Received February 23, 2019, accepted March 17, 2019, date of publication March 29, 2019, date of current version April 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2908223

Optimal Pricing and Service Selection in the Mobile Cloud Architectures

XIANWEI LI[1], CHENG ZHANG[2], BO GU[3], KYOKO YAMORI[4,5], AND YOSHIAKI TANAKA[5,6]

[1]School of Information Engineering, Suzhou University, Suzhou 234000, China
[2]Department of Computer Science and Communications Engineering, Waseda University, Tokyo 169-8555, Japan
[3]Department of Information and Communications Engineering, Kogakuin University, Tokyo 192-0015, Japan
[4]Department of Management Information, Asahi University, Mizuho-shi 501-0223, Japan
[5]Global Information and Telecommunication Institute, Waseda University, Tokyo 169-8555, Japan
[6]Department of Communications and Computer Engineering, Waseda University, Tokyo 169-0051, Japan

Corresponding author: Bo Gu (bo.gu@cc.kogakuin.ac.jp)

This work was supported in part by the Daze Scholar Project of Suzhou University under Grant 2018SZXYDZXZ01, in part by the National Science Foundation for Young Scientists of China under Grant 61702355, in part by the Key Projects of Natural Science Research in Anhui Colleges and Universities under Grant KJ2018A0448 and Grant KJ2018A0449, and in part by the Major Project of the Natural Science of Education Department of Anhui Province under Grant KJ2014ZD31. Xianwei Li left Waseda University, and is presently with the School of Information Engineering, Suzhou University, Suzhou, China. This work was conducted at Waseda University.

ABSTRACT With offloading the tasks that mobile users (MUs) running in their mobile devices (MDs) to the data centers of remote public clouds, mobile cloud computing (MCC) can greatly improve the computing capacity and prolong the battery life of MDs. However, the data centers of remote public cloud are generally far from the MUs, thus long delay will be caused due to the transmission from the base station to the public clouds over the Internet. Mobile edge computing (MEC) is recognized as a promising technique to augment the computation capabilities of MDs and shorten the transmission delay. Nevertheless, compared with the traditional MCC and MEC generally has a limited number of cloud resources. Therefore, making a choice on offloading task to the MCC or MEC is a challenging issue for each MU. In this paper, we investigate service selection in a mobile cloud architecture, in which MUs select cloud services from two cloud service providers (CSPs), i.e., public cloud service provider (PSP) and an edge cloud service provider (ESP). We use M/M/∞ queue and M/M/1 queue to model PSP and ESP, respectively. We analyze the interaction of the two CSPs and MUs by adopting Stackelberg game, in which PSP and ESP set the prices first, and then the MUs decide to select cloud services based on performances and prices. In particular, we study the relationship between PSP and ESP in the simultaneous-play game (SPG) scenario, in which they compete to set prices of their cloud services simultaneously. Our numerical results show that MUs prefer to select service from the edge cloud if the number of tasks they run is small. In another hand, more tasks will be offloaded to the remote public cloud if the number of tasks they run becomes large.

INDEX TERMS Pricing, mobile cloud computing, mobile edge computing.

I. INTRODUCTION

With the rapid growth of the technologies of wireless communications, mobile devices (MDs) have become a necessary part in our daily life. Cisco predicted that the number of mobile users (MUs) will be 5.2 billion in the year of 2019 and the total number of the worldwide MDs would be 75 billion by 2020 [1]. With the growing popularity of MDs, more mobile applications, like face recognition and natural language processing, have emerged and obtained a lot of attention [2]. However, it is a challenging issue for MDs to execute these mobile applications as both of their computing resources and battery lives are limited. Moreover, the gap between the limited capabilities of MDs and the demand to execute these resource-hungry applications is gradually increasing [3].

Mobile cloud computing (MCC) is recognized as a prospective technology to solve the aforementioned challenging issue by the way of offloading tasks to the data centers of the public clouds [4], such as Google Compute Engine and Amazon EC2. After the executions of tasks are finished in the public clouds, the final results will be returned to the MDs. Nevertheless, there is a long distance between the remote public clouds and the MUs. Therefore, long delay will be caused due to the transmission from

The associate editor coordinating the review of this manuscript and approving it for publication was Mianxiong Dong.

the base station to the public clouds over the Internet [5], which will have a bad impact on the MUs' experience and decrease the potential advantages of adopting the public cloud services [6].

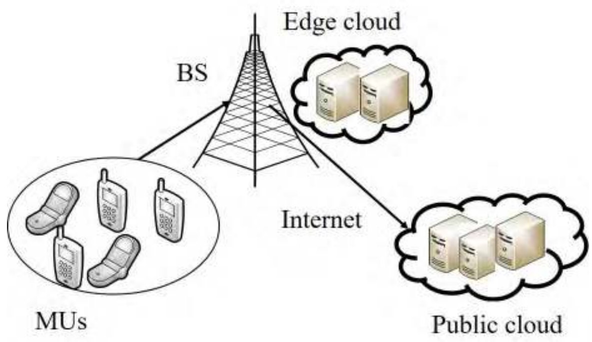


FIGURE 1. System architecture.

To eliminate these issues, mobile edge computing (MEC) has appeared and become an alternative solution to provide computing services for MDs [2], [7]–[9]. As illustrated in Fig.1, MEC can provision computing resources near to MUs such that not only the transmission delay can be significantly reduced but also the MDs' computation capabilities are enhanced and battery lives are prolonged [2], [10]–[12]. In MEC, the tasks of MDs can be offloaded to the edge cloud via cellular networks. With MEC, MUs can select services from the edge cloud or the remote public cloud for tasks processing. Nevertheless, as the edge cloud usually has less computing resources than the public cloud, the service rate of the edge cloud is slower than that of the public cloud [6], [7]. Therefore, it is challenging for each MU to make a choice as to which cloud service to select from the edge cloud and the remote public cloud to process its task.

Pricing is of great importance for resource management in public and mobile cloud markets; meanwhile, response time is considered as one of the critical performance indicators for *quality of service* (QoS) [13]. For the CSPs, how to price their cloud services is an important issue. If the prices are too high, some cloud users may be lost leading to the loss of their revenues. However, if the prices are too low, the CSPs may not get high revenues. In the meantime, guaranteeing the QoS of cloud services is of great importance, particularly for those cloud users when they are running delay-sensitive mobile applications like interaction gaming. The reason that the increased response time may have the possibility of preventing users from using cloud services or force them to choose cloud service from other CSPs. It is reported that Amazon may lose 1% in sales if every 100 ms of delay increases, and the traffic of Google may drop by 20% if 0.5 s increases in the search page generation time [14].

A great number of efforts have been denoted to studying resource management in the public or mobile cloud, such as [15], [16], and [17]. However, much of the existing work considered the public cloud service providers and mobile cloud service providers separately. To overcome this shortcoming, in our work, we investigate service selection in an mobile cloud architecture, in which one edge service provider (ESP) and one public service provider (PSP), both of whom are known as cloud service providers (CSPs), compete to provision services to MUs. Particularly, we investigate price competition between ESP and PSP whose objectives are to maximize their revenues by optimally determining the prices of their services. We model ESP by using an M/M/1 queue for the reason that this CSP has limited cloud resources, whereas we model PSP by an M/M/∞ queue as it has abundant cloud resources compared with ESP. Moreover, we study the simultaneous-play game (SPG) scenario that PSP and ESP determine the prices of their cloud services simultaneously. We analyze the interaction of the two CSPs and MUs by using the Stackelberg game, in which PSP and ESP determine the prices of their services first, and then MUs select cloud services to maximize their utilities according to the prices and performances. The backward induction method is employed to get the equilibrium of the game.

In summary, we make the following main contributions in this work.

- We investigate service selection in a mobile cloud architecture, in which MUs select cloud services from PSP and ESP. The objectives of the two CSPs are to optimally determine the prices of their services to maximize their revenues. Due to the reason that the two CSPs have different capacities of their cloud resources, we adopt M/M/∞ and M/M/1 queuing models to denote PSP and ESP, respectively.
- We consider the simultaneous-play game (SPG) competition scenario, in which PSP and ESP compete to simultaneously determine the prices of their cloud services.
- We conduct numerical results to evaluate our analysis. The analysis of our simulation results indicate that more MUs will offload their tasks to be processed in the edge cloud and the ESP can set higher prices if the number of tasks to be offloaded from MUs is small; however, more MUs will select the service of PSP with the number of tasks of MUs increasing. The numerical analysis also show that MUs prefer to choose the cloud services of ESP with the transmission delay of PSP increasing.

The rest part of this work is structured in the following. In the second section, we present a review and discussion of some related studies on resource management in cloud computing and mobile cloud computing. We introduce system models in the third section. We present the analysis of service selection in the mobile cloud market in Section IV. Section V presents the analysis of our numerical results to verify our theoretical analysis. Finally, we give the conclusions and some future research work.

II. RELATED WORK

We present a review and discussion of some work centered around resource management in public clouds and mobile clouds in this section.

In order to minimize energy consumption and delay cost of MDs, optimal resource management for offloading tasks of MUs has received a lot of research interests in mobile cloud architectures. In work [18], resource allocation for task offloading was explored in MEC. The authors tried to minimize the total energy consumed by the MDs by optimally allocating computing resources. In [19], the computation offloading in an MEC system with multiple MUs was investigated by Zhao *et al.*. They also tried to minimize energy consumption of MDs by optimizing both of radio and computational resources allocations. In [20], Chen and Hao investigated the problem of multiuser task offloading in MEC with the software defined unltra-dense network. Their goal is to optimally allocate computing resources in the edge cloud to minimize the delay costs, and meanwhile prolonging the battery life of MDs. In [5], the authors investigated communication offloading in MEC systems where MDs can harvest energy. In [21], Li *et al.* proposed a game theoretical approach for power control in an MEC system with multiple MUs by considering interference factor. In [22], Fang *et al.* studied the profit maximization problem in a mobile cloudlet platform. They proposed an online control algorithm by leveraging the Lyapunov optimization technique to the cope with the time-varying arrival rates of tasks from MUts. In [23], the authors aimed to improve the efficiency of radio transmission in an MEC enabled by non-orthogonal multiple access (NOMA). Nevertheless, these studies mainly study from the perspective of MUs without considering the pricing factor which is of great importance for ESPs.

Pricing is critically important for resource management in the public cloud and mobile cloud. A great number of research interests have been focused on the design of optimal prices in the literature. In [24], the authors introduced a new pricing mechanism to maximize the PSP's profit in a cloud cache. A pricing mechanism for resource management in a public cloud is proposed in [25] with the goal to maximize the PSP's profit. In [26], Fang and Li studied how to design the optimal price schemes in a monopoly public cloud market where only one CSP exists. In [16], the authors developed an framework to set price of the mobile cloud services, and analyzed the incentives for offloading CSPs to offload the tasks of MUs. They modeled the pricing scheme as a Stackelberg game, in which the CSPs are the followers and the MUs are the leaders. In [17], the authors studied the designing of optimal pricing strategy for task offloading in the MCC systems. Their objectives are to make the profits of the public CSP maximized while minimizing the energy consumption and delay costs of MUs. However, these aforementioned studies investigated the resource allocation of PSPs and ESPs separately, and they did not consider the competition between the PSP and ESP.

Competition of service providers has been extensively studied in cloud computing and wireless networks. In [27], the authors analyzed competition in a wireless network market consists of two network service providers competing to set the optimal prices of their network services in different time slots for maximizing their revenues. In [28], the authors studied competition of two service providers in an femtocell communication market where they compete to set the prices of the network services simultaneous for revenue maximization. In particular, designing the optimal pricing mechanisms to maximize the revenues of CSPs is especially important in a cloud market where the competition of several CSPs exists. Competition and cooperation of CSPs are studied by the work of [29], where the authors proposed a novel model. In [30], Feng *et al.* explored the problem of service provision in a public cloud market existing the competition of multiple CSPs. The authors analyzed how these CSPs should price their services such that their revenues can be maximized.

However, these previous studies considered either the public or mobile cloud ignoring the competition between PSP and ESP. With more MUs beginning to process their computation-intensive applications in the public cloud and edge cloud, there is no denying the fact that these MUs should make a choice as to which cloud services they choose. There are only few studies that considered service selection from the edge cloud and public cloud. In [31], Chen *et al.* considered a general MCC system consisting of an edge cloud and a remote public cloud. Their objective is to minimize the total costs of all MUs by optimally allocating the communication resources and finding the optimal offloading decisions of all the tasks. But how to select cloud services from the edge cloud and the remote public cloud is not fully studied. Especially, the pricing factor is not analyzed in the mobile cloud market where PSP and ESP compte to provide cloud services. Although Zhao *et al.* in [32] studied the price competition between the ESP and CSP, several problems remain to be further analyzed, such as how the arrival rate of task impact on the prices of the cloud services, and how the service rates of the two CSPs affect the choices of the MUs. Besides, the objective of [32] is different from our work.

III. SYSTEM MODELS

Consider a mobile cloud architecture where MUs can offload tasks for execution by selecting services from the edge cloud near to the Base Station (BS) or the public cloud, as depicted by Fig.1. The edge cloud and the remote public cloud belong to two different CSPs, respectively denoted by ESP and PSP. ESP can be network service providers, such as NTT Docomo in Japan or the China Mobile in China, and the PSP can be major cloud companies like Amazon or Google. The ESP and PSP compete to deliver cloud services to a number of MUs. We assume that the MUs submit their tasks with arrival rate λ (tasks/second) according to the Poisson process [30], [33]. We assume that each MU owns one MD, and each MD has one task to be offloaded. Each task can be an application (e.g. face recognition) or a function (e.g. an image compression). Hence, we use MUs and tasks interchangeably in the paper. The tasks from MUs are assumed to have high requirements for computing resources and be sensitive to delay, such as the face recognition applications. We model PSP by using the M/M/∞ queue reflecting its adequate cloud resources, whereas we model ESP as an M/M/1 queue as it has constrained cloud capacities [7], [33], [34]. For MUs, each of them will select service from the CSP or ESP according to the prices and performances to process their tasks. Let λ_p and λ_e denote the respective arrival rate of tasks at PSP and ESP, so that $\lambda_p + \lambda_e = \lambda$.

A. PSP'S MODEL

Queueing models have been widely adopted in the cloud computing systems [35], [36]. In comparison with ESP, PSP has abundant cloud resources to process the tasks of MUs [6], [32]. Therefore, its system can be modeled as an M/M/∞ queue [6], [32]. Similar assumption is widely adopted in many existing studies. For example, in [33], cloud broker is modeled by using an M/M/∞ queue, and the work of [37] modeled a system with multiple servers as an M/G/∞ queue. Denote α as the delay cost per unit time and μ_p (in the number of processed tasks per second) as the service rate of PSP. Different values of α reflect different types of applications and their sensitivity to delay [17]. A higher value of α means that the application has higher sensitivity to delay. When offloading tasks to the remote public cloud, extra transmission delay will be incurred. We assume that the transmission time from the BS to the public cloud is a fixed value, denoted by d_t [34]. For a MU, if it selects cloud services from the PSP, the total costs can be expressed as

$$C_p = \frac{\alpha}{\mu_p} + d_t + p_p \tag{1}$$

where $\frac{1}{\mu_p}$ is the mean delay that MUs experienced in M/M/∞ queue including processing delay and the time waiting in this queue, and p_p is the subscription price (in \$ per task) set by PSP.

Accordingly, the revenue of PSP is given as [30], [33], [38]

$$\pi_p = p_p \lambda_p \tag{2}$$

B. ESP'S MODEL

Compared with PSP, ESP has constrained cloud capacity; therefore, ESP has an M/M/1 queue of cloud resources [7], [34]. Let μ_e (in the number of processed tasks per second) denote the service rate of ESP. For an MU, if it selects cloud service from the ESP, the total costs that this MU will experience can be expressed as

$$C_e = \frac{\alpha}{\mu_e - \lambda_e} + p_e \tag{3}$$

where $\frac{1}{\mu_e - \lambda_e}$ is the mean delay in M/M/1 queue including processing delay and the time waiting, and p_e is the subscription price of ESP (in \$ per task).

Accordingly, the revenue of ESP is given as [30], [33], [38]

$$\pi_e = p_e \lambda_e \tag{4}$$

For convenient analysis, we summarize some main notations in Table 1.

Remark: How to model cloud computing platforms still remains in discussion. The system models in a lot of existing work have some gaps and constraints in comparison with the real cloud computing systems. Numerous studies have realized this point and discussed about it, such as [6], [15], [30], [33]. When selecting a proper system model, the ease to analyze and the convenience of getting closed form solutions should be considered. The work of [15] and [33] also modeled a CSP with limited cloud resources as an M/M/1 queue. In [30], the authors argued that modeling a cloud system by using an M/M/1 queue is enough. In [6], [32] and [34], the authors modeled the remote public cloud computing system as an M/M/∞ queue, as it has infinite cloud computing resources compared with the edge cloud system.

TABLE 1. Notations summary.

Notation	Description
λ	the total arrival rate of the tasks
λ_p	the arrival rate of the tasks at PSP
λ_e	the arrival rate of the tasks at ESP
μ_p	the service rate of PSP
μ_e	the service rate of ESP
α	the delay cost per unit time
d_t	the transmission delay of PSP
p_p	the price of the PSP
p_e	the price of the ESP
p_p^n	the price of PSP in equilibrium
p_e^n	the price of ESP in equilibrium
λ_p^e	the equilibrium arrival rate of PSP
λ_e^e	the equilibrium arrival rate of ESP
λ_p^n	the equilibrium arrival rate of PSP
λ_e^n	the equilibrium arrival rate of ESP
π_p	the revenue of PSP
π_e	the revenue of ESP
π_p^n	the revenue of PSP in equilibrium
π_e^n	the revenue of ESP in equilibrium

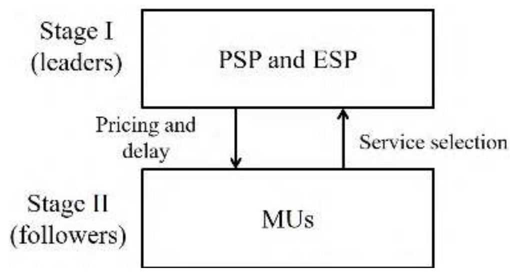


FIGURE 2. The Stackelberg game in mobile cloud architecture.

IV. SERVICE SELECTION IN THE MOBILE CLOUD ARCHITECTURE

Consider a mobile cloud architecture consists of PSP and ESP provisioning cloud services to a number of MUs. The MUs pay to select cloud services from the two CSPs to execute their tasks in the edge cloud or remote public cloud. The objectives of PSP and ESP are to maximize their revenues by setting the optimal price for its cloud service. We study the interaction of the two CSPs and MUs by adopting the Stackelberg game [39], in this context, PSP and ESP compete to provision their services with optimal prices first, and then MUs make decisions to select cloud services based on performances and prices of the two CSPs in stage two, as illustrated by Fig.2. This game can be solved by resorting to the backward induction method [40]. In particular, we consider the relationship of PSP and ESP in the simultaneous-play game (SPG) scenario, where they simultaneously start to provision cloud services with different performances and prices [41].

It is important to note that two competition scenarios exist, i.e., static and dynamic scenarios. In this paper, our study only considers the static scenario, and we leave the dynamic scenario as a future research direction. There are some related studies on dynamic scenario. For example, the work of [33] analyzed the dynamic scenario by using the evolutionary game. In [27], Zhang *et al.* investigated dynamic price competition in the duopoly network market, where two service providers compete to offer network services with different prices in different time slots to maximize their revenues.

A. WARDROP EQUILIBRIUM

In the cloud service selection game, a Wardrop equilibrium [33], [42], [43] is reached, as a large number of MUs must individually make a determination on which cloud service they will choose. This equilibrium meets with the following two Wardrop's principles [33], [42], [43]: the MUs experience the equal total costs on all the used cloud services, that is, the mean delay and cost are minimized.

Therefore, according to the Wardrop equilibrium, all the MUs will experience the same total costs, i.e., $C_1 = C_2$. Accordingly, we have

$$\frac{\alpha}{\mu_e} + d_t + p_p = \frac{\alpha}{\mu_e - \lambda_e^e} + p_e \tag{5}$$

where λ_e^e is the number of tasks arrived at the ESP in equilibrium.

From Eq.(5), we get

$$\lambda_e^e = \mu_e - \frac{\alpha}{p_p - p_e + d} \tag{6}$$

Accordingly, the number of tasks arrived at the PSP in equilibrium is

$$\lambda_p^e = \lambda - \mu_e + \frac{\alpha}{p_p - p_e + d} \tag{7}$$

where $d = \frac{\alpha}{\mu_p} + d_t$.

According to the arrival rates of tasks in equilibrium at PSP and ESP in the Eqs.(6) and (7), respectively, PSP and ESP compete to maximize their revenues by setting optimal prices of their services. Therefore, we can formulate the following game:

- **Players**: PSP and ESP,
- **Strategies**: PSP sets price p_p, and ESP sets price p_e,
- **Payoff**: the revenues of the two CSPs: $\pi_p^e = p_p\lambda_p^e$ and $\pi_e^e = p_e\lambda_e^e$.

The corresponding revenues of the PSP and ESP at the equilibrium are respectively expressed as

$$\begin{aligned}\pi_p^e &= p_p\lambda_p^e \\ &= p_p\left[\lambda - \mu_e + \frac{\alpha}{p_p - p_e + d}\right]\end{aligned} \tag{8}$$

$$\begin{aligned}\pi_e^e &= p_e\lambda_e^e \\ &= p_e\left[\mu_e - \frac{\alpha}{p_p - p_e + d}\right]\end{aligned} \tag{9}$$

B. SIMULTANEOUS-PLAY GAME (SPG)

We consider the simultaneous-play game (SPG) competition scenario, in which PSP and ESP simultaneously determine the prices of their services with the goal of maximizing their revenues [15], [43]. For PSP, we have the following revenue optimization problem:

Problem1:

$$\begin{aligned}&\max_{p_p} p_p\lambda_p^e \\ &\text{s.t.} \quad p_p \geq 0\end{aligned} \tag{10}$$

where λ_1^e is denoted by Eq.(7).

The revenue optimization problem of ESP can be formulated as:

Problem2:

$$\begin{aligned}&\max_{p_e} p_e\lambda_e^e \\ &\text{s.t.} \quad p_e \geq 0\end{aligned} \tag{11}$$

where λ_e^e is given in Eq.(6).

From solving **Problem1** and **Problem2** simultaneously, we can get **Proposition 1**, which is proved in the **Appendix**.

Proposition 1: A unique price pair (p_p^n, p_e^n) exists in the Equilibrium.

According to the **Proposition 1**, we can obtain the following **Corollary 1**.

Corollary 1: In the Equilibrium, PSP and ESP respectively obtain the following revenues:

$$\pi_p^n = p_p^n\lambda_p^n \tag{12}$$

$$\pi_e^n = p_e^n\lambda_e^n \tag{13}$$

V. PERFORMANCE EVALUATION

In this section, numerical results are conducted to evaluate the theoretical analysis of this paper. We present an analysis of the two CSPs' revenues, prices and MUs' arrival rates in equilibrium, to different parameters, such as the total task arrival rate of MUs, the service rates of the two CSPs and the transmission delay of PSP.

A. PARAMETER SETTING

Unless otherwise specified, we set the default values of parameters of the mobile cloud architecture as follows: service rate $\mu_p = 20$ tasks/s, $\mu_e = 20$ tasks/s, total tasks arrival rate $\lambda = 30$ tasks/s, the transmission delay of PSP $d_t = 0.4$ s and $\alpha = 1$. We set the parameters values by referring to [6], [34], in which these values are in line with they are experimentally measured [6]. We emphasize that the selected values of these parameters are just for demonstration. Other values of parameters also have the similar results, which means that the numerical analysis of this paper is not limited by the selected values of these parameters.

B. IMPACT OF TASK ARRIVAL RATE

We first present an analysis of the impact of arrival rate of tasks on PSP and ESP in equilibrium. Fig.3 shows the prices of the PSP and ESP in equilibrium versus the arrival rate of MUs' tasks with λ varying in the range [20, 30]. Fig.4 shows the prices of the PSP and ESP in equilibrium versus the arrival rate of MUs' tasks with λ varying in the range [30, 40). Fig.3 and Fig.4 show that ESP can set higher equilibrium prices than PSP under the condition that the number of being offloaded tasks is small; however, PSP can set higher equilibrium prices than ESP with the increase of the number of to be offloaded tasks. Fig.3 and Fig.4 indicate that the equilibrium prices of PSP and ESP increase rapidly with the increase of the number of MUs' tasks.

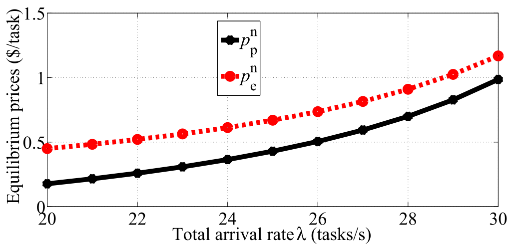


FIGURE 3. Equilibrium prices of PSP and ESP versus total task arrival rate varying in [20, 30].

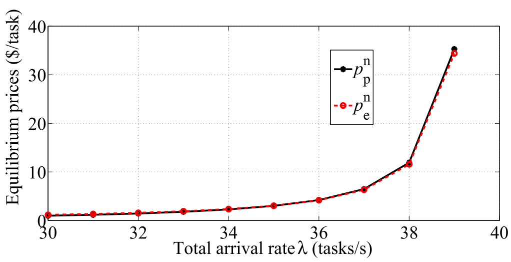


FIGURE 4. Equilibrium prices of PSP and ESP versus total task arrival rate varying in [30, 40).

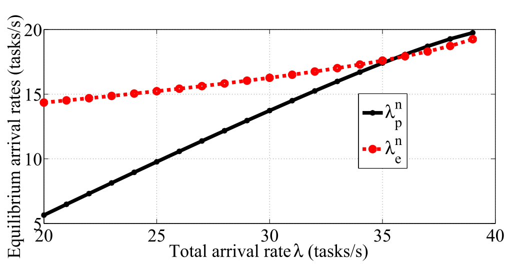


FIGURE 5. Equilibrium arrival rates of tasks of PSP and ESP versus total task arrival rate.

Fig.5 shows how the arrival rates of offloaded tasks at the PSP and ESP in equilibrium versus the arrival rate of MUs' tasks with λ varying in the range [20, 40). From this figure, we can observe that more MUs tend to select cloud services from the ESP if the number of tasks of MUs is small; however, more MUs will select cloud services from the PSP when the number of tasks becomes large. This figure demonstrates that MUs prefer to make use of the service from the public cloud to execute their tasks if the number of tasks that they run is large, even if the transmission delay of the public cloud is long. This is due to the reason that MUs will contend to use cloud services with the increase of the number of their tasks.

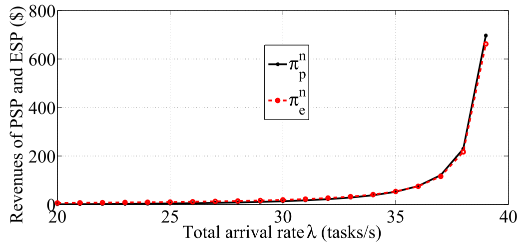


FIGURE 6. Comparison of revenue of PSP and ESP versus total task arrival rate.

Fig.6 compares the revenues of PSP and ESP versus the total task arrival rates. As we can see from this figure that the revenue of ESP is higher than that of the PSP, when the number of offloaded tasks is small; however, the revenue that PSP obtains is surpassing the revenue of ESP with the number of tasks becoming large. From this figure, it is obviously observed that the revenues of PSP and ESP increase rapidly with the number of tasks increasing.

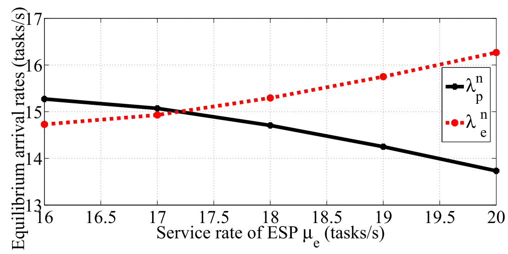


FIGURE 7. Equilibrium task arrival rates of PSP and ESP versus service rate μ_2.

C. IMPACT OF THE SERVICE RATE OF ESP

In this part, we present an analysis of the impact of ESP's service rate μ_e on task arrival rates, and the prices of PSP and ESP in equilibrium. We set the service rate of PSP as $\mu_p =$ 20 tasks/s and total task arrival rate $\lambda = 30$ tasks/s with the service rate μ_e varying from 16 to 20 tasks/s. Fig.7 shows the arrival rates of PSP and ESP in equilibrium varying versus the service rate of ESP μ_e. From this figure we can find that more MUs tend to adopt the cloud service of PSP when the service rate of ESP μ_e is small; however, the number of MUs that chooses the cloud service of ESP increases with the service rate of ESP μ_e increasing. Fig.7 suggests that the ESP can enlarge its cloud capacity to attract more MUs.

Fig.8 shows the prices of PSP and ESP in equilibrium versus the service rate of ESP μ_e. From this figure we have the observation that ESP can set higher prices in equilibrium than PSP with the service rate of ESP increasing μ_e. Fig.8 indicates that PSP has to decrease the prices of its

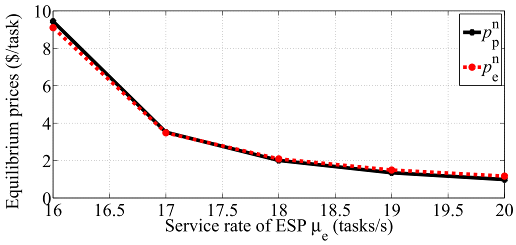


FIGURE 8. Equilibrium prices of PSP and ESP versus service rate μ_2.

cloud services when the ESP enlarges the capacity of cloud resources.

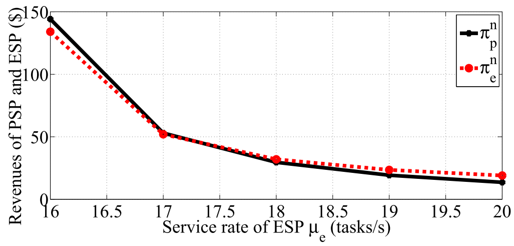


FIGURE 9. Comparison of revenue of PSP and ESP versus service rate μ_2.

Fig.9 shows the comparison of the revenue of PSP and ESP versus the service rate of ESP μ_e. As we can find from the figure that the revenue that ESP obtains is less than that of PSP when the service rate of ESP μ_e is smaller; however, the revenue that ESP obtains is growing more than that of PSP with the service rate of ESP increasing. Figs.7, 8, and 9 show that ESP can improve the competitive advantage by enlarging the capacity of its cloud resources.

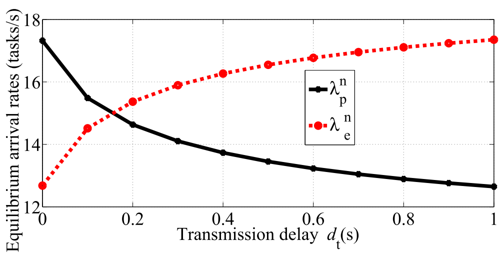


FIGURE 10. Equilibrium task arrival rates of PSP and ESP versus transmission delay d_t.

D. IMPACT OF THE TRANSMISSION DELAY

We then analyze the impact of transmission delay d_t on the arrival rates and revenues of PSP and ESP in equilibrium. We set $\mu_p = 20$ tasks/s, $\mu_e = 20$ tasks/s, and $\lambda = 30$ tasks/s. Fig.10 shows the impact of transmission delay d_t on the equilibrium arrival rates of PSP and ESP. From this figure we have the observation that more tasks will be offloaded to the edge cloud with the transmission delay increasing, which implies that the MUs prefer to choose the cloud services with better QoS.

VI. CONCLUSIONS AND FUTURE WORK

We have analyzed a mobile cloud architecture, where MUs make choices on selecting services from PSP and ESP. We analyzed the interaction between CSPs and MUs by adopting the Stackelberg game. Especially, we considered the relationship of the two CSPs in the SPG scenario, in which PSP and ESP simultaneously determine the prices of their cloud services. By leveraging the method of backward induction, we got a unique equilibrium in this game. The numerical results indicated that when the number of tasks becomes large, more MUs will select cloud services from PSP and ESP, and the two CSPs can achieve higher prices. In particular, more MUs will choose the services from the ESP if the number of tasks to be offloaded is small. Moreover, the ESP can improve its competition by enlarging the capacity of its cloud resources. Our numerical results on the transmission delay showed that MUs prefer to select service from the edge cloud to execute their tasks if PSP has long transmission delay.

There are some research directions that can be further discussed and studied. For example, the static price competition scenario can be extended to the dynamic case by using the evolutionary game. In the dynamic price competition scenario, the prices of the cloud services are different in different time slots, which is more practical. We can also analyze the cooperation of the PSP and ESP by leveraging the coalition game [44], where the PSP and ESP cooperate to maximize their revenues.

ACKNOWLEDGMENT

This work was conducted at Waseda University.

APPENDIX

PROOF OF PROPOSITION 1

It is obvious that the objective functions of **Problem1** and **Problem2** are convex functions; therefore, according to the first derivative of the objective functions,

$$\frac{\partial \pi_p^e}{\partial p_p} = \lambda - \mu_e - \frac{\alpha(p_e - d_t)}{(d + p_p - p_e)^2} = 0 \tag{14}$$

and

$$\frac{\partial \pi_e^e}{\partial p_e} = \mu_e - \frac{\alpha(d + p_e)}{(d + p_p - p_e)^2} = 0 \tag{15}$$

we have

$$p_p = p_e - d + \sqrt{\frac{\alpha(p_e - d)}{\lambda - \mu_e}} \tag{16}$$

$$p_e = p_p + d - \sqrt{\frac{\alpha(d + p_p)}{\mu_e}} \tag{17}$$

Therefore, the Nash Equilibrium prices of PSP and ESP in SPG scenario can be computed from the Eqs.(16) and (17), which are expressed as

$$p_p^n = \frac{-A + \mu_e \alpha + B}{D} \tag{18}$$

where

$$A = d(2\lambda^2 - 6\lambda\mu_e + 4\mu_e^2) \tag{19}$$

$$B = \sqrt{8\mu_e^3\alpha d + \mu_e^2\alpha - 4\lambda\alpha d\mu_e^2} \tag{20}$$

$$D = 2(\lambda - 2\mu_e)^2 \tag{21}$$

Substituting Eq.(18) into Eq.(17), we obtain

$$p_e^n = \frac{-A + \mu_p\mu_e + B}{D} + d - \sqrt{\frac{\alpha(Dd - A + \mu_e\alpha + B)}{D\mu_e}} \tag{22}$$

Substituting Eqs.(18) and (22) into Eqs.(7) and (6) respectively, we have

$$\lambda_p^n = \lambda - \mu_e + \sqrt{\frac{D\mu_e\alpha}{Dd - A + \mu_e\alpha + B}} \tag{23}$$

$$\lambda_e^n = \mu_e - \sqrt{\frac{D\mu_e\alpha}{Dd - A + \mu_e\alpha + B}} \tag{24}$$

After getting p_p^n and p_e^n, and λ_p^n and λ_e^n, we get the proof of **Proposition 1**.

REFERENCES

[1] *Cisco visual networking index: Forecast and methodology, 2014–2019*, Cisco Systems, San Jose, CA, USA.
[2] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795–2808, Oct. 2016.
[3] F. Liu *et al.*, "Gearing resource-poor mobile devices with powerful clouds: Architectures, challenges, and applications," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 14–22, Jun. 2013.
[4] X. Jin, Y. Liu, W. Fan, F. Wu, and B. Tang, "Energy-efficient resource management in mobile cloud computing," *IEICE Trans. Commmu.*, vols. E101–B, no. 4, pp. 1010–1020, Apr. 2018.
[5] Y. Mao, J. Zhang, Z. Chen, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3590–3605, Dec. 2016.
[6] V. Cardellini *et al.*, "A game-theoretic approach to computation offloading in mobile cloud computing," *Math. Program.*, vol. 157, no. 2, pp. 421–449, Jun. 2016.
[7] L. Liu, Z. Chang, X. Guo, S. Mao, and T. Ristaniemi, "Multiobjective optimization for computation offloading in fog computing," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 283–294, Feb. 2018.
[8] X. Lyu, H. Tian, C. Sengul, and P. Zhang, "Multiuser joint task offloading and resource optimization in proximate clouds," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3435–3447, Apr. 2017.
[9] Z. Zhou, J. Feng, L. Tan, Y. He, and J. Gong, "An air-ground integration approach for mobile edge computing in IoT," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 40–47, Aug. 2018.
[10] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.
[11] Z. Zhou, H. Liao, B. Gu, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "Robust mobile crowd sensing: When deep learning meets edge computing," *IEEE Netw.*, vol. 32, no. 4, pp. 54–60, Jul./Aug. 2018.
[12] Z. Zhou *et al.*, "When mobile crowd sensing meets UAV: Energy-efficient task assignment and route planning," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5526–5538, Nov. 2018.
[13] H. Khazaei, J. Misic, and V. B. Misic, "Performance analysis of cloud computing centers using $m/g/m/m + r$ queuing systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 5, pp. 936–943, May 2012.
[14] J. Liu, Y. Zhang, Y. Zhou, D. Zhang, and H. Liu, "Aggressive resource provisioning for ensuring QoS in virtualized environments," *IEEE Trans. Cloud Comput.*, vol. 3, no. 2, pp. 119–131, Apr./Jun. 2015.
[15] C. T. Do, N. H. Tran, D. H. Tran, C. Pham, M. G. R. Alam, and C. S. Hong, "Toward service selection game in a heterogeneous market cloud computing," in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manage. (IM)*, Ottawa, ON, Canada, May 2015, pp. 44–52.
[16] K. Wang, F. C. M. Lau, L. Chen, and R. Schober, "Pricing mobile data offloading: A distributed market framework," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 913–927, Feb. 2016.
[17] H. Shah-Mansouri, V. W. Wong, and R. Schober, "Joint optimal pricing and task scheduling in mobile cloud computing systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5218–5232, Aug. 2017.
[18] X. Lyu, H. Tian, W. Ni, Y. Zhang, P. Zhang, and R. P. Liu, "Energy-efficient admission of delay-sensitive tasks for mobile edge computing," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2603–2616, Jun. 2018.
[19] P. Zhao, H. Tian, C. Qin, and G. Nie, "Energy-saving offloading by jointly allocating radio and computational resources for mobile edge computing," *IEEE Access*, vol. 5, pp. 11255–11268, Jul. 2017.
[20] M. Chen and Y. Hao, "Task offloading for mobile edge computing in software defined ultra-dense network," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 587–597, Mar. 2018.
[21] N. Li, J.-F. Martinez-Ortega, and V. H. Diaz, "Distributed power control for interference-aware multi-user mobile edge computing: A game theory approach," *IEEE Access*, vol. 6, pp. 36105–36114, 2018.
[22] W. Fang, X. Yao, X. Zhao, J. Yin, and N. Xiong, "A stochastic control approach to maximize profit on service provisioning for mobile cloudlet platforms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 4, pp. 522–534, Apr. 2018.
[23] Y. Wu, K. Ni, C. Zhang, L. P. Qian, and D. H. K. Tsang, "Noma-assisted multi-access mobile edge computing: A joint optimization of computation offloading and time allocation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12244–12258, Dec. 2018.
[24] V. Kantere, D. Dash, G. Francois, S. Kyriakopoulou, and A. Ailamaki, "Optimal service pricing for a cloud cache," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1345–1358, Sep. 2011.
[25] J. Cao, K. Hwang, K. Li, and A. Y. Zomaya, "Optimal multiserver configuration for profit maximization in cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1087–1096, Jun. 2013.
[26] G. Fang and X. Li, "Optimal pricing for service provision in an IaaS cloud market with delay sensitive cloud users," in *Proc. Int. Conf. Comput. Netw., Electron. Automat. (ICCNEA)*, Xi'an, China, Sep. 2017, pp. 226–230.
[27] C. Zhang, B. O. Gu, K. Yamori, S. Xu, and Y. Tanaka, "Duopoly competition in time-dependent pricing for improving revenue of network service providers," *IEICE Trans. Commmu.*, vol. E96-B, no. 12, pp. 2964–2975, Dec. 2013.
[28] S. Ren, J. Park, and M. V. D. Schaar, "Entry and spectrum sharing scheme selection in femtocell communications markets," *IEEE/ACM Trans. Netw.*, vol. 21, no. 1, pp. 218–232, Feb. 2013.
[29] T. T. Huu and C.-K. Tham, "A novel model for competition and cooperation among cloud providers," *IEEE Trans. Comput.*, vol. 2, no. 3, pp. 251–265, Jul./Sep. 2014.
[30] Y. Feng, B. Li, and B. Li, "Price competition in an oligopoly market with multiple IaaS cloud providers," *IEEE Trans. Comput.*, vol. 63, no. 1, pp. 59–73, Jan. 2014.
[31] M.-H. Chen, B. Liang, and M. Dong, "Multi-user multi-task offloading and resource allocation in mobile cloud systems," *IEEE Wireless Commun.*, vol. 17, no. 10, pp. 6790–6805, Oct. 2018.
[32] T. Zhao, S. Zhou, X. Guo, Y. Zhao, and Z. Niu, "Pricing policy and computational resource provisioning for delay-aware mobile edge computing," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Chengdu, China, Jul. 2016, pp. 1–6.
[33] C. T. Do, N. H. Tran, E.-N. Huh, C. S. Hong, D. Niyato, and Z. Han, "Dynamics of service selection and provider pricing game in heterogeneous cloud market," *J. Netw. Comput. Appl.*, vol. 69, pp. 152–165, Jul. 2016.
[34] L. Liu, Z. Chang, and X. Guo, "Socially aware dynamic computation offloading scheme for fog computing system with energy harvesting devices," *IEEE Internet Things J.*, vol. 25, no. 10, pp. 1869–1879, Jun. 2018.
[35] W. Fang, Y. Li, H. Zhang, N. Xiong, J. Lai, and A. V. Vasilakosd, "On the throughput-energy tradeoff for data transmission between cloud and mobile devices," *Inf. Sci.*, vol. 283, pp. 79–93, Nov. 2014.
[36] X. Chang, R. Xia, J. K. Muppala, K. S. Trivedi, and J. Liu, "Effective modeling approach for IaaS data center performance analysis under heterogeneous workload," *IEEE Trans. Cloud Comput.*, vol. 6, no. 4, pp. 991–1003, Oct./Dec. 2018.
[37] N. G. Duffield and W. Whitt, "Control and recovery from rare congestion events in a large multi-server system," *Queueing Syst.*, vol. 26, no. 1, pp. 69–104, Oct. 1997.

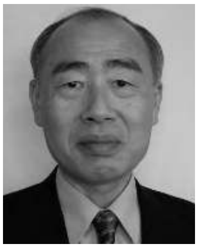
[38] F. Liu, Z. Zhou, H. Jin, B. Li, B. Li, and H. Jiang, ''On arbitrating the power-performance tradeoff in SaaS clouds,'' *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 10, pp. 2648–2658, Oct. 2014.
[39] Z. Zhou, L. Tan, B. Gu, Y. Zhang, and J. Wu, ''Bandwidth slicing in software-defined 5G: A stackelberg game approach,'' *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 102–109, Jun. 2018.
[40] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA, USA: MIT Press, 1991.
[41] Z. Zhou, M. Dong, K. Ota, R. Shi, Z. Liu, and T. Sato, ''Game-theoretic approach to energy-efficient resource allocation in device-to-device underlay communications,'' *IET Commun.*, vol. 9, no. 3, pp. 375–385, 2015.
[42] J. G. Wardrop, ''Some theoretical aspects of road traffic research,'' in *Proc. Inst. Civil Eng.*, vol. 1, no. 3, pp. 325–362, 1952.
[43] J. Elias, F. Martignon, L. Chen, and E. Altman, ''Joint operator pricing and network selection game in cognitive radio networks: Equilibrium, system dynamics and price of anarchy,'' *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4576–4589, Nov. 2013.
[44] Z. Zhou, H. Yu, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, ''Dependable content distribution in D2D-based cooperative vehicular networks: A big data-integrated coalition game approach,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 953–964, Mar. 2018.

XIANWEI LI was a doctor student in the Global Information and Telecommunication Institute at Waseda University from October 2013 to September 2016. Since 2011, he has been with the School of Information Engineering, Suzhou University. His research interests include resource allocation and power management in cloud computing, mobile edge cloud computing, the Internet of Things, and 5G communication networks. He is a member of IEICE.

CHENG ZHANG received the Ph.D. degree from Waseda University, Tokyo, Japan, in 2015. From 2008 to 2015, he was a Research Engineer with Sony Digital Network Applications, Japan and Hitachi Global Storage Technologies (HGST) Japan, Inc., where he researched and developed control algorithms for image stabilization module of Sony digital camera, and servo control algorithms for next generation high capacity HDD. He is currently an Assistant Professor of Graduate Program for Embodiment Informatics (Program for Leading Graduate School) with the Graduate School of Fundamental Science and Engineering, Waseda University. His research interests include machine control algorithm, embedded software, game theory, network economics, and machine learning. He received the IEICE Young Researcher's Award, in 2013. He is a member of IEICE.

BO GU received the B.E. degree from Tianjin University, Tianjin, China, in 2004, the M.E. degree from Peking University, Beijing, China, in 2007, and the Ph.D. degree from Waseda University, Tokyo, Japan, in 2013. From 2007 to 2011, he was a Research Engineer at Sony Digital Network Applications, Japan. In 2013, he joined the Department of Communications and Computer Engineering, Waseda University, as an Assistant Professor. Since 2016, he has been with the Department of Information and Communications Engineering, Kogakuin University. His current research interests include network economics, game theory, and network optimization. He is a member of IEICE.

KYOKO YAMORI received the B.A. degree in business administration and M.A. and Ph.D. degrees in information management science from Asahi University, in 1995, 1997, 2000, respectively, where she is currently a Professor with the Department of Management Information. She is also a Visiting Professor with the Global Information and Telecommunication Institute, Waseda University. She received the IEICE Switching System Research Award, in 2001, the IEICE Young Research's Award, in 2005, and the IEICE Best Paper Award, in 2005. She is a Senior Member of IEICE.



YOSHIAKI TANAKA received the B.E., M.E., and D.E. degrees in electrical engineering from The University of Tokyo, Tokyo, Japan, in 1974, 1976, and 1979, respectively. He became a staff at the Department of Electrical Engineering, The University of Tokyo, in 1979, and has been engaged in teaching and researching in the fields of telecommunication networks, switching systems, and network security. He is currently a Professor with the Department of Communications and Computer Engineering, Waseda University. He received the IEICE Best Paper Award, the IEICE Achievement Award, the IEICE Distinguished Achievement and Contributions Award, the Okawa Publication Prize, and the Commendation by Minister for Internal Affairs and Communications. He is an Honorary Member of IEICE.