Improving Data Accessibility in Vehicle Ad hoc Network

Sungjin Park and SangKeun Lee

Department of Computer Science and Engineering, Korea University
{sungjin75, yalphy}@korea.ac.kr

Abstract

Vehicle Ad-hoc Network (VANET) has recently received a lot of attention as the fundamental technology of Intelligent Transportation System. Among other characteristics of VANET, the rapid mobility particularly causes the problem of time constraint in delivery of data, thereby resulting in the low data accessibility. Therefore, the requested service is likely to suffer from poor performance. This paper proposes an efficient method to improve data accessibility in VANET by utilizing the data replica of Roadside Unit (RSU). Both data access pattern and driving pattern are utilized to decide which data item should be replicated in RSU. The conducted simulations clearly reveal that the proposed methodology outperforms other schemes in a variety of environments.

Keywords: *VANET, Data Replica, RSU, Driving pattern*

1. Introduction

VANET is one of the MANET with differentiated characteristics, such as traffic characteristics, the velocity of vehicle, driving pattern and data access pattern. As a main characteristic of VANET, high velocity causes the time constraint problem in delivery of data, which results in low data accessibility. Main factors of time constraint problem vary, such as heavy traffic and collision of data request in wireless network of VANET, and such as congestion, packet loss, and delay in wired network of VANET. The congestion in wired network is caused by delivery of many data requests and data simultaneously like bursts from each RSU. The delay occurs due to the multimedia data size as current trend of favorite information type of users like. The time constraint problem in wireless has been studied and is presupposed as appropriately controlled by optimized scheduling algorithm [1].

In the complicated and complex network environment, data replica in RSU can be a great option to solve the time constraint problem since the request on replicated data requires no communication cost. For the improving data accessibility with data replica in RSU, it is important to choose which data item is replicated due to the limited storage capacity of RSU. As characteristics of VANET, driving pattern and data access pattern help to choose the data item to be replicated. Generally, there are two types of driving patterns of vehicles on the road. One is the transient vehicles, which pass the road once and not repeatedly pass again on the same road. The other is the frequent vehicles (FVs), which pass the same road frequently [3]. These two types of vehicles reveal the different patterns of data access. The data access pattern of the transient vehicle is likely to request once the event data which disappear after a while. Nevertheless, the event data is requested intensively, since most transient vehicles request the same event with interest during the specific period. In case the RSU replicates the event data item based on static access frequency, this pattern may decrease the data accessibility due to the characteristic of the event data which disappears shortly and is never requested during next period. In this regard, giving a weighted value to the data item requested by FVs for selection of data item can produce good performance. Therefore, this paper proposes the algorithm to select the data item to be replicated based on data access pattern according to driving pattern.

To the best of our knowledge, in VANET environment this is the first paper to utilize the data replica in RSU for solving time constraint problem and to consider the data access pattern according to driving pattern for selection of data item to be replicated.

2. Related Work

The research on solving the time constraint issue for data access while vehicles are passing by RSU has rarely been studied. Zhang, et. al., [1] proposed scheduling algorithms for RSU that takes the problem of time constraint in VANET into account. However, they assume that all data are already stored in RSU, and they do not consider the status of wired network communication among connected RSUs at all.

In order to overcome the assumption and to solve the time constraint problem, we consider the data replica technique taking account of the status of wired network communication among RSUs. Data replica techniques have been extensively studied in [2]. It proposed SAF (Static Access Frequency) as a data replica technique to improve data accessibility based on the observation that mobility of nodes can lead poor data accessibility in MANET environment. SAF only takes data request frequency into account when choosing data item to be replicated. That is, the data item requested frequently is stored in the node which receives the request of query. It is the way to implement data replica at low cost and lower network load. However, this method does not consider the characteristics of VANET, such as high velocity, driving pattern and data access pattern.

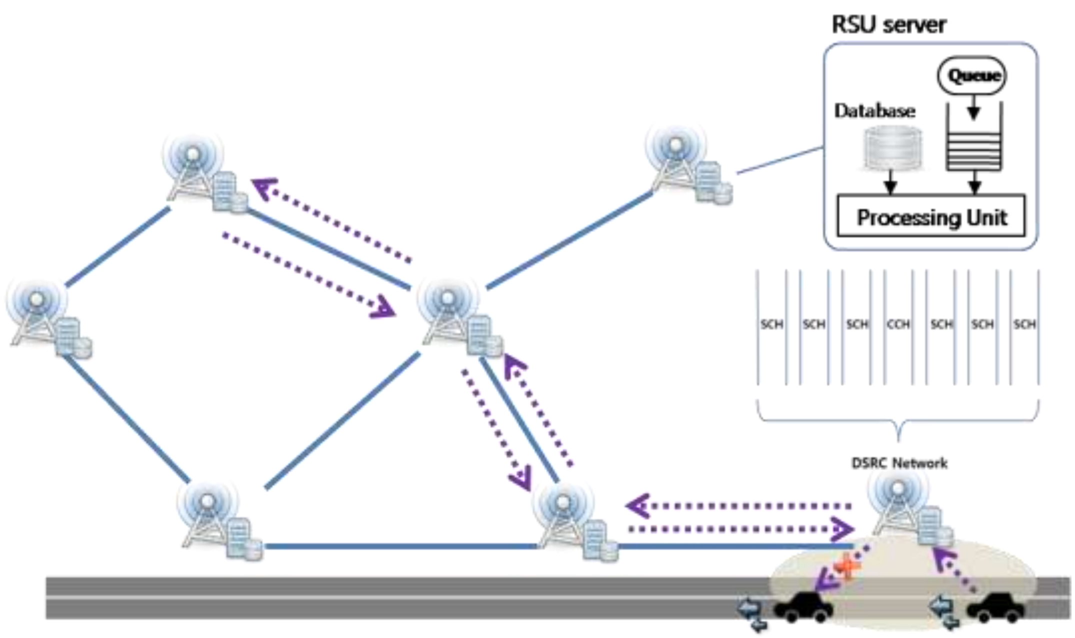

Figure 1. Data Access Model in V2I Communication

3. Modeling VANET Environment

In this paper, we research the roadside to vehicle communication based on Dedicated Short Range Communication (DSRC) compatible with IEEE 802.11p which is regarded as a fundamental technology of wireless communication standard in VANET, as shown in Figure 1. The wired network among RSUs can be high speed network than the wireless network. Wired network might have bottlenecks due to a few of lines with lower bandwidth than other lines. We make assumptions as following, similar to those in [2].

- RSUs: The set of all RSUs is denoted by $R = \{r_1, r_2, ..., r_m\}$, which m is the total number of RSUs. Each RSU has memory space of data replica as a storage server except the space for original data which the RSU owns. However, the memory space of RSU is limited with capacity of replica amount due to deployment with low cost and low management.
- Data items: The set of all data items is denoted by $D = \{d_1, d_2, ..., d_m\}$. We assume that each data item is owned by a particular RSU as the original data, so that the total number of data items is same with the total number of RSUs. The data items are not updated for the sake of simplicity, that is, we do not concern data consistency [2].
- Vehicles: The set of all vehicles is denoted by $V = \{v_1, v_2, ..., v_n\}$, which n is the total number of vehicles. We assign an additional identifier to frequent vehicles. The set of those vehicles is denoted by $FV = \{fv_1, fv_2, ..., fv_z\}$, which $FV \subset V$ and z is the total number of frequent vehicles. Each vehicle v_k has its own data access frequency to a data item $d_i \in D$.

4. Data Replica Allocation Algorithm in RSU

The replicated data enables the RSU to send the requested data immediately to the vehicle without communication with other RSUs which owns the original data. Even though the RSUs are connected with wired line, possibly it takes some time which results in low data accessibility due to congestion, delay, packet loss. It is important to choose which data item is replicated in RSU due to the limited storage capacity. As for the choice of data replica, driving pattern and data access pattern need necessarily to be taken into account.

4.1. Driving and Data Access Pattern

H. Dia [3] says that there are two different types of driving patterns such as vehicles to change route frequently and vehicles to change route rarely. According to the survey, 65% of drivers in Brisbane change the route less than 3 times per month. If the route is satisfied with the driver based on the preference, then the same route is likely to be repeated on the following driving, which forms driving pattern. With the above driving pattern, there are two types of vehicles on the road; the transient vehicles which rarely pass by due to frequent route change, the frequent vehicles (FVs) which pass by frequently on the road.

As for the data access pattern, the transient vehicles are likely to request queries related to a specific event with interest at the specific time upon passing by, such as wholesale advertisement on the road. Hence it shows the pattern which the specific event data are requested intensively within a period due to the characteristics of transient vehicles. However, the event data is likely to disappear and take place new event data after the period, and then the transient vehicles request queries the new event data with interest. Meanwhile FVs require same data continuously for their need, for examples, traffic information along the road, road condition information and fuel info which periodically change. Therefore FVs show the pattern to request queries for same data periodically.

4.2. Replica Allocation Methods

The transient vehicles request the event data intensively during specific period, so that the RSU stores the event data item based on estimation of data access pattern with previous period, in order to prepare the data request to arrive during the next period. However, it is useless if the event data disappears in the next period and has been never requested. In order to prevent this occasion, we propose the method to give a weighted value to the data item which FVs request in addition to data access frequency. This method is likely to increase cache hits so that we expect that the data accessibility is improved.

In order to choose data items to be replicated in RSU, we propose to calculate a replica value on each data item, which is defined as $RV(d_i)$. We rank the data items according to their replica values and select top k data items to be replicated. There are two approaches to compute replica values. One approach is Static Access Frequency (SAF) approach, which is our baseline, and the replica value of this approach is calculated as follows;

$$RV(d_i) = SAF(d_i) = \sum_{v_k \in V} DF(v_k, d_i) \quad , \tag{1}$$

where $DF(v_k, d_i)$ means the access frequency of vehicle v_k to data item d_i. The other approach is Frequent Vehicle Data Access Pattern (FVDAP) approach, which is our proposed method, and the replica value of this approach is calculated as follows;

$$RV(d_i) = FVDAP(d_i) = \sum_{v_k \in V} DF(v_k, d_i) + \sum_{fv_k \in FV} DF(fv_k, d_i) \times \sqrt[3]{TF(fv_k)}) \quad , \tag{2}$$

where $DF(fv_k, d_i)$ means the access frequency of frequent vehicle fv_k to data item d_i, and $TF(fv_k)$ is the total access frequency of frequent vehicle fv_k. The coefficient multiplying the frequency on the requested d_i from fv_k give an influence on increasing the value of $FVDAP(d_i)$ than the value of $SAF(d_i)$, even though the number of data access frequency on the data item is low. It results in making an effect on increasing probability to be higher ranking to the data item of $FVDAP(d_i)$ for data replica in RSU. With the heuristic approach, $\sqrt[3]{TF(fv_k)})$ is inferred as the coefficient in FVDAP approach.

4.3. RSU Operation

As for the data replica, RSU calculates $RV(d_i)$ to each data item respectively. According ranking based on value of $RV(d_i)$, RSU stores the data items with considering the capacity of storage. Upon receiving the data request, the RSU put the request into the queue and handle the queue in FIFO (First In First Out) way. The RSU deals each received data request in parallel. This results in minimizing of process time in RSU. With the request in queue, the RSU checks replicated data. If the match exists, then the RSU sends the requested data to the vehicle. Otherwise, the RSU forwards the data request to the target RSU owning the original data. The target RSU sends the requested data to the sender RSU.

5. Performance Evaluation

5.1. Simulation Environment

We developed an ns-3 [4] based simulator based on ns-3 [4] using VANET highway project [5] to evaluate the proposed data replica method. The simulation is based on 1500m *

1200m square street scenario. Every 300 meter has intersection of horizontal and vertical road, where each one way road has four lanes. RSUs are placed on the intersection. This simulation forces the vehicles to use only one road, because it is enough for vehicle to request data to one RSU. Other RSUs are aimed at providing the original data. It is designed for 400 vehicles to pass by the specific RSU. All vehicles with predefined velocity drive without collision among vehicles. In order to control the ratio of FVs on the road, some vehicles are generated with same ID and passed by the RSU repeatedly within the number of 5 ~ 20. The whole repetition of vehicles with same ID is adjusted according to the ratio of FVs on the road.

Our VANET model has 20 RSUs with their original data item. It is designed for vehicles to request the data once when passing by the RSU. The data access pattern of the transient vehicles follows uniform distribution among 20 data items to be requested randomly. Two event data are respectively requested with 20% probability when the transient vehicles pass. The data access frequency of FVs corresponds to the ratio of FVs on the road. For example, if the ratio of FVs on the road is 40%, then the data access frequency of FVs is 40%. Each FV has its own preference for one data item among predefined four data items. The data access pattern of the FV follows Zipf distribution [6]. In the Zipf distribution, the data access pattern is getting skewed with an increasing value of θ [6]. The data access probability of the data item is represented as follows;

$$P_i = \frac{1}{i^{\theta} \sum_{j=1}^{m} \frac{1}{j^{\theta}}}, \quad (3)$$

where $\theta > 0$, m is 20 as the number of data item. When $\theta = 0$, it becomes the uniform distribution. The more θ increases, the access probability of the preferred data item is severely skewed. Most of the system parameters and their default values are listed in Table 1. We measure the data accessibility in term of Service Ratio which is the ratio of number of requests served to the whole number of the data requests before vehicles pass the coverage of RSU.

Table 1. Parameter Setup

Parameter		Value	Parameter	Value
Wireless Environment		802.11p	Number of Data Item	20
Transmission Rate	Wireless	6Mbps	Data Size	2M
	Wired	100Mbps 10Mbps (Bottleneck)	Zipf Parameter θ	2 (default)
Wireless Coverage		300m	The number of vehicles	400
Wired Network Delay		30ms [7]	Vehicle Velocity	60 km/h (default)
The capacity of storage in RSU		3	Ratio of FVs on the road	60 % (default)

5.2. The Effect of Velocity, FV Ratio and Zipf Parameter

The problem which this paper arises is time constraint when vehicles request a data during passing-by a RSU. The problem becomes worse when the vehicles pass with high velocity than low velocity. Figure 2 shows the service ratio according to the velocity of vehicles. As expected, when the velocity becomes high, the service ratio gets poor performance.

Nevertheless, the service ratio with FVDAP shows better performance than those of No Cache, SAF.

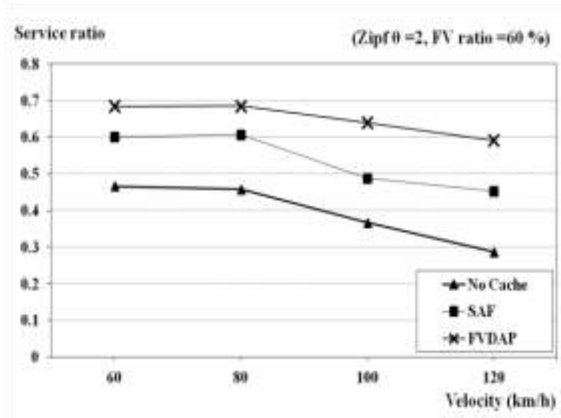

Figure 2. The Effect of Velocity

In addition, the ratio of fv_k on the road can make impact on the service ratio. Figure 3 shows the change of service ratio according to the ratio of fv_k. As the ratio of fv_k increases, the service ratio achieves high performance. Herein, two cases of fv_k ratio need to be taken into consideration. When fv_k ratio is 20% and 80%, the performance of the service ratio is same with SAF and FVDAP. In case of 20% of fv_k ratio, it results from the low weighted value for FVDAP. In other words, since the preferred data item gets very low request from fv_k, this does not make big impact on the value of $FVDAP(d_i)$. Therefore, the ranking of $FVDAP(d_i)$ has same data items to be replicated in RSU with one of $SAF(d_i)$. Similarly, in case of 80% of fv_k ratio, the ranking of $SAF(d_i)$ and $FVDAP(d_i)$ has same data item. It results from the reason why the preferred data item is the most requested due to high ratio of fv_k.

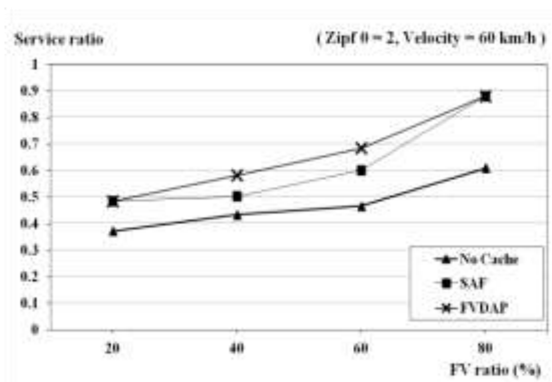

Figure 3. The Effect of FV Ratio

Lastly, Figure 4 shows the performance as an effect of the data access pattern, Zipf parameter θ. In the Zipf distribution, the bigger the parameter θ is, the more the preferred data item is intensively requested. With the regards, the service ratio becomes higher finely, as the parameter θ is bigger. The change of θ does not make big impact on the performance.

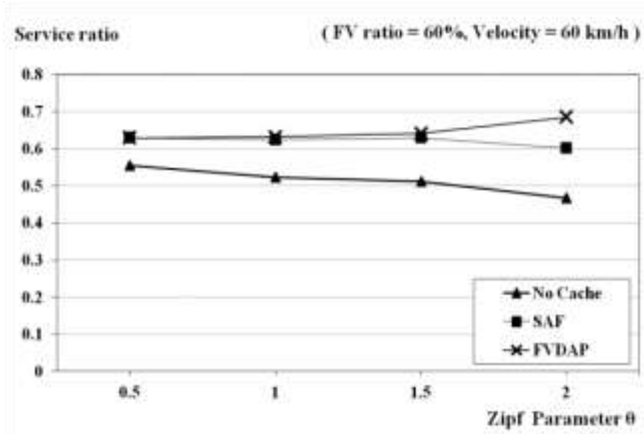

Figure 4. The Effect of Zipf Parameter

6. Conclusion

In the paper, we addressed the issue of time constraint in delivery of the requested data within the coverage of RSU during the time vehicles stay. We proposed a novel data replica method based on both driving pattern and data access pattern. Both patterns are utilized to decide which data items should be replicated in RSU. The conducted simulation shows that our methodology significantly outperforms other schemes in terms of service ratio. Furthermore, the proposed algorithm is adaptive to different FV ratio on the road. We plan to investigate the impact of different mobility pattern on the performance.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2011-0010325).

References

[1] Y. Zhang, J. Zhao and C. Cao, "On Scheduling Vehicle-Roadside Data Access", Proceedings of the 4th ACM International Workshop on Vehicular Ad Hoc Networks, **(2007)** September 10; Montreal, Quebec, Canada.
[2] T. Hara, "Effective Replica Allocation in Ad hoc Networks for Improving Data Accessibility", Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies, **(2001)** April 22-26; Anchorage, AK, USA.
[3] H. Dia, "An Agent-based approach to Modeling Driver Route Choice Behavior under the Influence of Real-time Information", Transportation Research Part C, vol. 10, **(2002)**.
[4] ns-3, http://www.nsnam.org.
[5] H. Arbabi and M. C. Weigle, "Highway Mobility and Vehicular Ad-Hoc Networks in ns-3", Proceedings of the Winter Simulation Conference, **(2010)** December; Baltimore, MD, USA.
[6] G. Zipf, "Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology", Cambridge, Addison-Wesley Press **(1949)**.
[7] W. Jiang and H. Schulzrinne, "Modeling of Packet Loss and Delay and Their Effect on Real-Time Multimedia Service Quality", Proceedings of the 10th International Workshop on Network and Operating System Support for Digital Audio and Video, **(2000)** June 26-28; Chapel Hill, North Carolina, USA.

Authors

SungJin Park received the B.S. degrees in electronic engineering from KyungHee University in 2002. He is a M.S. candidate in Dept. of Computer Science and Engineering, Korea University. He is working for Samsung Electronics as a Standard Research Engineer as well. His research interests include data management in mobile ad hoc network, vehicle ad hoc network, and in-vehicle network.

SangKeun Lee received the B.S., M.S., and Ph.D. degrees in computer science and engineering from Korea University in 1994, 1996, and 1999, respectively. He was a recipient of the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship in 2000. Since 2003, he has been an assistant/associate/full professor in the College of Information and Communications, Korea University. His research interests include data management in mobile/pervasive computing systems, location-based information systems, XML databases, and data management in mobile ad hoc networks.