Spatio-temporal distributed background data storage and management system in VANETs

Josip Balen
Faculty of Electrical Engineering, J. J. Strossmayer University of Osijek, Croatia
josip.balen@etfos.hr

***Abstract*—With the advances in the development of embedded technology, wireless communication and VANETs, and the awareness that vehicles possess a vast amount of unused resources, new services and applications have been proposed recently. The emphasis of this paper is on describing an idea, related work, and open research challenges of using parked vehicles as a spatio-temporal distributed storage system for large amounts of data. We have briefly surveyed different data storage and management techniques in VANETs in order to get directions for the implementation of our strategy. Open research challenges and system requirements, as well as future work are pointed out.**

I. INTRODUCTION AND MOTIVATION

The use of parked vehicles in Vehicular Ad Hoc Networks (VANETs) has gained attention in the recent years due to the awareness that they represent static roadside nodes which are available in large number, long-time stationary, well distributed among specific locations, and often a vast of unused resources. Therefore, parked vehicles have been included in VANETs as a complement to the driving vehicles and Road Side Units (RSUs). Moreover, new services and applications have been proposed recently, for improving Internet access throughout [1], extending RSUs' service coverage [2], and assisting Vehicle-to-Vehicle (V2V) communication [3].

Due to the dynamic and unbalanced nature of VANETs, topology management is one of the key issues while providing services. Here as well, VANETs benefit from leveraging parked vehicles since they remain static for a longer period of time and exist in large numbers. Furthermore, roughly 70% of vehicles are parked for an average of 23 hours per day [2], during the day in front of companies and in the city centers and during the night in front of home buildings and houses. Hence, by using parked vehicles as a backbone in VANETs it is possible to offer services and applications requiring certain stability and reliability like storage of large amounts of data.

Motivation for this paper also comes from the fact that parking lots are distributed throughout the urban areas at pre-known locations and occupied with parked vehicles on average more than 50% of the time [4]. Therefore, parked vehicles in urban areas represent a huge unused resource in terms of distributed storage and processing power that could be used as a spatio-temporal distributed data storage system. Furthermore, VANETs are ubiquitous, particularly in urban environments, hence users can easily access it and transfer data.

Another important point, as described in [5], is that persistent storage is becoming less expensive over time and it is expected that computers in vehicles will have multiple types of storage attached. Furthermore, vehicles are not considered as resource constrained since they have plenty of space to accommodate multiple hard drives even with today's technology and sizes. This will result in a vast amount of unexploited and wasted storage resources.

A centralized approach to using mobile network infrastructure and 3G/4G for connecting to online servers is often expensive especially when roaming. Furthermore, it is unlikely to expect the Internet uplink capacity to scale faster than users' bandwidth needs. Although transferring bigger amounts of data using a mobile network infrastructure is costly in time and money, due to the global access capabilities using mobile network infrastructure it should still be considered for low data rate applications.

We envision the two following scenarios for using spatio-temporal distributed background data storage and management systems in VANETs. In the first scenario, data is generated by vehicles or various infrastructure; in the second, data is transferred by users. The amount and size of data generated by vehicles and infrastructure like traffic cameras, traffic lights, traffic signs or weather stations is constantly increasing. We can expect vehicles and infrastructure to generate multimedia data like photos and videos, and therefore storage and management of large amounts of data will be required. In the second scenario there are two types of users: single people and companies. An example use case for single people would be a situation where

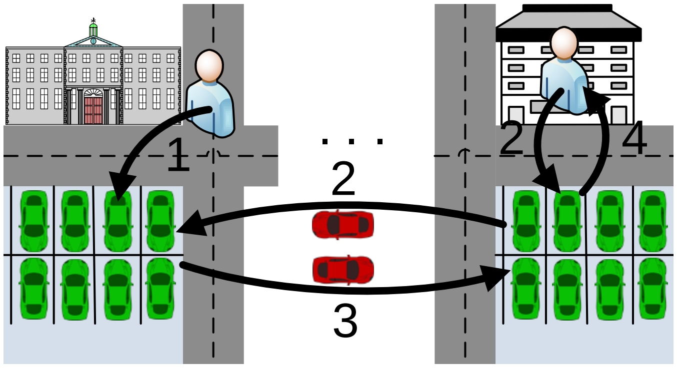

Figure 1. Spatio-temporal distributed data management example scenario. (1) User stores data in the VANET; (2) User starts data lookup; (3) Data is transferred from the source to the destination vehicle; (4) Data is transferred to the user.

they would like to temporarly store data from their smartphones and cameras in VANETs when they run out of space on their devices, as shown in Figure 1. Some interesting use cases for companies are mentioned in [5]. The authors propose using vehicles as data centers in parking lots of the companies, malls or airports. Drivers could rent their vehicle resources to the companies on a per-day, per-week or per-month basis. As a result parking lots could be turned into huge distributed data storage facilities.

In the rest of the paper, we survey different data storage and management techniques in VANETs in order to get directions for the implementation of our strategy. Furthermore, open research challenges are pointed out as well as system requirements and future work.

II. RELATED WORK

Using VANETs as a spatio-temporal distributed background data storage and management system has already been identified in the literature. We have briefly surveyed different strategies in order to get directions for the implementation of our strategy.

Location-based strategy: In [6], a location based data overlay strategy for Disruption-Tolerant Networks (DTNs) called Locus has been proposed. The basic idea is to maintain the locality of data by exchanging data between nodes that are close enough to the location where data was generated (home location). The authors argue that DHTs are not appropriate for the DTNs since the network model underlying DHTs assumes stable connectivity between all nodes participating in the overlay. Moreover, they point out that since a VANET is a highly dynamic network, a DHT based system will be unreliable and hence will have low probability of finding data in DTNs. Therefore, data objects are used as the primary entities. Data objects can contain any type of information that are tagged with an application-specific type indicator, the timestamp, and location of data creation. Data storage is based on an utility function that is based on the distance from the data location. It defines three zones: home area, drop-off zone, and the void zone. In order to keep data close to the home zone, data replication is used in the three following cases: (i) if a node is currently in the data's home area it will copy the data to surrounding nodes to increase the probability of keeping the replica in the home area; (ii) if a node is not in the drop-off zone the data is copied to the nodes closer to the home location; (iii) if a node is outside the data drop-off zone data should be replicated to the node that is moving towards the drop-off zone. Data access is done through queries and response messages. Since the primary matching condition of the query is location, nodes requesting data have to know the data location. Beside location, a querying node is sending its location and trajectory. After the data is found the node generates the response message and (based on received data and age of the query) it estimates the querying node's current location. This approach seems unrealistic, since a node's trajectory shouldn't be known due to privacy concerns. This could lead to errors in calculating the querying node's future location. Moreover, it is much easier (and loads the network much less) to keep the data close to the place where it has been generated just by transferring it to the nearby parked vehicles.

Parked vehicles domains strategy: Using parked vehicles as a temporary network and storage infrastructure is presented in [7]. The idea is to have a vehicular cloud consisting of independent clusters of parked vehicles where the data will be stored. In order to establish spatio-temporal network infrastructure for providing connectivity as well as data storage and management capabilities, the Virtual Cord Protocol (VCP) is used. VCP provides Distributed Hash Tables (DHT) services, integrated greedy routing with guaranteed delivery and inter-domain routing. VCP is based on sending Hello messages and including new vehicles in an existing domain. If existing domains are not accessible in a defined period of time, creation of new a domain is initiated. Data management is based on PUBLISH and LOOKUP operations with greedy routing for message forwarding. Upon reception of a publish message, data is mapped to hash values. Although in [6], it is argued that a DHT based system will be unreliable and have low probability of finding data in DTNs here is shown that with using DHT for data management it is possible to get a high rate of successful data access operations. Future work is need in order to build efficient inter-domain routing strategy and connect multiple disconnected domains of parked vehicles with a store-carry-forward paradigm.

Vehicular cloud computing: This concept is based on cloud computing and considers each vehicle as a computation node. The main difference between traditional and vehicular cloud computing is due to the dynamic environment of VANETs and in terms of resource availability. An exhaustive survey on vehicular cloud computing can be found in [8]. There are several potential architectures based on using vehicles' processing, storage and network resources: Network as a service (NaaS), Storage as a service (STaaS), Cooperation as a service (CaaS), Information as a service (INaaS), Entertainment as a service (ENaaS), Computing as a service and Pictures on wheels as a service. For our work, the most interesting architecture is STaaS. Since it is predicted that vehicles will have huge storage resources, [5] proposed to use parked vehicles as datacenters at the parking lots of malls, airports, and companies since there are often several hundreds of parked vehicles. However, they conclude that users cannot immediately benefit from storage like they can from computation power or network access since vehicles will eventually leave the parking lot and take data with them. Therefore, they propose creating multiple replicas of the same data or decomposing data into smaller block sizes that can be quickly delivered on request so that big files can be reconstructed. However, the concept is not farther expanded on.

III. SYSTEM REQUIREMENTS AND RESEARCH CHALLENGES

The main differences between traditional distributed data storage and data storage based on parked vehicles are in availability and wide distribution of nodes, dynamic environment and resource availability. Since parked vehicles can randomly leave and join parking lots, it is necessary to have a data management system on individual parking lot level. Furthermore, individual parking lots have to be connected with driving vehicles, and therefore it is necessary to have a data management system realizing a store-carry-forward paradigm on a more global level.

The desired spatio-temporal distributed background data storage and management system based on parked vehicles must be reliable and robust, enable data delivery in a timely manner in a big geographical area, cover sparse parts of a network, and prevent network overload. Furthermore, it has to provide on demand data storage and retrieval to users and applications. These requirements raise several research questions and challenges:

redundancy – keep in the network an optimal number of data replicas based on the type and priority of data in order to keep data safe,

storage locations – should be in proximity of the desired data delivery destination and should enable high data reachability. This could be based on prediction, user tracking (based on user request to store data on a different locations), or learning algorithms,

data validity – should be based on the type and priority of data and storage capacity,

buffer management mechanism – if storage capacity on an individual node is completely consumed data could be dropped or transferred to a different node,

area of relevance – define a geographic area based on possible data destination and data source,

timeliness – enable data delivery in a timely manner but at the same time prevent network overload. Data delivery delay directly depends on distance from the storage location. Expedited access for frequently queried data should be supported,

network bandwidth – since different QoS is handled on MAC layer, for transferring large amounts of data full network bandwidth could be used,

data transfer – fastest data transfer from source to destination could be achieved with either V2V greedy routing (for connected source and destination) or store-carry-forward routing (for disconnected source and destination).

There are several methods for meeting the above mentioned research challenges. First, all components could be predefined in the data management system. Second, based on network conditions all components could be dynamically handled. Third, applications and users could request specific quality of service.

In larger cities we can expect to have large scale vehicular networks with several hundred thousands of vehicles which represents huge distributed storage (in space and size) and pose new and unique challenges to efficient data management.

IV. CONCLUSION AND FUTURE WORK

We have scratched the surface of using parked vehicles as a spatio-temporal storage and management system for large amounts of data. Parked vehicles in urban areas represent a huge unused distributed storage and processing power resource that is currently wasted. Therefore, we consider using parked vehicles as a distributed data storage system for large amounts of data. The goal of our system is to use as little infrastructure (mobile network or RSU) as possible since it costs money and time and is a single point of failure, in contrast with infrastructure-less systems that are mostly cheap and more fault tolerant. The purpose of an efficient data management strategy in VANETs is to provide synergy between data storage and data lookup, and consequently enable reliable storage for large amount of data in VANETs. This overview will generate a productive future research with the implementation and evaluation of spatio-temporal data storage and management system in terms of data delivery efficiency and delay, network traffic, data source and destination distance by using Veins simulator with a realistic scenario.

REFERENCES

[1] R. Crepaldi, R. Beavers, B. Ehrat, M. Jaeger, S. Biersteker, and R. Kravets, “LoadingZones: Leveraging Street Parking to Enable Vehicular Internet Access,” in *7th ACM International Workshop on Challenged Networks (CHANTS 2012)*, Istanbul, Turkey: ACM, Aug. 2012, pp. 23–30.

[2] F. Malandrino, C. Casetti, C.-F. Chiasserini, C. Sommer, and F. Dressler, “The Role of Parked Cars in Content Downloading for Vehicular Networks,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4606–4617, Nov. 2014.

[3] C. Sommer, D. Eckhoff, and F. Dressler, “IVC in Cities: Signal Attenuation by Buildings and How Parked Cars Can Improve the Situation,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1733–1745, Aug. 2014.

[4] J. Balen, G. Martinovic, K. Paridel, and Y. Berbers, “PVCM: Assisting multi-hop communication in vehicular networks using parked vehicles,” in *4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT 2012)*, Oct. 2012, pp. 119–122.

[5] S. Olariu, T. Hristov, and G. Yan, “The Next Paradigm Shift: From Vehicular Networks to Vehicular Clouds,” in *Mobile Ad Hoc Networking*, Wiley, 2013, pp. 645–700.

[6] N. Thompson, R. Crepaldi, and R. Kravets, “Locus: A Location-based Data Overlay for Disruption-tolerant Networks,” in *5th ACM International Workshop on Challenged Networks (CHANTS 2010)*, Chicago, IL: ACM, 2010, pp. 47–54.

[7] F. Dressler, P. Handle, and C. Sommer, “Towards a Vehicular Cloud - Using Parked Vehicles as a Temporary Network and Storage Infrastructure,” in *15th ACM Int. Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc 2014): ACM International Workshop on Wireless and Mobile Technologies for Smart Cities (WiMobCity 2014)*, Philadelphia: ACM, Aug. 2014, pp. 11–18.

[8] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya, “A survey on vehicular cloud computing,” *Journal of Network and Computer Applications*, vol. 40, pp. 325–344, 2014.