https://doi.org/10.5392/JKCA.2019.19.04.079

복점 멀티미디어 클라우드 서비스 시장에서의 가격 경쟁

Price Competition in Duopoly Multimedia Cloud Service Market

이두호
강원대학교 소프트웨어미디어·산업공학부

Doo Ho Lee(enjdhlee@kangwon.ac.kr)

요약

최근 들어 다수의 클라우드 서비스 제공자가 클라우드 컴퓨팅 서비스를 제공함으로써 각 제공자는 더 많은 사용자를 확보하기 위해 치열한 경쟁을 벌이고 있다. 서비스 제공자별 컴퓨팅 자원의 구성 및 서비스 제공 부하가 다르기 때문에 사용자는 다양한 수준의 서비스 품질을 경험할 수 있다. 따라서 클라우드 서비스 시장에서 더 많은 사용자를 확보하여 수익을 최대화하기 위해서는 서비스 품질에 대한 가장 합리적인 가격을 결정하는 것이 매우 중요하다. 본 연구에서는 두 명의 서비스 제공자가 존재하는 멀티미디어 클라우드 서비스 시장에서 두 제공자 간 서비스 가격 경쟁에 대해 다룬다. 두 명의 클라우드 서비스 제공자가 최적의 가격을 결정하여 상호 경쟁하고 자신의 이익을 최대화할 수 있는 가격 산정 방법을 비협력 게임 이론으로 설명한다. 이를 위해 멀티미디어 클라우드 서비스의 제공 프로세스를 대기행렬 시스템으로 모형화하고, 분석 결과를 바탕으로 복점 멀티미디어 클라우드 서비스 시장에서 가격 경쟁 문제를 제안한다.

■ 중심어 : | 가격경쟁 | 멀티미디어 클라우드 서비스 | 복점 | 수익 최대화 | 균형가격 |

Abstract

As an increasing number of cloud service providers begin to provide cloud computing services, they form a competitive market to compete for users. Due to different resource configurations and service workloads, users may observe different response times for their service requests and experience different levels of service quality. To compete for cloud users, it is crucial for each cloud service provider to determine an optimal price that best corresponds to their service qualities while also guaranteeing maximum profit. To achieve this goal, the underlying rationale and characteristics in this competitive market must be clarified. In this paper, we analyze price competition in the multimedia cloud service market with two service providers. We characterize the nature of non-cooperative games in a duopoly multimedia cloud service market with the goal of capturing how each cloud service provider determines its optimal price to compete with the other and maximize its own profit. To do this, we introduce a queueing model to characterize the service process in a multimedia cloud data center. Based on performance measures of the proposed queueing model, we suggest a price competition problem in a duopoly multimedia cloud service market. By solving this problem, we can obtain the optimal equilibrium prices.

■ keyword : | Price Competition | Multimedia Cloud Service | Duopoly | Profit Maximization | Equilibrium Price |

* 이 논문은 2017년도 강원대학교 대학회계 학술연구조성비로 연구 하였음 (관리번호-620170025). 이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행 된 기초연구사업임 (NRF-2018R1D1A3B07040887).

접수일자 : 2019년 03월 19일
수정일자 : 2019년 04월 11일
심사완료일 : 2019년 04월 17일
교신저자 : 이두호, e-mail : enjdhlee@kangwon.ac.kr

I. Introduction

Cloud computing is a new paradigm for the provisioning of a variety of computing resources, such as infrastructure, platforms, and software applications, to reduce the cost of operating and managing hardware and software resources by converting the locations of computing resources to networks. Through cloud computing services, users can focus on their core business processes without the hindrance of ICT obstacles[1-4].

The main enabling technology for cloud computing is virtualization. Virtualization software creates a temporarily simulated or extended version of computing resources such as processors, operating systems, storage devices, and network resources. The simulated or extended version (virtual machine) will resemble an actual resource. There are numerous objectives of virtualization. First, this strategy seeks to utilize shared resources fully by applying partitioning and time-sharing. Second, it centralizes resource management. Third, it enhances cloud data center agility and provides the required scalability and elasticity for on-demand capabilities. Fourth, it can improve the testing and running of software diagnostics on different operating platforms. Fifth, it seeks to improve the portability of applications and the capability of workload migration. Sixth, it can provide the isolation required for a high degree of reliability, availability, and security. The seventh goal is to enable server consolidation, and the eighth is to provide self-management frameworks[5].

Among the various cloud-based software services currently available, multimedia services are crucial for cloud computing. As is well known, multimedia services such as the media retrieval, video-on-demand (VOD), free viewpoint video (FVV), and over-the-top (OTT) services typically require intensive computation and network resources, which are burdens to client devices, especially to resource-constrained mobile devices[6]. Various multimedia cloud services provide a way to resolve this problem. By migrating multimedia processing to the cloud, the hardware requirements on the user side are dramatically reduced. Users are able to access targeted cloud services without restrictions on time and/or place. The elastic and on-demand characteristics of resource provisioning in the cloud effectively satisfy the intensive resource demands of multimedia processing[7].

Given that a user's level of multimedia service demand may be met by any multimedia cloud service provider (CSP), a rational user will choose the one that maximizes the user's net reward, i.e., its utility obtained by choosing the multimedia cloud service minus the required payment. The utility of a user is not only determined by the importance of the task (i.e., the magnitude of the benefit received by the user when finishing this task), but is also closely related to the urgency of the task (i.e., how quickly it can be finished). The same task is able to generate more utility for a cloud user if it can be completed within a shorter period of time. Because diversity among CSPs leads to different net rewards, multiple multimedia CSPs form a market to compete for cloud users. Existing real-world measurement results [8] reveal that different CSPs complete tasks with different completion times, with a CSP possibly becoming less competitive with an inappropriate price setting. With different price settings, payments made to finish each benchmarking task are also different across different CSPs. As a consequence, CSPs are presented with a question: How can each multimedia CSP determine the optimal price to maximize its profit in such a competitive market, in which the demands from users are sensitive to both the finishing time and the payment related to the completion of a task? It turns out that answering this question is nontrivial. On the one hand, CSPs may wish to increase prices to generate more profit. On the other hand, increasing prices excessively in a competitive environment may risk losing potential users, which then results in reduced profits. Moreover, although reducing the price should intuitively be an effective means of attracting users, these users may overwhelm the CSP due to an unreasonably low price, which then leads to longer finishing times with regard to the tasks to be completed. Hence, the reduced utility will prohibit future users to choose this CSP.

This study explores price competition in a multimedia cloud service market formed by multimedia CSPs. More specifically, we present an analytical result of a duopoly multimedia cloud service market in which two multimedia CSPs are competing with each other. We use a tandem queueing network to model a multimedia cloud data center. Given that the pricing strategy of a CSP depends on its competitor, we take a game theoretic perspective to study the strategic situation. To the best of our knowledge, this is the first study to investigate price competition in a duopoly multimedia cloud service market. The topic of price competition has been of interest in numerous studies in the context of economic markets with multiple service providers. Petri et al.[9][10] studied the effects of risk in service-level agreements in service provider communities. Chen and Frank[11] presented an analysis of equilibrium prices in a monopoly market, and Chen and Wan[12] dealt with equilibrium prices in a duopoly market with varying levels of demand. Allon and Federgruen[13] analyzed a general market for an industry of competing service facilities. Firms differentiate themselves by their price levels and the waiting times experience by their customers as well as by different attributes not determined directly through competition. A simultaneous price competition game among multiple service providers was also considered in networking research. Anselmi et al.[14] studied a load balancing game with multiple network links, each of which was under the control of a profit-maximizing provider. Employing the theory of a processor-sharing queue, they discussed the existence of an oligopolistic equilibrium price for a network service. Feng et al.[15] studied a non-cooperative price competition model in an IaaS (infrastructure-as-a-service) cloud service market and derived equilibrium prices for both monopoly and duopoly markets. However, they modeled a very complex data center as a simple M/M/1 queueing system. More recently, Kilcioglu and Rao[16] introduced a price-quality competition game in not only a monopoly cloud service market but also a duopoly cloud service market. For more details on price competition in a duopoly market, readers are recommended to see [15][16], and references therein.

The rest of this paper is organized as follows: In section 2, we model a multimedia cloud data center as a queueing network and introduce the utility function of a cloud user. In section 3, we analyze price competition in a duopoly market and present Nash equilibrium prices for each multimedia CSP. In section 4, we briefly deals with price competition in a oligopoly market and shows the method to derive Nash equilibrium prices. Section 5 discusses certain characteristics of Nash equilibrium prices with several numerical experiments.

II. Model description and assumptions

In this section, we present our system model, including the data center architecture and its queueing model. The data center architecture characterizes the infrastructure of a multimedia cloud; therefore, the queueing model is built to identify the response time for service requests from users. We also determine the structure of user utility functions to describe duopoly price competition in the multimedia cloud service market.

1. Queueing analysis

Currently, the vast majority of clouds are in the form of data centers. [Figure 1] shows a simplified version of the architecture of a multimedia cloud data center. It is composed of a scheduler server, a number of computing servers, and a transmission server.

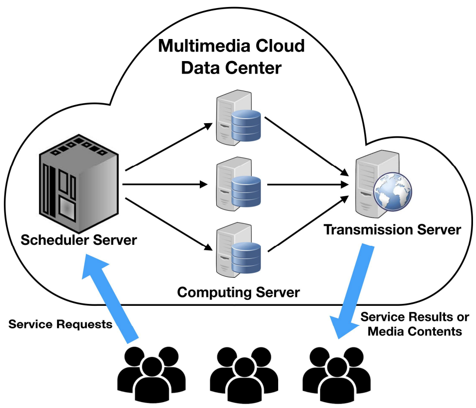

Figure 1. Multimedia cloud data center architecture

All of the servers in the data center are configured by multiple virtual machine instances in order to ensure more powerful resource capacity and higher resource efficiency levels. When requests which generated by multiple users arrive at the multimedia cloud data center, the scheduler server receives all of the requests and then distributes each request to a computing server. As an actual processor, the computing server utilizes the allocated computation resources and the associated media contents to serve the request. After processing, the service results or the requested media contents will be sent back to users. The transmission server acts as a gateway node which controls the overall traffic and directs the given packets to a specific destination. All servers in the data center are usually connected with reliable and high-speed communication links. Therefore, we assume that the latency for transferring requests is negligible and that no link connection errors occur between servers.

We now introduce a queueing model to determine the performance measures of the multimedia cloud data center, in this case the response time for a request. Based on the above architecture, a corresponding queueing model is given in [Figure 2]. The model is expressed as a three-phase serial queueing network (or a tandem queueing network) which consists of a schedule queue, computation queues, and a transmission queue.

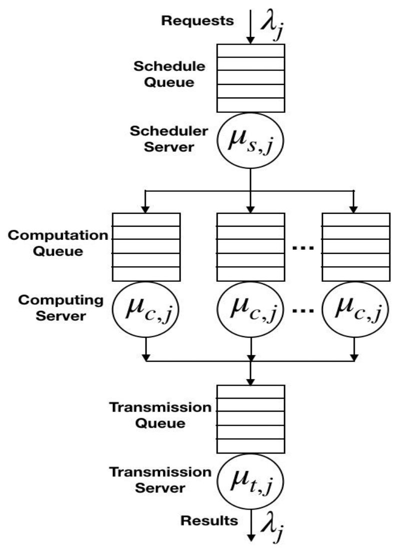

Figure 2. Queueing model of multimedia cloud data center

All arriving requests are initially buffered at the schedule queue on a first-in-first-out basis.

According to Cao et al.[17], a request arrival on a web server can be modeled as a Poisson arrival process. Thus, we assume that the request arrival for a multimedia cloud service follows a Poisson arrival process with a rate of λ. The requests are scheduled to be sent to one of the computing servers at a rate of μ_s by the scheduler server. The processing time for a request in the scheduler server is assumed to follow an exponential distribution. Therefore, the mean response time for a request in the schedule queue is given by $T^{sche}=(\mu_s-\lambda)^{-1}$. Suppose that there are N computing servers in the data center. Each computing server manages its corresponding computation queue to process requests. The scheduler server allocates requests to the computing servers in a round-robin fashion for proper load-balancing among the virtual machine instances. According to the decomposition property of a Poisson process[18], subflows resulting from the even splitting of a Poisson flow also follow a Poisson flow. Therefore, a request arrival to the i^{th} $(i=1,2,\cdots,N)$ computation queue is modeled as a Poisson arrival process with a rate of λ/N. The requests are sent to the transmission server at a rate of μ_c by each computing server. The processing time for a request in each computing server is assumed to follow an exponential distribution. Therefore, the mean response time of a request in the computation queue is given by $T^{comp}=(\mu_c-\lambda/N)^{-1}$. After being processed, service results or media contents are sent back to users through the transmission server at a rate of μ_t. We also assume that the processing time for a request in the transmission server follows an exponential distribution. Therefore, the mean response time of a request in transmission queue is given by $T^{tran}=(\mu_t-\lambda)^{-1}$. According to Jackson's theorem [19], the total response time in the data center is the sum of the response times in the three phases. This is formulated as $T^{dc}=T^{sche}+T^{comp}+T^{tran}$.

This study considers the multimedia cloud service market which is consisting of two CSPs. Hence, users must choose one of the two CSPs for their service requests to be served. Let T^{dc_j} be the total response time of a user's request after the user selects the j^{th} CSP $(j=1,2)$. Assuming that one CSP owns one data center, T^{dc_j} is expressed as

$$T_j^{dc}(\lambda_j)=\frac{1}{\mu_{s,j}-\lambda_j}+\frac{N_j}{N_j\mu_{c,j}-\lambda_j}+\frac{1}{\mu_{t,j}-\lambda_j}, \quad (1)$$

where λ_j is the effective arrival rate of a request at the j^{th} data center, N_j is the number of computing servers in the j^{th} data center, and $\mu_{s,j}$, $\mu_{c,j}$ and $\mu_{t,j}$ are the processing rates of the scheduler server, the computing server and the transmission server in the j^{th} data center, respectively. For analytical simplicity, this study assumes that no request is dropped during this process. In addition, for a data center to be stably operated, the following condition should hold: $0<\lambda_j<\min\{\mu_{s,j}, N_j\mu_{c,j}, \mu_{t,j}\}$.

Remark 1. Due to the recent virtualization technology such as a multiple-container and an advanced hypervisor technology, it is possible to generate the almost infinite number of virtual computing servers in a cloud data center. In this case, the computing server farm can be modeled as $M/M/\infty$ queueing system. Hence, the mean response time of a request in the computation queue is given by $T^{comp}=\lim_{N\to\infty}(\mu_c-\lambda/N)^{-1}=\mu_c^{-1}$. In consequence, the total response time of a user's request in (1) is simplified as follows:

$$T_j^{dc}(\lambda_j)=\frac{1}{\mu_{s,j}-\lambda_j}+\frac{1}{\mu_{c,j}}+\frac{1}{\mu_{t,j}-\lambda_j}.$$

2. User's utility

Introduced by Naor[20], the linear reward-cost utility function has been adopted in numerous works due to its ease of calculation and interpretation. In this work, we also employ the linear reward-cost utility. Let c be the average delay cost units per unit time, and p_j be the admission price charged by the j^{th} CSP. Assume that all users are identical and that each user obtains a reward or a service value of R units after being served. Then, a user's utility after being served by the j^{th} CSP can be expressed by the following linear equation:

$$U_j(\lambda_j)=R-TC_j=R-\left(p_j+cT_j^{dc}(\lambda_j)\right), \qquad (2)$$

where the user's total cost TC_j includes the delay cost and the payment to the j^{th} CSP. Because user's utility may be negative when the total cost exceeds the reward, we assume that the user will decide to use the multimedia cloud service if its utility is not negative.

III. Duopoly price competition

Since two CSPs compete with each other by setting prices to maximize their revenue, we can formulate their price competition as a non-cooperative game. In this section, we present simultaneous pricing (or parallel pricing) strategy and then shows the existence of a Nash equilibrium solution. A brief concept of the price competition in the duopoly multimedia cloud service market is given in [Figure 3].

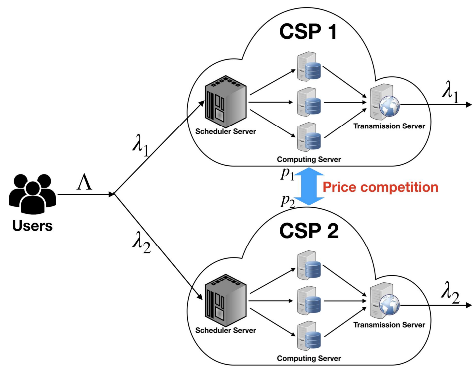

Figure 3. Price competition in duopoly multimedia cloud service market

We assume that the arriving users are individual optimizers. Then, given a particular admission price p_j of the CSP j, the equilibrium arrival rate of service requests λ_j satisfies the equilibrium conditions $U_j(\lambda_j)=0,\ j=1,2$.

Let π_j be the expected revenue of the j^{th} CSP. Each CSP aims to maximize π_j by choosing its admission price p_j, which clearly depends on the reaction of the other CSP and that of all cloud users. Let $\pi_j(p_1,p_2)$ denote the expected revenue of the CSP j if it chooses a price p_j, given the other CSP k's price p_k, $j\neq k$, and $j,k=1,2$. A pair of prices (p_1^*,p_2^*) is said to be a Nash equilibrium if it satisfies following conditions:

$$\begin{aligned}\pi_1(p_1^*,p_2^*)&\geq \pi_1(p_1,p_2^*),\ \forall p_1\geq 0,\\ \pi_2(p_1^*,p_2^*)&\geq \pi_2(p_1^*,p_2),\ \forall p_2\geq 0.\end{aligned} \qquad (3)$$

At a Nash equilibrium state, any CSP cannot increase the expected revenue by changing its admission price unilaterally. That is, the Nash equilibrium price is the optimal price a CSP can achieve in the market when the two CSPs are not cooperative. In addition, the expected revenue of both CSPs are maximized, and the market is balanced dynamically. The equilibrium prices can be obtained by a standard procedure of identifying the best response function of each CSP. Let $p_j = F_j(p_k)$ be CSP j's optimal admission price given the admission price p_k selected by CSP k. A Nash equilibrium solution in this duopoly multimedia cloud service market is then a pair of prices (p_1, p_2) such that $p_1 = F_1(p_2)$ and $p_2 = F_2(p_1)$, i.e., an intersection point of two best response functions.

Take CSP 1 as an example. The best response function F_1 can be found by assuming that CSP 2's admission price p_2 is given and by solving CSP 1's problem as follows:

$$\begin{aligned} &\max_{p_1,\lambda_1} p_1\lambda_1 \\ &s.t. \\ &\quad \lambda_1+\lambda_2=\Lambda, \\ &\quad U_1(\lambda_1)=R-\left(p_1+cT_1^{dc}(\lambda_1)\right)=0, \\ &\quad 0\le\lambda_1<\min\{\mu_{s,1},N_1\mu_{c,1},\mu_{t,1}\}, \\ &\quad p_1\ge 0, \end{aligned} \tag{4}$$

where Λ in the first constraint is the total arrival rate of all users to the market (see [Figure 3]). The second constraint indicates the equilibrium condition. The third constraint is the stability condition. Similarly, given the admission price p_1, the best response of CSP 2 that maximizes the revenue is given as follows:

$$\begin{aligned} &\max_{p_2,\lambda_2} p_2\lambda_2 \\ &s.t. \\ &\quad \lambda_1+\lambda_2=\Lambda, \\ &\quad U_2(\lambda_2)=R-\left(p_2+cT_2^{dc}(\lambda_2)\right)=0, \\ &\quad 0\le\lambda_2<\min\{\mu_{s,2},N_2\mu_{c,2},\mu_{t,2}\}, \\ &\quad p_2\ge 0. \end{aligned} \tag{5}$$

Using the first constraint of (4) and (5), we can rewrite the problem (5) as follows:

$$\begin{aligned} &\max_{p_2,\lambda_2} \left[p_1+cT_1^{dc}(\Lambda-\lambda_2)-cT_2^{dc}(\lambda_2)\right]\lambda_2 \\ &s.t. \\ &\quad p_1+cT_1^{dc}(\lambda_1)=p_2+cT_2^{dc}(\lambda_2), \\ &\quad 0\le\lambda_2<\min\{\mu_{s,2},N_2\mu_{c,2},\mu_{t,2}\}, \\ &\quad p_2\ge 0. \end{aligned} \tag{6}$$

The above optimization problem can be solved by differentiating the objective function with respect to λ_2 to determine the first-order condition for the value of λ_2 to be optimal value such as

$$p_2 = c\lambda_2\left[T_1^{dc^{(1)}}(\Lambda-\lambda_2)+T_2^{dc^{(1)}}(\lambda_2)\right], \tag{7}$$

where $f^{(1)}$ means the first derivative of a function f. Similarly, using the symmetric relation, the first-order condition for the value of λ_1 to be optimal given the admission price p_2 is obtained as follows:

$$p_1 = c\lambda_1\left[T_1^{dc^{(1)}}(\lambda_1)+T_2^{dc^{(1)}}(\Lambda-\lambda_1)\right]. \tag{8}$$

Combining (7) and (8), we have

$$p_1-p_2 = c(2\lambda_1-\Lambda)\left[T_1^{dc^{(1)}}(\lambda_1)+T_2^{dc^{(1)}}(\Lambda-\lambda_1)\right]. \tag{9}$$

From the first constraint in (6), we obtain

$$p_1-p_2 = c\left[T_2^{dc}(\Lambda-\lambda_1)-T_1^{dc}(\lambda_1)\right]. \tag{10}$$

Then, we finally get

$$(2\lambda_1-\Lambda)G(\lambda_1) = T_2^{dc}(\Lambda-\lambda_1)-T_1^{dc}(\lambda_1), \tag{11}$$

where $G(\lambda_1)=T_1^{dc^{(1)}}(\lambda_1)+T_2^{dc^{(1)}}(\Lambda-\lambda_1)$. Determining the optimal value of λ_1 is equivalent to finding the root of (11), but its procedure is very complex and long. Instead, one possible numerical method to find the solution is the bisection algorithm with logarithmic complexity. Let λ_j^* be the optimal equilibrium solution of λ_j. Then, λ_1^* is the root of (11) and λ_2^* is easily calculated by using the equation $\lambda_1^* + \lambda_2^* = \Lambda$. Using the second constraints of (4) and (5), we obtain the Nash equilibrium admission prices as follows: $p_j^* = R - cT_j^{dc}(\lambda_j^*),\ j = 1, 2$.

IV. Oligopoly Price Competition

This section extends the result of the duopoly market case to the oligopoly market case. Consider the multimedia cloud service market that consists of the number M of CSPs, where $M \geq 2$. Let $\pi_j(p_1, \cdots, p_j, \cdots, p_M)$ denote the expected revenue of the CSP j if it chooses a price p_j, given the other CSPs' prices p_k, $j \neq k$, and $j, k \in \boldsymbol{S} = \{1, \cdots, M\}$. A vector of prices $(p_1^*, \cdots, p_j^*, \cdots, p_M^*)$ is said to be a Nash equilibrium if it satisfies following conditions:

$$\pi_j(p_1^*, \cdots, p_j^*, \cdots, p_M^*) \geq \pi_j(p_1^*, \cdots, p_j, \cdots, p_M^*), \tag{12}$$

where $\forall p_j \geq 0$ and $j \in \boldsymbol{S}$. Like the duopoly market case, the equilibrium prices can be obtained by a standard procedure of identifying the best response function of each CSP. Let $p_j = F_j(\boldsymbol{p}_{(j)})$ be CSP j's optimal admission price, where $\boldsymbol{p}_{(j)} = (p_1, \cdots, p_{j-1}, p_{j+1}, \cdots, p_M)$, given the other CSPs' prices p_k, $k \in \boldsymbol{S}_{(j)} = \boldsymbol{S} - \{j\}$. A Nash equilibrium solution in this oligopoly market is then a vector of prices $(p_1, \cdots, p_M)$ such that $p_j = F_j(\boldsymbol{p}_{(j)})$, $j \in \boldsymbol{S}$. Take CSP j as an example. The best response function F_j can be found by assuming that the other CSPs admission price are given and by solving CSP j's problem as follows:

$$\begin{aligned} & \max_{p_j,\ \lambda_j} p_j \lambda_j \\ & s.t. \\ & \quad \sum_{n=1}^{M} \lambda_n = \Lambda, \\ & \quad p_j + cT_j^{dc}(\lambda_j) = p_k + cT_k^{dc}(\lambda_k),\ \forall k \in \boldsymbol{S}_{(j)}, \\ & \quad 0 \leq \lambda_j < \min\{\mu_{s,j}, N_j \mu_{c,j}, \mu_{t,j}\}, \\ & \quad p_j \geq 0. \end{aligned} \tag{13}$$

The solving procedure of (13) can use the method of Lagrange multipliers. The Lagrange function of (13), L_j, can be formulated as

$$\begin{aligned} L_j = & \, p_j \lambda_j - \alpha \left(\sum_{n=1}^{M} \lambda_n - \Lambda \right) \\ & - \sum_{k \in \boldsymbol{S}_{(j)}} \beta_k \left[p_k - p_j + c\left(T_k^{dc}(\lambda_k) - T_j^{dc}(\lambda_j) \right) \right], \end{aligned} \tag{14}$$

where α and β_k, $k \in \boldsymbol{S}_{(j)}$ are Lagrange multipliers of the constraints in (13). Hence, the optimal solution of (13) can be obtained by solving the following equations in group:

$$\begin{aligned} & \frac{\partial L_j}{\partial p_j} = \lambda_j + \sum_{k \in \boldsymbol{S}_{(j)}} \beta_k = 0, \\ & \frac{\partial L_j}{\partial \lambda_j} = p_j - \alpha \\ & \qquad + c \sum_{k \in \boldsymbol{S}_{(j)}} \beta_k \left(T_j^{dc^{(1)}}(\lambda_j) + T_k^{dc^{(1)}}(\lambda_k) \right) = 0, \\ & \frac{\partial L_j}{\partial \alpha} = \sum_{n=1}^{M} \lambda_n - \Lambda = 0, \\ & \frac{\partial L_j}{\partial \beta_k} = p_k - p_j \\ & \qquad + c\left(T_k^{dc}(\lambda_k) - T_j^{dc}(\lambda_j) \right) = 0,\ k \in \boldsymbol{S}_{(j)}. \end{aligned} \tag{15}$$

For each CSP j, $j \in \boldsymbol{S}$, in the oligopoly market, we derive the group of the above equations and we can obtain the optimal admission price of each CSP by solving the simultaneous equations. Let us take an example where $M=2$ (duopoly market):

Example 1. Setting $M=2$, CSP 1's equations are presented as follows:

$$\frac{\partial L_1}{\partial p_1}=\lambda_1+\beta_2=0,$$
$$\frac{\partial L_1}{\partial \lambda_1}=p_1-\alpha+c\beta_2\left(T_1^{dc^{(1)}}(\lambda_1)+T_2^{dc^{(1)}}(\lambda_2)\right)=0,$$
$$\frac{\partial L_1}{\partial \alpha}=\lambda_1+\lambda_2-\Lambda=0,$$
$$\frac{\partial L_1}{\partial \beta_2}=p_2-p_1+c\left(T_2^{dc}(\lambda_2)-T_1^{dc}(\lambda_1)\right)=0. \quad (16)$$

In the same manner, CSP 2's equations are presented as follows:

$$\frac{\partial L_2}{\partial p_2}=\lambda_2+\beta_1=0,$$
$$\frac{\partial L_2}{\partial \lambda_2}=p_2-\alpha+c\beta_1\left(T_2^{dc^{(1)}}(\lambda_2)+T_1^{dc^{(1)}}(\lambda_1)\right)=0,$$
$$\frac{\partial L_2}{\partial \alpha}=\lambda_1+\lambda_2-\Lambda=0,$$
$$\frac{\partial L_2}{\partial \beta_1}=p_1-p_2+c\left(T_1^{dc}(\lambda_1)-T_2^{dc}(\lambda_2)\right)=0. \quad (17)$$

Summarizing the results in (16) and (17), we have the following equations in group:

$$\lambda_1+\lambda_2=\Lambda, \quad (18)$$
$$p_1=c\lambda_1\left(T_1^{dc^{(1)}}(\lambda_1)+T_2^{dc^{(1)}}(\lambda_2)\right),$$
$$p_2=c\lambda_2\left(T_1^{dc^{(1)}}(\lambda_1)+T_2^{dc^{(1)}}(\lambda_2)\right),$$
$$p_2-p_1=c\left(T_1^{dc}(\lambda_1)-T_2^{dc}(\lambda_2)\right),$$

which are identical to equations (7)-(10) in section 3. For more details on the method of Lagrange multipliers, see Boyd and Vandenberghe[21].

V. Numerical Experiments

In this section, we present some numerical examples with which to explore the equilibrium arrival rates and the price competition in a duopoly multimedia cloud service market.

Example 2. This example deals with the optimal equilibrium arrival rates λ_1^* and λ_2^*, and their corresponding optimal prices p_1^* and p_2^*. Initially, we conduct a sensitivity analysis of the number of computing servers in CSP 2's data center, N_2 while assuming $\Lambda=10{,}000$, $\mu_{s,1}=\mu_{s,2}=6{,}000$, $\mu_{c,1}=\mu_{c,2}=110$, $\mu_{t,1}=\mu_{t,2}=5{,}200$, and $N_1=1{,}000$, $R=100$, and $c=1{,}000$. Varying the value of N_2 from 70 to 1,000, we record λ_1^* and λ_2^* in [Figure 4], and p_1^* and p_2^* in [Figure 5].

When $N_1>N_2$, we can confirm that $\lambda_1^*>\lambda_2^*$. This means that the CSP with a larger service capacity has more customers in a duopoly market. By the way, as N_2 approaches $N_1=1{,}000$, the two CSPs' data centers become identical in scale; therefore, the entire market is completely divided by two CSPs, and λ_1^* and λ_2^* approaches $\Lambda/2=5{,}000$ (see [Figure 4]).

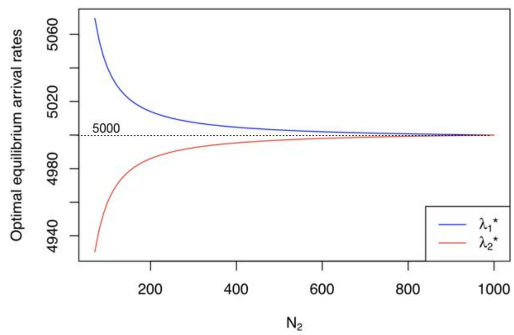

Figure 4. Optimal equilibrium arrival rates vs. the number of computing servers in CSP 2's data center

We can also find that, when $N_1 > N_2$, $p_1^* < p_2^*$. This is because CSP 1 achieves the economies of scale, it can provide cloud services at a lower price. In addition, as N_2 approaches $N_1 = 1{,}000$, $T_1(\lambda_1^*) = T_2(\lambda_2^*)$ and $p_1^* = p_2^* = 84.476$ (see [Figure 5]).

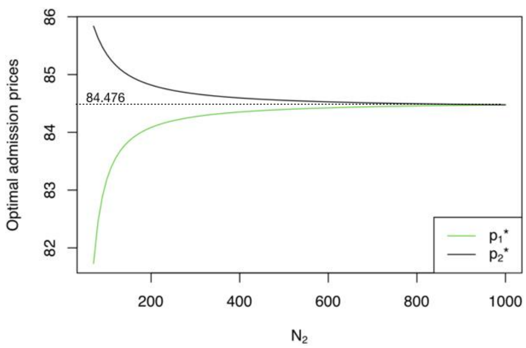

Figure 5. Optimal prices vs. the number of computing servers in CSP 2's data center

Example 3. This example deals with the optimal equilibrium arrival rates λ_1^* and λ_2^*, and their corresponding optimal prices p_1^* and p_2^* while varying the total request arrival rate Λ from 100 to 10,000. In this example, we set $N_1 = 800$ and $N_2 = 350$, and the other parameters are identically set to those in *Example 1.*

In [Figure 6], we find that both λ_1^* and λ_2^* increase as Λ increases. That is, as demand for multimedia cloud services grows, more customers are looking for two CSPs. We also find that $\lambda_1^* > \lambda_2^*$ at all values of Λ. Because the service capacity of CSP 1's data center is larger than that of CSP 2's data center $(N_1 > N_2)$, CSP 1 can get more customers. However, as Λ approaches 10,000, λ_1^* and λ_2^* approaches 5,000, respectively. Although the service capacity of CSP 1's data center is larger than that of CSP 2's data center, CSP 1's data center can serve a customer's request at a rate of up to 5,000. Then, the rest of service requests are precessed at a rate of up to 5,000 in CSP 2's data center. That is, when the service request rate reaches a maximum that two CSPs in the multimedia cloud service market can handle, this market is evenly divided by two CSPs.

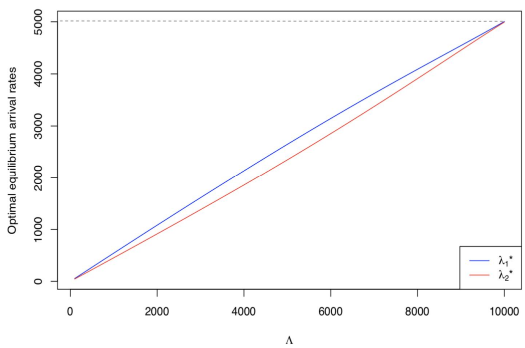

Figure 6. Optimal equilibrium arrival rates vs. the total service request rate

Next, we investigate the relation between the total service request rate and the optimal price for a service. As shown in [Figure 7], the optimal prices p_1^* and p_2^* are both decreasing functions of the total service request rate Λ. This is due to the fact that a Nash equilibrium admission price is defined as follows: $p_j^* = R - cT_j^{dc}(\lambda_j^*)$, $j = 1, 2$. As Λ increases, λ_j^* increases (see [Figure 6]), which causes the total response time for a service $T_j^{dc}(\lambda_j^*)$ to increase. Hence, p_j^* decreases. That is, a long service processing time means that service quality is degraded, so the price is lowered in terms of compensation. We also find that $p_1^* < p_2^*$. As mentioned in *Example 2*, CSP 1 achieves the economies of scale, so it can provide cloud services at a lower price.

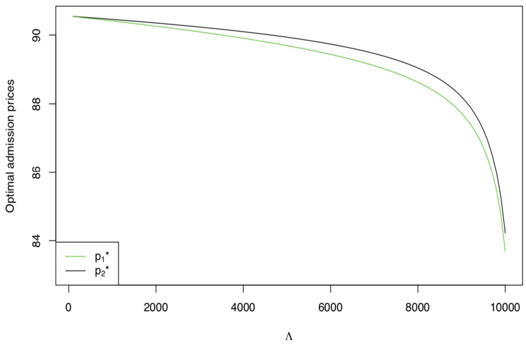

Figure 7. Optimal prices vs. the total service request rate

VI. Conclusion

In this paper, we studied the parallel pricing game between two service providers in a multimedia cloud service market. First, we modeled the process of providing the multimedia cloud service as a tandem queueing network. Second, We established the profit-maximization problem for each CSP and presented the procedure of determining the optimal and equilibrium arrival rates and prices. We also dealt with the pricing game in an oligopoly multimedia cloud service market. Finally, we demonstrated the price competition by conducting some numerical experiments.

This study can be extended as follows: i) the market consists of two CSPs, but one is a market leader and the other is a follower. Let CSP 1 and CSP 2 be a market leader and a follower, respectively. Then, CSP 1's pricing is made first and then CSP 2's pricing is made based on the information on CSP 1's pricing. This is called sequential pricing game (or Stackelberg price competition [22]). ii) we can assume that CSP 1 and CSP 2 are not competitive but cooperative through bargaining. Bargaining theory is categorized in a cooperative game theory. Hence, we can find a Nash bargaining solution of the cooperative pricing game between CSP 1 and CSP 2 in a duopoly (or oligopoly) multimedia cloud service market.

참 고 문 헌

[1] P. Mell and T. Grance, "The NIST definition of cloud computing," National Institute of Standards and Tchnology, Vol.53, No.6, p.50, 2009.

[2] C. N. Hoefer and G. Karagiannis, "Taxonomy of cloud computing services," In: GLOBECOM Workshops (GC Wkshps), 2010 IEEE. pp.1345-1350, 2010.

[3] P. Bhaskar, J. Admela, K. Dimitrios, and G. Yves, "Architectural Requirements for Cloud Computing Systems: An Enterprise Cloud Approach," Journal of Grid Computing, Vol.9, No.1, pp.3-26, 2011.

[4] W. T. Tsai, X. Sun, and J. Balasooriya, "Service-Oriented Cloud Computing Architecture," In: 7th International Conference on Information Technology, 2010.

[5] D. H. Lee, "An Efficient VM-Level Scaling Scheme in an IaaS Cloud Computing System: A Queueing Theory Approach," International Journal of Contents, Vol.13, No.2, pp.29-34, 2017.

[6] R. Shea, J. Liu, E. H. Ngai, and Y. Cui, "Cloud Gaming: Architecture and Performance," IEEE Network, Vol.27, No.4, pp.16-21, 2013.

[7] X. Nan, Y. He, and L. Guan, "Queueing Model Based Resource Optimization for Multimedia Cloud," Journal of Visual Communication and Image Representation, Vol.25, No.4, pp.928-942, 2014.

[8] A. Li, X. Yang, S. Kandula, and M. Zhang, "CloudCmp: Comparing Public Cloud Providers," In; 10th Annual Conference on Internet

Measurement (IMC), pp.1-14, 2010.
[9] I. Petri, O. Rana, and G. C. Silaghi, “SLA as a Complementary Currency in Peer-2-Peer Markets,” In: 7th International Conference on Economics of Grids, Clouds, Systems, and Services (GECON), pp.141-152, 2010.
[10] I. Petri, O. F. Rana, Y. Regzui, G. C. Silaghi, and Y. Rezgui, “Risk Assessment in Service Provider Communities,” Future Generation Computer Systems, Vol.41, pp.32-43, 2014.
[11] H. Chen and M. Frank, “Monopoly Pricing When Customers Queue,” IIE Transactions, Vol.36, No.6, pp.569-581, 2004.
[12] H. Chen and Y. W. Wan, “Price Competition of Make-to-Order Firms,” IIE Transactions, Vol.35, No.9, pp.817-832, 2003.
[13] G. Allon and A. Federgruen, “Competition in Service Industries,” Operations Research, Vol.55, No.1, pp.37-55, 2007.
[14] J. Anselmi, U. Ayesta, and A. Wierman, “Competition Yields Efficiency in Load Balancing Games,” Performance Evaluation, Vol.68, No.11, pp.986-1001, 2011.
[15] Y. Feng, B. Li, and B. Li, “Price Competition in an Oligopoly Market with Multiple IaaS Cloud Providers,” IEEE Transactions on Computers, Vol.63, No.1, pp.59-73, 2014.
[16] C. Kilcioglu and J. M. Rao, “Competition on Price and Quality in Cloud Computing,” In: 25th International Conference on World Wide Web, pp.1123-1132, 2016.
[17] J. Cao, K. Hwang, K. Li, and A. Y. Zomaya, “Optimal Multiserver Configuration for Profit Maximization in Cloud Computing,” IEEE Transactions on Parallel and Distributed Systems, Vol.24, No.6, pp.1087-1096, 2013.
[18] S. M. Ross, Stochastic Processes, Wiley, 1996.
[19] J. Medhi, Stochastic Models in Queueing Theory, Academic Press, 2003.
[20] P. Naor, “The Regulation of Queue Size by Levying Tolls,” Econometrica, Vol.37, No.1, pp.15-24, 1969.
[21] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
[22] en.wikipedia.org/wiki/Stackelberg_competition

저 자 소 개

이 두 호(Doo Ho Lee) 정회원

▪ 2006년 8월 : 동국대학교 산업시스템공학부 산업공학전공 학사
▪ 2008년 8월 : KAIST 산업 및 시스템공학과 석사
▪ 2012년 2월 : KAIST 산업 및 시스템공학과 박사
▪ 2012년 9월 ~ 2015년 2월 : 한국전자통신연구원 선임연구원
▪ 2016년 3월 ~ 현재 : 강원대학교 소프트웨어미디어・산업공학부 조교수

<관심분야> : 확률모형, 클라우드 컴퓨팅