begin by first introducing the structures of the messages. CAM is composed of five:

1. `ItsPduHeader`: general header, which indicate type of message, protocol version and station identifier,
2. `BasicContainer`: generation time and ITS station (ITS-S) position,

3. `HighFrequencyContainer`: containing data, which evolve quickly for vehicle (speed, acceleration etc.) and list of protected communication zone for road side unit like tolling zone,
4. `LowFrequencyContainer`: containing data, which evolve slowly for vehicle (exterior lights, path history and vehicle role for vehicle and tolling zone information for RSU).
5. `SpecialVechicleContainer`: containing data depending on vehicle's role to specify how vehicle is a special vehicle (light bar and siren for example).

DENM is also composed of five elements as presented:

1. `ItsPduHeader`: it is the same as for CAM,
2. `ManagementContainer`: containing event position, validity duration, etc.
3. `SituationContainer`: containing information about the type of event, information quality of the event (from 1 to 7, with 1 for very poor quality and 7 for the most reliable information about the event), etc.
4. `LocationContainer`: containing more information about the location of the event (paths to go to the event, type of road, etc.)
5. `AlacarteContainer`: containing more information depending on use case.

According to CAM and DENM structure, there are two ways that we can eliminate redundancy and make aggregations for data. First, we can make an aggregation of CAM data by creating a lighter CAM that contains only average data (speed, heading, acceleration, length, etc.) in a given observation interval and transmit only these aggregated data to the Big Data platforms. Second, we can also select the predefined useful data in the DENM, because some information (e.g. height and position of the longitudinal career left or right) are here to improve the impact reduction in pre- and post-crash use cases but are *a priori* not useful for other statistics.

3.2 Sensor pre-processing

In order to prevent Internet of Things (IoT), where many sensors are deployed, from sending useless data or too much amount of data too often, the pre-processing can be set to send data only when significant changes occur. For example, a temperature sensor does not need to communicate engine temperature every second, but only when it is hot enough to become dangerous. If sensor does not send data, it means that data has not evolved. Thresholds can be set to achieve this.

3.3 Implicit pre-processing

Fortunately, some pre-processing is implicitly made by VANET communication protocol: if a message is sent many times by multi-hop broadcasting (cf. 4.2) or by sending many times the same event when new neighbors appears, the protocol discards these messages, and reduces the redundancy.

4. Communication

4.1 Architecture

In VANET, two types of communications exist, which are commonly called V2X (Vehicle to X):

1. Vehicle to Vehicle (V2V), when data are sent from a vehicle to another to cooperate or to warn of an event;
2. Vehicle to Infrastructure (V2I / I2V), when data are sent from a vehicle to an RSU (and vice-versa) in order to share information about events.

According to [11], there are three layers in a Big Data network architecture:

1. Access Network. It is the user side, where devices communicate with the network infrastructure.
2. Internet Backbone. It is where data is transported to the data center. This means the communications from RSU to data center (and vice versa).
3. Inter- and intra- data center network. It is the network formed by the data centers: intra-data center when data center is inside the same building or farms and inter-data center when data centers are situated in different locations.

From a VANET point of view, the specific arrangement will be done at the access network. There are two main possibilities: use V2I communication to collect data or use cellular network. However, there is another alternative using a hybridization between these two technologies: vehicle uses V2I near an RSU and switches to the cellular network to guarantee QoS when RSU becomes too far.

4.2 Message dissemination in VANET

In VANET communication, there are two types of dissemination: single-hop and multi-hop. Single hop is used to communicate with the neighborhood and multi-hop to send message to a destination (can be both an ITS-S or a geographic area) [12]. Because of multi-hop mechanism, ITS-S will forward these messages, and they can become redundant for the observer. As mention before in 3.3, the known message of an ITS-S will not be forwarded again thanks to the implicit pre-processing. For example, in Figure 6, the red car sends a message, the blue car forwards it; however, the red car will not forward the message again, because it already knows this message. When we use multi-hop dissemination, denser is the traffic, faster is the dissemination. In [13], a traffic light sends an echo to a pre-defined distance. Vehicles receive the echo and reply with multi-hop dissemination. Traffic light collects the response and estimate traffic density with the delay of reply (shorter is the delay, heavier is the traffic density).

5. Storage

5.1 Storage mechanisms in Big Data

According to [4], data can be stored in structured, semi-structured or unstructured databases. Structured data are databases or tables like Excel tables, unstructured data are files, logs, video or web pages, and semi-structured data are JSON files with keys for example. If we have a lot of unstructured data, then Big Data approaches are indispensable, and so we need NoSQL (Not only SQL) technologies. In [4], the authors compare three main NoSQL databases. Firstly, Key-values Databases are constituted by a simple data model and data is stored corresponding to key-values; for example, Dynamo and Voldemort. Secondly, Column-oriented Databases store and process data according to columns other than rows for examples, BigTable and Cassandra. Finally, Document Databases use document as key; for example, MongoDB, SimpleDB, and CouchDB. To have an efficient analysis of these databases, parallel computing is needed, but it can be improved by using some models like MapReduce, Dryad, All-Pairs, or Pregel in addition to classical parallel models like OpenMP, MPI.

5.2 Structure of VANET data

We recall here that CAM and DENM are structured messages defined by a standard so they match structured data type. However, sensor data are implementation-dependent and so are semi-structured data because they are raw data that have an implicit structure to be able to retrieve data. Consequently, for messages such as CAM and DENM, we can use classical relational databases systems (RDBMS), however we need NoQSL databases for sensor data. Fortunately, NoSQL means Not Only SQL, and then we can use a NoSQL database to store both structured and semi-structured data [14]. Nevertheless, in the future VANET, we will need all these technologies because VANET will also include unstructured data like video, news, storm warning, etc.

5.3 Current Data storage in VANET

In VANET, there two levels of storage:

1. ITS-S,
2. ITS-C (ITS Central platform)

At the ITS-S level, ETSI defined a standard called Local Dynamic Map (LDM) [15]), which is used to retrieve events that occur near the ITS-S and to discover the neighborhood. This LDM is implementation-dependent; however, this standard only specifies requests between LDM and ETSI layers. Therefore, LDM mainly stores DENM and CAM. To access to the data, LDM supports filters in request like equals, not equals, greater than or equal to, like, not like etc.

At the ITS-C level, the storage method is implementation dependent. The method will be made in function of data type we need to store. If we only store sensors data, then semi-structured data will be used, but if we store only messages, then we will use RDBMS.

6. Conclusion

In this paper, we have seen that data in VANET can be considered as Big Data because VANET complies with the 4V: Variety with the number of stations that uses different data, Velocity with data that evolve quickly, Veracity to ensure a good cooperative ITS, and Volume with the number of station that produces huge amount of data. Some Big Data pre-processing mechanisms will be needed to improve data quality and utility. Data will be able to use VANET to reach data centers. Data can be stored in different way according to the situation.

References

[1] IEEE Draft Standard for Information Technology Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (November 2011).

[2] Connected Car Market by Hardware, Application, and Geography - Global Forecast to 2022, Report ID: 4090898, March 2017.

[3] Christian Radüge Richard Viereckl Jörg Assmann. “The bright future of connected cars”. In: Strategy& (2014).

[4] Min Chen, Shiwen Mao, and Yunhao Liu. “Big data: a survey”. In: Mobile Networks and Applications 19.2 (2014), pp. 171–209.

[5] LeMO project. URL: http://cordis.europa.eu/project/rcn/ 211234_fr.html.

[6] TransformingTransport project. URL: https://transformi ngtransport.eu/.

[7] AutoMat project. URL: http://www.automatproject.eu.

[8] ETSI EN 302 637-2; Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service. European Standard. ETSI, Nov. 2014.

[9] ETSI EN 302 637-3; Intelligent Transport Systems (ITS); Vehicular Com-

munications; Basic Set of Application; Part 3: Specifications of Decentralized Environmental Notification Basic Service. European Standard. ETSI, Nov. 2014.
[10] Scoop@f project. URL: https://ec.europa.eu/inea/en/ connecting-europe-facility/cef-transport/projectsbycountry/multi-country/2014-eu-ta-0669-s.
[11] Xiaomeng Yi et al. “Building a network highway for big data: architecture and challenges”. In: IEEE Network 28.4 (2014), pp. 5–13.
[12] ETSI EN 302 636-4-1; Intelligent Transport Systems (ITS); Vehicular Communications; GeoNetworking; Part 4: Geographical Addressing and forwarding for point-to-point and point-to-multipoint communications; Sub-part 1: Media-Independent Functionality. European Standard. ETSI, July 2014.
[13] S. Gashaw and J. Härri. “V2X data dissemination delay for vehicular traffic density estimations”. In: 2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM). June 2015, pp. 1–6. DOI: 10.1109/WoWMoM.2015.7158211.
[14] Rolf Sint et al. “Combining unstructured, fully structured and semi-structured information in semantic wikis”. In: Fourth Workshop on Semantic Wikis–The Semantic Wiki Web 6th European Semantic Web Conference Hersonissos, Crete, Greece, June 2009. Citeseer. 2009, p. 73.
[15] ETSI EN 302 895; Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Local Dynamic Map (LDM). European Standard (Telecommunications series). ETSI, Sept. 2014.

Acknowledgements

This work is partially supported by the EC SCOOP project (INEA/CEF/TRAN/2014/1042281).

Biography

Emilien Bourdy obtained M.S. in Computer Science specialized in development of distributed application from University of Reims Champagne-Ardenne (France) in 2014. He joined the SCOOP project (European C-ITS deployment project) in 2015 as a research engineer about on testing and evaluation of C-ITS components. He started his PhD on 2017 working on Big Data algorithms for vehicular network to model driver behavior.

REPORT OF LEADING SIG ACTIVITIES

- Special Interest Group on "Communication softwares for Vehicular AdHoc Networks"
Coordinator : Prof. Hacene Fouchal, Hacene.Fouchal@univ-reims.fr

- Special Interest Group on "NFV and SDN technologies"
Coordinators: Dr. Adlen Ksentini, adlen.ksentini@eurecom.fr

- Special Interest Group on "Security in Software Communication"
Coordinators: Prof. Jalel Ben-Othman, Jalel Ben-Othman
Dr. Yessica Saavedra Yessica Saavedra
Address theoretical, conceptual and technological aspects of communication

- Special Interest Group on "Big Data"
Coordinator : Dr. Periklis Chatzimisios, peris@it.teithe.gr
Address new trend on Big Data and Communication software

- Special Interest Group on "Designing Future Optical Wireless Communication Networks-DETERMINE"
Coordinator: Dr. Scott Fowler, scott.fowler@liu.se

•

TC OFFICERS AND NEWSLETTER EDITORS

TC Officers

Names	**Affiliation**	**Email**
Hacène Fouchal	Université de Reims Champagne-Ardenne, France	Hacene.Fouchal@univ-reims.fr
Adlen Ksentini	IRISA, France	Adlen.Ksentini@irisa.fr
Abdellatif Kobbane	University of Rabat Morocco	abdellatif.kobbane@um5.ac.ma

Editor-In-Chief

Names	**Affiliation**	**Email**
Lynda Mokdad Hacène Fouchal	University of Paris-Est, Créteil Université de Reims Champagne-Ardenne	Lynda/mokdad@u-pec.fr Hacene.Fouchal@univ-reims.fr