

A Channel Selection Algorithm Using Reinforcement Learning for Mobile Devices in Massive IoT System

Honami Furukawa*, Aohan Li*, Yozo Shoji[†], Yoshito Watanabe[†], Song-Ju Kim[‡],
Koya Sato*, Yiannis Andreopoulos[§] and Mikio Hasegawa*

*Department of Electrical Engineering, Tokyo University of Science, Tokyo, Japan

[†]Social-ICT System Laboratory,

National Institute of Information and Communications Technology, Tokyo, Japan

[‡]Graduate School of Media and Governance, Keio University, Fujisawa, Japan

[§]Department of Electronic and Electrical Engineering, University College London (UCL), London, UK

*h-furukawa@haselab.ee.kagu.tus.ac.jp, {aohanli, hasegawa}@ee.kagu.tus.ac.jp, k_sato@ieee.org

[†]{shoji, yoshito-watanabe}@nict.go.jp, [‡]songju@sfc.keio.ac.jp, [§]i.andreopoulos@ucl.ac.uk

I. INTRODUCTION

It is necessary to develop an efficient channel selection method with low power consumption to achieve high communication quality for distributed massive IoT system. To this end, Ma *et al.* [1] proposed an autonomous distributed channel selection method based on the Tug-of-War (ToW) dynamics. The ToW-based method can achieve equivalent performance to UCB1-tuned [2], [3] with low computational complexity and power consumption, which is recognized as a best practice technique for solving multi-armed bandit (MAB) problems. However, Ref. [1] only considered fixed IoT devices with simplex communication.

In this paper, We have extended the ToW-based channel selection to V2X duplex communications in a distributed Massive IoT system, which is called ToW-based channel selection algorithm (ToWCS). We evaluate the proposed ToWCS algorithm in terms of the average delivery time. Besides, we compare the proposed ToWCS algorithm with the existing methods, the ϵ -greedy algorithm and random frequency hopping algorithm.

II. SYSTEM MODEL

We consider a distributed massive IoT system with M IoT devices in area S . Various IoT devices operate in such a network, e.g., high-speed moving cars, low-speed moving pedestrians, and fixed IoT devices. IoT devices perform communication in a distributed manner. The spread of the information follows the store-carry-forward (SCF) mode that each IoT device repeats channel selection and data transmission to spread information to IoT devices in area S . To realize a high-quality communication services, we aim at avoiding channel congestion while reducing the average delivery time by dynamically selecting appropriate channels. Hence, we propose a ToWCS algorithm that can select channels with low computational complexity and power consumption in this paper. The number of channels is assumed to be K . At

each sensing period, each IoT device senses the status of the selected channel, which is used to update the access channel.

III. TOWCS ALGORITHM

In this section, we present the proposed ToWCS algorithm in detail. In the ToWCS algorithm, IoT devices select channels in a distributed manner. The policy that we use to select an access channel for each IoT device at the t th decision-making is expressed as

$$k^* = \operatorname{argmax}_{k \in K} Q_k(t), \quad (1)$$

where $Q_k(t)$ is the accumulated Q value for channel k at the t th decision-making, which can be expressed as

$$Q_k(t) = \sum_{i=1}^n Q_{i,k}(t), \quad (2)$$

where n is the number of sensing times of the selected channel at the t th decision-making. $Q_{i,k}(t)$ is used to estimate the reward for the i th sensing, which is expressed as follows,

$$Q_{i,k}(t) = \alpha Q_{i,k}(t-1) + \Delta Q_{i,k}(t), \quad (3)$$

where α ($0 \leq \alpha \leq 1$) is the forgetting factor for reward estimation. By adding α , the influence of past reward estimation can be well controlled, making it more applicable for the dynamic communication environment. $\Delta Q_{i,k}(t)$ is the increment of $Q_{i,k}(t)$, which can be expressed as follow,

$$\Delta Q_{i,k}(t) = \begin{cases} +1, & \text{when ACK returned,} \\ -\omega_i(t), & \text{otherwise.} \end{cases} \quad (4)$$

$\omega_i(t)$ is derived from (5),

$$\omega_i(t) = \frac{p_{i,1st}(t) + p_{i,2nd}(t)}{2 - \{p_{i,1st}(t) + p_{i,2nd}(t)\}}, \quad (5)$$

TABLE I
SIMULATION PARAMETERS

Parameters	Value
Area of massive IoT system S	2000 m×2000 m
Speed of the cars, the bicycles, and the pedestrians [m/s]	15, 5, 2
The number of IoT devices	2000
Number of channels K	3
Communication distance [m]	100

TABLE II
FOUR SCENARIOS

	Cars	Bicycles	Pedestrians	Fixed-devices
Scenario 1	0%	25%	25%	50%
Scenario 2	25%	25%	25%	25%
Scenario 3	50%	0%	0%	50%
Scenario 4	50%	25%	25%	0%

where $p_{i,1st}(t)$ and $p_{i,2nd}(t)$ are the two highest reward probabilities among that obtained from all channels with $p_{i,k}(t)$ expressed as

$$p_{i,k}(t) = \frac{R_{i,k}(t)}{N_{i,k}(t)}, \quad (6)$$

where $N_{i,k}(t)$ and $R_{i,k}(t)$ denote the cumulative number of trial that channel k was selected and the cumulative number of reward obtained until the i th sensing at the t th decision-making, respectively. $N_{i,k}(t)$ and $R_{i,k}(t)$ are given by

$$N_{i,k}(t) = \begin{cases} 1 + \beta N_{i,k}(t-1), & \text{when } k = k^*, \\ \beta N_{i,k}(t-1), & \text{otherwise,} \end{cases} \quad (7)$$

$$R_{i,k}(t) = \begin{cases} 1 + \beta R_{i,k}(t-1), & \text{when } k = k^* \text{ and ACK returned,} \\ \beta R_{i,k}(t-1), & \text{otherwise,} \end{cases} \quad (8)$$

where β ($0 \leq \beta \leq 1$) is a forgetting factor for the cumulative number of trial $N_{i,k}(t)$ and the cumulative number of reward $R_{i,k}(t)$ [4].

IV. SIMULATION

In this section, we evaluate the performance of the proposed ToWCS algorithm by the store-carry-forward routing simulation. We firstly compare the ToWCS algorithm with random frequency hopping algorithm and the ϵ -greedy algorithm in terms of the average delivery time. Then, we evaluate the average delivery time of the proposed ToWCS algorithm under different combinations of forgetting factors α and β . The parameters of the simulation are shown in Table I and the percentage of the number of each device is shown in Table II.

From Fig. 1, it can be seen that the ToWCS algorithm can reduce the average delivery time the most than random frequency hopping algorithm or the ϵ -greedy algorithm in both scenarios. In addition, Fig 2 shows that adding both forgetting factors α and β can further reduce the average delivery time.

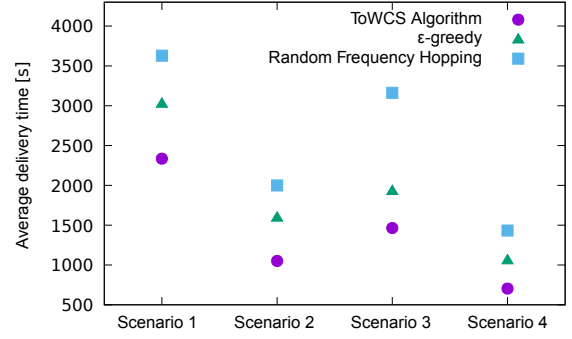


Fig. 1. Comparison of the ToWCS algorithm to the ϵ -greedy algorithm and random frequency hopping in terms of the average delivery time.

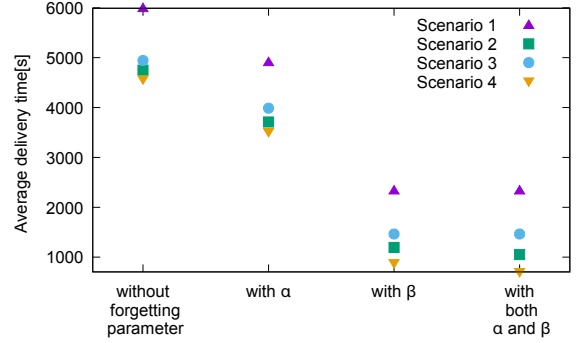


Fig. 2. Performance evaluation in terms of forgetting factors under four scenarios summarized in Table II.

V. CONCLUSION

In this paper, we evaluate the proposed ToWCS algorithm in terms of average delivery time. The results of V2X simulation with duplex communication show that the proposed ToWCS algorithm can reduce the average delivery time compared to other methods. Besides, the average delivery time can be further reduced by adding forgetting factors to ToWCS algorithm. Therefore, ToWCS with forgetting factors is considered to be an effective channel selection method in V2X with duplex communication.

REFERENCES

- [1] J. Ma, S. Hasegawa, S.-J. Kim, and M. Hasegawa, "A Reinforcement-Learning-Based Distributed Resource Selection Algorithm for Massive IoT," *Applied Sciences*, vol.9, no.18, 3730, 2019.
- [2] S. -J. Kim, M. Aono, and E. Nameda, "Efficient Decision-Making by Volume-conserving Physical Object," *New Journal of Physics*, vol.17, 083023, 2015.
- [3] S. -J. Kim, and M. Aono, "Amoeba-Inspired Algorithm for Cognitive Medium Access," *IEICE NOLTA*, vol.5, pp.198–209, 2014.
- [4] S. Hasegawa, R. Kitagawa, T. Ito, T. Nakajima, S. -J. Kim, Y. Shoji, M. Hasegawa, "Performance Evaluation of Machine Learning Based Channel Selection Algorithm Implemented on IoT Sensor Devices and Its Application to Wireless Sensor Network for Building Monitoring System," *ICAHC2020*, Fukuoka, Japan, 2020.