# Autonomous Data Transmission Control Based on Node Density for Multiple Spatio-temporal Data Retention

| | |
|---|---|
| 著者 | Nobayashi Daiki, Tsukamoto Kazuya, Ikenaga Takeshi |
| journal or publication title | 2018 IEEE 7th International Conference on Cloud Networking (CloudNet) |
| year | 2018-10 |
| URL | http://hdl.handle.net/10228/00007394 |

doi: info:doi/10.1109/CloudNet.2018.8549521

# Autonomous Data Transmission Control Based on Node Density for Multiple Spatio-temporal Data Retention

Daiki Nobayashi[1], Kazuya Tsukamoto[1], and Takeshi Ikenaga[1]
[1] Kyushu Institute of Technology, Fukuoka, Japan
nova@ecs.kyutech.ac.jp, tsukamoto@cse.kyutech.ac.jp, ike@ecs.kyutech.ac.jp

***Abstract*—Although countless services can now be accessed via the Internet, some specific services such as local traffic conditions and limited-time sale advertisements are strongly dependent on geographical locations and times. Information of this type, which is commonly referred to as spatio-temporal data (STD), is not readily available online. Since the paradigm of local production and consumption of STDs can be effective for location-dependent applications, we previously proposed an STD data retention system that uses vehicular networks to create a mobile cloud. Unfortunately, effective data retention is difficult when multiple STDs exist in the same area because channel interference will result from the increased number of data transmissions. To resolve this issue, we herein propose an autonomous data transmission control method based on node density for multiple STD retention that facilitates a highly reliable coverage rate while limiting the individual data transmissions for each STD. Then, through simulations, we show that our proposed method is effective for simultaneously retaining multiple STDs.**

***Index Terms*—Data Retention, Autonomous Data Transmission Control, Locally Produced and Consumed of Data**

## I. Introduction

Recent advances in machine-to-machine (M2M) and Internet of Things (IoT) technologies have resulted in an explosive increase in devices with wireless communication interfaces. In fact, the Organisation for Economic Co-operation and Development (OECD) has reported that more than 50 billion units of M2M devices will be in service by 2020 in [1]. In most current IoT service schemes, service providers (SPs) embed sensors into everyday objects in order to collect and forward data to cloud servers or data centers for analysis. Those SP analysis results are then reprocessed to provide or improve user services. However, this upsurge in IoT devices has resulted in prodigious data collection amounts, thus requiring enlarged data centers equipped with expensive high-performance central processor units (CPUs) to handle the analyses, as well as large-capacity data storage systems to retain the collected/analyzed data. The increased amounts of transmitted data also put heavy stresses on network traffic processing capabilities.

Another issue of ongoing concern is that data generated from IoT devices often relate only to the specific locations and timeframes in which they are generated. Such data, which include local road traffic and accident information, limited-time sale advertisements in shopping areas, etc., are referred to as *spatio-temporal data* (STD). However, in addition to being collected by remote Internet servers for later analysis and use, such data have strong potential for direct provision to and use by clients in certain locations. Accordingly, as a **local production and consumption** paradigm, we have proposed a new information infrastructure for IoT devices that uses a novel network architecture to collect, maintain (retain), and deliver local STDs without connecting to the Internet [2][3][4].

978-1-5386-6831-318$31.00 ©2018 IEEE

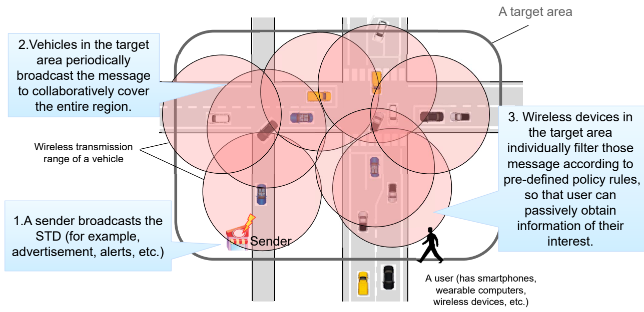



Fig. 1: Overview of data retention system (DRS).

Previously, we proposed a data retention system (DRS) that uses vehicular networks for local production and consumption of STDs [2]. Figure 1 shows the overview of that system. However, when the STD are retained within a specific area on vehicular ad-hoc networks (VANETs), channel contention occurs between InfoHubs if the vehicle density is high and all vehicles share the same wireless resource. Such contention results in packet losses and transmission delays that cause communication quality to deteriorate. On the other hand, if vehicle density is low, the sparsity of vehicles makes it difficult to thoroughly disseminate STDs within that area. To solve these problems, we previously introduced an adaptive data transmission control method based on DRS node density and evaluated its effectiveness. However, since this DRS did not consider cases where multiple information sources exist simultaneously within the same space, the number of data transmission increases with the upswing in the number of STDs within the same specific area, which makes it challenging to avoid data collisions.

In this paper, we propose an autonomous data transmission control method for multiple STD retention based on node density that achieves effective data transmission by predicting the number of neighboring nodes and the number of data in flight.

Furthermore, to satisfy the different quality requirements of each STD in terms of the transmission period, the proposed method uses adaptive data aggregation to provide high coverage rates while minimizing data transmissions for each STD. Finally, using simulations, we show that our proposed method is effective for retaining multiple STDs.

## II. Our Previous Work: DRS for Single STD

In this section, we first describe the assumptions and system requirements behind the DRS and then give an overview of its adaptive data transmission method.

### *A. Assumptions and System Requirements*

As in the previous study [2], we begin by assuming that the information source is located at the center of the target area. Information regarding the target area such as radius $R$ and data transmission interval $d$ is merged with the data and transmitted as the STD. This STD is spread throughout the target area by a network composed of InfoHubs (hereafter referred to as nodes). The goal is to ensure a user can always retrieve data transmitted from the information source, in real-time and passively, regardless of his/her location within the target area. Note that because the STD is distributed and managed without using existing infrastructures such as the Internet, the load is reduced, and data fault tolerance is improved.

As in our previous study [2], the system requirements of our current proposal define the coverage rate based on the probability that the user can passively receive data everywhere within the target area. The coverage rate is expressed as follows:

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

where $S_{DT}$ denotes the size of target area, and $S_{TA}$ denotes the size of total area in which a user can obtain the data transmitted from any of the nodes within the transmission interval.

### *B. Adaptive Data Transmission Control based on Node Density*

The DRS employs an adaptive transmission control method in which the transmission probability is dynamically changed based on the neighboring node density [2]. In this method, nodes around the target area are classified into two types based on distance from the center of the target area. The specific conditions are described below:

$$\begin{cases} 0 \le Distance \le R + r: & Target\ Area \\ Otherwise: & no\ belonging \end{cases}$$

where $Distance$ denotes the distance between the node and the target area center. This distance is calculated from Global Positioning System (GPS) information. $r$ is the wireless communication range. In this method, nodes outside the target area also cooperate to facilitate effective data retention.

On the other hand, the data are transmitted based on the following procedure. When a node $i$ receives STD from another node, it first checks the transmission intervals of $d$ seconds included in the STD. Then, the node randomly determines the next transmission time in $s_{(i,t)}$ seconds. Note that the actual transmission time is determined at the start time of the $t$-th cycle. The random determination within $d$ seconds allows the node to avoid data transmission collisions.

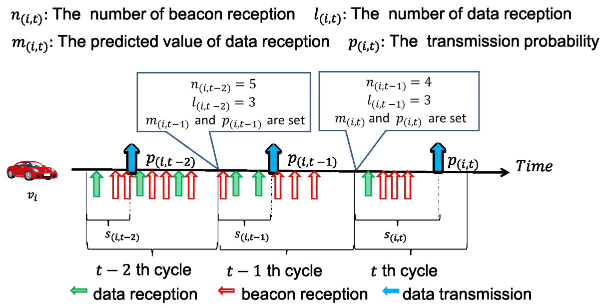



Fig. 2: Outline of transmission probability decision.

Next, each node calculates the number of neighboring nodes by the number of beacons (including the node ID) that are transmitted periodically from the neighboring nodes. As shown in Fig. 2, the number of neighboring nodes $n_{(i,t)}$ is calculated by using receiving beacons within cycle $t-1$. Note that the processing of probabilistic transmission control can be divided into the following two cases:

**case1** $n_{(i,t-1)} \le 3$:
Transmission probability $p_{(i,t)}$ is set to 1. Since the node's own transmission coverage cannot be completely covered by that of the neighboring nodes, it has to transmit, i.e., $p_{(i,t)}$ is set to 1.

**case2** $n_{(i,t-1)} \ge 4$:
$p_{(i,t)}$ is determined based on the number of neighboring nodes and the number of received data. However, since such high node density inherently poses transmission collision risks, data transmission should be limited to the minimum number of nodes required to maintain the high coverage. Consequently, in situations where the locations of neighboring nodes are radically asymmetrical and could become imbalanced, the transmission coverage may not be complete, even if there are a large number of neighboring nodes.

To solve these abovementioned problems, we define $m_{(i,t)}$ as the estimated value of the number of received data during the $t$-th cycle and adjust the transmission probability based on $m_{(i,t)}$, given by

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

where $l_{(i,t-1)}$ is the number of received data in the previous cycle (actual value) and $\alpha$ is the moving average coefficient. Furthermore, the node adjusts its transmission probability so that the number of data transmissions in the $t$-th cycle becomes the given target value $\beta$.

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

In this case, the initial value of transmission probability at the first cycle is set to $\frac{\beta}{n_{(i,t-1)}+1}$. This means the average

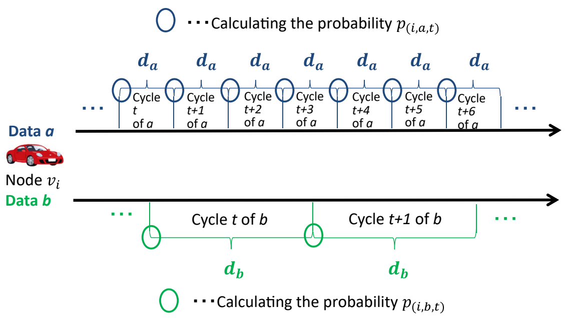



Fig. 3: Calculation timing of data $x$ transmission probability.

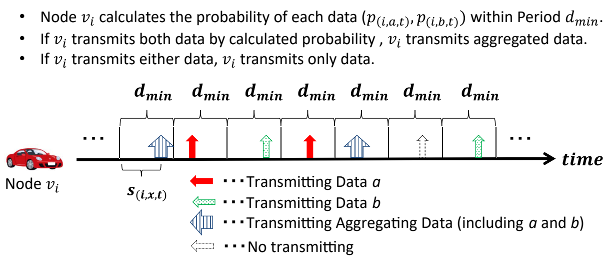



Fig. 4: Transmission timing.

transmission probability of all nodes (including itself and a number of neighboring nodes) is set to control the number of data transmissions as $\beta$. 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, $p_{(i,t)}$ is set to $p_{(i,t-1)}$, and the transmission probability range varies from $\frac{\beta}{n_{(i,t-1)}+1}$ to 1.

## III. Our Proposed Method: DRS for Multiple STDs

Although the above-mentioned DRS could effectively retain a single STD because it does not address the possibility of multiple information sources existing simultaneously within the same space, data transmissions increase with increases in the number of STDs. This makes it hard to avoid data collisions. To counter this problem, we will now describe our proposed method of extending the DRS to accommodate multiple STDs.

### A. Objective of Aggregating Transmission of STDs

Since a DRS is not set up to retain multiple (different) STDs, when a node transmits independently for each STD $x$ ($x \in X$, X indicates data set generated in the target area), the number of data transmissions increases in proportion to the number of data. As a result, the number of channel collisions between the nodes increases and STD retention becomes difficult. Furthermore, STDs transmitted from each device (source) have different required qualities such as target areas and transmission cycles. Therefore, while our proposed system must ensure efficient data retention, it is also necessary to consider the required qualities of each data $x$.

To solve these problems and satisfy the above-mentioned requirements, we propose a method to adaptively transmit multiple different data simultaneously. First, data are adaptively selected for inclusion in a bundle based on the differences between the transmission cycles for each data, after which the bundled data are transmitted. This process is intended to reduce the number of data transmissions in the target space and improving the coverage ratio for all STDs.

### B. Aggregating STD transmissions

In order to aggregate and transmit multiple data with different transmission periods, it is first necessary to decide the data transmission timing. Therefore, we begin by focusing on the data transmission period $d_x$ set for each data $x$. If a node were to transmit all the data at one time while following the most extended transmission period among the stored data, it would be impossible to satisfy the data quality requirement within a short transmission period. For example, suppose that the node has data $a$ and $b$ with data transmission periods $d_a = 2$ and $d_b = 6$, respectively. Then, if the node aggregates and transmits the two data simultaneously on $d_b$, the aggregated data are transmitted every 6 seconds, and thus the coverage rate of $a$ decreases significantly. On the other hand, if the node selects the shortest transmission period among the stored data, the required quality of all data can be satisfied. However, in this situation, data with an extended transmission period are transmitted excessively and data collisions increase. Furthermore, it is difficult to control the number of data transmissions within the transmission period set to $\beta$.

Therefore, in our proposed method, a node first determines the transmission probability of each data $p_{(i,x,t)}$ by using adaptive transmission control based on node density, as described in Section III. Next, the shortest transmission period $d_{min}$ among the stored data is set as the reference data transmission period (Figure 4). As a result, the node transmits data every $d_{min}$ but determines the data to be transmitted at that time based on the probability $P_{add(x)}$ weighted $p_{(i,x,t)}$ set for each $x$.

$$P_{add(x)} = p_{(i,x,t)} * \frac{d_{min}}{d_x}. \tag{4}$$

The weight is for preventing excessive data transmission when data $x$ is transmitted at $d_{min}$ intervals. When multiple data are transmitted at the timing of $d_{min}$, the node aggregates the data and transmits them simultaneously.

## IV. Performance Evaluation

In this section, we report on a simulation-based performance evaluation of our proposed method. To show the effectiveness of our proposed method, we utilize the *naive* and *individual* methods for comparison purposes. The naive method stipulates that the transmission probability $p_{(i,x,t)}$ of all nodes in the simulation area will always be set to 1. The individual method uses the DRS for single data[2].

### A. Simulation Model

We evaluated our proposed method on the Vehicles in Network (Veins) [9] simulation platform. The Veins platform implements the IEEE 802.11p specification for wireless communications on OMNeT++ [7] and can simultaneously generate a VANET mobility model by using the Simulation of Urban Mobility (SUMO) [8] package.

Figure 5 shows a simulation topology in which a stable number of neighboring nodes was maintained. In this topology,

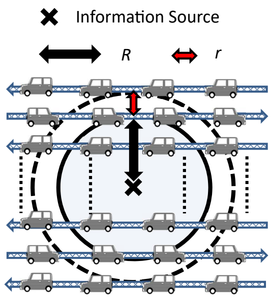
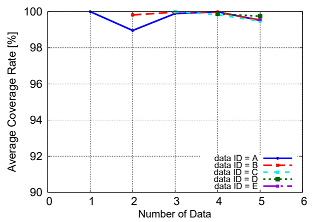
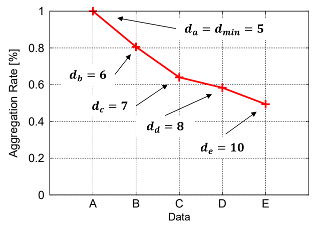


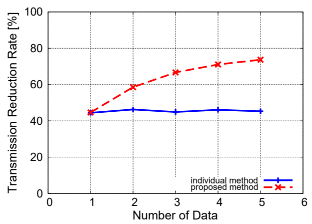

Fig. 5: Simulation topology



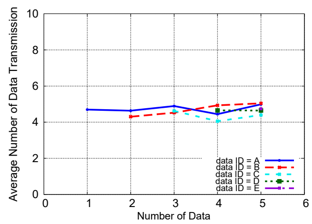

(a) Proposed method coverage rate.




(b) Aggregation rate.




(c) Transmission reduction rate.




(d) Average number of data transmissions for nodes in the target area.

Fig. 6: Simulation results.

rows of vehicles placed at set distances from each other run alternately from east and west every 200 m. The number of neighboring nodes was set to 10. Furthermore, the communication range of the node, $R$, $r$, beacon interval, $alpha$, and $\beta$ were set to 300 m, 750 m, 250 m, 1 m, 0.5, and 4, respectively. To retain multiple data within the target area, we varied the number of data from 1 to 5. The data IDs were set from $a$ to $e$, and their data transmission periods were set to from 5 to 10 s. Note that all data were generated from the same point and the size of all data was set to 300 bytes. The simulation time was 300 s, and the number of simulation trials was 10.

### *B. Simulation Results*

Figure 6a shows the average coverage rate of each data in the steady state. This steady state is the period from 70 to 90 s after the start of the simulation because data retention has already been completed. From these results, it can be seen that the proposed method achieved a coverage rate in excess of 99 % for all data, even if the number of data increases. Figure 6b shows the average aggregation rate of each data compared with the number of transmissions for the independent method (transmission without aggregation). This result shows that the probability control provided by our proposed method is useful because each data is aggregated with values close to $\frac{d_{min}}{d_x}$. Furthermore, Figure 6c shows the reduction rate of the proposed and independent methods in relation to the total number of transmissions compared with that of the naive method. Since the independent method calculates the transmission probability for each data autonomously, there is no reduction effect on the data increase. On the other hand, since our proposed method aggregates multiple data prior to transmission, it is possible to reduce the total number of data transmissions as the data increases. Finally, Figure 6d shows the average number of data transmissions in the target area within the period $d_x$. This result shows how close the number of data transmissions of each data is to its target value $\beta$. From this result, we can see that the proposed method can approximate the number of transmissions of each data to the value of $\beta = 4$. That is, even though the number of data increases, the proposed method can effectively retain each data within the target area.

## V. Conclusion

In this study, as part of efforts to achieve a "local production and consumption" paradigm, we proposed an autonomous data transmission control method for multiple STDs that is based on node density retention. Through simulations, we clarified that our proposed method could satisfy the specific quality requirements of various STDs while maintaining effective data retention for multiple STDs. In the future, we plan to extend the proposed method to address other quality issues.

## Acknowledgement

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI via a Grant-in-Aid for Scientific Research (B) (No. 18H03234) and the research results have been achieved by the "Resilient Edge Cloud Designed Network (19304)," the Commissioned Research of National Institute of Information and Communications Technology (NICT), Japan.

## References

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

[2] H. Teshiba, et al., Adaptive Data Transmission Control for Reliable and Efficient Spatio-Temporal Data Retention by Vehicles, Proc. of The Sixteenth International Conference on Networks (ICN2017), Apr. 2017

[3] T. Higuchi, et al., Regional InfoHubs by Vehicles: Balancing Spatio-Temporal Coverage and Network Load, Proc. of First International Workshop on Internet of Vehicles nad Vehicles of Internet, pp. 25-30, Jul. 2016

[4] D. Nobayashi, et al., Spatio-temporal Data Retention System with MEC for Local Production and Consumption, Proc. of IEEE COMPSAC 2018, Jul. 2018.

[5] 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.

[6] 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.

[7] "OMNeT++," [Online]. Available from: https://www.omnetpp.org/ (accessed 2018.2.7).

[8] "SUMO," [Online]. Available from: http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/ (accessed 2018.2.7).

[9] "Veins," [Online]. Available from: http://veins.car2x.org/ (accessed 2018.2.7).