On-demand Transmission Interval Control Method for Spatio-Temporal Data Retention

Shumpei Yamasaki, Daiki Nobayashi, Kazuya Tsukamoto, Takeshi Ikenaga, Myung Lee

Abstract With the development and the spread of Internet of Things (IoT) technologies, various types of data are generated for IoT applications anywhere and anytime. We defined such data that depends heavily on generation time and location as Spatio-Temporal Data (STD). In the previous works, we have proposed the data retention system using vehicular networks to achieve the paradigm of "local production and consumption of STD." The system can provide STDs quickly for users within a specific location by retaining the STD within the area. However, the system does not consider that each STD has different requirements for the data retention. In particular, the lifetime of the STD and the diffusion time to the whole area directly influence to the performance of data retention. Therefore, we propose a dynamic control of data transmission interval for the data retention system by considering the requirements. Through the simulation evaluation, we found that our proposed method can satisfy the requirements of STD and maintain a high coverage rate in the area.

1 Introduction

With the development of IoT (Internet of Things) technology, numerous IoT devices and new applications are spreading. The majority of the existing works employ centralized architecture [1], and computers with high performance CPU and large storage capacity are required to process the enormous application data generated from IoT devices. Therefore, the current network infrastructure is difficult to cope with exponentially increasing data traffic.

Some data generated by IoT devices depend on location and time, such as traffic, weather, disaster, and time-limited store advertisement, and such data are referred to as Spatio-Temporal Data (STD). STDs may not be stored to a remote

Shumpei YAMASAKI
Kyushu Institute of Technology, Japan, e-mail: p232087s@mail.kyutech.jp

Daiki NOBAYASHI
Kyushu Institute of Technology, Japan, e-mail: nova@ecs.kyutech.ac.jp

Kazuya TSUKAMOTO
Kyushu Institute of Tecnology, Japan, e-mail: tsukamoto@cse.kyutech.ac.jp

Takeshi IKENAGA
Kyushu Institute of Technology, Japan, e-mail: ike@ecs.kyutech.ac.jp

Myung LEE
CUNY, City College, USA, e-mail: mlee@ccny.cuny.edu

server connected to the Internet, but they may be sufficient to utilize in the location. M.Beuchert, et al. [2] proposed the platform which makes easy to manage and analyze STDs for location-based applications. The paradigm of "local production and consumption of STD" can solve the problem of current network infrastructure.

H.Teshiba, et al. [3] proposed the data retention system using vehicular network to achieve the paradigm of "local production and consumption of STD" without using existing network infrastructure. This system focuses on the vehicle mobility and the possibility that vehicles are commonly equipped with a high performance CPU and radio communication I/F. Moreover, the system retains STDs within specific area and time by utilizing vehicles as relay nodes (InfoHubs) of data diffusion. It realizes the following three things.

- Collecting and spreading real-time information
- Improving fault tolerance by distributed information management
- Offloading the existing network infrastructure

In addition, the previous works [3][4] proposed the control method of data transmission probability based on node density to avoid collision of radio communication between vehicles. This method improved the performance of the retention system. However, although each STD has different requirements in terms of its diffusion completion time and its lifetime, the previous methods did not consider these requirements at all. Thus, it could be ineffective data retention.

In this paper, we propose a new method that dynamically controls the data transmission intervals by considering the demands for diffusion completion time. Furthermore, the proposed scheme deletes the STD immediately after its lifetime expires. The effectiveness of the proposed method is evaluated by simulation experiments.

In section 2, we review related works. In section 3, we describe the STD retention system proposed in the previous work. Our proposed method is explained in section 4 and section 5 provides the effectiveness of the proposed method through simulation experiments. Finally, section 6 is our conclusion and future problems.

2 Related Work

F. Li, et al. [5] discussed the problem of data diffusion and sharing of vehicular network, and proposed a protocol that uses Geocast Routing based on location within a specific area and sends information from a source to all nodes (Vehicles).

Literature [4][6], Floating Content [7], Locus [8], etc. have been proposed. In literature [6], a node heading to a retention area is specified by switching navigation information of each node, and data is effectively transmitted. In the system of Floating Content and Locus, each node has a list of maintaining data, and exchanges the data list with neighboring nodes. When a node does not have the data, it sends a transmission request to the neighboring node to get the data. Since whether a node performs transmission or not is determined by the transmission probability according to the distance from the center, data acquisition decreases when the node is far from the center, and data collision may occur frequently when the node is biased to near the center.

In order to solve the problem of the related work, in the literature [3], a STD retention system was constructed with the aim of periodically transmitting the data to all receivers existing in the retention area at the set transmission interval using the broadcast based on the positional information. Then, in the next section, the technique devised in the construction of the STD retention system is explained.

3 Spatio-Temporal Data Retention System

In this section, we first describe the assumptions, goals and requirements of the system. Then, we explain the control method of data transmission probability proposed in the previous work [4].

3.1 Assumptions

The STD retention system assumes that each node obtains their position using GPS. Each node broadcasts a beacon containing its own ID at regular intervals. The nodes also broadcasts the data. We assumed that the data includes the information of the retention area (center coordinate, retention area radius R) and the data transmission interval d. Each node determines whether it is located outside or inside the retention area based on its own position information and the retention area information included in the data.

3.2 System Objectives and Requirements

The goal of this system is to periodically spread and retain data to the whole retention area. Therefore, data can be automatically received when the system user enters the retention area. In addition, the system can reduce server load and improve the fault tolerance because of distributed management of STDs without using existing network infrastructure. As a requirement of the system, the *Coverage rate* is defined. It shows the probability which system users can automatically receive STDs when entering the retention area.

$$Coverage\ Rate = \frac{S_{DT}}{S_{TA}}. \tag{1}$$

S_{DT} is the area where can receive data transmitted by nodes within a certain period. Also, S_{TA} is the whole retention area.

The previous work proposed a method of controlling transmission probability based on node density in order to suppress useless data transmission. Section 3.3 shows this technique.

3.3 A data transmission control based on node density

In this system, the radio coverage, the distance from the center coordinates of the retention area to nodes are defined as r and *Distance*, respectively. The transmission target area is defined as follows.

$$\begin{cases} 0 < Distance \leq R + r. & : \text{transmission target area} \\ otherwise. & : \text{unsent area} \end{cases} \tag{2}$$

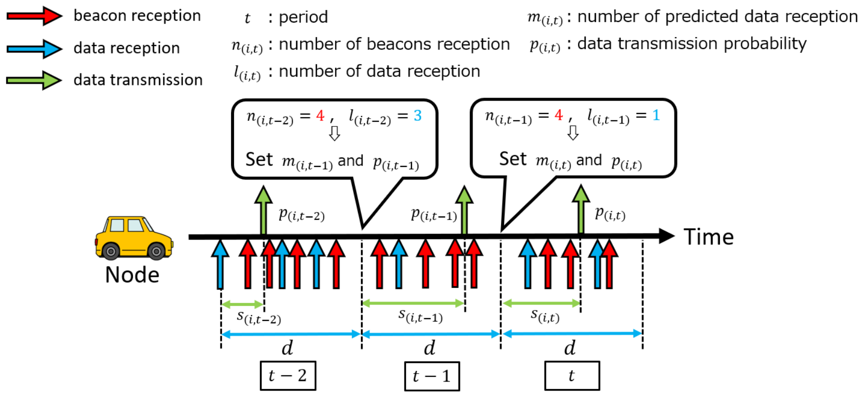

Fig. 1 Setting data transmission probability

Based on the time when data is received for the first time, data is transmitted periodically with the data transmission interval d as one cycle. At this time, data transmission is performed for some nodes which exist outside the retention area in order to improve the coverage rate. The actual transmission timing is transmitted according to the time $s_{(i,t)}$ randomly determined by each node v_i (i: a unique ID given to the node) within d, thereby suppressing the collision of radio channels between nodes. Next, all nodes transmit beacons containing their node IDs at b [s] intervals. By doing so, it is possible to know the number of neighboring nodes around itself. Based on the number of neighboring nodes $n_{(i,t)}$ and the number of beacons received in the $t-1$ cycle, the transmission probability is set in every t cycle (Fig. 1).

When the number of neighboring nodes is 3 or less, the node sets transmission probability $p_{(i,t)} = 1$ in order to transmit data.

When the number of neighboring nodes is 4 or more, $p_{(i,t)}$ is calculated according to the number of neighboring nodes and the surrounding data transmission status. This is to prevent excessive data transmission and data collisions. Although literature [3] shows concrete equations to adjust the data transmission probability, we do not explain the detailed algorithm in this paper due to the lack of space.

In our work, we use this data transmission probability control based on node density to maintain the STD retention after the STD diffusion process (our focus) is completed, as in the literature [3].

3.4 Problems of the previous work

The previous work set the allowable waiting time so that nodes deliver data to users. In other words, users can receive the data if they wait as long as the allowable waiting time in the retention area of the data. In the previous work, the allowable waiting time is set to data transmission interval d of nodes, and the operation was evaluated. However, the previous work does not assume the concrete constraint of the allowable waiting time, and does not mention how to set the data transmission interval.

Some STDs with strict constraints, such as a short lifetime and diffusion completion time, will be retained as well as other STDs with loose constraints (e.g.

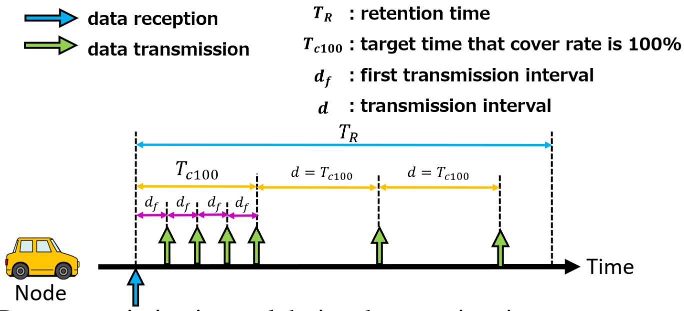

Fig. 2 Data transmission interval during the retention time

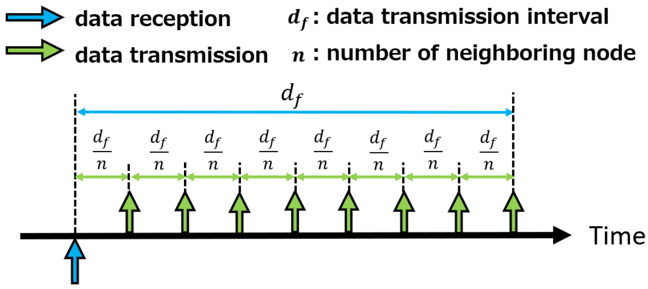

Fig. 3 Random Time Selection($n = 8$)

weather or advertising information). For example, they are radio resource information including available frequency bands, channel interference information, and so on. If the transmission interval of STD is longer than its lifetime, the STD cannot be completely diffused within the area until the expiration of the lifetime (i.e., ineffective retention), and a user cannot get STDs. Therefore, these STDs need to set the transmission interval d as $0 < d < (lifetime\ of\ STDs)$.

When handling the STD retention with severe time-constraint, we need to consider the diffusion completion time-aware data transmission. So, we will propose a method of controlling the data transmission interval in order to satisfy the requirement of diffusion completion time, in the next section.

4 Proposed Method

We propose an on-demand transmission interval control method based on short lifetime and diffusion completion time of STD.

4.1 Decision policy of data transmission interval

First, we define the data lifetime as *retention time* T_R. We also define the target allowable waiting time for the diffusion completion time as T_{c100}. Since the requirements for T_{c100} differ depending on the content, the possible range of T_{c100} could be $0 < T_{c100} \leq T_R$. We assume when T_R of a STD is short, some users cannot obtain the STD due to the expiration of T_R if T_{c100} is set to a large value but less than T_R. For example, when the available frequency band changes every moment (T_R is short), T_{c100} is need to set to a small value. To solve this problem, we also need to consider the number of neighboring nodes and the size of the retention area for the decision of initial data transmission for the first T_{c100} seconds. If the number of neighboring nodes are small and the retention area is large, the time required to complete the data diffusion tends to be large.

In this paper, we define the initial data transmission interval as $d_f(< d = T_{c100})$ and propose a decision method (Fig. 2). After T_{c100} seconds is over, each node sets the transmission interval to $d = T_{c100}$ in order to reduce the unnecessary data transmissions.

4.2 Dynamic decision of initial data transmission interval

4.2.1 Consideration for random waiting time and Hop Count

Assuming that the number of neighboring nodes is n, the number of received data for d is n. Then, each of nodes first checks the transmission interval d included in the received data, and randomly decides the transmission timing $s(0 < s \leq d)$ [s]. Since the average of the random time s is $0.5d$, we assume that each of n nodes re-broadcasts STDs at every intervals of $s = \frac{d_f}{n}$ after receiving the STD (Fig. 3). Furthermore, we define the minimum number of nodes transmitting STD in order to diffuse STD in all direction as γ. In other words, nodes in all direction can receive STD after $\frac{d_f}{n} * \gamma$ seconds. Therefore, the number of possible hops until the expiration of the target time T_{c100} [s] can be expressed by the following equation.

$$\frac{T_{c100}}{\frac{d_f}{n} * \gamma} = \frac{nT_{c100}}{\gamma d_f}. \tag{3}$$

If x [m] is the extended distance of the radius of the data retention area by one data transmission, we need to set the initial transmission interval d_f by following Eq. (4), while satisfying STD diffusion within the data retention area (Eq. (3)).

$$x \times \frac{nT_{c100}}{\gamma d_f} = R$$
$$d_f = \frac{n}{\gamma} T_{c100} \times \frac{x}{R}. \tag{4}$$

4.2.2 Consideration for expansion of retention area

In this work, we assume that nodes are uniformly located in the retention area. One hop distance x changes according to location of nodes, so we consider the following three cases.

1. O_case：In the case of Optimistic
2. E_case：In the case of Expected value
3. P_case：In the case of Pessimistic

First, (1) O_case is a state in which a node exists on the edge of the radio coverage r of some node (Fig. 4). In the case of this node arrangement, since the data is

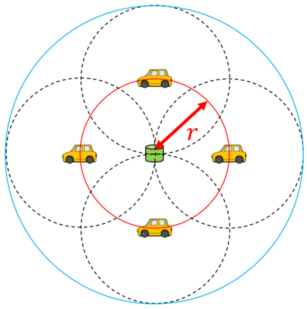

Fig. 4 Node placement for O_case

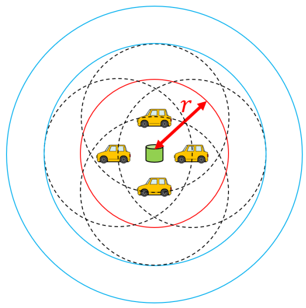

Fig. 5 Node placement for P_case

always spread with the radius of r by one data transmission. As a result, the required number of hops that can cover the retention area becomes the minimum value. Next, (2) E_case shows that nodes exist uniformly in the radio coverage r. In this case, the expected distance x becomes $\frac{1}{\sqrt{2}}r$. Finally, (3) P_case is defined as the distance $\frac{r}{2}$, which is smaller than the expected value (Fig. 5). In this case, the required number of hops becomes twice compared with that of the optimistic case.

The following initial transmission intervals d_fs (d_{f_o} (O_case), d_{f_e} (E_case), and d_{f_p} (P_case)) are determined by following Eq. (5) to satisfy 100% coverage within T_{c100} seconds.

$$\begin{cases} d_{f\_o} = \frac{n}{\gamma} T_{c100} \times \frac{r}{R}. & (x = r) \\ d_{f\_e} = \frac{1}{\sqrt{2}} \frac{n}{\gamma} T_{c100} \times \frac{r}{R}. & (x = \frac{1}{\sqrt{2}} r) \\ d_{f\_p} = \frac{1}{2} \frac{n}{\gamma} T_{c100} \times \frac{r}{R}. & (x = \frac{1}{2} r) \end{cases} \tag{5}$$

4.2.3 STD elimination procedure

If the STD retention is maintained after the expiration of its lifetime T_R, the STD wastes the wireless and storage resources. To solve this problem, in our proposed method, each node checks the start of data transmission (T_s) and the lifetime (T_R) included in the received STD. Then, if $T_n - T_s > T_R$ where T_n shows a present time, the data is discarded, thereby preventing unnecessary data distribution.

5 Simulation Evaluation

In this section, we evaluate the effectiveness of our proposed method by network simulation. We first show simulation models in section 5.1, and present the simulation result and discussion in section 5.2. Note that the comparative method is referred to as the pre_method.

5.1 Simulation Models

We evaluate the proposed method using the network simulator OMNeT++, the traffic simulator SUMO, and Veins implemented in IEEE 802.11p. Figure 6 shows the simulation model. Nodes exist uniformly regardless of location. The lane interval i_l is 200 m, the vehicle interval i_v is 100 m or 200 m ($n = 16$ or $n = 8$) and the speed of nodes is 40 km/h. Vehicles appear intermittently and are set to maintain any ve-

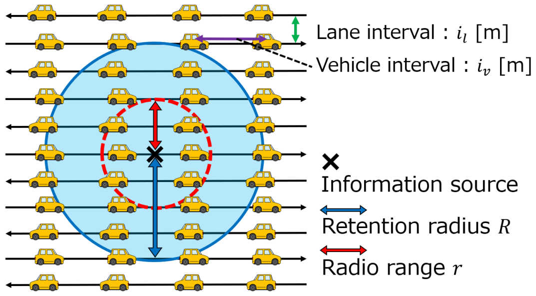

Fig. 6 Simulation model

Table 1 Initial transmission interval by number of neighboring node d_f[s]

	O_case	E_case	P_case	pre_method
$n = 8$	1.000	0.7071	0.5000	5.000
$n = 16$	2.000	1.4142	1.000	5.000

hicle density, and the traveling direction of the nodes alternates between right and left. Moreover, we set the radio range r to 300 m. A beacon transmission interval was set to 1 s, a coefficient for moving average $\alpha = 0.5$, the target number of data transmissions in the t-th cycle γ is 4. Furthermore, retention radius R of all data, a retention time T_R, and T_{c100} is set to 600 m, 4 s, 1 s, respectively. In this simulation, the nodes transmit 10 data at 0.5 second intervals and they retain the data within the retention area. 10 data are sequentially transmitted from the transmission node at 5.5 second intervals, and are accumulated. The packet size of data is 300 bytes.

In this case, the number of neighboring nodes is set to $n \geq 8$, and we set the initial transmission intervals of d_{f_o}, d_{f_e}, and d_{f_p} according to Table 1. As in literature [3], we set the transmission interval d of pre_method to 5 s. We evaluate our proposed method from the viewpoint of coverage rate (Eq. (1)) and the average number of transmissions from 0 to T_{c100}.

5.2 Simulation results and discussions

Figures 7 and 8 show that the coverage rate is varied with the time series when the number of neighboring nodes is set to 8 and 16, respectively. From these results, the pre_method tends to increase the coverage rate as the number of neighboring nodes increases, but the coverage rate of pre_method at $n = 8$ and $n = 16$ becomes 46.87% and 61.73% at the T_{c100}, respectively. This is because the transmission interval d is set to 5 s ($d = 5$ is longer than T_{c100}). Thus, the initial transmission interval d_f should be set at $0 < d_f < T_{c100}$. The coverage rate of pre_method approaches up to 100% at the time of 5 seconds. However, since the retention time of the STD (T_R) is 4 s, it can be said that data transmission after T_R is out of demand.

On the other hand, the expected values of d_{f_o}, d_{f_e} and d_{f_p} show high coverage rate at T_{c100}. However, for O_case, the coverage rate decreases to 94.34% when n=16 although the coverage rate is 99.78% when n=8. Because the expected

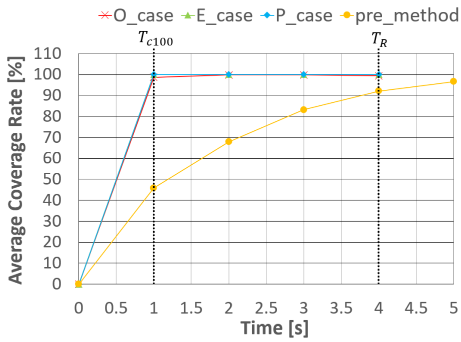

Fig. 7 Coverage rate ($n = 8$)

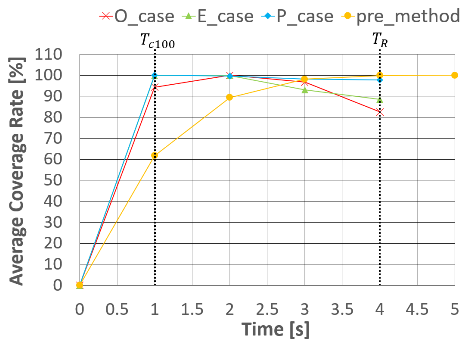

Fig. 8 Coverage rate ($n = 16$)

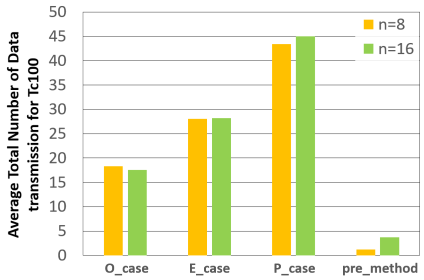

Fig. 9 Average number of data transmissions by method

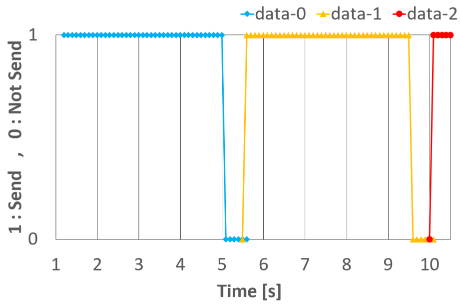

Fig. 10 Time variation of data transmission

value of distance from the center of retention area to first gamma nodes transmitting data approaches $\frac{1}{\sqrt{2}}r$ (E_case) as the number of neighboring nodes increases. In other words, the data transmission interval to satisfy the requirement of T_{c100} becomes shorter than O_case (d_{f_o}) assumed that first γ nodes exist on r [m] from the center of retention area. Next, the coverage rate at d_{f_e} was maximum 100% and minimum 99.75%, and the coverage rate at d_{f_p} was maximum 100% and minimum 99.99%. This indicates that the coverage rate achieved over 99% regardless of the number of neighboring nodes. It can be seen that the coverage increases as the initial transmission interval decreases. In other words, it indicates that the trend is $d_{f\_o} > d_{f\_e} > d_{f\_p}$.

Figure 9 shows the average number of data transmissions within T_{c100}. In the proposed method, the average number of data transmissions is that O_case is minimum and P_case is maximum. Since a large number of data transmissions may cause data collisions, it is better that the number of data transmissions is small. Although O_case is the smallest number of data transmission (Fig. 9), the coverage rate is not always more than 99% at T_{c100}. O_case may not be able to satisfy the request because the initial transmission interval is long, especially when the number of neighboring nodes is large. E_case and P_case can always satisfy the coverage of 99% or more at T_{c100}. In terms of the number of data transmission, E_case is 35.41% less than

P_case. Thus, E_case can satisfy the requirement while suppressing the number of data transmission.

Finally, Figure 10 shows the retention time for each data. In this figure, the transmission of each data is indicated by "1", and the non-transmission is indicated by "0." STD1 is retained from 1 s to 5 s, and then the retention of STD2 maintains at 5.5 after 0.5 s interruption until 9.5 s. In this way, we can demonstrate that the proposed method can preserve reliable data retention within T_R.

6 Conclusion

In this paper, we focused on the STD retention system as a way of achieving the paradigm of "local production and consumption of STD." Although the previous method determines the data transmission probability, how to diffuse the STD in the initial phase has not been considered at all. Therefore, in this paper, we propose an on-demand transmission interval control method to complete STD diffusion until the requirement. Through simulations, we showed that the proposed scheme can reliably achieves almost 100% coverage rate, while significantly limiting the increase in the number of STD transmissions. Moreover, we can demonstrate that the STD retention can be controlled in response to the STD lifetime. However, this work does not assume that vehicles move randomly. If the vehicles move randomly, the density of the vehicles in the retention area changes according to the location. Therefore, the STD may not be evenly diffused in the retention area, and we will need to consider that each node changes the data transmission interval by itself.

Acknowledgements This work was partially supported by JSPS KAKENHI Grant Number 18H03234, NICT Grant Number 19304, and USA Grant number NSF 17-586.

References

1. S.Wang, X.Zhang, Senior Member, IEEE, Y.Zhang, Senior Member, IEEE, L.Wang, J.Yang and W.Wang, Senior Member, IEEE, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications, " IEEE Access, vol.5, pp.1-3, 2017.
2. M.Beuchert, S.H.Jensen, O.A.Sheikh-Omar, M.B.Svendsen and B.Yang, "aSTEP: Aau's Spatio-TEmporal Data Analytics Platform," 2018 19th IEEE International Conference on Mobile Data Management (MDM), Aalborg, pp. 278-279, 2018.
3. H.Teshiba, D.Nobayashi, K.Tsukamoto and T.Ikenaga, "Adaptive Data Transmission Control for Reliable and Efficient Spatio-Temporal Data Retention by Vehicles, " The Sixteenth International Conference on Networks, pp.46-52, 2017.
4. T.Higuchi, R.Onishi, O.Altintas, D.Nobayashi, T.Ikenaga and K.Tsukamoto, "Regional InfoHubs by vehicles: balancing spatio-temporal coverage and network load, " In Proc.IoV-VoI'16, pp.25-30, 2016.
5. F.Li, and Y.Wang, "Routing in vehicular ad hoc net-works: A survey, " IEEE Vehicular Technology Magazine, Vol.2, Issue 2, pp.12-22, 2007.
6. I.Leontiadis, P.Costa and C.Mascolo, "Persistent content based information dissemination in hybrid vehicular networks, " In Proc.IEEE PerCom, pp.1-10, 2009.
7. J.Ott, E.Hyyti, P.Lassila, T.Vaegs and J.Kangasharju, "Floating Content: Information Sharing in Urban Areas, " In Proc.IEEE PerCom, pp.136-146, 2011.
8. N.Thompson, R.Crepaldi and R.Kravets, "Locus: A Location-based Data Overlay for Disruption-Tolerant Networks, " In Proc.ACM CHANTS, pp.47-54, 2010.