Scientific
Research
Publishing

Effective Life and Area Based Data Storing and Deployment in Vehicular Ad-Hoc Networks

Hirokazu Miura[1,2], Hideki Tode[2], Hirokazu Taki[1]

[1]Faculty of Systems Engineering, Wakayama University, Wakayama, Japan
[2]Department of Computer Science and Intelligent Systems, Osaka Prefecture University, Sakai, Japan
Email: miurah@sys.wakayama-u.ac.jp

Abstract

In vehicular ad-hoc networks (VANETs), store-carry-forward approach may be used for data sharing, where moving vehicles carry and exchange data when they go by each other. In this approach, storage resource in a vehicle is generally limited. Therefore, attributes of data that have to be stored in vehicles are an important factor in order to efficiently distribute desired data. In VANETs, there are different types of data which depend on the time and location. Such kind of data cannot be deployed adequately to the requesting vehicles only by popularity-based rule. In this paper, we propose a data distribution method that takes into account the effective life and area in addition to popularity of data. Our extensive simulation results demonstrate drastic improvements on acquisition performance of the time and area specific data.

Keywords

1. Introduction

In an Intelligent Transportation System (ITS) [1], a number of applications for safety, comfort and convenience have been proposed. Many of them rely on distributing data, e.g., on the current traffic conditions [2], or on free parking spaces around the current location of the vehicle [3]. Entertainment and information services such as multimedia communication and messaging or advertising from roadside to vehicle are also very attractive applications [4]. In ITS, information sharing based on inter-vehicle communication is effective for improving data availability. ITS involves two categories of communication, vehicle-to-infrastructure and vehicle-to-vehicle

communication. In vehicle-to-infrastructure communication, a distributed database stored at fixed sites, such as roadside units, is queried by the moving vehicles via the wireless network infrastructure. Then, the vehicle obtains data from the database and carries it in the direction of travel. In vehicle-to-vehicle communication, two vehicles can communicate with each other when their distance is smaller than a wireless communication range which can be connected by a local area wireless protocol such as IEEE 802.11 [5], Bluetooth [6], and so on. These protocols provide broadband but short-range peer-to-peer communication, and enable vehicular ad-hoc networks (VANETs) [7]. In VANETs, a mobile user discovers the desired information from the vehicles it encounters or from distant vehicles by multi-hop transmission relayed by intermediate moving vehicles [8]. This direct communication between individual vehicles can significantly increase passenger comfort. These are so called application layer store-carry-forward approach and categorized in Disruption Tolerant Networks (DTNs) [9].

In the above approach, as the amount of data items carried by a vehicle becomes larger, the performance such as the ratio of desired data reception is more improved. This is because a vehicle has many chances that it can encounter another vehicle which has desired data. Storage resource in a vehicle, in other words, the amount of saved data is limited in general. Therefore, attributes of data that should be carried by vehicles are an important factor in order to efficiently disseminate desired data. So far, there has been Roadcast study as a typical popularity aware content sharing scheme in VANETs. Roadcast consists of two components called popularity aware content retrieval and popularity aware data replacement. The popularity aware content retrieval scheme finds the most relevant and popular data for user's query. The popularity aware data replacement ensures that different data is deployed inside a vehicular network according to its popularity. Roadcast achieves that more popular data tends to be shared with other vehicles so that the query delay and the query hit ratio can be improved. The existing methods such as Roadcast deploy the data to vehicles randomly according to its relative popularity. In VANETs, however, there are different types of data which depend on the time and location, and some information has an effective life or a deployment area. Such kind of data cannot be deployed adequately to the requesting vehicles only by popularity-based rule, and the data out of effective scope may not be a high valuable for requesting user even if it is obtained. Existing method based on popularity cannot achieve the system which takes into account the effective scope of the data. Therefore, in this paper, we propose a data distribution method that takes into account the effective life and area in addition to popularity of data.

The rest of this paper is organized as follows. Related work is discussed in Section 2. Section 3 describes our system model. Section 4 presents the effective life and area based data storing and deployment. Performance evaluations are shown in Section 5. Finally, we conclude the paper in Section 6.

2. Related Works

2.1. VANET

Vehicular networks represent an interesting application scenario not only for traffic safety and efficiency but also for more commercial and entertainment support. So far, however, most of vehicular network researches focus on routing issues [3] [7] [10]. They all assume the consumer related information is known beforehand so that the sender can route the content to its destination. For example, VADD studies how to choose the best routing path based on the traffic information. Other researches in vehicular networks have focused on content distribution [7] [11] [12]. The efficient discovery and distribution of information is a challenging problem especially in a dynamic environment such as vehicular network. Literature [7] introduces data pouring and buffering techniques to disseminate data along the roads. This paper studies content sharing, where each vehicle queries useful data from its encountered neighbor vehicles. Different from destination aware routing and dissemination [7], how to disseminate the most suitable data to neighboring vehicles is the main focus of this paper.

In the last couple of years there has been an increasing interest in in-network aggregation mechanisms for vehicular ad hoc networks [13]. This technology aims at reducing redundant information and improving communication efficiency by summarizing information that is exchanged between vehicles. Our deployment takes into consideration the number of replicas so that data with high density in the network is not stored in vehicle's storage, but not summarizing information.

On the other hand, content retrieval through intermittent contact opportunities in vehicular networks is also an important technique. In literature [4], content retrieval is studied in a small area, where vehicles in adjacent lanes exchange information as they pass through one another. The scheduling issues of content retrieval at the road

intersection are analyzed in [14]. To accelerate content retrieval, randomized network coding is proposed [11]. However, the network coding based data diffusion brings in large amount of redundant data which may not be useful but taking much communication bandwidth and memory space. In Roadcast, content retrieval [15] is based on users query request and how to efficiently share content with future encountered vehicles based on local information is studied.

Studies in data replacement start from cache replacement. In literature [16], several replacement algorithms for web cache are studied. Later, these replacements are improved by adding the popularity factor. However, all these works are based on web cache which is in a centralized environment. Roadcast differs from the existing works in that it is a distributed replacement algorithm and it aims to optimize the network-wide content sharing performance.

2.2. Roadcast

Roadcast [15] has been proposed as a system for sharing information on VANET. In Roadcast, popular data is distributed to many vehicles so that it can satisfy many users request in the future. Roadcast achieves these objectives with two techniques. One is popularity aware content retrieval and the other is popularity aware data replacement. First, the popularity aware content retrieval scheme makes use of information retrieval (IR) techniques to find the relevant data towards user's queries. However, different from the traditional IR techniques, the factor of data popularity is considered and the relevance of the data to queries is re-ranked, so that more popular data is more likely to be shared with other vehicles. Second, in Roadcast, the downloaded data is stored as replica which can be shared with other Roadcast users. When the local memory is full, some data objects have to be replaced. The proposed data replacement algorithm ensures that the data replications with different popularity can have different life time so that popular data can have more, while not too many, copies in the network. Roadcast considers the popularity of data as a most important factor to improve the query hit ratio.

2.3. Issues of Roadcast

In Roadcast, the popularity of data is considered as the most important factor. However, effective life and area of the content is not taken into account. In general, various kinds of information are shared in VANET. This information include not only entertainment information such as MP3 music or video but also restaurant and parking information, sale advertisement of shops located on the roadside, and so on. Such information may be delivered only for specific areas or may have limited valid time. In other words, the value of the information may be reduced or become invalidout side effective life and area. When such content is disseminated, not only popularity of the content but also effective life and area should be taken into account. In the paper, we propose the data deployment based on the effective life and area as well as popularity of data.

3. System Model

In vehicular ad-hoc networks (VANETs), moving vehicles carry data and exchange it as they pass each other. In this section, we describe our system model. In our system, a vehicle obtains data from the wireless network infrastructure. A vehicle with obtained data moves on a road and encounters another vehicle on the opposite lane. Then the vehicle exchanges the obtained data each other. The vehicle obtains the desired data by repeating this behavior. **Figure 1** shows the system model.

In **Figure 1**, vehicle A moves to the right side and vehicle B moves to the left side, respectively. When the vehicle A passes near the data source, the data source deploy data in the vehicle. Then, the vehicle A moves and carries the obtained data. If the vehicle A encounters the vehicle B which requested the data in the vehicle A, the data is exchanged between A and B.

4. Effective Life and Area Based Data Storing and Deployment

4.1. Replacement Algorithm

In vehicular ad-hoc networks (VANETs), moving vehicles carry data and exchange it as they pass each other. Storage resource in a vehicle, in other words, the amount of saved data is limited in general. Therefore, attributes of data that should be carried by vehicles are an important factor in order to efficiently disseminate de-
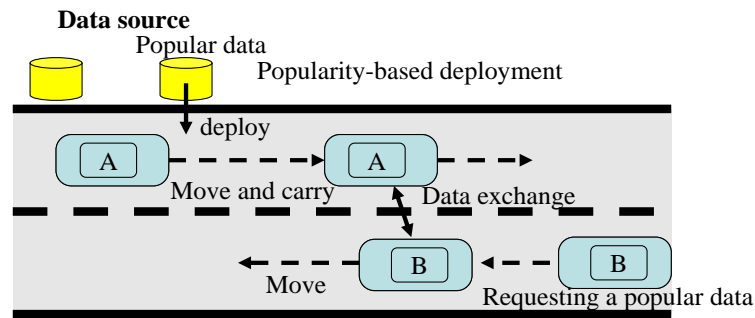
Figure 1. System model.

sired data. In this paper, we focus on effective life and area as well as popularity of data, and propose a replacement algorithm based on these attributes. The proposed method decides the data to be replaced according to the following operations.

In this method, the data which has the largest w_i, defined as Equation (1), has to be replaced from the storage of the vehicle, if new data is input when the storage capacity is full.

$$w_i = \frac{1}{f_i} * \left\{ \alpha g_i + (1-\alpha) c_i \right\} * T_i * D_i \tag{1}$$

$$(0 \leq \alpha \leq 1)$$

where f_i, g_i and c_i are popularity, generation number of data and the replication count of data i, respectively. T_i and D_i are calculated based on effective life and area of its data.

Popular data should not be replaced if it has not been disseminated yet, in order to give more opportunities to disseminate the data with higher popularity. In our method, therefore, each data source decides whether it deploys a specific object or not with a probability which is predefined by its popularity. When the total number of vehicles is V and the request probability of data i is P_i, the expected number of vehicles which obtain data i is VP_i. This means that the data is deployed randomly in a VANET so that the number of replicated data is linear to its popularity. Furthermore, in our method, it can be expected that data with low popularity is also deployed to vehicles because of the probabilistic manner. In our system, it is also necessary to collect popularity information of all data which is expected to be requested from users.

The generation number g_i corresponds to the number of vehicles which are transferred from its original source. The replication count c_i represents the number of copies which are generated by the same vehicle. If the data has large g_i or c_i, it can be expected that many copies of the data exist in the VANET. Therefore, this system is willing to replace the data which has higher g_i and c_i to avoid deployment of redundant copies and store other kinds of data. α is the coefficient that decides which g_i or c_i is more important.

T_i represents a temporal effectiveness of data i and is obtained by normalizing the age of data i to adjust the scale of the other elements. T_i is calculated by the following equation.

$$T_i = \text{elapsed time from the data } i \text{ generation time/effective life time of data } i$$

If $T_i < 1$, the age of data i is within the effective life, otherwise the data i grow stale.

D_i indicates a spatial effectiveness of data i, similarly with temporal effectiveness. D_i is expressed by the following.

$$D_i = \text{distance from original location of data } i / \text{radius of the effective area for data } i$$

There are different possible types of functions for T_i and D_i function. In this paper, we use above function as one example in which the effectiveness decays linearly with time and distance.

4.2. Scheduling Method of Data Deployment

In VANET where vehicles move at high speed, since the time for the data exchange is less, the amount of data which can be transmitted and received at a time is also limited. Thus, the performance of our system depends on the scheduling method of data deployment. In the paper, we proposed the following scheduling method.

When a vehicle encounters the other vehicle, our proposal prefers to deploy the data which has the smallest w_i, defined as Equation (1). It can be expected that popular data within effective life time and area is distributed faster.

5. Performance Evaluation

In this section we describe the simulation model and compared model for our system evaluation, and then present the simulation results.

5.1. Simulation Setup

To investigate the performance of our deployment, we evaluate probability of data reception, *i.e.*, the percentage of data items which requesting vehicles successfully obtained. In our evaluation, we implement our deployment on the ONE simulator (ver.1.4.0) [17] and use a map of the Helsinki area (**Figure 2**). There are totally 100 roadside units on this map. A data object is generated by each roadside unit. When the effective life of the data is expired, the roadside unit generates a new data object. Some keywords are assigned to each data object or each user's query, according to Zipf-like distribution [18]. The simulation parameters are summarized in **Table 1**.

5.2. Compared Method

In this evaluation, we compare the performance of the proposed method with popularity-only which decides the data to be deployed according to its popularity. In the popularity-only, the data which has the largest w_i', defined in Equation (2), has to be replaced from the storage of the vehicle if new data is input when the storage capacity is full. We assume this method works as conventional method.

$$w_i' = \frac{1}{f_i} * \left\{ \alpha g_i + (1-\alpha) c_i \right\}$$

$$(0 \le \alpha \le 1)$$

(2)

5.3. Simulation Results

Figure 3 shows the query hit ratio with several size of buffer memory. The query hit ratio is the possibility that the query can be served by local vehicle or roadside unit. **Figure 4** shows the query delay with several size of buffer memory. The query delay is defined as the average delay from initiating the query to receiving the re-
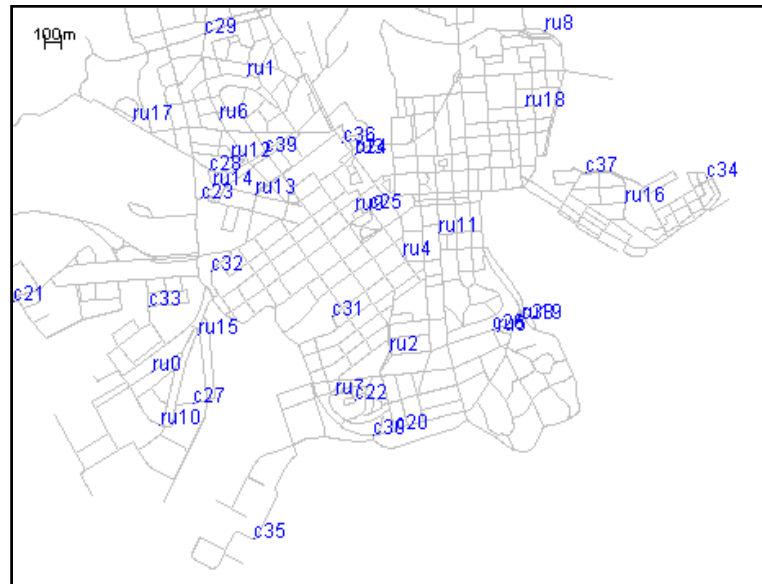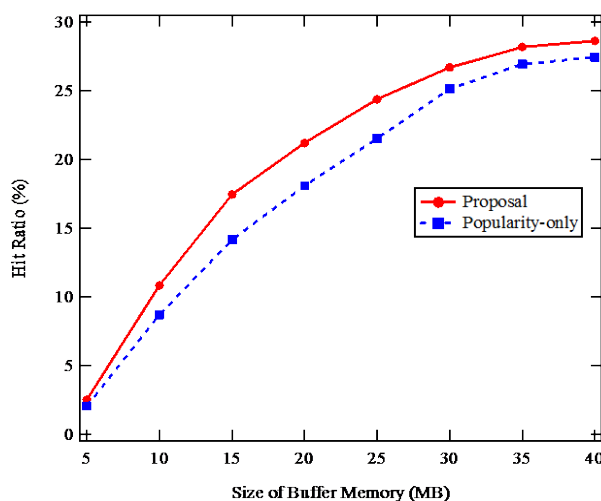

Figure 2. Simulation map.

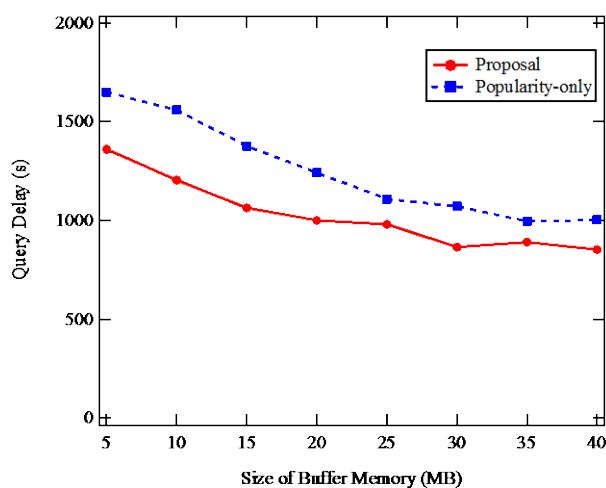**Figure 3.** Query hit ratio.

Figure 4. Query delay.

quired data. From **Figure 3**, our proposal consistently outperforms popularity-only in query hit ratio, the improvement is up to around 3%. As shown in **Figure 4**, when the memory size is small, both schemes have a relatively higher query delay. When the memory size increases, the query delay decreases. This is because as the memory size increases, vehicles are able to buffer more data objects. Hence, there will be more data replicas and the queries can be served by these replicas quickly. From these figures, we also observe our proposal slightly outperforms popularity-only.

Furthermore, we also evaluate the performance of our proposal in terms of effective data. **Figure 5** shows query hit ratio of the effective data which means percentage of the data within its effective life and area in successfully obtained data. **Figure 6** shows its query delay. From these results, we can observe that our proposal significantly improved these performances on valid data acquisition.

Figure 7 shows the query hit ratio with different Zipf parameters. On the Internet, popularities of data follow Zipf-like distribution. This distribution indicates that the number of data with high popularity is small and many data have low popularity. When the Zipf parameter is large, the number of data with low popularity is large and the popularity of few data with high popularity becomes enhanced. Therefore, Popularity-only shows better performance than our proposal. However, when the Zipf parameter is small, access pattern tends to be quite uniformly distributed, and different keywords have similar popularity. From **Figure 7**, when the content access is close to uniform distribution, our proposal has much advantage.
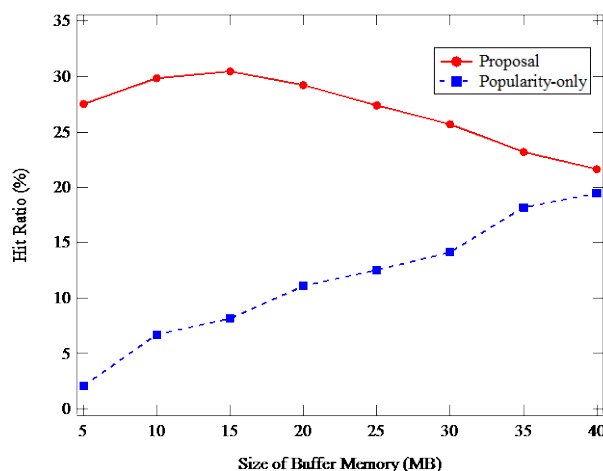
Figure 5. Hit ratio of effective data.

Table 1. Simulation parameters and their values.

Parameter	Value
Simulation Time	12 h
Number of Vehicles	150
Simulation Area	4.5 km × 3.4 km (Helsinki area)
Communication Range	200 m
TransmissionSpeed	250 KBps
Data Size	500 KB - 2 MB
Size of Buffer Memory	5 MB - 40 MB
Keyword Set Size	40
Number of Keywords in Data	2 - 8
Number of Keywords in Query	3 - 5
Effective Life of Data	0.5 h - 3 h
Effective Range of Data	500 m - 2 km
Zipf Parameter	0.5
Vehicle Speed	10 m/s - 15 m/s
α value	0.5

5.4. Effective Data Ratio in the Effective Area

We also evaluate effective data ratio in its effective area according to elapsed time from data generation. We set the parameters as shown in **Table 2** and the other parameters are same as shown in **Table 1**.

In **Figure 8**, x-axis shows the elapsed time from data generation and y-axis shows the ratio of the vehicles which have data in the vehicles in effective area. In this case, effective life of data is 1h. In the popularity-only, even if the effective life is expired, the amount of data in effective areas increases and ineffective data is still deployed. After a while, the deployed ineffective data starts to decrease. On the contrary, in our proposal, although effective data ratio starts to decline before the expiration of effective life, more valid data is distributed to more vehicles in effective area. **Figure 9** and **Figure 10** are results in the case where effective life is 0.5 h and 1.5 h, respectively. These figures show the similar results as **Figure 8**. From these results, it is possible for our
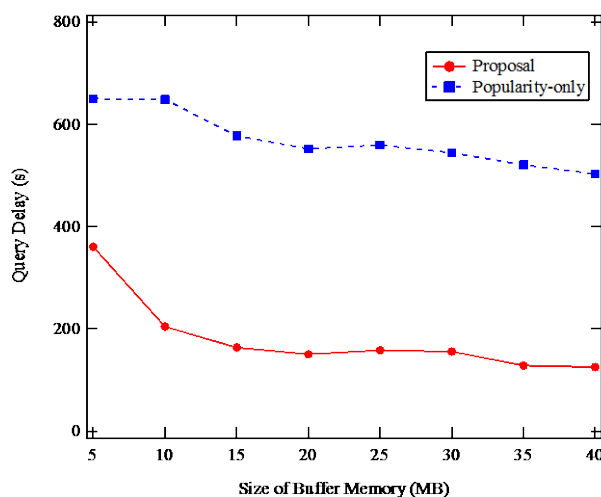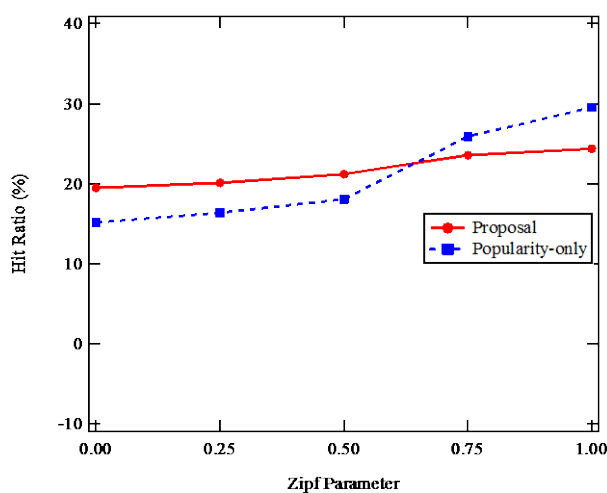
Figure 6. Query delay of effective data.

Figure 7. Impact of Zipf parameter.

Table 2. Simulation parameters.

Parameter	Value
Size of Buffer Memory	15 MB
Effective Range of Data	500 m
Effective Life of Data	0.5 h, 1 h, 1.5 h

proposal to distribute more valid data to the effective area within the effective life of data. In the proposed method, data is gradually deleted from vehicles' buffer before its expiration. Thus, our proposal can suppress that the vehicle buffer is occupied by invalid data, and release the useless buffer resources of the vehicle for other more valid data. Therefore, area-limited information can be distributed to most effective vehicles by our proposal method.

5.5. Impact of Replacement Policy

In this section, we compare the performance on each factor of replacement policy. **Figure 11** represents the result of data acquisition performance in the case of popularity only, popularity and effective area, popularity and
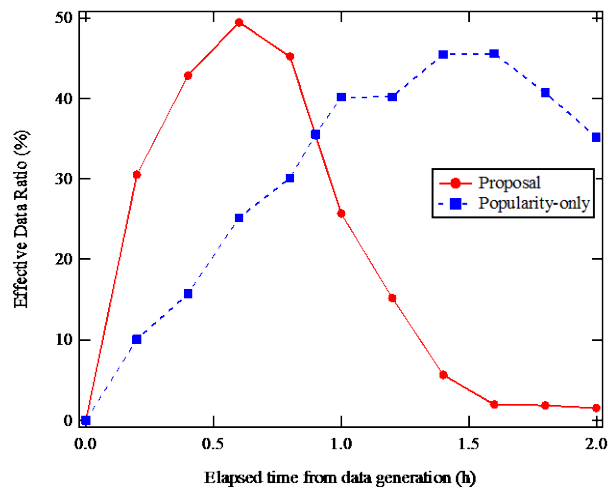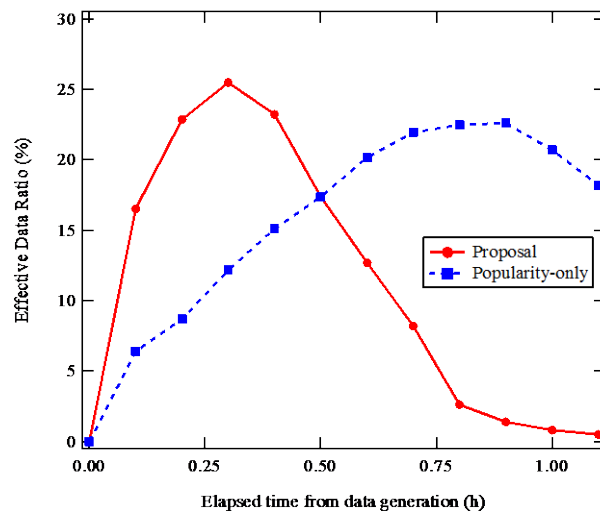
Figure 8. Effective data ratio (Effective life: 1.0 h).

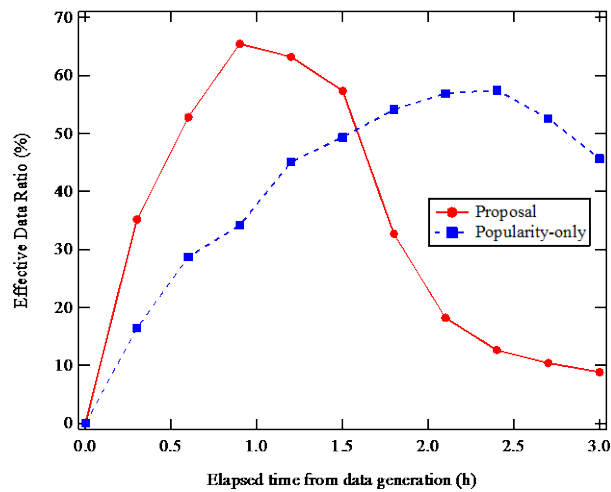**Figure 9.** Effective data ratio (Effective life: 0.5 h).

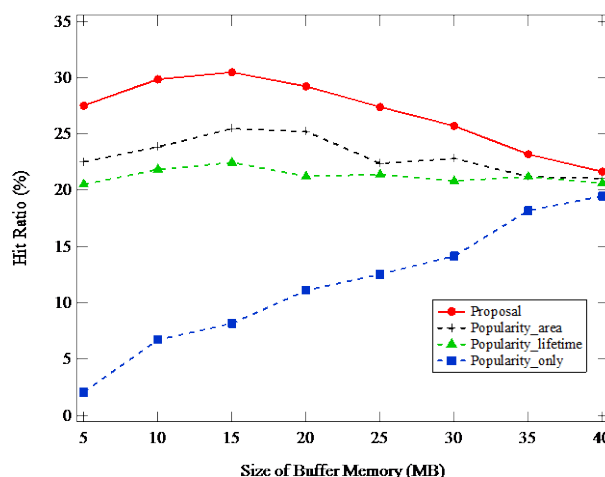**Figure 10.** Effective data ratio (Effective life: 1.5 h).

Figure 11. Impact of replacement algorithms.

effective life, and considering all factors (proposal). As the size of buffer memory becomes large, there is no difference in the buffered data, and all policy shows similar performance. However, in the case where the size of buffer memory is small, it is possible for the proposed method to consider the expiration time and scope of data at the same time, and achieve performance improvement.

5.6. Impact of Parameter α

In this section, data acquisition performance is evaluated when a parameter α in Equation (1) is changed. **Figure 12** shows the result in the case of 100, 150 and 200 vehicles. From this result, by setting α to around 0.3 to 0.4, the performance can be improved. Therefore, it is important to balance between the number of copies and the number of generations to obtain a better performance.

5.7. Impact of Deployment Scheduling Method

In VANET where vehicles move at high speed, since the time for the data exchange is less, the amount of data which can be transmitted and received at a time is also limited. In this section, in order to investigate the differences in performance of the scheduling of deployment, we evaluate the performance of our proposal in comparison with the following scheduling method. In this evaluation, we have adopted our proposed replacement to both scheduling methods.

(**FIFO**) The data to be transmitted is decided by FIFO. Neither the expiration time nor scope of the data is taken into account. The data is transmitted according to the order of old data which the vehicle stored.

The above method (FIFO) is evaluated in the same manner as **Figure 5** and the result is shown in **Figure 13**. By our proposed scheduling, distributing the locally valid data that depends on the time and place shows better performance. The effect of the scheduling method depends on the time for data exchange. It is also expected that this effect can increase as the number of transmitted data at a time is small.

6. Conclusions

In this paper, we propose a deployment method which can help a user get the useful data as much as possible through intermittently connected VANET. In VANETs, store-carry-forward approach may be used for data sharing, where moving vehicles carry and exchange data when they go by each other. In this approach, attributes of data that have to be stored in vehicles are an important factor in order to efficiently distribute desired data. In VANETs, there are different types of data which depend on the time and location. Such kind of data cannot be deployed adequately to the requesting vehicles only by popularity-based rule, and the data out of effective scope, such as effective life and area, may not be a high valuable for requesting user even if it is obtained. Existing method based on popularity cannot achieve the system which takes into account the effective scope of the data. We focus on effective life and area as well as popularity of data, and propose a data replacement algorithm of
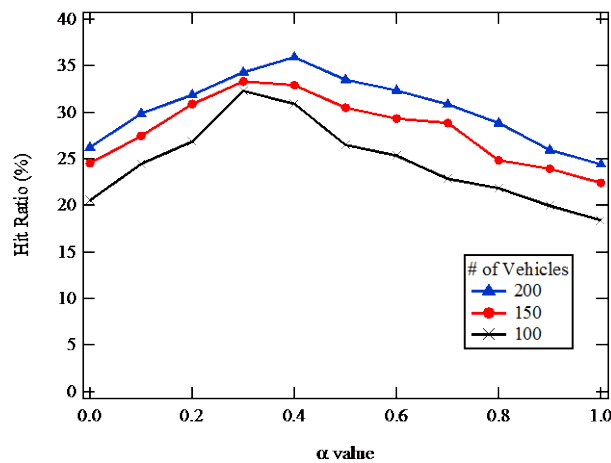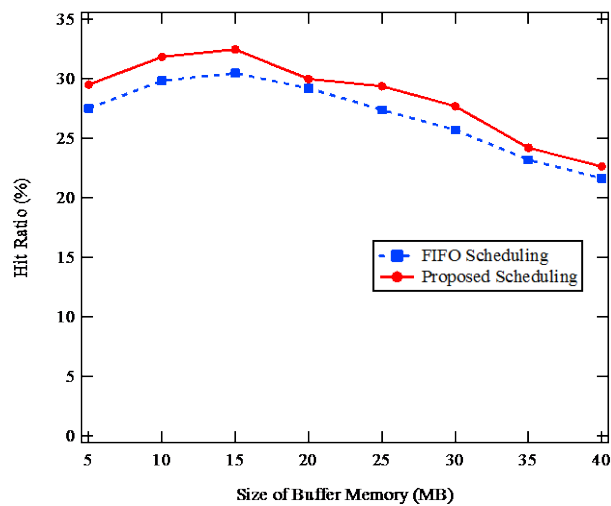
Figure 12. Impact of *α* value.

Figure 13. Evaluation of deployment scheduling.

full buffer inside a vehicle and also scheduling method of data deployment from vehicle based on these attributes.

From our simulation results, it is possible for our proposal to distribute more valid data to the effective area within the effective life of data. In the proposed method, data is gradually deleted from vehicles' buffer before its expiration. Our proposal can suppress that the vehicle buffer is occupied by invalid data, and release the useless buffer resources of the vehicle for other more valid data. Thus, our proposal can greatly improve acquisition performance of the time and area specific data.

In the future, it is necessary to set up a proper effective life and area according to the kind of content.

References

[1] Figueiredo, L., Jesus, I., Tenreiro Machado, J.A., Rui Ferreira, J. and Martins de Carvalho, J. L. (2001) Towards the Development of Intelligent Transport Systems. *Proceedings of IEEE Intelligent Transport Systems Conference*, Oakland, 25-29 August 2001, 1206-1211.

[2] Zhao, J., Zhang, Y. and Cao, G. (2007) Data Pouring and Buffering on the Road: A New Data Dissemination Paradigm for Vehicular Ad Hoc Networks. *IEEE Transactions on Vehicular Technology*, **56**, 3266-3277.

[3] Wu, H., Fujimoto, R., Guensler, R. and Hunter, M. (2004) Mddv: A Mobility-Centric Data Dissemination Algorithm for Vehicular Networks. *Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks*, Philadelphia, 1 October 2004, 47-56.

[4] Guo, M., Ammar, M. and Zegura, E. (2005) V3: A Vehicle-to-Vehicle Live Video Streaming Architecture. 3*rd IEEE International Conference on Pervasive Computing and Communications*, Kauai, 8-12 March 2005, 171-180.

[5] (2003) Standard Specification for Telecommunications and Information Exchange between Roadside and Vehicle Systems—5GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY) Specifications. ASTM E2213-03.

[6] Bluetooth SIG. http://www.bluetooth.org/

[7] Zhao, J. and Cao, G. (2006) VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks. *IEEE Transactions on Vehicular Technology*, **57**, 1-12.

[8] Wisitpongphan, N., Bai, F., Mudalige, P. and Sadekar, V. (2007) Routing in Sparse Vehicular Ad Hoc Wireless Networks. *IEEE Journal on Selected Areas in Communications*, **25**, 1538-1566.

[9] Zhang, Z. (2006) Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges. *IEEE Communications Surveys & Tutorials*, **8**, 24-37.

[10] Burgess, J., Gallagher, B., Jensen, D. and Levine, B.N. (2006) Maxprop: Routing for Vehicle-Based Disruption-Tolerant Networks. *Proceedings of the* 25*th IEEE International Conference on Computer Communications*, Barcelona, 23-29 April 2006, 1-11.

[11] Lee, U., Park, J.-S., Yeh, J., Pau, G. and Gerla, M. (2006) Code Torrent: Content Distribution Using Network Coding in VANET. *Proceedings of the* 1*st International Workshop on Decentralized Resource Sharing in Mobile Computing and Networking*, Los Angeles, 25 September 2006, 1-5.

[12] Fiore, M., Casetti, C. and Chiasserini, C.-F. (2005) On-Demand Content Delivery in Vehicular Wireless Networks. *Proceedings of the* 8*th ACM International Symposium on Modeling*, *Analysis and Simulation of Wireless and Mobile Systems*, Montréal, 10-13 October 2005, 87-94.

[13] Dietzel, S., Petit, J., Kargl, F. and Scheuermann, B. (2014) In-Network Aggregation for Vehicular Ad Hoc Networks. *IEEE Communications Surveys & Tutorials*, **16**, 1909-1932.

[14] Zhang, Y., Zhao, J. and Cao, G. (2007) On Scheduling Vehicle Roadside Data Access. *Proceedings of the* 4*th ACM International Workshop on Vehicular Ad Hoc Networks*, Montréal, 10 September 2007, 9-18.

[15] Zhang, Y., Zhao, J. and Cao, G. (2009) Roadcast: A Popularity Aware Content Sharing Scheme in Vanets. 29*th IEEE International Conference on Distributed Computing Systems*, Montreal, 22-26 June 2009, 223-230.

[16] Cao, P. and Irani, S. (1997) Cost-Aware WWW Proxy Caching Algorithms. *Proceedings of the USENIX Symposium on Internet Technologies and Systems*, Monterey, 8-11 December 1997, 193-206.

[17] ONE Simulator. http://www.netlab.tkk.fi/tutkimus/dtn/theone/

[18] Breslau, L., *et al*. (1999) Web Caching and Zipf-Like Distributions: Evidence and Implication. *Proceedings of IEEE* 18*th Annual Joint Conference of the IEEE Computer and Communications Societies*, New York, 21-25 March 1999, 126-134.