2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)

Data Management Portfolio for Improvement of Privacy in Fog-to-cloud Computing Systems

Prajak Chertchom
Faculty of Information Technology
Thai-Nichi Institute of Technology
Bangkok, Thailand
prajak@tni.ac.th

Shigeaki Tanimoto
Faculty of Social Systems Science
Chiba Institute of Technology
Chiba, Japan
shigeaki.tanimoto@it-chiba.ac.jp

Tsutomu Konosu
Faculty of Social Systems Science
Chiba Institute of Technology
Chiba, Japan
tklab@it-chiba.ac.jp

Motoi Iwashita
Faculty of Social Systems Science
Chiba Institute of Technology
Chiba, Japan
iwashita.motoi@it-chiba.ac.jp

Toru Kobayashi
School of Engineering
Nagasaki University
Nagasaki, Japan
kobayashi-toru@nagasaki-u.ac.jp

Hiroyuki Sato
Information Technology Center
The University of Tokyo
Tokyo, Japan
schuko@satolab.itc.u-tokyo.ac.jp

Atsushi Kanai
Faculty of Science and Engineering
Hosei University
Tokyo, Japan
yoikana@hosei.ac.jp

***Abstract*—With the challenge of the vast amount of data generated by devices at the edge of networks, new architecture needs a well-established data service model that accounts for privacy concerns. This paper presents an architecture of data transmission and a data portfolio with privacy for fog-to-cloud (DPPforF2C). We would like to propose a practical data model with privacy from a digitalized information perspective at fog nodes. In addition, we also propose an architecture for implicating the privacy of DPPforF2C used in fog computing. Technically, we design a data portfolio based on the Message Queuing Telemetry Transport (MQTT) and the Advanced Message Queuing Protocol (AMQP). We aim to propose sample data models with privacy architecture because there are some differences in the data obtained from IoT devices and sensors. Thus, we propose an architecture with the privacy of DPPforF2C for publishing data from edge devices to fog and to cloud servers that could be applied to fog architecture in the future.**

Keywords—Fog Architecture; Data Privacy for Fog-to-Cloud (DPPforF2C); Cloud Computing; IoT; Fog Computing; Data Management; Data Portfolio; Data Privacy

I. INTRODUCTION

The "internet of things" (IoT), which plays an important role in all aspects of our lives by connecting devices and people and machines to machines, is growing rapidly. According to Forbes, the IoT will have a market value of about $520 billion in 2021, more than double the $235 billion spent in 2017 [1].

Many IoT technologies such as those for smart sensing devices, wearable devices, smartphones, cars, cameras, smart buildings, gas pumps, shopping carts, airplanes, and smart agriculture can connect via the internet to use various services in the cloud. In addition, cloud services are currently provided by big-name companies such as Amazon, IBM, Google, and Microsoft [2, 3]. Figure 1 shows an overview of IoT devices; the diagram presents the connecting of many devices via a network, and their communication models now rely on the centralized, server/client paradigm to connect, authenticate, and authorize different nodes in a network.

The biggest challenges of IoT devices connecting to cloud services are that they are now relying on current communication models to authenticate, authorize, and connect to different nodes in a network.

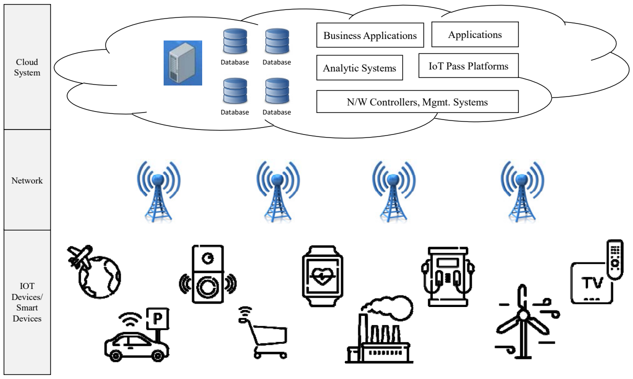

Fig 1. Overview of IoT devices and cloud

Furthermore, billions of devices join centralized network systems such as clouds. The problem now is a bottleneck for services dealing with an increasing number of smart devices and emerging applications in terms of latency because the centralized architecture of cloud services cannot smartly synthesize and manage computation, storage, and networking resources provisioned at the network edge [4]. Thus, "fog computing" is now promoted as a new architecture by the "OpenFog Consortium." Fog computing uses edge devices to communicate a substantial amount of

978-1-7281-2627-2/19/$31.00 ©2019 IEEE
DOI 10.1109/IIAI-AAI.2019.00179

distributed computation, communication, control, and storage closer to the end users and routed over the internet [5, 6].

Likewise, with regards to the challenges of security concerns, there is a massive amount of time-sensitive data, data transmission times, and latency times for destination devices in receiving a response from cloud services.

In this paper, we study and propose a set of data portfolios such as secured plane data and common/pricey data that can practically support and facilitate intermediaries in requests from IoT devices seeking resources from cloud servers.

The paper is organized as follows. Section 2 describes related work on "cloud computing and IoT," "communication protocols," and "fog computing and IoT" for understanding how data are managed and processed at the edge of the network and how they are routed through a central data center in the cloud. Section 3 proposes an architecture and framework for analyzing fog data transmission, data portfolios, and the filtering policies for the fundamental decisions of storing data through fog computing. This paper pays particular attention to common/privacy data of IoT as a large application domain over the fog architectural foundation.

II. Identifying Fog-to-Cloud Architecture

This section depicts primitives and prior work suggested by other researchers for handling distributed data storage for fog computing.

A. *Cloud computing and IoT*

Cloud computing is a familiar term that covers services based on the "data center" approach. Services are setup to enable us to rent computer system resources (computing power, database storage, applications, and other IT resources with pay-as-you-go pricing). The infrastructure of cloud computing is expensive, and operation requires many resources such as servers, redundant power, cooling systems, and backup batteries [7]. IoT uses cloud computing to enable better collaboration and the transmission and reception of data between devices and the cloud.

We can divide IoT devices into three types according to the usage of data transmission.

1. IoT devices that store and transmit data only.
2. IoT devices that wait to receive orders and work according to the order received.
3. IoT devices that do both.

Basically, with IoT and cloud computing models, the data (streaming or static, organized or unorganized) are collected from IoT devices and are then stored in a cloud to perform a lot of computation tasks. To sum it up, the greater the generation of data from IoT, the larger the infrastructures cloud systems require to deal with the huge amount of data and to perform real-time analysis.

B. *Fog computing and IoT*

By 2025, the International Data Corporation (IDC) predicts that 49% of the world's stored data (175 zettabytes) will reside in public cloud environments [8]. Thus, routing all the information through a centralized data center into the cloud has been a challenging problem in distributed systems. Whereas IoT devices do not have large storage resources to do advanced analytics and machine-learning tasks, they use cloud servers to do all these things. Fog computing was introduced by Cisco Systems in 2015 and is designed for distributed computation. It uses edge devices to ease the wireless data transfer by using an IoT network paradigm. In addition, fog computing extends cloud computing and services to the edge of the network by providing distributed computation, communication, control, and storage closer to the end users [9].

Furthermore, fog computing supports the IoT concept, in which most of the devices used by humans on a daily basis are connected to each other, and data are processed and responses are returned in a timely manner. Despite managing data and processing data at the edge of the network, fog computing has to overcome challenges such as data management, privacy, security, regulatory standards, and illegal implications [10]. Figure 2 shows an overview of the IoT-fog-cloud and describes how fog nodes are a new architecture that enables a quick response time, reduces the latency of the network and traffic, and supports the backbone bandwidth, leading to a better quality of service (QoS).

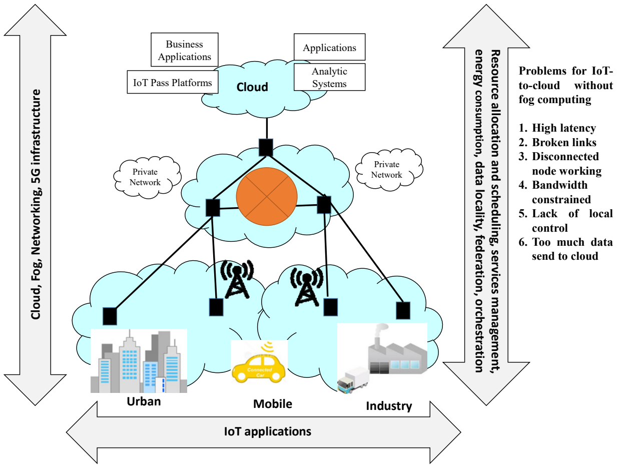
Business Applications
Applications
IoT Pass Platforms
Cloud
Analytic Systems
Private Network
Private Network
Cloud, Fog, Networking, 5G infrastructure
Resource allocation and scheduling, services management, energy consumption, data locality, federation, orchestration
Problems for IoT-to-cloud without fog computing
1. High latency
2. Broken links
3. Disconnected node working
4. Bandwidth constrained
5. Lack of local control
6. Too much data send to cloud
Connected Car
Urban
Mobile
Industry
IoT applications

Fig 2. Overview of IoT-cloud infrastructure without fog computing, and problems (adapted from [18])

From Figure 2 and Table 1, we can summarize that IoT devices depend on decentralized IoT networks and cloud systems. In summary, there are many problems as shown in Figure 2 such as high latency, broken links, disconnected nodes working, bandwidth constraints, lack of local control, and too much data being send to the cloud.

TABLE I [11] Fog Nodes Extend the Cloud to the Network Edge

	Fog Nodes Closest to IoT Devices	Fog Aggregation Nodes	Cloud
Response time	Milliseconds to sub seconds	Seconds to minutes	Minutes, days, weeks
Application examples	M2M communication Haptics2, including telemedicine and training	Visualization Simple analytics	Big data analytics Graphical dashboards
How long IoT data are stored	Transient	Short duration: perhaps hours, days, or weeks	Months or years
Geographic coverage	Very local: for example, one city block	Wider	Global

III. Analyzing Fog Data Portfolios and Filtering Policies for Privacy Protection

A. Data life cycle model

With the rapid growth of IoT devices, organizing the processing of information and data wherever it is appropriate at the edge of a network is challenging. Moreover, processing must be done as quickly as possible for IoT devices used in manufacturing industries where all machines are connected to a network and are required to react to incidents in a timely manner. For example, for oil and gas pipeline projects that generate terabytes of data every day, fog nodes were applied for computing data to bridge the gap between the cloud and IoT devices [12].

In this section, to investigate an appropriate data management method (i.e., a data portfolio, privacy for personal use and general use), we first find sources for data management challenges in fog-to-cloud systems from literature reviews and try to use a work break down structure (WBS) in defining the attribute service categories and their sub-coordinated attributes.

Sinaeepourfard et al. [13] proposed the Smart City Comprehensive Data LifeCycle (SCC-DLC) model for implementation in a smart city with fog-to-cloud (F2C) resource management as shown in Figure 3.

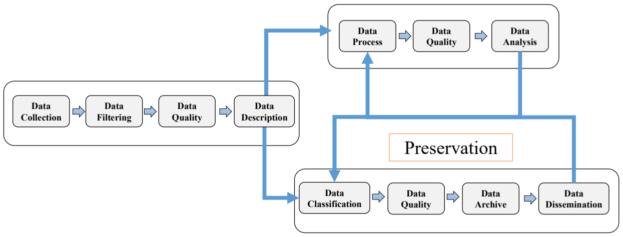

Fig 3. Smart City Comprehensive Data LifeCycle (SCC-DLC) model [13]

From Figure 3, as seen in the first block at fog nodes at the extreme edge, they collect, explore, and discover new data sources that may extend the available data scopes from devices and then filter data by applying methods for data optimization. In addition, the rest of the responsibilities for fog devices are checking data quality, checking timing information, and preserving and processing data.

In the previous network and device design paradigm, network, computation, and storage are separated. Leading companies such as Cisco, IBM, and Dell include these three parts in the fog layer to enable processing and forwarding completely within a single system as shown in Figure 4 [14]. This unified platform as shown in Figure 4 explains the paradigm shift in which cloud computing simply becomes decentralized to its edge networks layer.

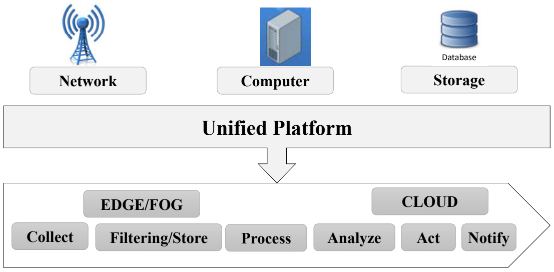

Fig 4. Paradigm shift with fog computing

B. MQTT, AMQP, and CoAP

Message Queuing Telemetry Transport (MQTT) and the Constrained Application Protocol (CoAP) are both designed to be used in lightweight environments. They are suitable for low-power and network-constrained devices. For IoT system architectures such as machine-to-machine networks, MQTT is the best protocol. In addition, MQTT provides publish/subscribe semantics (on the same socket) that help to program on the IoT device side. Moreover, MQTT is a good protocol for remote/cloud communication. CoAP provides functions for sending commands to IoT nodes; thus, for controlling IoT devices with a smartphone/web browser, CoAP is a good protocol [15]. The Advanced Message Queuing Protocol (AMQP) and MQTT are quite similar protocols in this context; however, AMQP offers more security when used in mobile communications over unstable network environments. AMQP is more reliable and scalable compared to MQTT when they are used with edge nodes and with low-speed wireless access [21].

In our study, the F2C environment suits very dynamic networks in which each node may be a different device such as a mobile, wearable, or sensor device with small capacity. Thus, when collecting data directly from physical devices, data have to be aggregated in pieces that reside in various nodes so that they can be accessed as needed by services running in nearby nodes. In summary, in fog computing, devices need a mechanism for defining the visibility of data as well as for replicating data to guarantee that nodes needing a piece of data are allowed to access it on a per object basis since not all pieces of data require the same visibility [16].

C. Data portfolio at fog node (DPPforF2C)

Using MQTT and CoAP for communicating between nodes, we also take privacy into consideration when sending data from nodes to nearby nodes and to the cloud. We require a logical flow of data management privacy on the basis of the 7 principles for F2C as shown in Figure 5.

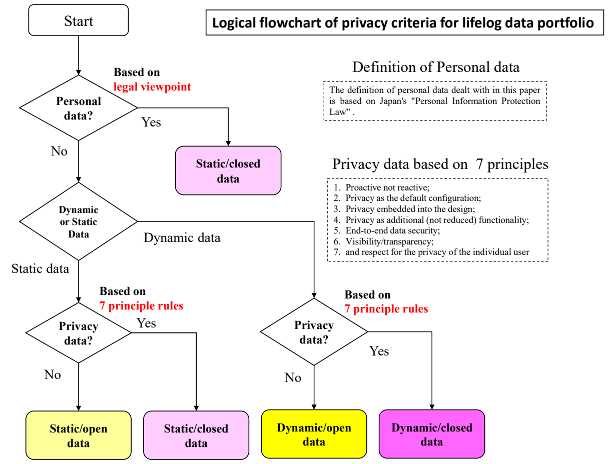
Start
Logical flowchart of privacy criteria for lifelog data portfolio
Based on
legal viewpoint
Personal data?
Yes
No
Static/closed data
Definition of Personal data
The definition of personal data dealt with in this paper is based on Japan's "Personal Information Protection Law" .
Privacy data based on 7 principles
1. Proactive not reactive;
2. Privacy as the default configuration;
3. Privacy embedded into the design;
4. Privacy as additional (not reduced) functionality;
5. End-to-end data security;
6. Visibility/transparency;
7. and respect for the privacy of the individual user
Dynamic or Static Data
Dynamic data
Static data
Based on
7 principle rules
Privacy data?
Yes
No
Based on
7 principle rules
Privacy data?
Yes
No
Static/open data
Static/closed data
Dynamic/open data
Dynamic/closed data

Fig 5. Logical flowchart of privacy criteria for data portfolio [17]

Thus, from the literature reviews in Section II and III, we use a WBS for systematic organization of data and divide the data portfolio categories for F2C into two main categories: common data and privacy data, as shown in Figure 6.

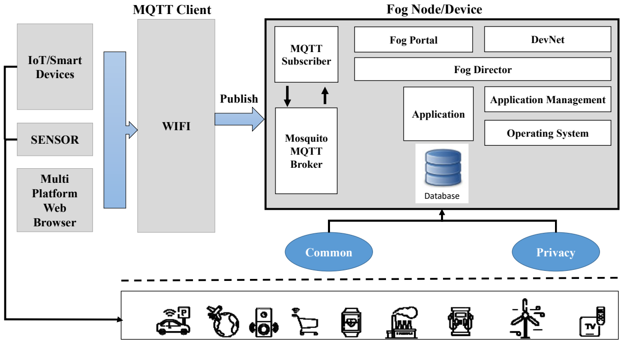

Fig 6. Proposed concept of two main categories of data portfolio for F2C

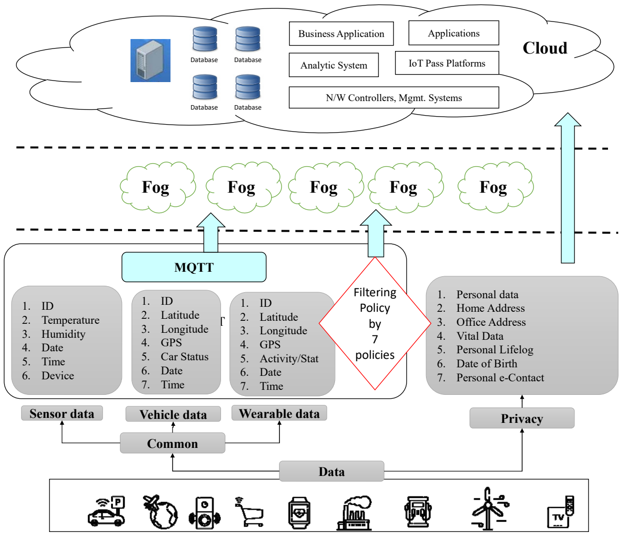

Fig 7. Proposed concept of data portfolio and privacy for F2C

Next, from Figure 7, we describe the data portfolio and privacy for fog-to-cloud (DPPforF2C). For setting the criteria of common data and privacy data, we used previous rules as guidance to classify which attributes should be common or private. From Figure 7, part of the common hierarchy consists of sample data attributes that should be transmitted with MQTT from edge devices to the fog. The common layer has 6–7 or more attributes based on particular devices and applications. The results of using appropriated data at the right time will help in making decisions, for example, in cases where medical emergency information is transmitted using IoT devices. While part of the privacy hierarchy consists of privacy data, for example, personal data, home addresses, office addresses, vital data, personal lifelogs, dates of birth, and personal e-contacts, these data should be stored at edge devices or sent to cloud services upon a privacy agreement between a service provider and users. For our work, it may be difficult to specify all common and private data in all jurisdictions with one proposed solution; however, we attempt to propose a framework and logical conditions for managing F2C data to protect IoT data while complying with basic privacy policies.

IV. Analysis of Privacy Protection Architecture for Fog-to-Cloud (DPPforF2C)

The value of F2C management architecture will be measured by the services it can support for privacy and security. Whether the F2C framework will be successful for business or not depends on a security and data privacy solution [19]. The architecture of the fog is a miniature version of cloud computing architecture. It differs from the cloud in that it can store data close to the "ground" and perform short-term analytics at the edge. Thus, many attacks such as denial of service attacks, malware injection, and authentication attacks can affect the fog. In addition, in some countries, for example, the USA and Thailand, there are still no national rules, regulations, or laws regarding the security and privacy of edge devices for the collection and use of personal data [20].

A. Proposed architecture 1 (Fog filtering)

The data from edge devices can be divided into time sensitive data, less time sensitive data, and data that are not time sensitive such as personal data. As shown in Figure 8, the data generated from IoT devices are almost used for actions in real time. Thus, time sensitive data (sensor data, vehicle data, machine data) should be sent (by MQTT) and processed at the nearest fog by applying the 7-rules policy, and the remaining data that are not time sensitive should be sent to an aggregate fog node for analysis and then sent to the cloud on the basis of the 7 rules and the agreement of the user.

B. Proposed architecture 2 (Edge filtering)

As shown in Figure 9, the 7-rules privacy logic may not be applicable at the fog layer because fog computations should be processed within a fraction of a second. Thus, in this case, the 7 privacy rules should be applied at edge devices before sending data. Moreover, when dealing with privacy data from privacy edge devices, we recommend AMQP for more reliable and advanced clustering messaging infrastructures over an ideal WLAN [21].

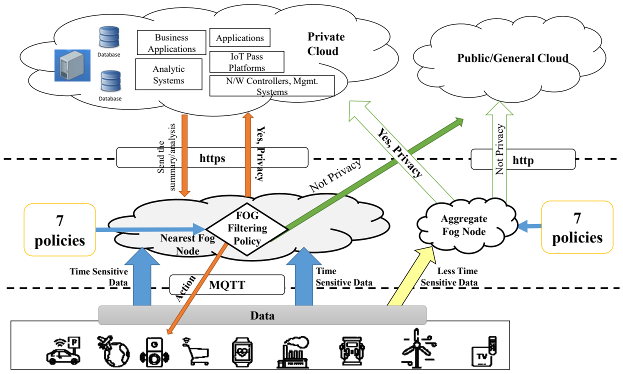

Fig 8. Fog filtering architecture of data privacy for F2C

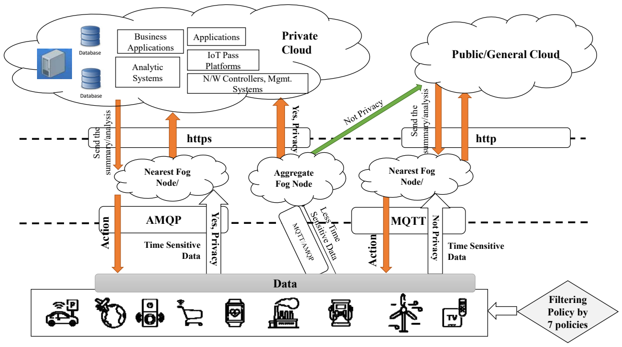

Fig 9. Edge filtering architecture of data privacy for F2C

C. *Proposed architecture 3 (Dedicated edge)*

As shown in Figure 10, in some cases, dedicated devices such as food quality sensors and customer-facing cameras that are built or programmed to do only specific procedures, we may use MQTT and apply the 7 privacy rules at the fog level because generally these devices do not contain private or personal data. Furthermore, some privacy devices such as surveillance cameras that are used to authenticate faces should be secured by using AMQP, and at the fog layer, data should be transmitted with a dedicated fog-to-cloud system.

In addition, from these three figures (8–10), all privacy data such as personal data should be analyzed locally at devices instead of being sent to the fog for analysis or should be send to a cloud service on the basis of user agreement. Furthermore, the fog architecture and its nodes must use the same design, control, and policy as the cloud system, or in other areas of an IT environment, a dedicated fog and secured protocol must be used in cases where data are private or highly secure as shown in Figure 10.

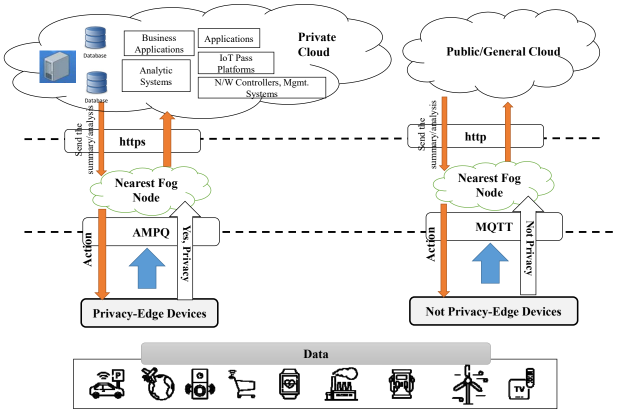

Fig 10. Dedicated edge architecture of data privacy for F2C

V. Discussion

Fog computing is a recent innovation in network computing for businesses regularly processing large amounts of data. A startup using fog infrastructure might have its server located near users and is ideal for businesses handling sensitive data. Likewise, fog computing is advantageous in that it can reduce data movement across the network and in that it consumes less bandwidth. All of these advantages result in reducing latency, congestion, cost, and bottlenecks resulting from cloud computing. However, there are many challenges at the national and organizational level for using fog computing for IoT systems. The future of using fog computing should take the following points into account.

1. Connectivity and security: when applying fog computing to businesses, the volume of data produced by IoT devices should be considered along with the dynamics of their applications and systems. Connections must be made securely and in real time. In addition, IoT data must be secured and protected.
2. Privacy: private fog computing may enable third parties and businesses to profit by tracking others through data sent from their smart devices. Thus, for better privacy, confidential data should be stored in private local servers or in a private fog system. For more security, only data that can be shared on the fog and cloud should be sent.
3. Intelligent Analysis & Actions: as fog and cloud computing are increasingly used in factories and smart cities, real-time analytics must be accurate for timely decision making and actions. In addition, fog nodes should be able to be programmed according to customer needs.

VI. Conclusion and Future Work

We proposed an architecture of data transmission and a data portfolio with privacy for fog-to-cloud (DPPforF2C). Specifically, we proposed a practical data model with privacy from a digitalized information perspective at fog nodes. In addition, we also proposed an architecture for implicating the privacy of DPPforF2C used in fog computing.

Technically, we designed a data portfolio based on MQTT and AMQP. We aimed to propose sample data models with privacy architecture because there were some differences in the data obtained from IoT devices and sensors. Thus, we proposed an architecture with the privacy of DPPforF2C for publishing data from edge devices to fog and to cloud servers that could be applied to fog architecture.

Our future work is to examine a framework for publishing data from fog to cloud servers. Future experiments will use real data from smart devices with algorithms for further evaluations on security and privacy. In addition, the final results will be used to classify a fog privacy framework for data organization challenges and data presentation challenges.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP 19H04098.

REFERENCES

[1] L. Columbus, "IoT Market Predicted To Double By 2021" [Online] Available: https://www.forbes.com/sites/louiscolumbus/2018/08/16/IoT-market-predicted-to-double-by-2021-reaching-520b/#6e6c76431f94. [Accessed: 15- Feb- 2019].

[2] S. Singh and N. Singh, "Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce," 2015 International Conference on Green Computing and Internet of Things (ICGCIOT), Noida, 2015, pp. 1577-1581. doi: 10.1109/ICGCIOT.2015.7380718

[3] A. Banafa, "Three Major Challenges Facing IoT" [Online] Available: https://IoT.ieee.org/newsletter/march-2017/three-major-challenges-facing-IoT.html. [Accessed: 18- Feb- 2019].

[4] J. Ni, K. Zhang, X. Lin and X. S. Shen, "Securing Fog Computing for Internet of Things Applications: Challenges and Solutions," in IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 601-628, Firstquarter 2018. doi: 10.1109/COMST.2017.2762345

[5] C. Puliafito, E. Mingozzi and G. Anastasi, "Fog Computing for the Internet of Mobile Things: Issues and Challenges," 2017 IEEE International Conference on Smart Computing (SMARTCOMP), Hong Kong, 2017, pp. 1-6. doi: 10.1109/SMARTCOMP.2017.7947010

[6] M. Chiang and T. Zhang, "Fog and IoT: An Overview of Research Opportunities," in *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854-864, Dec. 2016. doi: 10.1109/JIOT.2016.2584538

[7] T. Mengistu, A. Alahmadi, A. Albuali, Y. Alsenani and D. Che, "A "No Data Center" Solution to Cloud Computing," *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*, Honolulu, CA, 2017, pp. 714-717. doi: 10.1109/CLOUD.2017.99

[8] D. Reinsel, J. Gantz and J. Rydning, "The Digitization of the World From Edge to Core" [Online] Available: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf. [Accessed: 18-Feb- 2019].

[9] A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong and J.P. Jue, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," in Journal of Systems Architecture, 2019.

[10] A. Aljumah and T. A. Ahanger, "Fog computing and security issues: A review," *2018 7th International Conference on Computers Communications and Control (ICCCC)*, Oradea, 2018, pp. 237-239. doi: 10.1109/ICCCC.2018.8390464

[11] "Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are" [Online] Available: https://www.cisco.com/c/dam/en_us/solutions/trends/IoT/docs/computing-overvie w.pdf [Accessed: 20- Feb- 2019].

[12] P. Joshi, "The 4 Computing Types for the Internet of Things" [Online] Available: https://www.IoTforall.com/4-computing-types-for-IoT/ [Accessed: 20- Feb- 2019].

[13] A. Sinaeepourfard, J. Garcia, X. Masip-Bruin, E. Marin-Tordera, X. Yin and C. Wang, "A data lifeCycle model for smart cities," 2016 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2016, pp. 400-405. doi: 10.1109/ICTC.2016.7763506

[14] "Cisco IOx Data Sheet" [Online] Available: https://www.cisco.com/c/en/us/products/collateral/cloud-systems-management/iox/datasheet-c78-736767.html?dtid=osscdc000283 [Accessed: 20- Feb- 2019].

[15] G. C. Hillar, "MQTT Essentials-A Lightweight IoT Protocol," Packt Publishing Ltd, 2017, pp.17-20

[16] A. Salis, "Data management challenges in fog-to-cloud systems," [Online] Available: https://www.mf2c-project.eu/data-management-challenges-in-fog-to-cloud-systems/ [Accessed: 20- Feb- 2019].

[17] P. Chertchom, S. Tanimoto, H.Ohba, T. Kohnosu, T. Kobayashi, H. Sato and A. Kanai, "A Lifelog Data Portfolio for Privacy Protection Based on Dynamic Data Attributes in a Lifelog Service," In International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2017, pp. 107-120, Springer, Cham.

[18] L. Bittencourt, R. Immich, R. Sakellariou, N. Fonseca, E. Madeira, M. Curado, L. Villas, L. D. Silva, C. Lee, and Om.r Rana. "The internet of things, fog and cloud continuum: Integration and challenges," in Internet of Things, 2018

[19] X. Masip-Bruin, E. Marín-Tordera, G. Tashakor, A. Jukan and G. Ren, "Foggy clouds and cloudy fogs: a real need for coordinated management of fog-to-cloud computing systems," in *IEEE Wireless Communications*, vol. 23, no. 5, pp. 120-128, October 2016. doi: 10.1109/MWC.2016.7721750

[20] Roberta Rottigni, "Why IoT Data Protection Has Become More Important than Ever" [Online] Available: https://readwrite.com/2018/11/27/why-IoT-data-protection-has-become-more-important-than-ever/ [Accessed: 25- Feb- 2019].

[21] J. E. Luzuriaga, M. Perez, P. Boronat, J. C. Cano, C. Calafate and P. Manzoni, "A comparative evaluation of AMQP and MQTT protocols over unstable and mobile networks," 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, 2015, pp. 931-936.doi: 10.1109/CCNC.2015.7158101