# Adaptive Data Transmission Control for Spatio-temporal Data Retention over Crowds of Vehicles

Daiki Nobayashi, *Member, IEEE* Ichiro Goto, *Non-member, IEEE* Hiroki Teshiba, *Non-member, IEEE* Kazuya Tsukamoto, *Member, IEEE* Takeshi Ikenaga, *Member, IEEE* Mario Gerla, *Fellow, IEEE*

***Abstract*—Some specific services for Internet of Things, such as real-time map and providing local weather information, depend strongly on geographical time and location. We refer to the data for such service as spatio-temporal data (STD). When STD is used in a query response system similar to conventional Internet services, users not only need to acquire data actively as required, they must also have functions for retrieving data available STD. Therefore, we propose an STD retention system that uses vehicles as information hubs (InfoHubs) for disseminating and retaining the data in a specific area. In our system, InfoHubs diffuse, maintain, and advertise STD over places and times where the STD are strongly dependent, thereby allowing users to receive such data passively within the specific area. Additionally, because STD are associated with a particular space, the system can reduce search costs. We also propose an adaptive transmission control method that each vehicle effectively operates its wireless resources autonomously and STD are retained and distributed efficiently. Finally, we evaluated our proposed method using simulations and clarified that our proposed system is capable of achieving a coverage rate of nearly 100% for STD while reducing the number of data transmissions compared to existing systems.**

***Index Terms*—Vehicular networks, Information Hubs, Spatio-temporal data retention, Adaptive data transmission control.**

## I. Introduction

With the development and growth of machine-to-machine (M2M) and Internet of Things (IoT) technologies, the number and types of devices equipped with various wireless interfaces have proliferated and IoT devices have become increasingly ubiquitous. In the current Internet paradigm, most data are first gathered by remote servers connected to networks such as cloud servers and data centers, after which they are provided to applications as required. However, according to an Organization for Economic Co-operation and Development (OECD) report [1], the number of M2M devices will grow up to 50 billion by 2020, and the enormous quantities of small data packets flowing to the Internet will skyrocket. In fact, Cisco Systems expects the amount of traffic in 2020 to be 2.7 times more than that of 2015 [2], and 500 billion devices are expected to be connected to the Internet by 2030 [3]. To efficiently distribute so much data, it is necessary to install routers with high packet transfer performance in the Internet infrastructure and to increase link bandwidth levels, which means that the load on Internet infrastructure will increase significantly. Furthermore, to store and analyze these data effectively, the acquisition of large-capacity storage modules and high-performance central processing units is essential.

From the viewpoint of data content, some applications, such as those that provide weather and traffic information, are strongly dependent on time and location. In this paper, we define data generated by IoT devices for those applications as spatio-temporal data (STD). Because such data are highly dependent on the time and places where they are generated, the most effective way to use STD is to provide STD directly to the users who are at that location rather than accumulating STD on remote servers for later dissemination. For example, we consider the case of an ambulance driving to a destination during an emergency. If a traffic accident occurs at a certain point along the way and a traffic jam results, the driver will need to search for alternate routes. However, when using an existing Internet architecture, such as a cloud service, the following problems result:

- The ambulance driver has no way of grasping what accidents may be occurring without sending queries to a server. (In other words, the ambulance driver must actively and continuously send queries to the server before he or she can learn what is going on down the road.)
- In order to search for alternate routes, the ambulance driver needs to know the real-time traffic situation of other routes at that time. (The Internet does not effectively handle multiple and simultaneous searches that are based on location and time.)

On the other hand, if it were possible to produce an architecture in which events occurring in a particular location

D. Nobayashi and T. Ikenaga are with the Department of Electrical and Electronics Engineering, Faculty of Engineering, Kyushu Institute of Technology, Japan, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan, e-mail: nova@ecs.kyutech.ac.jp, ike@ecs.kyutech.ac.jp

I. Goto is with the Department of Electrical and Electronics Engineering, Graduate School of Engineering, Kyushu Institute of Technology, Japan, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan, e-mail: goto.ichiro959@mail.kyutech.jp

H. Teshiba is with Department of Computer Science and Electronics, Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan, 680-4 Kawazu, Iizuka-shi, Fukuoka 820-8502, Japan, e-mail: teshiba@infonet.cse.kyutech.ac.jp

K. Tsukamoto is with Department of Computer Science and Electronics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan, 680-4 Kawazu, Iizuka-shi, Fukuoka 820-8502, Japan, e-mail: tsukamoto@cse.kyutech.ac.jp

M. Gerla is with the Computer Science Department, University of California at Los Angeles, Los Angeles, CA 90095, USA.

Manuscript received XX XX, XX; revised XX XX, 201X.

(STD/content) can be directly and passively provided to users in the vicinity of that location, the following advantages can be obtained:

- The users can passively obtain STD that exist in that space, so they do not need to actively search for relevant data.
- Because various types of STD are ubiquitous everywhere and anytime, users can combine these data as needed to create new content.

In other words, since the "locally produced and timely consumed" paradigm of STD use is effective for location-dependent applications, there is a crucial need for a novel network architecture that can achieve data retention within a specific area.

In this paper, we focus on vehicular ad hoc networks (VANETs) as an important network infrastructure type that can achieve the required level of STD retention. Modern vehicles have three remarkable features that are not found in mobile devices, such as smartphones. First, data can be collected by and analyzed within individual vehicles because they are equipped with significant amounts of data storage, battery power, and high-level computational resources, such as On-board Units [4] and high-performance car navigation systems with communication interfaces. In addition, with the advancement of vehicle-to-everything (V2X) technology such as semi-autonomous vehicle platooning, we can expect that vehicles equipped with high-performance computing resources will appear in the near future. Second, V2X communications can be easily realized because modern vehicles are required to be equipped with wireless interfaces for short-distance communication, as outlined in the IEEE 802.11b/g/n, and 11p amendment, which covers dedicated short-range communications. It should also be noted that some vehicles have wireless interfaces for wide-area fourth- or fifth-generation communication. Third, the enormous numbers of highly mobile vehicles operating all over the world can provide a ready-made foundation from which data can be collected and disseminated efficiently [5].

Hence, it is clear that the potential for spatio-temporal information communication within a VANET has made it possible for us to support the development of a new promising network infrastructure. In our study, we use vehicles as a mobile vehicular cloud for STD, within which each vehicle acts as a regional information hub (InfoHub) to disseminate and maintain STD within some pre-defined areas. The STD are ultimately received by users, not the vehicles. For example, a user passively receives multiple STD packets retained in the user's area, and these packets can be used for various activities. STD management by InfoHub vehicles has the following advantages:

- Since the users passively receive the STD provided by the InfoHubs, there is no need for them to be aware of the existence of the data and no need to search for the data.
- The workload imposed on Internet cloud servers and data centers can be reduced.

Novel network architectures for data retention will be among the important technologies necessary for achieving the local production and consumption paradigm. However, the radio resources used by vehicles, as well as other wireless networks, may become congested due to the increased data traffic. Since all vehicles in a VANET utilize the same communication channel, frame (data) collisions and certain levels of interference are inevitable. In networks with large numbers of vehicles (dense traffic environments), each vehicle could suffer multiple and frequent frame collisions, thus leading to a decline in communication quality. Conversely, in networks with small numbers of vehicles (sparse traffic environments), each vehicle must accelerate its data transmission activities due to the scarcity of vehicles available for data transmissions. With these points in mind, it is clear that the use of adaptive data transmission control based on utilizing the capabilities of InfoHub vehicles in response to vehicle density could provide an indispensable component for distributed data management.

With the above points in mind, we propose an adaptive data transmission control method for STD retention that can accelerate the local production and consumption of STD. More specifically, our proposed method can alleviate channel interference and achieve a high coverage rate that represents the percentage of the range in which data are transmitted by vehicles to the specified geographic range (defined by the application) during a certain period. In our proposed method, vehicles adaptively change their data transmission probabilities based on the density of neighboring vehicles in order to maintain STD retention within a pre-defined area and thus allow area users to efficiently obtain local STD. We performed evaluations using a traffic simulation model that provided stable and random vehicle densities to approximate an actual traffic environment. Note that this study is based on improvements and further evaluations of previous work contained in [6].

The remainder of this paper is organized as follows. In Section II, we discuss other works related to data retention in ad hoc networks. In Section III, we outline our STD retention system, and then Section IV describes our proposed adaptive transmission control method in detail. Section V provides the simulation model, simulation results, and discussion. Finally, we provide conclusions in Section VI.

## II. Related Works

Li et al. discussed various VANET-related problems, such as data dissemination and data sharing enabled by the high mobility of vehicles [7] and proposed a geocast routing protocol, which is basically a location-based multicast routing, to deliver data from a source vehicle to all other vehicles within a target area. Maihofer et al. [8] proposed an abiding geocast in which data are delivered to all vehicles within a target area and then maintained within those vehicles as long as they remained in the network. They provided three solutions for retaining the geocast data within the target area: *server approach*, *election approach*, and *neighbor approach*. We provide an overview of these approaches in the following paragraph.

In the server approach, a pre-defined fixed server within the target area is used to store and periodically transmit data to other vehicles within the target area based on a geocast routing protocol. However, since the server sends data and exchanges location information among all vehicles within the target area, it is susceptible to overloading. Recent studies to support vehicle-to-vehicle or vehicle-to-infrastructure communications with mobile edge computing [9][10] have not addressed the dissemination and use of the STD for those locations. In the election approach, only the elected vehicles maintain the data and periodically send the data to other vehicles within the target area. In both of these approaches, broadcasting from a restricted number of vehicles can result in STD retention performance degradation.

In contrast, the neighbor approach, which functions without a dedicated or elected server vehicle, has been studied recently due to its high feasibility, and a number of systems have been proposed. These include that of [11], floating content [12], Locus [13], and our previous work [6][14]. In the method proposed in [11], a vehicle exchanges navigation information with neighboring vehicles, identifies other vehicles that are moving towards the target area, and then delivers the data to them. In the floating content and Locus systems, each vehicle has a list of data and exchanges its list for the lists of the other vehicles it encounters. If any vehicle has data that are not stored in a neighboring vehicle, the neighboring vehicle can then acquire the missing data from the vehicle that has the data. In this situation, the vehicle that has the data decides what data to send based on the transmission probability, which changes dynamically depending on the distance from the location where the data were generated. More specifically, the transmission probability decreases as the vehicle moves away from the center of the target area, which means that some outlying recipients will be less likely to receive the data. However, if numerous vehicles are present near the center of the target area, data collisions tend to occur frequently in VANETs because each vehicle attempts to send data with high transmission probability at the same time. In particular, the floating content method has been actively studied. For example, Manzo et al. used Luxembourg's traffic model (LuST) [15] to evaluate the effectiveness of floating content in urban areas, while Rizzo et al. focused on data distribution using this method and considered the use of software-defined networks (SDNs) to achieve it. This method is a hybrid of the server and neighbor approaches. More specifically, the server becomes an SDN controller and the data are collected from vehicles and analyzed to determine the most appropriate delivery method. However, the processing load of the SDN controller in this method is high, and the fault tolerance is still low.

Meanwhile, in contrast to the floating content and Locus methods, our previous work [6][14] aimed at delivering data to all vehicles and users within a target area at pre-determined intervals using a geolocation-based broadcasting method. In those methods, the transmission probability for periodical data dissemination is determined based on the location information of all neighboring vehicles. Thus, those methods require the vehicles to perform complicated calculations.

In our previous research, such as [14], we focused on a VANET-based system that disseminates and maintains STD within a target area by adaptively controlling the data transmission probability in response to the vehicle density, which is estimated solely by the number of received data transmissions. In our newly proposed method, rather than the location information of all vehicles, as in [14], the process used for determining transmission probability is simplified because only the message information is employed. More specifically, the methods proposed in earlier studies [14] require accurate location information for all vehicles in order to calculate the distance between vehicles, which is quite difficult in terms of computational overhead in a practical environment. Therefore, the method proposed in this paper uses only the number of data transmissions to decide the transmission probability, which means that no complex information (e.g., location information) is required. We refer to our new VANET-based system as an *STD retention system*.

Our STD retention system is similar to a vehicular cloud computing (VCC) [16][17]. Lee et al. proposed the method for maintaining the STD in VCC by using information centric networks (ICNs) that assume the STD "name" is known to everybody. Hence, in the case of a traffic jam at a certain intersection, or an ambulance in a certain area, queries to STD are georouted to the specified location, and the first data owner that receives the interest query returns the response. The primary advantage of such VCC-ICN collaboration is that there is no widespread STD dissemination with in the target area, and remote users can acquire STD in the area using the ICN. In contrast, in our approach, the STD are disseminated within a certain area in order to allow users to access any STD regardless of the service type ("name") provided by the provider. The major difference between these methods is that our approach passes the data generated at that location directly to the user for consumption in the same area. Therefore, in our approach, users do not request data from the vehicular network (no query). Instead, they passively receive data transmitted by vehicles serving as InfoHubs. In other words, our system is specific to the use of STD at a specific location and provides one solution to the use of STD for vehicular clouds.

## III. STD Retention System

In this section, we describe the assumptions underlying the STD retention system, the objectives and a use case of our system design, and its requirements.

### A. Assumptions

In our system, we assume that STD is generated by various types of sensor devices like IoT devices or data providers, and STD consists of one packet. Furthermore, STD packets are assumed to include not only data for an application but also parameters for data retention, such as a central coordinate, a radius, and a data transmission interval by a data origin. The data origin allows the STD to retain anywhere by freely specifying the center of the retention area. However, this paper assumes that the STD retains in a location relatively close to the data origin (Must be less than specified retention radius). If the data origin specifies the center of the retention area that is too far away, the STD must be transferred to the center point. This situation is outside the scope of this paper's subject. Moreover, each data origin can set different parameters for each STD packet.

We assume that InfoHub vehicles are equipped with an onboard wireless interface that meets not only the IEEE 802.11p but also the IEEE 802.11b/g/n and Bluetooth. Since each vehicle can obtain location information using its Global Positioning System (GPS) receiver and has a unique identifier (ID), such as the serial number of the on-board unit or the vehicle registration numbers issued in each country, it can estimate the number of neighboring vehicles based on the received beacon messages broadcast by other InfoHub vehicles. Moreover, each vehicle performs an operation to determine whether it is within the data retention target area.

### B. System Objectives and a Use Case

The objective of this system is to achieve data retention, especially for STD packets that are specific to a target area, in order to construct a novel network architecture that can effectively deliver STDs to the users in their location. Such packets typically contain weather data, traffic reports, real-time three-dimentional (3D) maps for automotive navigation, and local store advertisements. To achieve this, we focus on vehicles with InfoHub characteristics. In this system, since STD are distributed and periodically transmitted within the target area, users in this area receive myriad data passively. However, users can also take advantage of the data at that location by selecting only the desired data (e.g., using an application). Thus, a user who enters a particular area can obtain all STD for that area very quickly. Furthermore, since multiple vehicles have the same data, a high level of fault tolerance can be achieved. Finally, since the STD are only stored on vehicles, no burden is imposed on Internet (cloud) servers.

Herein we consider a use case for real-time mapping using this system. On this real-time map, we assume that the following information can be plotted: the flow of people, the volume of traffic, the existence of accidents or inoperative vehicles, road surface conditions, building conditions, and temporary road closures due to construction, as well as similar information. Such information is continuously changing. Typically, information types that change slowly over long intervals can be managed by a cloud service, and users can obtain such information via the Internet. On the other hand, information types that can be used most effectively by consumers located around the location where the data are generated. Figure 1 shows an example of a real-time map using accident information. If the users can obtain information on the accident while they are still two blocks away, it can immediately be utilized for researching alternative routes. However, in a query-response system such as general web services or the dynamic maps [18] that upload real-time information from unmanned aerial vehicles (UAVs) or vehicles on to edge servers and clouds to generate maps, it is difficult for potential consumers to immediately learn that an accident has occurred because they need triggers (actions) to cause them to retrieve the information. In our proposed data retention system, the InfoHubs would diffuse, maintain, and

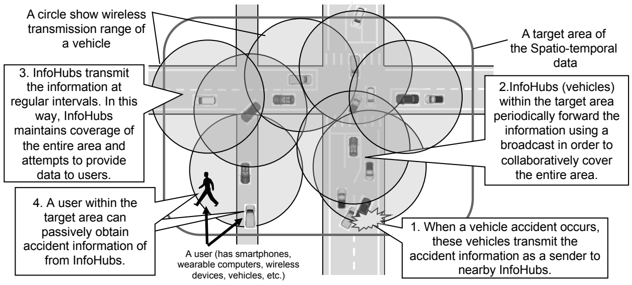



Fig. 1: Use Case of STD Retention System.

advertise information about the accident to vehicles within the target area, and users could receive that information passively[1]. Therefore, **users do not need to be aware of the existence of the data, and there is no need to search for the data.** As a result, users can leverage the information immediately upon receipt of the data. In other words, this system can successfully achieve the paradigm of "local production and consumption of spatio-temporal data."

### C. System Requirements

In this paper, the coverage rate is a performance metric that indicates how fast users can receive STD. To facilitate rapid data delivery, the entire target area should be covered within the transmission range of the InfoHub vehicles. That is, users should be able to obtain the STD from neighboring vehicles via one-hop broadcast communication. Additionally, we assume that the transmission range is less than the target area radius, and we calculate the coverage rate at predetermined intervals. The coverage rate formula is as follows:

$$CoverageRate = \frac{S_{DT}}{S_{TA}} \tag{1}$$

where $S_{TA}$ is the size of the target area and $S_{DT}$ is the size of the total area where the user can obtain data transmitted from any InfoHub vehicles within the transmission interval.

Figure 2 shows coverage rate examples. The black dots are vehicles, and the gray circles indicate their wireless range. A high level of coverage rate, as shown in Figure 2a, means that users can automatically receive STD from anywhere within the target area. In contrast, a low coverage rate, as shown in Figure 2b, means that users may not receive STD when no other InfoHub vehicle is near the user's travel path. Moreover, since the slope of the change in the coverage rate indicates the STD dissemination speed, the coverage rate can reveal the system's responsiveness.

Since the proposed system requires the rapid acquisition of STD from anywhere within the target area, each vehicle within that area needs to transmit data as frequently as possible. However, high vehicle density levels result in frequent data transmissions, which inevitably results in data collisions that

[1]Note that the STD retention system does not assume currently to provide detailed and continuous data in a brief period, such as inter-vehicle communication for autonomous cars. The retention system only provides the fact (data) of the accident to the user, and how to utilize the data is up to the user.

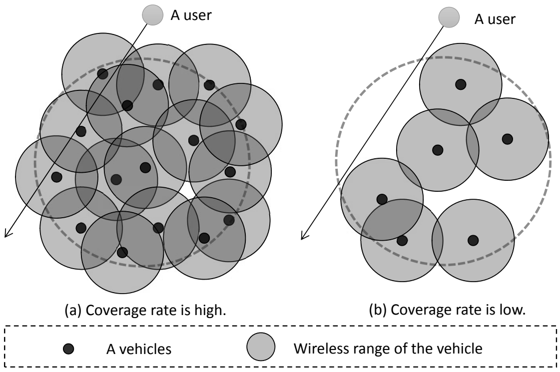



Fig. 2: Examples of high and low coverage rates.

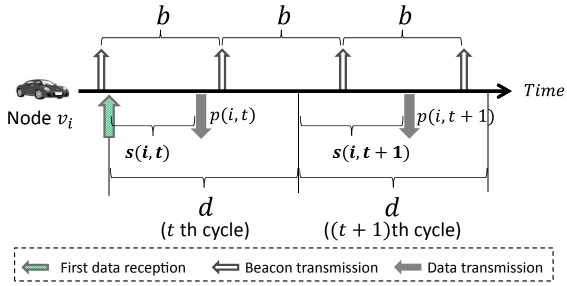



Fig. 3: Data transmission procedure.

can adversely affect the coverage rate. In contrast, when the vehicle density level within the target area is low, all vehicles should transmit data as often as possible to boost the coverage rate. Thus, the transmission probability needs to be changed flexibly in response to the density level of neighboring vehicles.

## IV. Node-Density-Aware Transmission Control

In this section, we describe our proposed transmission control method, which is based on the number of neighboring vehicles (neighboring vehicle density), and which aims at effective retention of the STD within a target area. In our transmission control method, the vehicles aim to operate as autonomously as possible without redundant data exchange between vehicles. Therefore, the number of neighboring nodes is determined by measuring the data (radio wave) that the vehicle receives itself, rather than the accurate number based on the GPS information. In this paper, hereafter, we define InfoHub vehicles as nodes. This method aims to disseminate STD by utilizing the appropriate number of nodes within the target area. Consequently, our STD retention system can maintain a high coverage while reducing the total number of data transmissions to the minimum number necessary. Here, it should be noted that our proposed method is an improvement to our previous control method [6], and the differences will be described as appropriate in the paragraph below. In addition, the subsequent description focuses on retaining one data to facilitate the explanation. If there are multiple data, a series of processes operate independently of each data.

### A. Data Transmission Timing

In our method, after a node receives data from another node or a sender, it needs to re-transmit the received data, as necessary, to ensure STD retention within the target area. However, to minimize transmission collisions, the transmission timing of each node is different.

Figure 3 shows the data transmission procedure. In our proposed system, each node periodically transmits a beacon message, but data are only transmitted when necessary. Here, this beacon message is newly defined in our proposed method, not a beacon that is used for IEEE 802.11p or inter-vehicle communication[2]. The beacon broadcast interval is fixed at $b$ seconds. Data are transmitted based on the following procedure. When a node $v_i$ receives data from another node, it first checks the transmission interval of $d$ seconds, which is included in the data. The subscript $i$ indicates a unique node ID, $i \in V$, where $V$ is the set of all nodes. Then, that node randomly determines the next transmission time $s_{(i,t)}$, where $t$ is the number of cycles of $d$ since the first data reception. The random determination within $d$ seconds allows the node to avoid data transmission collisions. The interval $d$ differs between applications, and we assumed that $d$ is set by a sender. Furthermore, $s_{(i,t)}$ is calculated at the beginning of cycle $t$.

### B. Adaptive Transmission Control Method

*1) Definition of data transmission area.:* If all nodes transmit data at different timing intervals, data collisions can be completely avoided. However, when the number of neighboring nodes within the transmission coverage area is larger than the number of transmission slots, collisions inevitably occur. Accordingly, we designed a new transmission control method in which the transmission probability changes dynamically based on the neighboring node density, thereby providing a high coverage rate with the minimum number of data transmissions. In our method, nodes within the target area are classified into two types based on distance from the center of the target area (data origin point), as shown in Figure 4. The specific conditions are as follows:

$$\begin{cases} 0 \le x \le R + r & (data\ transmission\ area) \\ otherwise & (out\ of\ area) \end{cases}$$

where $x$ is the distance between the node and the center of the target area, $R$ is the radius of the target area, and $r$ is the additional area where the node operates to retain data. The distance $x$ is calculated from both GPS information and the data origin point, which is included in the STD packet. The area radius $R$ is also included in the STD, and $r$ is determined by our proposed method. To realize a high coverage ratio in the target area, $r$ is assumed to extend beyond $R$ because the nodes outside the target area need to cooperate as well. This

[2]Note that beacons used in vehicle-to-vehicle communication or IEEE 802.11p can also be used to calculate the number of neighboring nodes for our proposed STD retention system. In this case, the overhead of beacon transmissions can be reduced because a transmission of our beacon is not required.

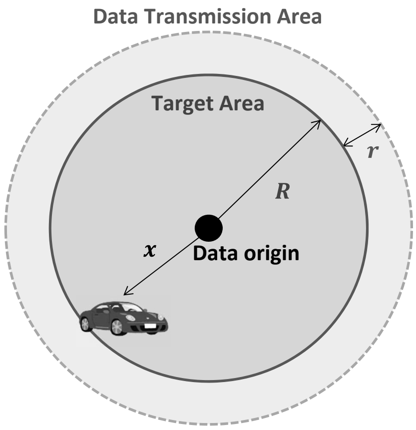



Fig. 4: Target area and data transmission area

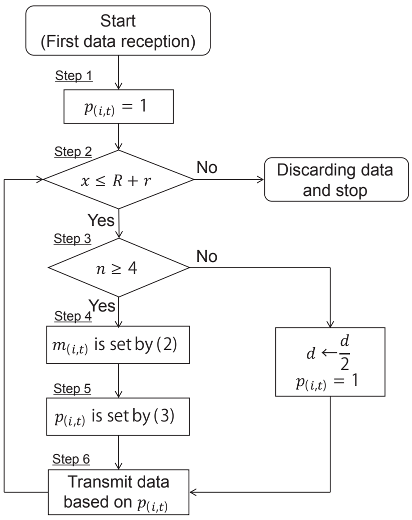



Fig. 5: Transmission probability decision process.

is one of significant difference from the previous method, as discussed in detail in Section IV-C. In Section IV-B2, we show how the transmission probability is determined in each area.

*2) Operation of data transmission area:* To provide a high coverage rate, the nodes in the data transmission area autonomously adjust their own transmission probability levels based on the density of neighboring nodes. Figure 5 is a flowchart of the transmission probability decision process. The transmission probability is $p_{(i,t)}$, where $i$ is the unique node ID and $t$ is the number of cycles. The transmission probability is always set at the beginning of each cycle.

Step 1: In the first step, when a node initially receives data from other nodes, the transmission probability during the first cycle, that is, $p_{(i,1)}$, is set to 1. The node always transmits new data to ensure that the other nodes in the area have the data. This allows us to increase the coverage rate quickly.

Step 2: The node calculates $x$ between itself and the center of the target area. When the node leaves the data transmission area ($x > R+r$), the node discards the data because it can no longer contribute to data retention within the target area. This reduces the storage load in the node. When the node remains in the data transmission area ($x \leq R+r$), it proceeds to Step 3 to determine the data transmission probability.

Step 3: In subsequent cycles ($t \geq 2$), $p_{(i,t)}$ is determined based on the number of neighboring nodes $n_{(i,t-1)}$. Here, when the number of neighboring nodes is four or more, the node's transmission range has the potential to be completely covered by that of all neighboring nodes. For example, when the neighboring four nodes are located to a node's north, south, west, and east (the ideal arrangement), the node's potential transmission coverage area is already completely enclosed by that of the other nodes, and the node need not transmit the data. Therefore, the data transmission probability $p_{(i,t)}$ is determined based on the number of neighboring nodes $n_{(i,t)}$ as described below.

If $n_{(i,t-1)} \leq 3$, then $p_{(i,t)}$ is set to 1 and $d$ is set to half. This point is also one of significant difference from the previous method. When the node density is low, since the transmission time $s_{(i,t)}$ is randomly determined within the transmission interval $d$, the data transmission interval in one vehicle ($s_{(i,t+1)} - s_{(i,t)}$) may reach up to $2d$, which means that the number of neighboring nodes is small and the data transmission interval becomes long. As a result, opportunities for data exchange with neighboring nodes decrease, and it becomes difficult to both disseminate and retain the data within the area. Therefore, the transmission interval $d$ is set to half and the node always transmits with the data transmission probability set to 1.

If $n_{(i,t-1)} \geq 4$, then $p_{(i,t)}$ is determined based on the number of neighboring nodes and the number of received data packets. In our proposed system, only the minimum number of nodes required to maintain a high coverage rate should transmit the data. However, in situations where the neighboring nodes' locations are radically asymmetrical and have the potential to become imbalanced, transmission coverage may be incomplete even if there is a large number of neighboring nodes. In such cases, the node proceeds to Step 4.

Step 4: To solve the abovementioned problems, we defined $m_{(i,t)}$ as the estimated value of the number of received data during $t$-th cycle and adjusted the transmission probability based on the $m_{(i,t)}$. The predicted value $m_{(i,t)}$ is given as in the following equation

$$m_{(i,t)} = \alpha * l_{(i,t-1)} + (1-\alpha) * m_{(i,t-1)} \quad (2)$$

where $m_{(i,t-1)}$ is the predicted value of the previous cycle, $l_{(i,t-1)}$ is the number of received data packets in the previous cycle (actual value), and $\alpha$ is the moving average coefficient. A node proceeds to Step 5 to decide the $p_{(i,t)}$ based on $m_{(i,t)}$.

Step 5: The node adjusts its transmission probability so that the number of data transmission in the $t$-th cycle becomes the given target value $\beta$. If $m_{(i,t)}$ is less than $\beta$, the node can then predict that the number of data transmission is unlikely to cover the area. Therefore, it must increase its transmission probability. However, if $m_{(i,t)}$ is more than $\beta$, then the node needs to decrease its transmission probability because excessive data transmissions will occur in the next cycle. At the start of the $t$-th cycle, each node estimates $m_{(i,t)}$

and then adjusts its transmission probability. The transmission probability is adjusted as follows:

$$p_{(i,t)} = \begin{cases} p_{(i,t-1)} + \frac{\beta - l_{(i,t-1)}}{n_{(i,t-1)}+1} & (0 < m_{(i,t)} < \beta) \\ p_{(i,t-1)} & (m_{(i,t)} = \beta) \\ p_{(i,t-1)} - \frac{l_{(i,t-1)} - \beta}{n_{(i,t-1)}+1} & (\beta < m_{(i,t)}) \end{cases} . \quad (3)$$

In this case, the initial value of transmission probability during the first cycle is set to $\frac{\beta}{n_{(i,t1)}+1}$. This means that the average transmission probability of all nodes (including itself and the number of neighboring nodes $n_{(i,t-1)}$) is set to control the number of data transmission as $\beta$. If $m_{(i,t)}$ is less than $\beta$, then all $n_{(i,t-1)} + 1$ nodes increase their individual data transmission probabilities by $\frac{\beta - l_{(i,t-1)}}{n_{(i,t-1)}+1}$ because their estimates will show that the number of data transmissions will not reach the $\beta$ value. However, if $m_{(i,t)}$ is more than $\beta$, then the individual nodes decrease their transmission probabilities by $\frac{l_{(i,t-1)} - \beta}{n_{(i,t-1)}+1}$ because they can predict that excessive transmissions will occur. If $m_{(i,t)}$ is equal to $\beta$, then $p_{(i,t)}$ is set to $p_{(i,t-1)}$ because the current data transmission probability is appropriate. Note that if the value of $\frac{\beta - l_{(i,t-1)}}{n_{(i,t-1)}+1}$ or $\frac{l_{(i,t-1)} - \beta}{n_{(i,t-1)}+1}$ is less than zero, then $p_{(i,t)}$ is set to $p_{(i,t-1)}$, and the transmission probability range is varied from $\frac{\beta}{n_{(i,t1)}+1}$ to 1.

Step 6: Finally, the node broadcasts the data based on $p_{(i,t)}$. Then, when the transmission timing $s_{(i,t+1)}$ in the next transmission period $d$ arrives, the node returns to Step 2 and then repeats the processing sequence.

### C. Improvement and Comparison

Table I compares our proposed method with our previous work [6][14] based on the STD retention system and the floating content method [12].

First, with a focus on data retention, the proposed method, previous control method, and previous geographical control method provide STD directly to the users by distributing and retaining the data with vehicles within the target area. However, the floating content method returns the data to the position where the data were created and stores the data there using vehicles. The former objective is to encourage users to utilize the STD that they received by passive reception. The floating content and conventional vehicular cloud methods have query-response architectures that require users to request information.

Next, we look at data transmission overhead. In the floating content method, each node effectively maintains data within a specific area by adjusting the transmission probability based on the distance to the source when the node transmits the data. However, when a node encounters a neighboring node, it exchanges the list of data held in addition to the beacon and then exchanges data with the neighboring node to obtain data that it does not have. As a result, both the process and amount of data transmitted may be large, which means that channel resources may not be effectively utilized. In contrast, the geographical control method uses the location and node movement information to effectively retain the STD. Since the node exchanges this information using the beacon, the amount of data transmission increases. In addition, it is conceivable that the calculation cost may increase to determine the transmission probability based on the position and movement of neighboring nodes. Finally, the proposed and previous control methods only transmit beacons and STD to achieve effective data retention by simple transmission control. However, the previous control method has an excessive data transmission problem because all nodes in area $r$ (within the data transmission area but outside the target area) have a transmission probability $p_{(i,t)}$ of 1. In addition, the previous method has a slightly higher computational complexity because each node uses its movement direction to determine its data transmission probability. By improving the previous control method, the method proposed herein achieves simple and effective retention by controlling the nodes in the $r$ area, as well as the nodes in the $R$ area.

## V. Performance Evaluation

In this section, we report on a simulation-based performance evaluation of our proposed method.

### A. Simulation Models

We evaluated our proposed method using the Vehicles in Network Simulation (Veins) [19] framework, which simultaneously implements both the IEEE 802.11p specification for wireless communications and the VANET mobility model. As a result, Veins can combine the Objective Modular Network Testbed in C++ (OMNeT++) network simulator [20] with the Simulation of Urban MObility (SUMO [21]) road traffic simulator.

In our simulations, all nodes were equipped with an interface compliant with IEEE 802.11p. In this paper, we assume that the user receiving the data is a vehicle and can receive the data using 802.11p interfaces. Furthermore, the coverage rate represents the percentage of the area where the vehicle can receive data transmitted at IEEE 802.11p within the retention area. We used the "Simple Path Loss Model" introduced in OMNeT++ as the wireless channel model. The Rx sensitivity for data reception was fixed to -89 dBm. In our simulation, the Tx power of each vehicle was set based on the communication range of wireless interfaces. For example, if the communication range is 300m, the Tx power should be set to approximately 10.4 dBm (11mW). As a result, it is possible to evaluate the simulation considering the vehicle's communication range and the size of the retention area. Furthermore, the channel frequency was set to 5.87 GHz, which is used as an unsafe application based on ITS-G5[22][23]. The transmission rate was set to 6 Mb/s based on the default value for the configured frequency. The size of the beacon for our STD retention system was set to 14 bytes. The beacon contains the media access control address (6 bytes) for a unique node ID and a timestamp filed (8 bytes). The data size was set to 1000 bytes assuming one-frame IoT data.

Next, the transmission interval $d$ was set to 5s. The user stays in the retention area for 5s, and then the user can receive the data. If $d$ is short, the total number of data transmissions per unit of time increases. However, if data retention with a transmission interval of 5s (a heavy load condition) is

TABLE I: Comparison between our proposed method and related work.

| Property | Proposed method | Previous control method [6] | Previous geographical control method [14] | Floating content method [12] |
|---|---|---|---|---|
| Data retention | Yes | Yes | Yes | No |
| Overhead related to data transmission | Low | Middle | High | High |
| Type of data transmitted | Beacons and STD | Beacons and STD | Beacons with geographical information and STD | Beacons, lists of STD, and STD |

achieved, our system will be also able to provide effective data to the user under various environments. Furthermore, the interval $b$ of the beacon for our STD retention was set to 1s because it is necessary to grasp the number of the neighboring nodes in a short period based on the simulation in advance. In these simulation evaluations, IEEE802.11p beacons are not transmitted because they may interfere with the pure evaluation of the proposed method. These parameters are set as the high load examples for the use of the proposed method. The moving average coefficient $\alpha$ of our proposed method was set at 0.5 in order to average the estimated number of neighboring nodes. The number of simulation trials was set at ten. A different mobility pattern of the traffic model was used for each trial.

As comparison methods, we used a naive method, our previous control method based on [6], and a geographical control method [14]. In the naive method, the transmission probability $p_{(i,t)}$ of all nodes is always set to 1 in the simulation area, so that it provides the case with the largest consumption of wireless resources within the retention system. We expect that the naive methods can obtain results close to Floating Content, which always exchanges data when passing neighboring nodes.

To show the effectiveness of our proposed method, we used three simulation topologies. The first topology in which a stable number of neighboring nodes is maintained, as shown in Figure 6. In this topology, rows of vehicles are placed at set distances and run alternately from the east and west directions every 200 m. The vehicle that reaches the edge of the road is lost, but the new vehicle is generated from the opposite edge of the road. The number of neighboring nodes is kept constant in order to evaluate the fundamental performance of the proposed method. Incidentally, this topology is not realistic. The second topology used a random traffic model that is produced by assuming actual vehicle movements, as shown in Figure 7. In this topology, nodes with randomly generated starting and endpoints operate on a road grid with a spacing of $h$ meters. The vehicle randomly determined the driving road from the candidate for the destination direction at the intersection. A traffic signal was installed at each intersection. The third topology used an urban traffic model using the Luxembourg SUMO Traffic (LuST) model [15] in order to evaluate the proposed method under a more practical mobility environment. LuST also reproduces vehicle generation and signal behavior. In Sections V-B to V-G below, we compare our proposed method with the previous control method and the naive method and discuss the simulation results using both first and second topologies. After that, we compare our proposed

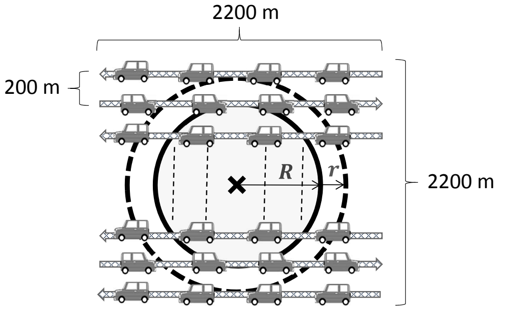



Fig. 6: Simulation topology 1.

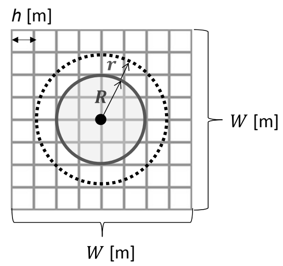



Fig. 7: Simulation topology 2.

method with the previous geographical control method using an ideal transmission probability combined with the random topology discussed in Section V-H. Finally, we compare our proposed method with the naive method using LuST model in Section V-I. [3]

In this paper, in order to clarify the performance of the STD retention function, we set the center point of STD retention as the location of the data origin. The STD starts to retain at the beginning of the simulation and retains for a period of time. Therefore, in this paper, we focus on the data retention of a single piece of data is targeted, but the retention system of multiple and different types of data is proposed separately [24].

[3]Here it should be noted that we do not introduce the obstacle in our simulation evaluation. In our proposed STD retention system, each vehicle autonomously controls data transmission according to the detection status of beacons and data. In an environment where beacons and data do not reach due to obstacles, each vehicle responds autonomously and adaptively, such as increasing the transmission probability. Therefore, the presence or absence of obstacles will not significantly affect the performance of data retention. Thus, we did not introduce any obstacles in our simulation to evaluate the feasibility of our STD retention system's behavior, the maintenance of high coverage, and the reduction in the number of data transmissions.

### B. Node Density Impact

In this simulation, we investigated how node density affects the coverage rate, transmission reduction rate, and the total number of data transmissions. We used the simulation topology of shown in Figure 6 and set $\beta = 4$ because the minimum number of nodes necessary to provide total transmission coverage over the target area is four. Furthermore, the communication range of the node was set to 300 m, the target area radius $R$ was set to 750 m, and $r$ was set to 250 m (which is the distance that a wireless communication can reach in one hop). In this environment, we evaluated the performance of our proposed method in cases where the node density was changed by varying the distance between nodes from 100 to 300 m. As an additional comparison method, we also show the results of the proposed method without the support of nodes in the $r$ area. As a result, the average number of nodes within the transmission range of a given node (i.e., 300 m) also varied from approximately five to 16.

Figure 8a shows the average coverage rate and 95% confidence interval in the steady-state. This steady-state denotes the period from 70 to 90 s after the start of simulation because data retention has already been completed. Since the cycle period is 5 s, the coverage rate is the average value measured over four cycles (i.e., 20 s). From these results, it can be seen that the naive method achieved a coverage rate of almost 100% for all node conditions because the method sends data every cycle regardless of the number of neighboring nodes. From Figure 8b, we can see that the total number of data transmissions was highest for the naive method.

The coverage rates of both the previous method and proposed methods exceeded about 99.5%, but while these schemes were inferior to the naive method, they could still deliver data throughout the target area. Figure 8c shows the reduction rate in the total number of data transmissions compared with that of the naive method. These results show that the proposed method reduced the maximum number of transmissions by up to 60%, thus indicating that it retained data more efficiently than the previous method. The vehicles within $r$ of the previous method always transmits data when they are moving toward the center of the target area. On the other hand, the vehicles within $r$ of our proposed method determine the data transmission probability based on the number of neighboring nodes and the number of data transmissions as well as the vehicles within $R$. As a result, the proposed method can reduce the number of data transmissions while maintaining a high coverage rate as the previous method. In addition, the number of data transmissions in Figure 8b show that data transmissions by the proposed method were almost constant even when the number of neighboring nodes increased. These results show that probabilistic transmission control of the proposed method works effectively.

However, the coverage rate of the proposed method without the support of the $r$ area was lowest because nodes outside the target area do not assist with data retention. In other words, those results indicate that it is difficult to retain data using only nodes within the target area. The impact of the $r$ area is discussed in Section V-F.

### C. Impact of $\beta$ Value

Next, we investigated the impact of the value of $\beta$. In this simulation, the number of neighboring nodes was fixed at approximately 16 (with a distance between nodes of 100 m), and the $\beta$ value was varied from 2 to 12. Other simulation conditions were the same as described in Section V-B.

As shown in Figure 9a, the coverage rate of the naive approach was always 100% regardless of the $\beta$ value, because the transmission probability of all nodes was always 1. Other methods had increasing coverage rates as the $\beta$ value increased, and the coverage was more than 99% when $\beta$ was 4. In particular, when the $\beta$ value was low, the performance levels of the proposed and previous methods were different due to the behavior of the nodes in area $r$. Since the transmission probability of those nodes in the previous method is always 1, the coverage rates are improved even if the $\beta$ values were low. However, since in the proposed method those nodes transmit based on (3), the number of transmissions did not increase, and the coverage rate was reduced.

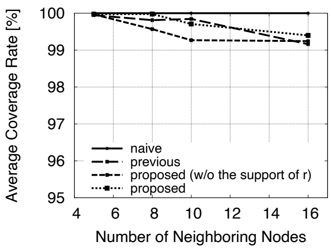
Focusing on the number of data transmissions in Figure 9b, we can see that the naive method continuously transmitted data regardless of the $\beta$ value. Since the proposed and previous methods increase the transmission probability with increasing $\beta$, their data transmissions increased. When $\beta$ was more than 4, the proposed method could reduce the number of data transmissions compared with the previous method. In Figure 9c, which compares the transmission reduction rates for the three methods, the proposed method had a reduction effect of about 1.5 times compared with the previous method.

Additionally, in all the graphs of Figure 9, the performance of the proposed method without the support of the $r$ area nodes appeared to be good because the average number of neighboring nodes in this scenario was 16. In other words, when the node density is very high, it is possible to achieve high levels of performance using only the nodes within the target area.

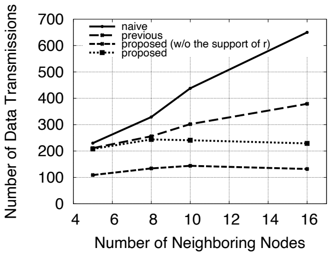
### D. Location-aware Analysis

Next, we investigated the data transmission status of nodes in different target area positions in order to obtain more detailed performance information. The simulation environment was the same as that in Section V-C. To perform this evaluation, we separated the target area into two different subareas: (a) *an edge area,* comprising nodes that can receive data transmitted from nodes in the $r$ area and (b) *a center area,* comprising nodes that cannot receive data transmitted from the $r$ area. The radius of the center area was 450 m. The edge area was defined as the area within a radius of 750 m but outside the center area. Our evaluation considered how the location differences impact the number of data transmissions.

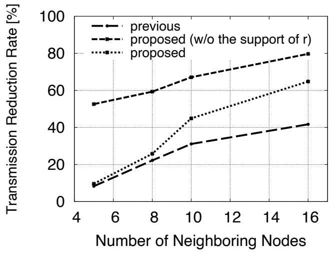
Figure 10a and 10b show the average number of data transmissions in the center area and the edge area, respectively. From Fig. 10a, it can be seen that the number of data transmissions in the center area of all methods could be controlled to nearly the $\beta$ value. On the other hand, from Fig. 10b, the average number of data transmissions in the edge area of the previous method was clearly larger than the $\beta$ value




(a) Average coverage rate.

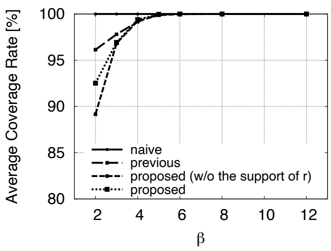



(b) The number of data transmissions.

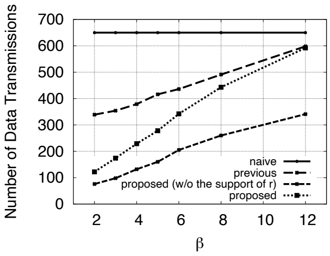



(c) Transmission reduction rate compared to the naive method.

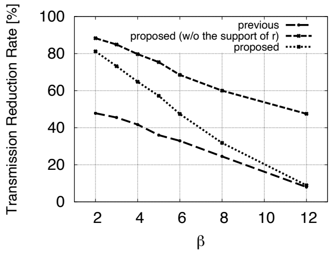
Fig. 8: Simulation results with changes in node density.




(a) Average coverage rate.

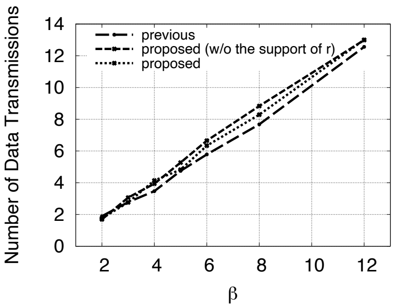



(b) Number of data transmissions.

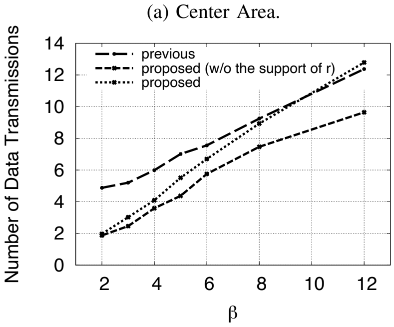



(c) Transmission reduction rate compared to naive method.

Fig. 9: Impact of $\beta$ values.




(a) Center Area.




(b) Edge Area.

Fig. 10: Average number of data transmissions for nodes in center and edge areas.

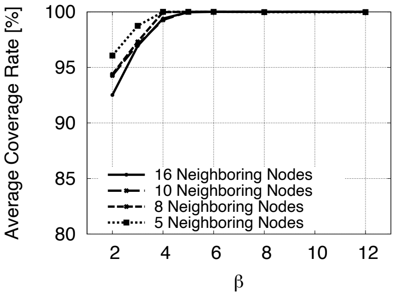
when the value of $\beta$ was low. This is because the nodes in the edge area could receive data from those in the $r$ area, which means they could receive many data transmissions. Because multiple nodes in the $r$ area were always transmitting data, redundant data transmissions occurred. Next, the number of data transmissions in the edge area of the proposed method without the support of nodes in the $r$ was insufficient to achieve $\beta$, especially for high $\beta$ values. This is because the density of nodes in the edge area was insufficient and data transmissions from nodes in the $r$ area were not provided at all. Finally, we can see that nodes in edge area of the proposed method adjusted the number of data transmissions to nearly the $\beta$ value. This indicates that, in order to reduce the number of data transmissions while maintaining a high coverage rate in the target area, the cooperation of nodes in the $r$ area is indispensable, and that these nodes need adaptive data transmission control.

### *E. Impact of $\beta$ and Node Density*

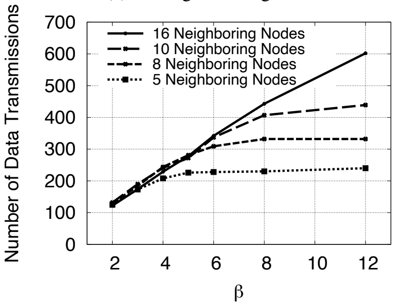
Next, we investigated the performance of the proposed method while varying both the $\beta$ value and the number of neighboring nodes. In this simulation, the $\beta$ value ranged from 2 to 12, and the number of neighboring nodes was set to 5, 8, 10, and 16. Figure 11 shows the average coverage rate and the number of data transmissions for the proposed method. From this figure, we can see that, the proposed method achieved high coverage rates regardless of the number of neighboring nodes when the $\beta$ value was more than 4. Therefore, we concluded that the $\beta$ value should be at least 4 in the proposed method. Furthermore, from Figure 11b, it can be confirmed that the number of data transmissions became saturated when the $\beta$ value exceeded the number of neighboring nodes. In other words, more effective data retention can be achieved by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3066438, IEEE Transactions on Mobile Computing



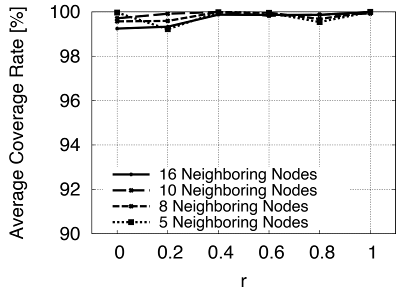



(a) Average coverage rate.

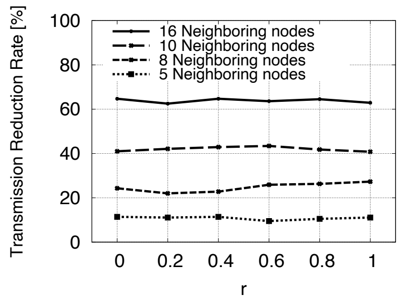



(b) Number of data transmissions.

Fig. 11: Simulation results for various $\beta$ values and node density conbinations.




(a) Average coverage rate.




(b) Transmission reduction rate compared to naive method.

Fig. 12: Simulation results for different $r$ areas and neighboring node numbers.

adjusting $\beta$ according to the number of neighboring nodes. This evaluation will be another topic of our future work.

### *F. Impact of r Area Size*

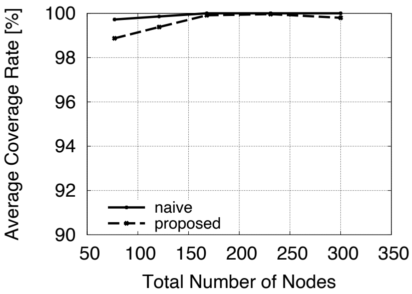
To investigate the effect of the $r$ area size, we evaluated the performance of the proposed method for various values of $r$. In this simulation, the communication range of each node was set to 300 m, the target area radius $R$ was set to 750 m, and the value of the $\beta$ value was set to 4. The model topology shown in Figure 6 was used. The $r$ area was varied as a ratio from 0 to 1 according to the ratio of $r$ to the wireless communication distance. The coverage rate and the transmission reduction rate (in comparison to the naive method) in $R + r$ for various numbers of neighboring nodes are shown in Figures 12a and 12b, respectively.

Figure 12a shows that all coverage rates are improved as the value of $r$ increased. This result indicates that the data retention in the target area are improved due to the cooperation of more nodes outside of the target area. In particular, the coverage rates reached 99% when the value of $r$ exceeded 0.75. In contrast, the data reduction rate was constant for increasing $r$ in Figure 12b. This indicates that the nodes in the $r$ area properly controlled the number of data transmissions by probabilistic data transmission as well as the vehicles in $R$, regardless of increases in the number of nodes located in the $r$ area. These results indicate that the cooperation of nodes within the wireless communication range but outside the target area is required to achieve effective data retention.

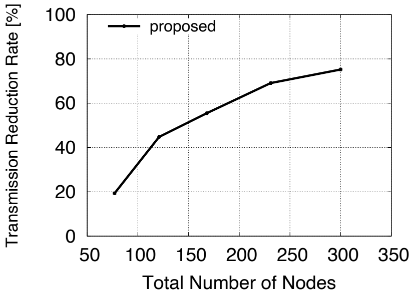
### *G. Performance Evaluation with Random Topology*

Next, to demonstrate the effectiveness of the proposed method in environments where the number of neighboring nodes dynamically changes because of node mobility, we conducted a performance evaluation using the random topology shown in Figure 7. In this simulation, $R$ was 750 m, $r$ was 250 m, the wireless communication distance was 300 m, the $\beta$ value was 4, the width $W$ of the simulation field was 2200 m, and the distance $h$ between each road was 200 m. The numbers of randomly generated nodes were 77, 121, 168, 231, and 300, depending on the number of nodes in the topology of Figure 6. The naive method was only used for comparison to simplify the discussion. This is because the evaluation using the first topology up to SectionV-F revealed the effectiveness of the proposed method compared to the previous method. Furthermore, the number of simulation trials was set at 100.

Figures 13a and 13b show the average coverage rate and the transmission reduction rate, respectively. When the number of nodes was 77, the coverage rate of the proposed method is varied greatly in each trial and the average value was as low as 99%. Because of the random movement of nodes, some nodes may not have had sufficient neighboring nodes. This situation improved when the number of nodes increased. The coverage of the proposed method was still approximately 0.5% lower than that of the naive method. However, Figure 13b shows that the proposed method succeeded in reducing the number of data transmissions by approximately 70% when the number of nodes was 300. Moreover, since the coverage rate of the proposed method becomes approximately 100% when the number of nodes was 168, the proposed method could be

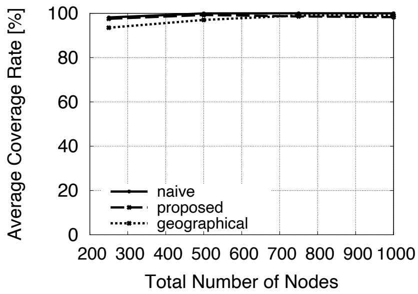



(a) Average coverage rate.

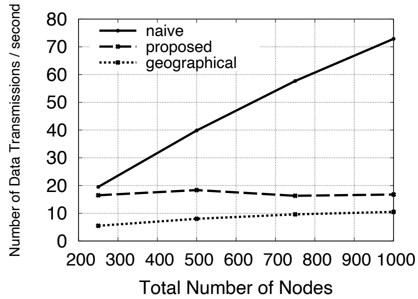



(b) Transmission reduction rate compared to naive method.

Fig. 13: Simulation results for different numbers of nodes in random topology.




(a) Average coverage rate.




(b) Number of data transmission per a second.

Fig. 14: Comparison of simulation results to geographical method on random topology.

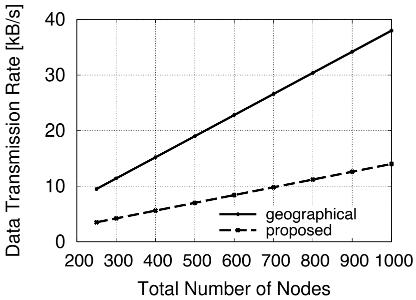



Fig. 15: Beacon packet payload results. The transmission rate was calculated as beacon packet size times number of nodes.

realized when 100 or more vehicles were within the 2,000m$^2$ target area where $R$ was 750m and $r$ was 250m ($r + R$ = 1,000m). This number of vehicular nodes is necessary for a target area of this size.

### H. Comparison to Geographic Control Method

We compared the performance of the proposed and geographical control methods with ideal transmission probabilities based on geographical information [14] on a random topology. The geographical method controls the transmission probability based on the exact position of each node. In this simulation, $R$ was 500 m, $r$ was 125 m, the wireless communication distance was 150 m, the $\beta$ value was 4, the width $W$ of the simulation field was 2000 m, and the distance $h$ between each road was 250 m. The number of randomly generated nodes was varied from 250 to 1000, because this scenario assumes a more realistic node-dense environment, such as the Manhattan model. Furthermore, the number of simulation trials was also set to 10 in this scenario.

Figure 14a shows the average coverage rate, while Figure 14b shows the number of data transmissions per second for the naive, proposed, and geographical methods. The coverage rate showed that while the proposed method could attain a performance level equivalent to that of the geographical method, that level decreased slightly as the number of nodes increased. The proposed method was unable to cover some areas because it controls the data transmission probability based on the node density and the number of data transmissions. However, the difference between the proposed and geographic methods was less than 1%, so users would not experience a significant variance in their data reception performance.

Finally, Figure 15 shows the beacon payload transmission rates for an increasing numbers of nodes. The beacon in the proposed method contains the media access control (MAC) address (6 bytes) for a unique node ID and a time stamp field (8 bytes), while the beacon message of the geographical method contains position information (x, y, z coordinates, which we assumed to be $8 \times 3 = 24$ bytes) in addition to the MAC address and the node ID. In other words, the beacon payload size for the proposed method is 14 bytes, while it is 38 bytes for the geographic method. As can be seen from Figure 15, the beacon transmission volume in the geographic method is increased with the increasing number of nodes, eventually consuming a significant amount of wireless resources. Thus, regarding the effective use of wireless resources, the proposed

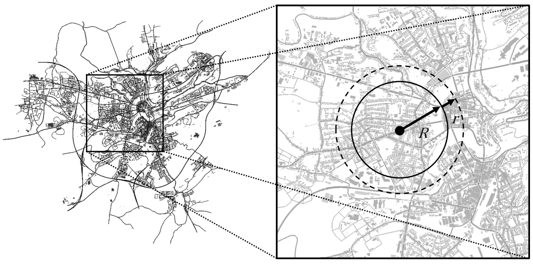



Fig. 16: The urban map of Luxembourg.

TABLE II: The number of nodes at each time.

| Time | The whole urban area | The retention area |
|---|---|---|
| 5:00 a.m. | 724 | 54 |
| 6:00 a.m. | 1674 | 191 |
| 7:00 a.m. | 3902 | 481 |
| 8:00 a.m. | 5345 | 700 |

and geographical methods can be said to have a trade-off relationship. However, the proposed method operates using a simple mechanism that does not require complex computations and can still achieve performance levels close to those possible with the geographic method, which employs ideal transmission control with a higher computational overhead.

### *I. Performance Evaluation using the Urban Model LuST*

Finally, we compared the performance of the proposed and naive methods on the urban traffic model LuST in order to demonstrate the effectiveness of the proposed method in a more realistic environment. In this simulation evaluation, the STD is retained at any point in Luxembourg, as shown in Figure 16. The LuST's mobility and traffic light model is used "shortest path with rerouting and activated traffic lights." $R$ was 750 m, $r$ was 250 m, the wireless communication distance was 300 m, the $\beta$ value was 4, and $d$ was 5 s as well as previous evaluations. We evaluated the coverage rate and the number of data transmissions when the data origin retains STD at 5, 6, 7, and 8 a.m. for 2 minutes. Table II shows the number of nodes in the whole urban area and the retention area $(R+r)$ at each time. A naive method in which the coverage ratio is maximum and the number of transmissions is maximum is used as a comparison method.

Figure 17a shows the average coverage rate per $d$ at each time. The coverage rate of both methods at 6 a.m. was approximately 93%, but the one reached almost 100% after 7 a.m. The number of nodes in the retention area at 6 a.m. was 54; it was insufficient to retain STDs throughout the retention area. From this result, it can be seen that a certain number of nodes is necessary to retain the STD in a realistic environment. On the other hand, Figure 17b shows the total number of data transmissions at each time. The total number of data transmission of the naive method increased over time, i.e., as nodes increased. The total number of data transmission of the proposed method did not increase much after 6 a.m. In the proposed method, since the node operates transmission control based on the neighboring node density, the number of data transmissions can be reduced even if the number of nodes increases.

Furthermore, Figures 17c and 17d show changes in the coverage rate and the number of data transmissions overtime at 7 a.m. Since there is a sufficient number of nodes within the retention area, the STD is spread over the whole retention area in the first cycle of $d$ after data transmission of the data origin. The number of data transmissions of the naive method is approximately 500 in each cycle of $d$. On the other hand, in the proposed method, a large amount of data is transmitted for data diffusion in the first cycle, but the number of data transmission can be suppressed to approximately 100 after the data diffusion is completed.

From the above results, it is clear that the proposed method can achieve high coverage while suppressing the number of data transmissions even in a more realistic traffic mobility environment.

## VI. Conclusion

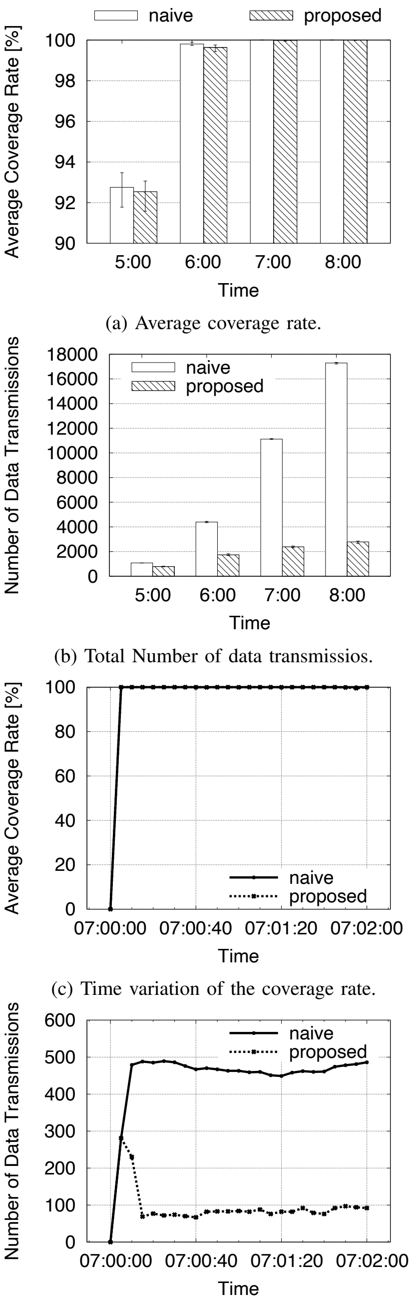
Our objective in this study was to achieve STD retention within a target area by allowing users to receive all STD passively from anywhere within a target area without the need to conduct any kinds of searches. To achieve this, we proposed an STD retention system that utilizes a VANET constructed from InfoHub vehicles. We also proposed an adaptive transmission control method that is used to set the data transmission probability based on the density of neighboring vehicles. In our proposed method, each vehicle first estimates the number of neighboring nodes based on the received beacon messages. Then, each node sets its probability adaptively by considering both the number of neighboring vehicles and the number of data transmissions during a previous time cycle. Through simulations, we clarified that our proposed method could control data transmissions in response to vehicle density changes. Although only one type of data was addressed in our study, it is obvious that various types of data must coexist in such environments. Therefore, in future work, we will extend the method to encompass the handling of various types of data simultaneously and to retain multiple pieces of continuous data. We will compare and analysis the performance related to data retention between our proposed method and other relared works for the spatio-temporal data utilization using vehicular networks, such as vehicular cloud networks. Furthermore, we will evaluate the effectiveness of the STD retention system using a realistic evaluation environment, including the obstacle's effect and multi-agent system.

## Acknowledgment

This work supported in part by the Japan Society for Promotion of Science (JSPS) KAKENHI Grant Number 18H03234 and 20K11792, and the National Institute of Information and Communication Technology (NICT).

## References

[1] OECD, "M2M Communications: Connecting Billion of Devices," OECD Digital Economy Papers, No. 192, 2012.




(a) Average coverage rate.

(b) Total Number of data transmissios.

(c) Time variation of the coverage rate.

(d) Time variation of Total Number of data transmissios.

Fig. 17: Simulation results using the urban traffic model LuST.


[2] Cisco,"Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 20162021 White Paper," https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, 2017.

[3] Cisco,"Internet of Things -At a Glance," https://www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/at-a-glance-c45-731471.pdf, 2019.

[4] S. M. Tornell, S. Patra, C. T. Calafate, J.-C. Cano and P. Manzoni, "A novel On-Board Unit to Accelerate the Penetration of ITS Services," In Proc. of 2016 IEEE 13th Annual Consumer Communications & Networking Conference (CCNC), 2016.

[5] R. Kumar and M. Dave, "A Review of Various VANET Data Dissemination Protocols," International Journal of u- and e- Service, Science and Technology, Vol. 5, No. 3, 2012.

[6] H. Teshiba, Daiki Nobayashi, Kazuya Tsukamoto and Takeshi Ikenaga, "Adaptive Data Transmission Control for Reliable and Efficient Spatio-Temporal Data Retention by Vehicles," Proc. of The Sixteenth International Conference on Networks (ICN2017), IARIA, 2017.

[7] F. Li, and Y. Wang, "Routing in vehiclar ad hoc networks: A survey," IEEE Vehicular Technology Magazine, Vol. 2, Issue 2, pp. 12–22, 2007.

[8] C. Maihofer, T. Leinumuller, and E. Schoch, "Abiding Geocast: Time-stable Geocast for Ad Hoc Networks," In Proc. ACM VANET, pp. 20-29, 2005.

[9] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. Zhang, "Mobile-Edge Computing for Vehicular Networks: A Promising Network Pradigm with Predictive Off-Loading," IEEE Vehicular Technology Magazine, Vol. 12, Issue. 2, pp. 36-44, 2017.

[10] S. Tabatabai, I. Mohammed, A. Al-Fuqaha, and J. Qadir, "Opportunistic Selection of Vehicular Data Brokers as Relay Nodes to the Cloud," In Proc. of 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), 2020.

[11] I. Leontiadis, P. Costa, and C. Mascolo, "Persistent content-based information dissemination in hybrid vehicular networks," In Proc. IEEE Pervasive Computing (PerCom) 2009, pp. 1-10, 2009.

[12] J. Ott, E. Hyyti, P. Lassila, T. Vaegs, and J. Kangasharju, "Floating Content: Information Sharing in Urban Areas," In Proc. IEEE Pervasive Computing (PerCom) 2011, pp. 136-146, 2011.

[13] 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.

[14] T. Higuchi, R. Onishi, O. Altintas, D. Nobayashi, T. Ikenaga, and K. Tsukamoto, "Regional InfoHubs by Vehicles: Balancing Spatio-Temporal Coverage and Network Load," Proc. of First International Workshop on Internet of Vehicles and Vehicles of Internet, pp. 25-30, Jul. 2016.

[15] L. Codeca, R. Frank, and T. Engel, "Luxembourg SUMO Traffic (LuST) Scenario: 24 Hours of Mobility for Vehicular Networking Research," in Proceedings of the 7th IEEE Vehicular Networking Conference (VNC15), 2015.

[16] E. Lee, E.K. Lee, M. Gerla, S. Y. Oh, "Vehicular cloud networking: architecture and design principles," IEEE Communications Magazine, Vol. 52, Issue 2, pp. 148-155, 2014.

[17] M. Whaiduzzamana, M. Sookhaka, A. Gania, and R. Buyya, "A survey on Vehicular Cloud Computing Journal of Network and Computer Applications," Vol. 40, pp. 325-344, 2014.

[18] M. Maiouak and Tarik Taleb, "Dynamic Maps for Automated Driving and UAV Geofencing, IEEE Wireless Communications," Vol. 26, Iss. 4, pp. 54-59, 2019.

[19] "Veins," [Online], Available from: http://veins.car2x.org/.

[20] "OMNeT++," [Online], Available from: https://omnetpp.org/.

[21] "Simulator of Urban Mobility (SUMO)," [Online], Available from: https://http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/.

[22] "European Telecommunications Standards Institute. ETSI ES 202 663: Intelligent Transport Systems (ITS); European profile standard for the physical and medium access control layer of Intelligent Transport Systems operating in the 5 GHz frequency band, v1.1.0 edition, 2010. " [Online], Available from: https://www.etsi.org/deliver/etsi_es/202600_202699/202663/01.01.00_60/es_202663v010100p.pdf

[23] V. Mannoni, V. Berg, S. Sesia, and E. Perraud, "A Comparison of the V2X Communication Systems: ITS-G5 and C-V2X," 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), 2019.

[24] D. Nobayashi, K. Tsukamoto and T. Ikenaga, "Autonomous Data Transmission Control Based on Node Density for Multiple Spatio-temporal Data Retention," Proc. of IEEE International Conference on Cloud Networking (CloudNet) 2018, Oct. 2018.

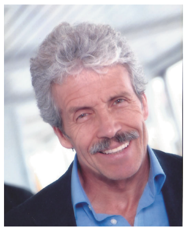
**Daiki Nobayashi** received the B.E., M.E. and D.E. degrees from Kyushu Institute of Technology, Tobata, Japan in 2006, 2008, and 2011, respectively. From April 2010 to March 2012, he was a Japan Society for the Promotion of Science (JSPS) Research Fellow. Since March 2012, he has been an Assistant Professor in Department of Electrical and Electronic Engineering, Faculty of Engineering, Kyushu Institute of Technology. His research interests include network architectures and protocols, wireless LANs, multi-hop wireless networks, wireless sensor and actuator networks. He is a member of IEEE and IEICE.

**Ichiro Goto** received the B.E. degrees in Department of Electrical and Electronics Engineering from Kyushu Institute of Technology, Japan in 2019. He is now a graduate student of Kyushu Institute of Technology. His current research interests include wireless networks and vehicular networks. He is a member of the IEICE.

**Hiroki Teshiba** received the B.E. degrees in Department of Computer Science and Electronics from Kyushu Institute of Technology, Iizuka, Japan in 2016. He was a graduate student of Kyushu Institute of Technology majoring in Department of Advanced Informatics. His current research interests include wireless networks and vehicular networks. He is a member of the IEICE.

**Kazuya Tsukamoto** received the DE degree in computer science from the Kyushu Institute of Technology (KIT), Japan, in 2006. From April2006 to March 2007, he was a Japan Society for the Promotion of Science (JSPS) research fellow at KIT. In 2007, he was an assistant professor in the Department of Computer Science and Electronics, KIT, and then has been an associate professor in the same department since April 2013. His research interests include performance evaluation of computer networks and wireless networks. He is a member of the ACM, the IPSJ, the IEICE, and the IEEE.

**Takeshi Ikenaga** received B.E., M.E. and D.E. degrees in computer science from Kyushu Institute of Technology, Iizuka, Japan in 1992, 1994 and 2003, respectively. From 1994 to 1996, he worked at NEC Corporation. From 1996 to 1999, he was an Assistant Professor in the Information Science Center, Nagasaki University. From 1999 to 2004, he was an Assistant Professor in the Department of Computer Science and Electronics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology. Since March 2004, he has been an Associate Professor in the Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Kyushu Institute of Technology. His research interests include performance evaluation of computer networks and QoS routing. He is a member of IEEE and IEICE.



**Mario Gerla** is Professor of Computer Science at UCLA. He holds the Jon Postel Chair in Networking. He received his Ph.D. degree from UCLA. At UCLA, as a graduate student, in the early 70's, he was part of the team that developed the ARPANET models and protocols that formed the basis of the modern INTERNET. After a period in Industry, he joined the UCLA Faculty in 1976. Since then, he has been involved in the design and evaluation of several network protocols such as clustering, multicast and TCP Westwood. As lead of the ONR MINUTEMAN project, he designed scalable airborne Internet concepts for tactical scenarios. His current interests include vehicular networks, mobile computing and privacy and security in mobile systems. Dr. Gerla is IEEE Fellow, serves on the IEEE TON Scientific Advisory Board and was recognized with the ACM Sigmobile Outstanding Contribution Award in 2015.