IoT Platform with Distributed Brokers on MQTT

Tetsuya Yokotani, Shuichi Ohno, Hiroaki Mukai, and Koichi Ishibashi

***Abstract*—Various wide-area Internet of Things services have been deployed. In most of these IoT services, a significant number of tiny data blocks are transferred across wide-area networks. Therefore, the transfer mechanisms should be simplified. One promising candidate for use as a transfer mechanism is MQ Telemetry Transport (MQTT). In this paper, an architecture for a distributed MQTT broker, referred to as a virtual ring approach, is proposed. This architecture complies with the IoT Data Exchange Platform, as discussed in ISO/IEC JTC 1/SC 41. The operations of this distributed broker architecture using a virtual ring network for real-time communication is also described, along with the superiority of the architecture based on a performance analysis using queuing models.**

***Index Terms*—IoT, IoT platform, data exchange platform, MQTT, standardization.**

I. Introduction

Communication technologies used by Internet of Things (IoT) services have been widely discussed [1]. When IoT services are deployed in a wide area, the network must support co-existence between IoT and any legacy services, and provide an efficient transfer of data for the IoT services. In most IoT services, a significant number of tiny data blocks from the sensors are transferred to the actuator across the network. Therefore, the framework for lightweight protocols with small overhead and simple communication sequences should be specified. For this purpose, the IoT Data Exchange Platform (IoT DEP) was proposed in ISO/IEC JTC1/SC41, which is an international standardization committee, and has been summarized in several articles [2]–[4].

In concepts of the IoT DEP, the networks overlaid among the interworking points for IoT services are specified. The end points, i.e., end devices and servers, access these interworking points using Information Centric Network (ICN) technologies [5]. These end and interworking points are implemented as a middleware module and are incorporated in conventional communication facilities through a socket interface.

In this paper, MQ Telemetry Transport (MQTT) [5] is assumed as the access protocol between an end point and an interworking point. Interworking points act as MQTT brokers. In this paper, operations among these brokers are proposed and evaluated using a queuing analysis. In addition, virtual ring topologies among these interworking points and cyclic communications using shared memories for real-time communication are proposed.

Manuscript received September 4, 2020; revised February 20, 2021.

Tetsuya Yokotani is with Department of Electrical and Electronics Engineering, College of Engineering, Kanazawa Institute of Technology, Nonoichi, Ishikawa, Japan (e-mail: yokotani@neptune.kanazawa-it.ac.jp).

Shuichi Ohno is with Graduate Program in Electrical Engineering and Electronics, Graduate School of Engineering, Kanazawa Institute of Technology, Nonoichi, Ishikawa, Japan (e-mail: b1629552@neptune.kanazawa-it.ac.jp).

Hiroaki Mukai and Koichi Ishibashi are with Department of Information and Computer Science, College of Engineering, Kanazawa Institute of Technology, Hakusan, Ishikawa, Japan (e-mail: mukai.hiroaki@neptune.kanazawa-it.ac.jp, k_ishibashi@neptune.kanazawa-it.ac.jp).

II. Summary of IoT DEP

IoT DEP was proposed in ISO/IEC JTC 1/SC 41 in 2018, and has been discussed as an international standard, i.e., ISO/IEC 30161, “Internet of Things (IoT) - Requirements of IoT data exchange platform for various IoT services,” the architecture of which is shown in Figure 1. End points, e.g., end devices and servers, access the edge of an IoT DEP network, which is an interworking point accommodating these end points using ICN technologies. IoT DEP networks are overlaid onto the Internet and are virtualized for IoT services. In addition, interworking points are associated with conventional communication facilities, e.g., IP routers. Communication between edges for IoT services is conducted through virtual paths among the interworking points.

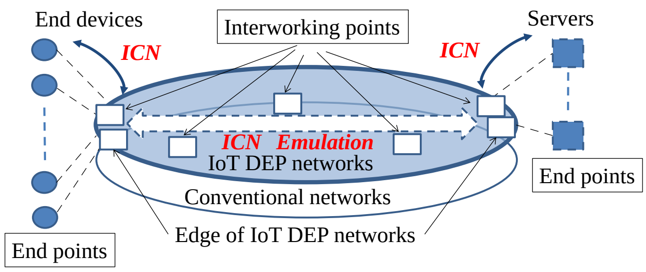

Fig. 1. The architecture of IoT DEP.

ICN technologies include various mechanisms. These mechanisms can be categorized into synchronous and asynchronous mechanisms [5]. In synchronous mechanisms, the request to obtain data and a response corresponding to this request are paired, as represented by a content-centric network (CCN) [6], [7]. In a CCN, a request corresponds to a packet of “interest,” and a response corresponds to a packet of “data.” By contrast, in an asynchronous mechanism, a request and a response are invoked independently, as represented by MQTT [8], [9]. In MQTT, data are provided by a “publish” packet, and are obtained by a “subscription” packet. These packets are invoked asynchronously.

Because ICN technologies do not require complicated communication sequences, e.g., IoT DEP provides lightweight access through such mechanisms as access sequences of a Domain Name System (DNS), three-way-handshake procedures of TCP, or a large protocol overhead, e.g., HTTP.

IoT DEP is implemented as a middleware module in each IoT termination point, e.g., end points and interworking

doi: 10.18178/ijfcc.2021.10.1.572

points, as shown in Fig. 2. Therefore, it acts as an application layer protocol through a socket interface.

Note that IoT DEP specifies the required framework for an efficient transfer of data for IoT services with a co-existence of legacy services and specifies communication between end points and an interworking point based on ICN technologies. However, it specifies that communication among interworking points depends on the implementation under conditional compliance with the specified requirements.

A mechanism of the detailed operations among interworking points based on the requirements specified in the IoT DEP is proposed herein.

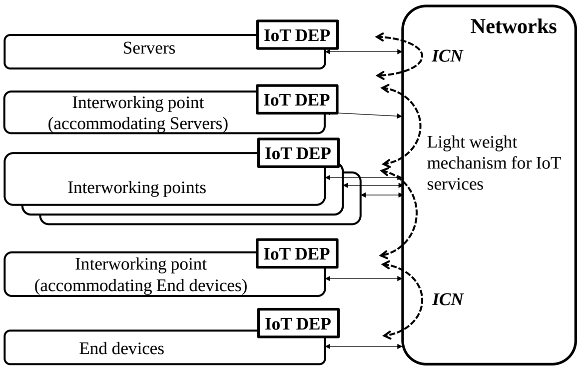

Fig. 2. Position of the IoT DEP associated with networks.

III. IoT Communication Using ICN Technologies

IoT communications are categorized into three types, as shown in Figure 3. End devices, e.g., various sensors, generate data and report to the servers with a notification, as shown in Case 1. The servers are invoked to obtain data and the end response required by the data according to the requests from the servers, as shown in Case 2. Finally, the servers invoke control to the end devices, e.g., actuators, as shown in Case 3.

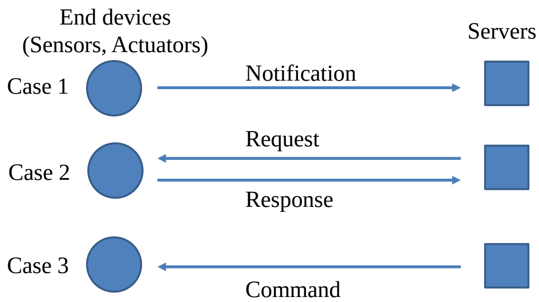

Fig. 3. Communication types among IoT end points.

In IoT services, most communication types are similar to Case 1 because a significant number of sensors should be installed to monitor various situations. Therefore, MQTT provides simpler communication sequences than the sequences of a CCN, because a CCN specifies sequences based on Case 2 [14]. Communication operations among the interworking points in IoT DEP based on MQTT are proposed in the following section.

IV. New Architecture for Larch-Scale Deployment

When IoT services are deployed across wide-area networks, many interworking points in the IoT DEP networks should cooperate with each other. In the case of MQTT between these points, a problem of cooperation among distributed MQTT brokers occurs. Various approaches have been discussed to solve this problem [10] – [13]. One solution is to broadcast communication among the brokers, which is referred to as a "flooding approach." However, with this approach, the traffic volume may be increased on the networks. Therefore, based on MQTT, a new architecture for large-scale deployment using IoT DEP, referred to as a virtual ring approach, is proposed in this section.

In this architecture, interworking points, as shown in Fig. 2, are connected as a logical ring, as shown in Fig. 3. This ring network is virtualized by lower layer protocols, e.g., VLAN. This architecture does not require specific routing protocols and differs from conventional ideas regarding the use of distributed brokers. As shown in Figure 4, the ring network is recognized by a VLAN. Interworking points, e.g., distributed brokers, includes access control and shared memory blocks. An access control block controls data on the ring, such as multiplexing, copying, and terminating. These operations are described in the next section. End points, e.g., end devices and servers, are connected to these interworking points according to the MQTT protocols. Data controlled by the MQTT protocol are referenced among the shared memory in a loop, as shown in Fig. 5. In this figure, two VLANs are provisioned. Each interworking point owns a VLAN, and specifies the initiation and termination points to avoid infinite looping.

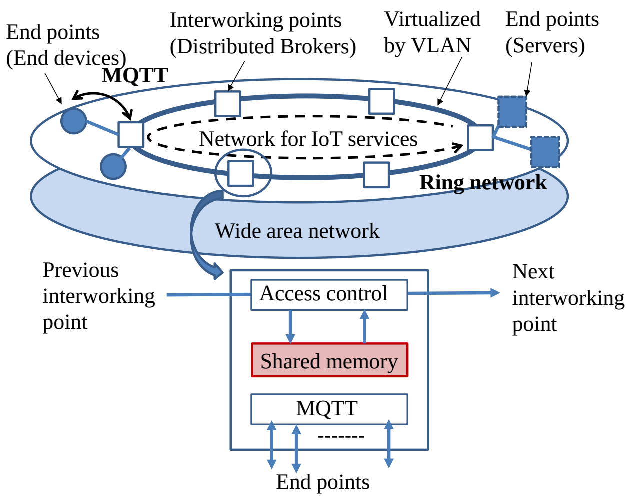

Fig. 4. Architecture of the proposed scheme for distributed brokers.

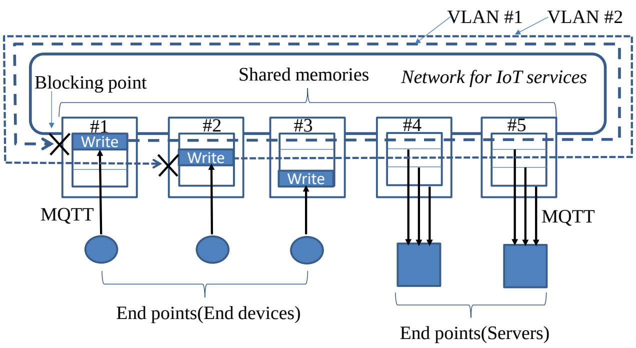

Fig. 5. Communication among shared memory in distributed brokers

In Fig. 5, end devices generate and transfer data to distributed Brokers according to the MQTT protocol. Data are stored in dedicated areas of the shared memory for each end device, and then transferred to other shared memory in distributed brokers in the ring. The transferred routes are identified using VLAN. In this figure, VLAN #1 is provisioned from Broker #1, and is blocked at the ingress point of this broker. By contrast, servers can refer to all of the areas in their shared memory.

V. Detailed Operations among Interworking Points

In the detailed operations among interworking points, distributed brokers of MQTT in the ring network are described as follows. These operations follow the architecture of communication using the shared memory, e.g., [15]. This architecture has been applied to real time communication of the industrial fields [16], [17].

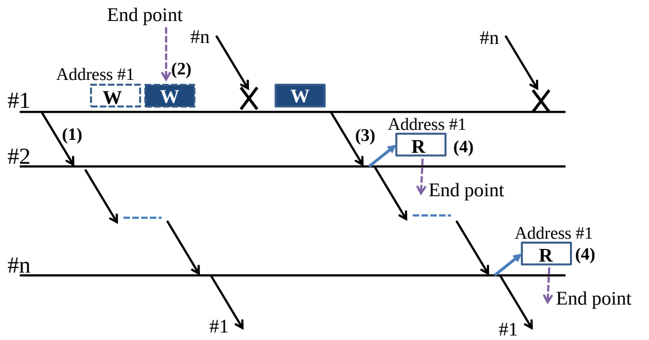

Fig. 6. The transfer mechanism among shared memory.

#1	Address #1
#2	Address #2
	:
#n	Address #n

Unified structure of shared memory

#1	#2		#n
W	R	·········	R
R	W	·········	R
:	:	:	:
R	R	·········	W

Write and Read areas for each Brokers

Fig. 7. The structure of the shared memory.

Each end point transfers information according to the MQTT protocols, as shown in Figure 4, to the shared memory in a distributed Broker, which accommodates this end point. The transfer mechanism among the shared memory is shown in Fig. 6. The structure of the shared memory is shown in Fig. 7.

The steps shown in Fig. 6 are as follows. In the ring network, frames are transferred at regular intervals among the distributed brokers (Step (1)). These frames are booked at the ingress point of the originating broker, i.e., Broker #1 in Fig. 6. When an end point generates information, this information is written in the dedicated address of the shard memory, shown in Figure 7 (Step (2)), The shared memory is divided into parts, which are identified based on the dedicated address for each broker, as shown in Fig. 7. These parts are categorized into a write or read area. This information is transferred by the next routed frame (Step (3)). This information is written in the read areas in other brokers. The end points accommodated by these brokers can read information stored in these areas (Step (4)).

These operations can update information in all parts in the shared memory within a fixed interval.

VI. Performance Evaluation

This section aims to clarify the performance of a new architecture, i.e., a virtual ring approach, and compare it with a conventional architecture, a flooding approach, using a queuing model.

The virtual ring and flooding approach can be modeled as the multiple queuing model, as shown in Figs. 8 and 9, respectively.

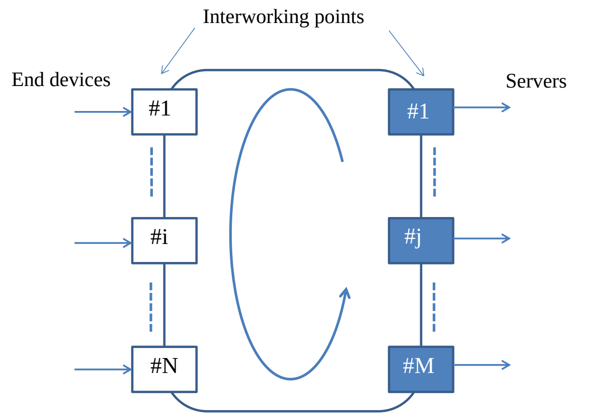

Fig. 8. The model on performance evaluation of the Virtual ring approach.

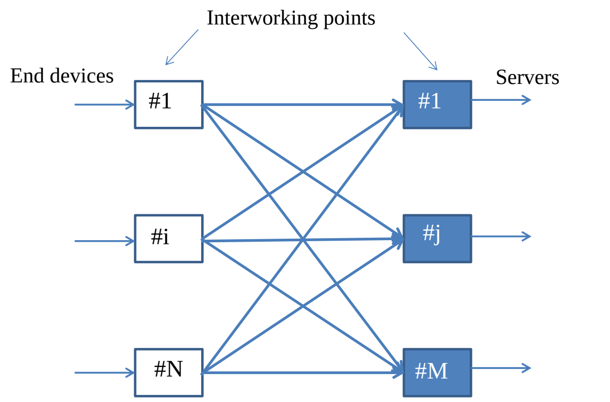

Fig. 9. The model on performance evaluation of the flooding approach.

In these figures, the numbers of interworking points that accommodate end devices and servers are denoted as M and N, respectively. This evaluation, shown in Figure, focuses on Case 1. Each interworking point accommodating the end devices receives data as packets generated randomly by the devices, the receiving rate of which is specified as follows.

$$\lambda_i \; (i = 1, \dots, N)$$

The average transmission time on a packet at this interworking point is as follows.

$$b_i \; (i = 1, \dots, N)$$

In the flooding approach, each transmission capacity between interworking points, which accommodate the end device and the server, divides the total capacity of the virtual ring approach into sizes of M × N.

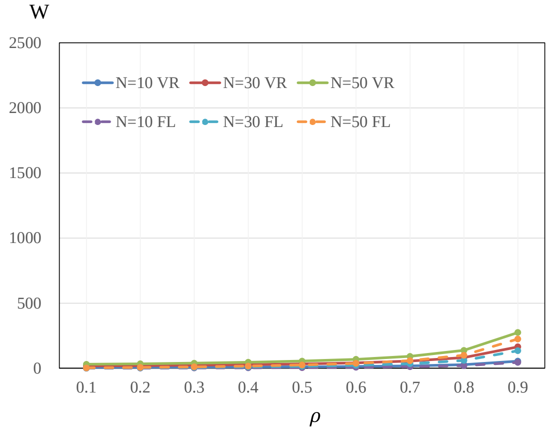

Fig. 10. The average delay in the small-scale case (M=1)

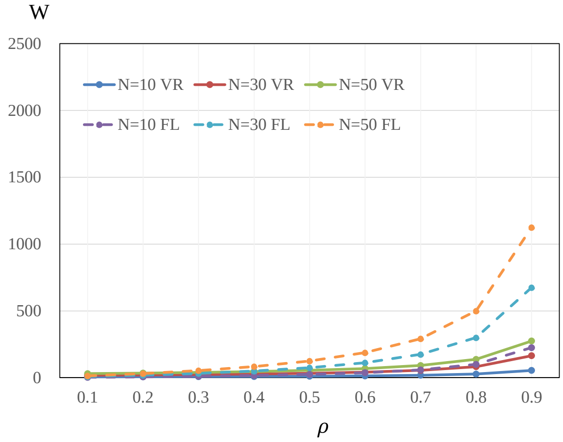

Fig. 11. The average delay in the small-scale case (M=5)

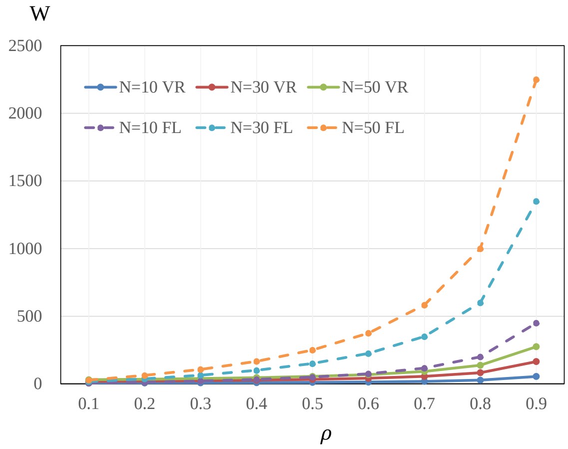

Fig. 12. The average delay in the small-scale case (M=10)

In this section, the average delays between these interworking points when applying these approaches are compared using queueing models.

In the virtual ring approach, because packets from an end device can be transferred through the interworking point accommodating this end device, when circulated frames arrive at this interworking point, models with multiple queue access can be applied as token passing mechanisms and polling systems [18]. In this approach, when a frame arrives at the interworking point accommodating the end devices, it is assumed that all information in this interworking point is transferred by this frame, which is referred to as an exhaustive policy. The average delay in a symmetrical case is derived from Eq. (1) [19].

$$W = \frac{\delta^2}{2r} + \frac{N\lambda b^{(2)} + r(N-\rho)}{2(1-\rho)} \tag{1}$$

The notations in Eq. (1) are specified as follows:

r: the average transmission delay per cycle
δ^2: the variance of transmission delay per cycle
λ: the arrival rate of packets at an interworking point from accommodated end devices
N: the number of the interworking point accommodating end devices on the ring
b: the average of transfer time of a packet, corresponding to the length of a transferred packet
$b^{(2)}$: the second moment of transfer time of a packet, corresponding to the length of a transferred packet
ρ: the utilization rate on a ring, corresponding to $N\lambda b$

In the flooding approach, the M/G/1 queueing model can be applied. According to the Pollaczek–Khintchine formula [20], the average transfer delay of packets in the symmetrical case is derived from Eq. (2) as follows:

$$W = \frac{r}{N} + \frac{NM\lambda b^{(2)}}{2(1-\rho)} \tag{2}$$

In Eq. (2), the transfer capacity between end points is divided into a capacity of $M \times N$ in the virtual ring approach. Because the communication is conducted using a direct route, the transmission delay per cycle, r, is divided into N.

Numerical examples are shown in the following to compare between the virtual ring approach (VR) and the Flooding approach (FL). The relative values of the average transfer delay are shown in these graphs for the case of b = 1. In these graphs, the length of the transferred information is fixed, and r is set to 0.1.

In small-scale cases, N is relatively small, as shown in Figs. 10, 11, and 12. By contrast, in large-scale cases, N is a relatively large number, as shown in Figs. 13, 14, and 15.

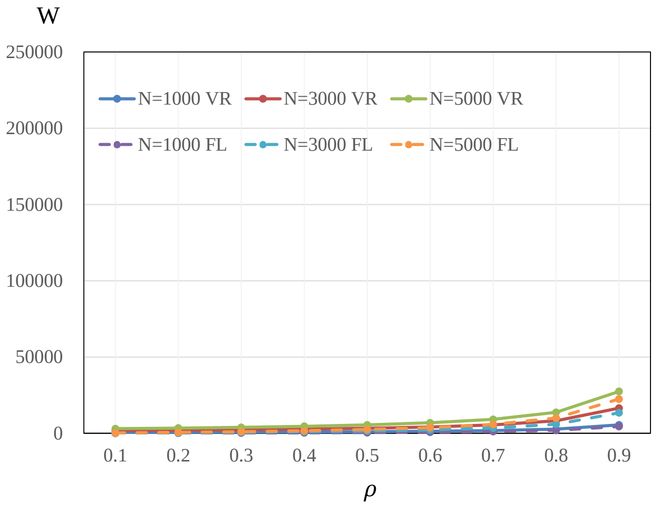

Fig. 13. The average delay in the large-scale case (M=1)

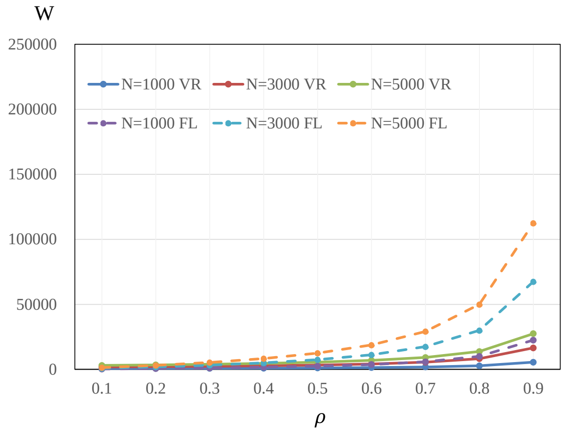

Fig. 14. The average delay in the large-scale case (M=5)

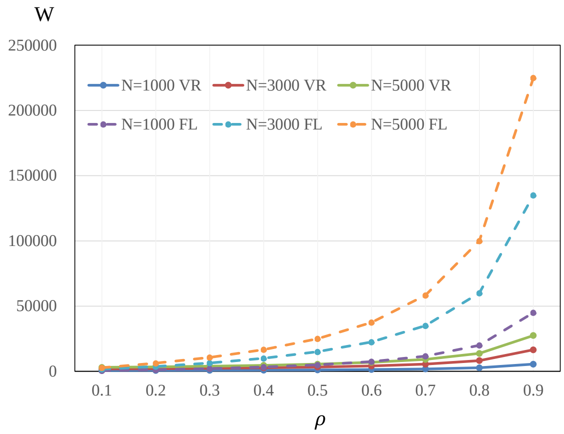

Fig. 15. The average delay in the large-scale case (M=10)

At both scales, the characteristics are mostly the same. As the number of interworking points accommodating servers is increased, the virtual ring approach is superior to the flooding approach. However, under a heavy load, the virtual ring approach is greatly superior to the flooding approach.

In general, for the deployment of IoT services over wide area networks, a significant number of end devices are provided. Moreover, multiple IoT services are overlaid in parallel. Therefore, it is assumed that M is a relatively large number.

As a result, it can be concluded that the virtual ring approach is particularly suitable to a large-scale deployment.

VII. Conclusions

In this paper, a framework of IoT DEP was introduced, which is a communication platform for various IoT services, and has been standardized in ISO/IEC JTC 1/SC 41 based on the authors' own promotion. In addition, detailed operations in the IoT DEP are proposed. Specifically, a virtual ring approach used to connect the interworking points in this platform was proposed and compared with a flooding approach, which is based on conventional technologies. It was then concluded that a virtual ring approach is superior to a flooding approach based on a queuing analysis.

As the next step, the virtual ring approach will be implemented as a prototype system.

Conflict of Interest

The authors declare no conflict of interest".

Author Contributions

Tetsuya Yokotani proposed the basic concept and detailed mechanisms in this research and described manuscript.

Shuichi Ohno performed numerical analysis.

Hiroaki Mukai and Koichi Ishibashi reviewed results and suggested modifications.

Acknowledgment

The research leading to these results has been supported by the EU-JAPAN initiative by the EC Horizon 2020 Work Programme (2018-2020) Grant Agreement No.814918 and Ministry of Internal Affairs and Communications "Federating IoT and cloud infrastructures to provide scalable and interoperable Smart Cities applications, by introducing novel IoT virtualization technologies (Fed4IoT)".

References

[1] T. Yokotani, "Network technologies for IoT," in *Proc. JAC-ECC 2018*, Tutorial paper, 2018.
[2] T. Yokotani, "IoT use cases analysis and possibility of adopting ICN technologies for these IoT use cases," in *Proc. 2018 WSCE*, pp. 7-12, 2018.
[3] T. Yokotani and K. Kawai, "Concepts and requirements of IoT networks using IoT Data Exchange Platform toward International standards," in *Proc. IEEE CSCN*, Technical Session 1, 2019.
[4] T. Yokotani and K. Kawai, "A proposal of IoT DEP as a platform for efficient IoT service deployment on wide area networks," in *Proc. JAC-ECC 2019*, IEICE Invited Paper 1, 2019.
[5] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *IEEE Communication Magagine*, vol. 50, no. 7, pp. 26-36, 2012.
[6] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "A survey of green information-centric networking: Research issues and challenges," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 3, pp. 1455-1472, 2015,
[7] A. Yokotani, S. Ohzahata, R. Yamamoto, and T. Kato, "A dynamic cache size assignment method with bandwidth reservation for CCN," in *Proc. ICOIN 2019*, pp. 2-15, 2019.
[8] Y. Sasaki and T. Yokotani, "MQTT over VLAN for reduction of overhead on information Discovery," in *Proc. ICOIN 2019*, pp. 1-14, 2019.
[9] MQTT: The Standard for IoT Messaging. (2012). [Online]. Available: http://mqtt.org
[10] R. Kawaguchi and M. Bandai, "Development of geographically distributed broker system in MQTT," in *Proc. IEIEC Society Conference 2018*, 2018.
[11] P. Chen and D. Li, "NCAF: A reduced flooding mechanism for route discovery in ICN," in *Proc. 9th International Conf. on Communication and Networking in China*, 2014.
[12] R. B. Masahiro and F. S. Takeuchi, "A proposal of distributed coordination proxy for scalable MQTT brokers," pp. 541-547, in *Proc. DICOMO 2016*, 2016.
[13] T. Rausch, S. Nastic, and S. Dustdar, "EMMA: Distributed QoS-aware MQTT middleware for edge computing applications," in *Proc. IEEE IC2E'18*, 2018.
[14] T. Yokotani and K. Kawai, "Concepts and requirements of IoT networks using IoT data exchange platform toward international standards," Technical Session 1, in *Proc. IEEE CSCN*, 2019.
[15] D. Buntinas, G. Mercier, and W. Gropp, "Implementation and evaluation of shared-memory communication and synchronization operations in MPICH2 using the Nemesis communication subsystem," *Parallel Computing*, vol. 33, no. 9, pp. 634-644, 2007.
[16] K. Oya, K. Tanaka, T. Sato, M. Yoshida, and N. Todoroki, "Communication method of ring network for industrial system," C-015, FIT 2018, 2018
[17] P. Danielis, J. Skodzik, V. Altmann, E. B. Schweissguth, F. Golatowski, D. Timmermann, and J. Schacht, "Survey on real-time communication via ethernet in industrial automation environment," PF-000167, IEEE ETFA 2014, 2014.
[18] H. Takagi, *Analysis of Polling Systems*, The MIT press, 1986.
[19] H. Takagi, "General report: Multiple queueing models with cyclic services," *Operations Reseach*, pp. 46-56, vol. 41, no. 2, 1996.
[20] L. Kleinrock, *Queueing System: Volume 1, Thory*, John Wiley & Sons, 1975.

Tetsuya Yokotani obtained his B.S., M. S, and the Ph. D degrees on information science from the Tokyo University of Science in 1985, 1987, and 1997, respectively. He joined Mitsubishi Electric Corporation in 1987. Since then, he has researched high-speed data communication, optical access and home network, and system performance evaluation based on queuing theory, and has promoted development and standardization of these related systems, in Information Technology R&D Center. In 2015, he moved to Kanazawa Institute of Technology as a professor of College of Engineering. He is a chair of Technical committee of CQR (Communication Quality and Reliability) in IEEE Communication Society. He is a fellow member of IEICE. He is a member of IEEE Communication Society and IPSJ.

Shuichi Ohno obtained his B.E. from Kanazawa Institute of Technology in 2020. Currently, he is in graduate program in electrical engineering and electronics, Graduate School of Engineering, Kanazawa Institute of Technology, His interests include communication protocols on IoT networks and traffic control on wide area IoT networks.

Hiroaki Mukai received his B.E. and M.S. degrees in electrical and electronic technology and his Ph.D. degree in engineering from Chiba University, Japan in 1988, 1990 and 2012 respectively. He joined Mitsubishi Electric Corporation, Japan in 1990. He was engaged in research and development of electrical communication systems from 1990 to 2017.Since 2017, he has been with Kanazawa Institute of Technology, where he is a professor. His current research interests include network security and network architectures for IoT

Koichi Ishibashi received the B.S. and M.S. degrees from Osaka City University in 1989 and 1991 and a Ph.D. degree in communications and integrated systems from Tokyo Institute of Technology in 2017, respectively. He joined the Mitsubishi Electric Corporation in 1991. Since then, he has been engaged in R&D of internetworking equipment, mobile networking, and ad hoc network systems. Moreover, in 2019, he moved to the Kanazawa Institute of Technology as an associate professor. His current research interests include routing technologies for wireless sensor networks and network architectures for IoT.