Future Generation Computer Systems 102 (2020) 173–189

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Two-stage game theoretical framework for IaaS market share dynamics

Mona Taghavi [a,*], Jamal Bentahar [a], Hadi Otrok [a,b]

[a] Concordia Institute for Information System Engineering, Concordia University, Montreal, Canada
[b] Department of ECE, Khalifa University, Abu Dhabi, United Arab Emirates

HIGHLIGHTS

- We propose a Stackelberg game to capture the user demand preferences.
- A differential game is proposed for IaaS providers to compete over service quality.
- Two-stage game allows the new IaaS Providers to have a share in the market.
- Experiments proved the best strategy is to increase service cost and quality.

ARTICLE INFO

Article history:
Received 11 October 2018
Received in revised form 19 July 2019
Accepted 5 August 2019
Available online 6 August 2019

Keywords:
Dynamic game
Cloud service quality
Cloud computing
Market competition
Multi-tenancy
Customization
Sensitivity analysis

ABSTRACT

In this paper, we consider the problem of cloud market share among Infrastructure as a Service (IaaS) providers in a competitive setting. The public cloud market is dominated by few large providers, which prevents a healthy competition that would benefit the end-users. We argue that to make the cloud market more competitive, new providers, even small ones, should be able to inter this market and find a share. This problem of deeply analyzing the cloud market and providing new players with mechanisms allowing them to have a market share has not been addressed yet. In fact, to make the cloud market open and increase the cloud service demand, we show in this paper that the cloud providers have to compete not only over price, but also quality. Most of the research performed in the cloud market competition focus only on pricing mechanisms, neglecting thus the cloud service quality and user's satisfaction. However, to be consistent with the new era of cloud computing, Cloud 2.0, providers have to focus on providing value to businesses and offer higher quality services. As a solution to the aforementioned problem, we propose a conceptual, user-centric game theoretical framework that includes a two-stage game: 1) to capture the user demand preferences (optimal capacity and price), a Stackelberg game is used where IaaS providers are leaders and IaaS users are followers; and 2) to enhance the service ratings given by users in order to improve the provider position in the market and increase the future users' demand, a differential game is proposed, which allows IaaS providers to compete over service quality (e.g., QoS, scalability and adding extra features). The proposed two-stage game model allows the new IaaS providers, even if they are small, to have a share in the market and increase user's satisfaction through providing high quality and added-value services. To validate the theoretical analysis, experimental results are conducted using a real-world cloud service quality feedback, collected by the CloudArmor project. This research reveals that due to the fact that service customization tends to enhance the customers loyalty in today's subscription cloud economy, the best strategy for small IaaS providers is to increase the service cost and improve the quality of their added-value solutions to prevent customers' defection. This not only elevates the provider's profit, but also increases the quality equilibrium that leads to a higher user satisfaction. Consequently, higher satisfaction enhances the provider's rating and future users demand.

1. Introduction

1.1. Motivations

The rising demand in the cloud infrastructure service market has tempted a large number of technology providers to

* Corresponding author.
E-mail address: m_tag@encs.concordia.ca (M. Taghavi).

https://doi.org/10.1016/j.future.2019.08.003

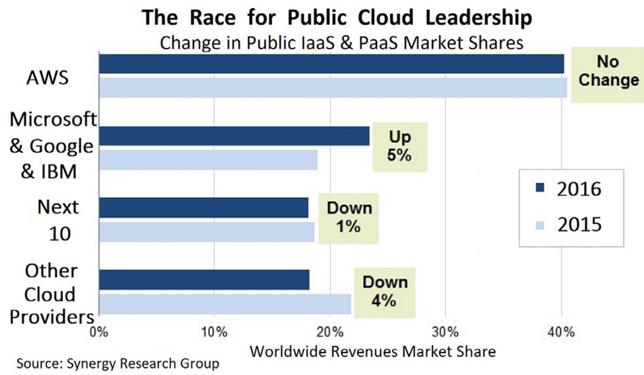


Fig. 1. Cloud revenue race among IaaS providers.

participate and compete in the market [1]. However, today's cloud market is dominated by only few large providers. As reported by the Synergy Research Group 2017,[1] Amazon, Microsoft, Google, and IBM gained ground in the market at the expense of smaller IaaS providers. The medium sized IaaS providers lost 1% of the market and a large number of small IaaS providers collective market share dropped by 4%, as illustrated in Fig. 1. Such a dominated market prevents a healthy competition. It also hinders compatibility with private clouds and prevents offering personalized added-value services by resellers [2]. Lack of these services may threaten the wide adoption of cloud computing in many industries. Thus, for the growth of the cloud computing industry, there is an increasing need to open the market to the new and smaller providers and create a more competitive environment.

Cloud IaaS has been in the center of attention for years and several research proposals about the technology itself have been lunched. Nonetheless, there is an urgent need to explore and address the business issues surrounding cloud computing, while considering the technical characteristics of such a paradigm. Nowadays, online market and rating platforms made it easy for users to compare a wide range of infrastructure services and for IaaS providers to establish their own credibility. In this paper, we argue that each IaaS provider entering the market needs to distinguish itself from the already established players and compete over both price and quality. As outlined in [3], today's market of Cloud 1.0 is price-focused. For that reason, there are extensive research that considered pricing competition and proposed optimal pricing strategies in order to maximize the final revenue of cloud providers [4,5]. However, there is a large number of modern business applications for which a price-focused service model will not be adequate. Often, users hesitate to move their critical business process to the cloud since the first-generation cloud obscured its operations detail behind its low pricing models [3]. Hiding the details blurs the vision of customers about the trade-offs that the IaaS provider has made in order to offer computing at such a low price.

The new era of cloud computing, Cloud 2.0, has been emerged to focus on providing value to small and medium enterprises (SME) as well as large enterprise markets at higher costs as well as higher quality [3,6–9]. For the revolution of Cloud 2.0 to take place for IaaS, two transformations need to occur: (1) IaaS providers must be prepared to provide value to businesses that entices them out of their built-in IT resources and applications; and (2) customers must demand a combination of fast, secure, and reliable IaaS from the providers to meet their end users' expectations [10]. In fact, data security and privacy are highly important in the context of Cloud 2.0 where cloud, fog and IoT must be consolidated and application providers are granted privileges to use and process the data [11]. In this context, to ensure the availability and delivery of low-latency services, Cloud 2.0 can be integrated with fog and edge computing to deal with the massive data volumes being produced by devices and users [12].

As an example of a cloud provider moving towards this revolution, SITA[2] is an IaaS provider that offers mobility-friendly on-demand hosting and application services specifically designed for the air transport industry. SITA has connected more than 160 airports which enabled the organization to host applications accessing to airports systems, such as terminals, gates and parking. A research conducted by Microsoft Cloud and Hosting Study[3] also confirmed the Cloud 2.0 movement by showing that 89% of companies are willing to pay additional fees for cloud management services. Despite the large number of pricing competition models, to the best of our knowledge, no one tackled the issue of the cloud providers competition from the perspective of service quality and end-users satisfaction. The only study about quality competition has been conducted by Fan et al. [13] who considered market competition among a software as a service provider and a traditional software provider. Their research focus on marketing advantages of bundling software in a service, neglecting the tight competition among cloud providers themselves and the user satisfaction effect on providers' revenue.

Considering the initiatives of Cloud 2.0 movement, this paper promotes a healthy market competition through rigor economical and theoretical models. To build a practical roadmap, we propose to empower new and small providers by considering two key features of Cloud 2.0:

1. High quality services: Considering the increasing number of clouds deployed in private data centers, the classic approach, such as the one used by Amazon, to build a cloud in which hardware and software developments are insourced, is no longer efficient and hardly deployable. Instead, clouds are being built out of commercial technology stacks with the aim of enabling the infrastructure providers to access the market rapidly and compete while providing high-quality services. However, finding cost-efficient component technologies offering high reliability, continues support, adequate quality, and easy integration is highly challenging. Unlike most of the research performed in the cloud market competition focusing only on pricing mechanisms, we model the competition from the perspectives of cloud service quality and user's satisfaction by focusing on added-value and superior quality services. Enabling small or new providers to access the market and offer personalized added-value services within our proposed model is part of this feature of Cloud 2.0 that enhances compatibility with private clouds.
2. Long-term commitment: The success of modern business applications relies on the reliability of services, such as incident response, security hardening, SLA assurance, software updates, and performance tuning. In fact, 80 percent of downtime is caused by service provisioning problems. Traditionally, these services have been delivered by the IT departments, and simply deploying remote servers in the cloud does not solve the services problem. Because services in the cloud will most likely be outsourced, they must be delivered while considering the customer's needs in a long-term commitment vision. Moving towards this long-term commitment strategy will drive providers to better

[1] www.srgresearch.com.

[2] https://www.sita.aero/.

[3] http://partner-l1.microsoft.com/hosting-cloud-research-report-2017.

focus on customer satisfaction to enjoy higher benefits. Our simulations also confirmed that providing added-value services along with customization could increase long-term commitment which is indeed very profitable.

1.2. Problem statement and contributions

In this paper, we consider the problem of IaaS cloud market share taking into consideration the need for new cloud providers to be in the market and the requirements of Cloud 2.0. We propose a conceptual, user-centric two-stage game theoretical framework that can help the IaaS providers and users optimize the service quality with a balanced profit. The first stage of our conceptual framework uses our Stackelberg game [14] to identify the user demand preferences and set the optimal price and capacity for the IaaS provider. The Stackelberg game model focuses on interaction among a single IaaS provider with a group of users to appropriately capture the demand elasticities and set the price and allocate resources for each Virtual Machine (VM) to meet the Service Level Agreement (SLA) and match the customers interests. However, our Stackelberg model does not consider the competition among the IaaS providers to provide higher quality services. Therefore, in the second stage, we formulate this competition through a differential game with service quality features as the main competitive factors. Most of the studies on strategic interactions among the cloud participants are grounded in static frameworks [15,16]. These models overlook the strategic issues arose when providers interact repeatedly over time. Thus, to tackle the limitation of static frameworks, we introduce a non-cooperative dynamic differential game that captures the important dimension of time.

The designed differential game takes multi-tenancy property into account, which leads to define competitive advantages for both the large and small IaaS providers. The large providers (the market leaders) make their profit through a virtuous cycle reflected through the following causal associations: (1) the more customers an IaaS provider gets, the more infrastructure and the better resource provision with robust cloud features (e.g., higher availability and more storage) it can afford; (2) the more infrastructure, the better economies of scale and the cheaper prices for IaaS; and (3) the lower prices and the better their quality, the more customers the provider can get. Meanwhile, the small IaaS providers have fewer users and limited resources. Thus, by targeting a specific industry or local region, they can have tenants who share the same scheme with similar requirements such as complying with data and security regulations, national and international standards or dealing with compatibility issues. This enables them amalgamate their needs by customizing their services to add value to the users' business solutions. Providing personalized cloud services can further drive customer loyalty [17]. To reflect the above arguments and take them into account, we introduce three main competition factors including ratings by users that reflect customers satisfaction, low cost QoS provisioning, and customization or added-value services.

In summary, our main contribution is a two-stage game theoretical model that:

- Allows new and small IaaS providers to compete against the existing and large ones and have a market share, which enables a productive cloud market industry that benefits the end-users. To the best of our knowledge, our work is the first that investigates this competition in the cloud computing context.
- Maximizes users satisfaction modeled using users' ratings by providing a continues service quality development. It is the first research that models a dynamic competition considering the quality of service among IaaS providers.
- Captures user preferences and demand elasticities for optimal price and resource allocation. To ensure the continued validity of the optimality in the presence of changing internal or external factors, a post-optimality analysis is provided.

The proposed model can help new born IaaS providers identify their users' needs and potential markets, anticipate their competitive advantage, formulate their valuation model and create new service provisioning scenarios. We implement our model using a real-world dataset containing users' ratings over cloud service quality features, obtained from the CloudArmor project.[4] Finally, it is worth mentioning that because the problem of making the cloud market competitive by analyzing how small providers can get a market share has not been addressed yet, no benchmark has been found for the purpose of comparison.

2. Related work

Small and medium businesses can take the advantage of cloud computing in several ways [18]. Cloud computing offers scalable services that businesses can use on demand as much as they need to. The competitive market of cloud services provides a variety of options in pricing and quality. The users can always shift their host provider to another provider offering more opportunistic service or lower price. Due to this opportunistic characteristics, this industry is predicted to reach $270 billion in 2020 [19]. Cloud economics plays a significant role in shaping the future of cloud computing industry. The economics of the cloud computing can have two dimensions [20]: (1) intra-organization that deals with internal factors such as labor, power, hardware and so on; and (2) inter-organization that refers to market competition factors between organizations such as price, quality of service, and reputation. A third dimension can also be considered where providers can adopt a cooperation strategy by forming coalitions or federations among data centers [21]. In such federations, different challenging problems have been addressed including virtual network provisioning [22] and trust management [23]. This paper deals with the second dimension. In this section, we present the work related to market share modeling from economics and marketing literature followed by the work done related to cloud services quality and pricing strategies.

2.1. Market share dynamics

Most of the proposals in the literature about market share are static [24]. A non-static approach has been taken by Breton et al. [25], who studied dynamic equilibrium advertising strategies in a duopoly market. They defined a model to formulate the market share dynamics for two competitors and obtained a feedback differential Stackelberg equilibrium. Gutierrez et al. [26] analyzed the dynamic strategic interactions between a manufacturer and a retailer in a distribution channel for innovative products. The underlying assumption was that the retail demand for such a product is influenced by word-of-mouth from past adopters. This influence creates a trade-off between immediate and future sales and profits of the manufacturer. The obtained equilibrium dynamic pricing showed that in some cases, far-sighted retailer is more profitable. The above mentioned studies utilize differential game to help businesses optimize their sale and advertisement channels regardless of the customer satisfaction, while this paper considers the technical characteristics of infrastructure cloud computing environment to distribute a

4 https://cs.adelaide.edu.au/~cloudarmor/ds.html.

fair market share among IaaS providers and fulfill the users' requirements.

Only few proposals have explored the users' ratings impact on business owners profit [27]. Nonetheless, their importance in marketing strategies has been recognized [28]. Duan et al. [29] studied video sales and movie recommender systems and found that users' ratings reflect movies quality, but they do not persuade the users to buy. In fact, they increase the users' awareness by word-of-mouth that is central to the efficacy of providers and increases their sales directly. Completing their study, our research proves that cloud service quality significantly affects the overall users' ratings, and further shows how cloud providers can take the advantage of those ratings to enhance reputation and increase profit.

2.2. The competition among cloud service participants

Game theory has been successfully applied in the cloud computing area, for instance for resource allocation and pricing mechanisms, where the interactions of players have to be taken into account [20]. A user-provider interactive approach is taken by Hadji et al. [30], where a Stackelberg game is designed to consider constrained pricing with limited resources offered by a cloud service provider and the optimal user demands. Xu et al. [5] optimized a pricing policy for cloud service providers to better compete with each other under the evolution of the cloud market. Forming a Stackelberg game, the authors applied a reinforcement learning (Q-learning) to find out an optimal policy for the leader provider. Following the leader, the optimal policy for followers will be uncovered. In the same line of research, Shen et al. [31] used a Stackelberg game to model the interactions among data providers, service providers, and users. The authors studied the optimization problem of the players' profits using deep learning in a context of data markets. However, price is the only utility factor considered in these studies and the importance of QoS and user satisfaction is somehow neglected.

Zhao et al. [32] investigated the impact of the two factors of energy consumption as well as SLA violations on degrading the cost-efficiency of data centers and the cloud providers' revenue. The authors developed online VM placement algorithms as an optimization problem of maximizing revenue from VM migration and achieved promising results. The research conducted by Kilcioglu et al. [33] calibrated a static model for price–quality trade-off in two cases of monopoly and duopoly price competitions where the IaaS marketplace is referred to as commoditized from the perspective of economic competition. The reason is that cloud providers use similar physical hardware which cannot be differentiated from each other and profit margins should become null. The conducted experiment explained the price cutting behavior of the current market trend and also how providers are able to make a profit despite predictions that the market should be totally commoditized. Conversely, this paper emphasizes a different approach aligned with the vision of Cloud 2.0. Commoditization for young and small competitors is not profitable and these providers cannot survive in the market of Cloud 2.0 due to their lower number of users and higher expenses. We advocate smaller providers to differentiate themselves from the established large providers in the market by providing added-value services to their customers.

The only study on cloud service quality that inspired our research is performed by Fan et al. [13] who considered market competition among a software as a service provider and a traditional software provider as a differential game. This research analyzes a short and long-term competition for price and dynamic quality between the two firms. The authors found that the cost of software implementation can significantly affect the equilibrium price while quality improvement has a more robust effect. Our work differs from this research in many points: (1) we focus on internal competition among IaaS providers considering the technical advantages and challenges specific to IaaS, specifically when a new provider enters the market to compete with big and dominant providers; (2) the user demand is formulated based on the user preferences and the two proposed game models prioritize the user satisfaction considering price, capacity and quality optimization; and (3) our model contains a continuous game loop where the players enter two different games and can evaluate post-optimality analysis to choose the right game, the right stage, and the right time to enter and to stay. Our post-optimality analysis also informs the players about the changes to the optimum values as they change over time.

3. Framework overview

The race to maximize the revenue, specifically for the new entrants to the cloud market, entails formulation of non-cooperative games. We form two key competing players representing each a group of the same type: (1) a small and fresh provider, and (2) a large and reputed provider. Conventional game theoretic frameworks modeling competition among players highlighted static models. A dynamic model enables the dimension of time, highly important to recognize the competitive decisions that change over time. Differential games that are dynamic in nature provide powerful tools to model competition in continuous time. In these games, critical state variables, such as demand and market share, are changing over time according to specific differential equations. Differential games have been widely applied in various domains, for instance to analyze competition in dynamic advertising and pricing [26].

The proposed conceptual model is specifically designed for the IaaS market. It is worth mentioning that game theory models could be applied not only to IaaS, but also to SaaS and PaaS. In fact, game theory, as a formal and mathematical framework that studies strategic interactions among players for decision making has been successfully applied in the cloud computing area, for instance for resource allocation and pricing mechanisms. However, since the technicalities and concerns differ from one layer to another, we argue that one model should focus on one particular layer. In this paper, the focus is on IaaS, and the designed games target the IaaS layer with its specific characteristics. Among which, multi-tenancy at the infrastructure level with performance segmentation [34] is specifically considered and modeled to drive competitive advantages in our model. Another significant aspect to be considered in this context is that multi-tenancy in other layers is highly different in nature and scale and goes beyond the infrastructure. Thus, at the SaaS level, multi-tenancy allows the database schema to be shared to support customization of the business logic, workflow and user-interface layers. For instance, Salesforce.com has 72,500 customers who are supported by 8 to 12 IaaS multi-tenant instances, where each one of these instances supports 5000 tenants who share the same database schema at the SaaS level. This difference in the multi-tenancy scalability between the two layers affects the pricing models and the competition strategies consequently. Our model proposes customization as a competitive advantage for IaaS, while it is considered as a must in the SaaS layer. Therefore, our mathematical model fits only IaaS providers. Other technical characteristics that make our model appropriate for IaaS but not SaaS are the important IaaS decision variables, namely: VM request size and preserved capacity, and rate of IaaS demand increase and drop, which differ categorically from the distribution of demands on software.

Considerable research efforts have been undertaken examining the physical and hardware aspects of IaaS. The main objective of this paper is to contribute to the business aspects of IaaS market and explore the business issues surrounding cloud computing while considering its technical characteristics. Thus, we do not use hardware and network measurements such as CPU, memory, hard disk, and transfer rates between VMs, but instead, we measure the price, service quality factors (based on SLA) and end-user satisfaction level, which are related to and relevant for Cloud 2.0.

This research tackles the problem of maximizing the IaaS providers' revenue through two interactive games in a cycle with 11 steps, as presented in Fig. 2. The first stage is the cloud market identification and demand provisioning for a new IaaS provider. The box on the top including the service selection Stackelberg game illustrates six interactive steps among an IaaS provider as the leader and the IaaS users as the followers. In the first and second steps, the IaaS provider k_J announces its price, quality and the average rating obtained so far. Then, the users decide the amount of their request (step (3)). The IaaS provider predicts the future user rating, plans the optimal capacity and offers the actual price under guaranteed SLA (steps (4 and 5)). In the final step (6), the user provides its rating. A brief explanation of this model needed in the rest of this paper is provided in Section 4 and more details can be found in our previous work [14]. This game produces two outcomes: optimum price and capacity of VM (P^*, ϕ^*).

After setting the price and capacity (step (7)), the provider enters into the second game (called differential competition game) which is proposed in this paper. During step (8), the IaaS provider has to compete with all the existing IaaS providers to enhance its service ratings through a justified amount of quality increments. The IaaS quality factors include functional and non-functional attributes such as QoS, adding new features, scalability and security. As our objective is to analyze the entrance of new providers to the cloud market, we denote by k_1 a typical small and new IaaS provider, and by k_2 a typical well-established IaaS provider competing against k_1. The outcome of the differential game (Game 2) is the required amount of quality improvement during the time interval $[0 - T]$ for a given T.

In the meantime, the optimality of the obtained values has to be analyzed since the game is dynamic and the values of the variables are changing. The users request (in terms of VM) arrival rate l that depends directly on the number of end users, will be used to assess the optimality of VM's price and capacity in step (9). Thus, if the variation of l remains less than a threshold, no changes are required and players shall stay in the second game (step (10): No). However, if the variation exceeds the threshold, a new optimal price has to be calculated using the new value of l (step (10): Yes). In the event that the price deviates from a certain threshold (the game sensitivity analysis), the game players have to go back to the Stackelberg game (Game 1) and start over the game, which includes computing the provider and users' best responses (step (11): Yes). Otherwise, the two IaaS providers only need to recompute their own best responses and obtain a new price and capacity through Game 1 (step (11): No).

The choice of the different techniques of our approach is not random, but strongly motivated by the appropriateness of the game theoretical models to tackle the problems addressed. The choice of game theory is motivated by the need to model the strategic interactions among different types of providers and by the fact that the cloud market includes selfish and utility-maximizer agents whose strategic actions influence each other. In fact, the first game part, namely the Stackelberg game, is the de facto approach to model strategic interactions such as the ones we model in this paper, where some players are leading the market, and where the actions of the other players are constrained by the first move of the leaders. In this context,

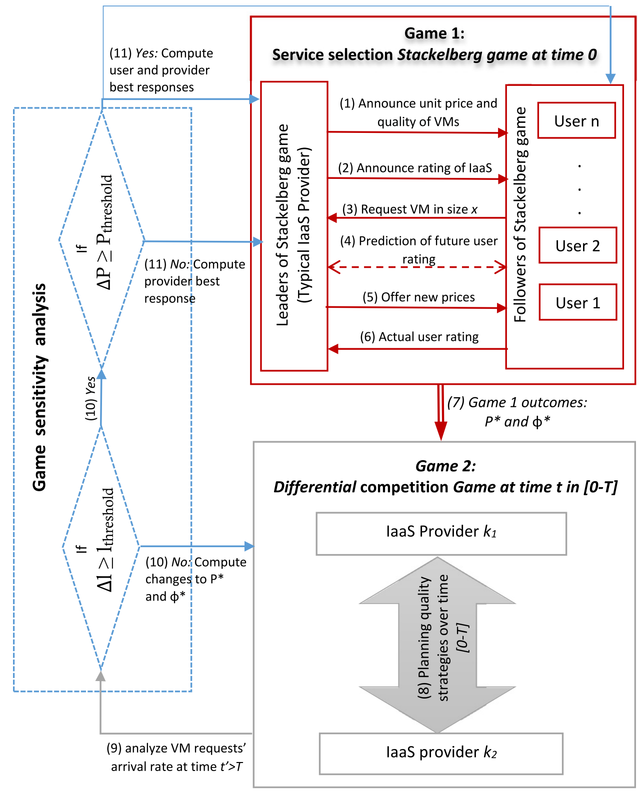


Fig. 2. Hierarchical Stackelberg and differential games' framework.

increasing the utility of cloud service consumers can decrease the profit of providers if not strategically planned. Stackelberg game provides a powerful framework to model and analyze this situation in which cloud providers can see their customers' expectations (i.e. quality–price tradeoff), and then optimize their price and preserved the capacity to earn user satisfaction as well as high ratings. The second part, namely the differential game is motivated by the dynamic and time-sensitive aspect of the competition among cloud providers, modeled as a conflict. Differential games have been proved to be highly appropriate for the modeling and analysis of conflicts in the context of dynamic systems that cannot be captured using static approaches where state variables evolve over time according to given differential equations. Our non-cooperative differential game allows us to model the maximization of the total discounted IaaS providers payment over the planning horizon [0 -T] as an optimal control problem. Moreover, the game allows us to capture the competitive factors and their effects on the providers' profit through the important dimension of time.

Remark

In concrete cases, IaaS providers have multiple types of VMs with different prices and performances. However, this situation does not limit the applicability of our model. We only considered one type of VM for the sake of simplicity and presentation in our mathematical modeling. Since there is no interdependence of VM factors in our model, a real IaaS provider who provides more than one type can use the model for every single type of VM considering the competitors who provide the same type.

Thus, in our game 1 (the service selection Stackelberg game), players will play repeated games in parallel; each game will consider one VM type. Thus, the leader will announce in parallel different prices and qualities, each price and quality are

Table 1
Notations used in the service selection Stackelberg game.

Decision variables	
x_i	VM request size of user i for the IaaS
ϕ	VM preserved capacity
P	Price per VM of the IaaS
Input parameters	
$i = 1, 2, \ldots n \in N$	Index of n users in the set N
B_i	User i budget
R_i	Rating utility of IaaS provider from user i
r_i	Service rating of user i for the IaaS provider
α_i	IaaS price elasticity for user i
β_i	IaaS rating elasticity for user i
γ_i	Size and number of VMs elasticity for user i
l	VM requests' arrival rate
μ	Constant scale of user IaaS demand
Q	Guaranteed QoS as stated in SLA
C_0/C	Fixed/marginal cost of the infrastructure
$\lambda_{i1}/\lambda_{i2}/\lambda_k$	Lagrange multipliers

announced in one individual game. The users (i.e. the followers) will react to each of those individual games, and a same user can also play several games in parallel. Each individual game 1 will produce the optimum price and capacity of one VM type, which will be used in game 2 (differential competition game). The same scenario continues then in this second game. Consequently, the overall situation can be seen as many parallel and independent two-stage games, each of which is considering one VM type. For example, let us consider Provider 1 providing VM type 1 and compare it with Provider 2 who has VM type 1. In parallel, Provider 1 with VM type 2 can compete with the same or another provider providing type 2.

4. IaaS selection Stackelberg game

In the first game, we formulate a Stackelberg game that models the cloud service market interactions between a typical IaaS provider as a leader and the service users as followers. The users observe the price and ratings to adjust their demand accordingly. In quest of the users' demands, the IaaS provider makes a decision on its pricing strategy and optimal capacity. The game parameters are provided in Table 1. We define the user demand as a Cobb–Douglas function to model demand elasticities and variations specific for each user in terms of price and rating [35]. It is assumed that the user will have the opportunity to verify the IaaS provider rating that reflects the actual user satisfaction level. The user demand function is defined as follows:

$$D_i(x_i, P, r_i) = \mu\ x_i^{\gamma_i}\ P^{-\alpha_i}\ r_i^{\beta_i} \tag{1}$$

where α_i, β_i and γ_i, $i = 1, 2, \ldots, n$ are elasticities (variations) of the IaaS price P, rating r_i and VM size x_i respectively. The price elasticity α_i is dependent on the user i because it reflects the price the user is willing to pay for the provided infrastructure. Market users have different requirements and their satisfaction levels differ accordingly. Thus, users do not react evenly to the same price or rating. It is the combination of these factors that produces different values of α_i, β_i and γ_i. The user aims to maximize its payoff as follows:

$$\begin{aligned} &\text{maximize} \quad UP(x_i) = D_i(x_i, P, r_i) - P \\ &\text{subject to} \quad Px_i \le B_i, x_i \ge 0 \end{aligned} \tag{2}$$

The IaaS provider is required to process users requests on time to maintain its reputation through the user ratings. Thus, it is highly desirable to consider VMs processing rate that shows VM capacity of the IaaS provider, denoted as ϕ. A large processing rate requires increasing the number and capacity of VMs, meaning a higher cost for ϕ that includes fixed cost of C_0 and marginal cost of C. Thus, the total cost for VM capacity ϕ is $C_0 + C\phi$.

Following previous literature in cloud computing [13], we model the arrival of VM requests as a Poisson process with mean arrival rate l. The average delay for a request in an M/M/1 queue can be defined as $\frac{1}{\phi - l}$. The IaaS provider aims to optimize its profit by increasing the price and ratings given by the users, and minimizing the costs. Thus, the IaaS provider's (i.e., the leader) optimization objective is:

$$\begin{aligned} &\text{maximize} \quad PP(P, \phi) = \sum_{i=1}^{n} (P - \ \phi\ C)\ D(x_i^*, P, r_i) + \sum_{i=1}^{n} R_i - C_0 \\ &\text{subject to} \quad \frac{1}{\phi\ -\ l} \le Q \\ &\qquad\qquad\quad \phi > 0, P > 0 \end{aligned} \tag{3}$$

R_i is the rating utility that is affected positively if the given rating is above the average user rating, and is affected negatively if the rating is below that average. Details of user rating prediction and rating utility calculation can be found in our previous work [14]. As usual, we use backward induction to find the equilibrium point of our Stackelberg game. Consequently, the followers' problem is solved first to get the response function of the users, and then the leader's decision problem is computed considering the possible reactions of the followers in order to maximize the provider's profit.

4.1. User best response

Because the objective function shown in Eq. (2) is continuous and concave with regard to x_i, we use the Lagrange multipliers, λ_{i1} and λ_{i2}, with Kuhn–Tucker conditions to obtain the solution. So, we will have a new objective function as follows:

$$L_{UP} = D_i(x_i, P, r_i) - P - \lambda_{i1}(x_iP - B_i) + \lambda_{i2}x_i \tag{4}$$

with the following conditions:

$$\lambda_{i1}(x_iP - B_i) = 0 \tag{5}$$

$$\lambda_{i2}x_i = 0 \tag{6}$$

$$\lambda_{i1}, \lambda_{i2}, x_i \ge 0$$

The only coupling point between users is x_i, so we take the derivative with respect to x_i.

$$\frac{\partial L_{UP}}{\partial x_i} = \frac{\partial D_i(x_i, P, r_i)}{\partial x_i} - \lambda_{i1}P + \lambda_{i2} = 0 \tag{7}$$

We have two cases: (1) $x_i = 0$: regardless of the value of λ_{i1} and λ_{i2}, this means the user is not demanding any services, so its utility will be null; and (2) $x_i > 0$: from slackness complementary condition in Eq. (6), we can conclude that $\lambda_{i2} = 0$; so we have:

$$x_i = \left(\frac{\lambda_{i1}P^{\alpha_i+1}}{r_i^{\beta_i}\gamma_i\mu}\right)^{\frac{1}{\gamma_i-1}} \tag{8}$$

By substituting x_i from Eq. (8) in Eq. (5) we obtain λ_{i1}:

$$\lambda_{i1}^{\frac{1}{\gamma_i-1}} = \frac{B_i r_i^{\beta_i}\gamma_i\mu}{P^{\frac{\alpha_i+1}{\gamma_i-1}+1}} \tag{9}$$

The final response x_i from user i is attained by replacing Eq. (9) in Eq. (8).

$$x_i^* = \frac{B_i(r_i^{\beta_i}\gamma_i\mu)^{\frac{\gamma_i-2}{\gamma_i-1}}}{P} \tag{10}$$

The above obtained x_i^* is optimal where Eq. (5) slacks and $\lambda_{i1} > 0$. However, we claim that it is reasonable to consider slackness rather than binding, since having $\lambda_{i1} = 0$ is an extreme case where the user cares only about the price and does not consider the previous ratings or quality.

4.2. IaaS provider best response

Using Lagrange multiplier λ_k, we model the objective optimization in Eq. (3) as follows:

$$L_{PP}(P,\phi,\lambda_k) = PP(P,\phi) - \lambda_k(\frac{1}{\phi - l} - Q) \tag{11}$$

The Kuhn–Tucker condition for our model is:

$$\frac{\partial PP(P,\phi)}{\partial \phi} - \lambda_k \frac{\partial(\frac{1}{\phi - l} - Q)}{\partial \phi} = 0 \tag{12}$$

$$\frac{\partial PP(P,\phi)}{\partial P} = 0 \tag{13}$$

$$\lambda_k(\frac{1}{\phi - l} - Q) = 0 \tag{14}$$

where $\lambda_k \geq 0, P, \phi > 0$. To find the optimal capacity ϕ, we first assume that Eq. (14) binds and $\lambda_k > 0$. Referring to Eq. (12) we have:

$$CD_i(x_i^*, P, r_i) - \lambda_k(\frac{-1}{(\phi - l)^2}) = 0$$

$$\lambda_k = -CD_i(x_i^*, P, r_i)(\phi - l)^2 \tag{15}$$

Knowing that $C > 0$ and $D_i(x_i^*, P, r_i) > 0$, we obtain a negative λ_k in Eq. (15) that contradicts the defined constraint $\lambda_k \geq 0$. Therefore, $\lambda_k = 0$ and Eq. (14) slacks, which means the IaaS provider should provide a higher VM capacity than what is promised in SLA. Any assigned capacity can be optimal as long as the following condition holds:

$$\phi^* = \frac{1}{Q} + l + \epsilon \tag{16}$$

ϵ represents a very small amount. By solving Eq. (13) we get the optimal price as follows:

$$P^* = \phi^* C(\frac{\alpha_i + \gamma_i}{\alpha_i + \gamma_i - 1}) \tag{17}$$

5. Differential competition game

In the previous section, we formulated a static game involving a typical IaaS provider and a typical IaaS user. However, the IaaS provider does not act alone in the market. After identification of the users' requirements, the IaaS provider needs to plan a suitable strategy against its competitors. To model dynamic competition among the IaaS providers over a period of time, we design a differential game with continuous strategies over a finite horizon time T. The game is between a typical new and small IaaS provider k_1 with n customers, and a typical large and established IaaS provider k_2 with m customers, $m \gg n$. The list of employed notations is given in Table 2. For the sake of simplicity, we use k when we refer to both providers. Considering the cost, technical reality and multi-tenancy characteristics of cloud computing, each IaaS provider faces different challenges to compete with high quality services. In the next section, we explain which of those can lead the way of IaaS providers.

5.1. IaaS architecture and competitive advantage

To scale the economical benefits and optimize resource utilization, multiple VMs are initiated on the same physical server simultaneously. Multi-tenancy implies multiple customers of a services set. For instance, multiple business units within a large organization with resources and data that should remain separate through a logical segmentation of the shared infrastructure by using software-defined technologies. The segmentation options available to be considered are: physical separation, logical separation, data separation, network separation, and performance separation. In the performance separation scheme, the infrastructure is shared but the capacity or QoS is guaranteed while no other separation scheme ensures such a quality. As discussed earlier, despite the tremendous momentum of the cloud computing, many firms are reluctant to move to the cloud due to the performance concerns. For that reason, we only consider the option of "performance separation" in our model, which remains a key component in the Cloud 2.0 movement. Considering the same scheme, each provider may take its own advantage to compete.

Competitive advantages of large IaaS providers (k_2): Tenants with guaranteed performance require consistency and predictability which are challenging for IaaS providers since infrastructure is shared by many tenants. For the sake of performance isolation, it is not enough to use host-based virtualization technologies since the bandwidth between VMs of the same tenant can change significantly over time. This variation depends on the network load and usage peak from other tenants. The larger number of customers, the less variation is expected from the overall average demand. A large IaaS provider that serves a large number of customers and operates within different industries and geographically dispersed locations can avoid the cost of overbooking. New scheduling algorithms allow multiple workloads on the same cluster of customers to access a common data pool along with hardware and software resources. Thus, larger providers can balance the workloads of the same clusters and would achieve the required performance with less preserved capacity [34]. On the other hand, a smaller IaaS provider with fewer customers has to provide a larger amount of reserved capacity to meet the variation. Another advantage of large infrastructure is energy saving cost of data centers. Large tenant clusters enable providers to shut down the idle servers and migrate tasks to other VMs running on active servers.

The multi-tenancy architecture is not visible to the user, however the user can observe its effects. The visible effects, such as higher availability and scalability, are directly reflected in our model through user rating (r_i) that ultimately increases the user demand and provider's revenue. Further, we consider the invisible effects of multi-tenancy for the IaaS providers (ζ_{k_2}). ζ_{k_2} is mainly considered as a discounted cost for the large provider due to its larger infrastructure as explained earlier.

Competitive advantages of small IaaS providers (k_1): Security regulations vary specially when the IaaS provider has to operate in diverse national and international markets. Thus, customers require to customize the security settings according to different country's regulations. The same thing can happen with regulated industries. Mobile realm is another example where customization is highly desired. Organizations are more and more adopting mobile applications and require to integrate them into their cloud infrastructure without introducing network risks. The lack of personalized infrastructure services may not allow organizations to gain the maximum potential value of their cloud investments. Despite the importance of customization, it is disputed that large providers are not willing to offer customized service-oriented architecture or application programming interfaces to SMEs [36].

Customized use of the cloud in a multi-tenancy environment is costly and hard to realize, unless the customers residing in a cluster share the same scheme. Having similar requirements (e.g., same regulations) enables the small IaaS provider to support added-value options for each application type. Since data sources for multiple tenants are in the same database, by using a simple data aggregation, the small IaaS provider can develop applications specific for its group of customers. Taking the explained example of SITA, the provider created a large network of organizations working in the air industry and expanding their offers to applications built on top of that network.

This competitive factor is not embraced in user ratings collected through online platforms and neither relates to cost. However, it definitely has a strong positive effect on the users' demand. Thus, we use ϑ_{k_1} to reinforce the formulated demand.

5.2. Differential game formulation

IaaS providers establish their credibility and gain their share through on-line market platforms where users can express their ratings. We make the following realistic assumption that is satisfied in market platforms in general.

Assumption 1. The quality of an IaaS has a significant effect on users' rating, so that improving the service quality will ultimately lead to the increase of the average users' rating.

Assumption 1 is the basis of our game design and is very important for the validity of our game. Thus, we will validate this assumption using statistical analysis in Section 7.

Conventionally, IaaS providers may compete over two factors: price and quality. Several research proposals showed that due to the cost of changing price and being inconvenient for the customers, prices do not change frequently [13]. Thus, we consider the optimum price as computed in the first game in Eq. (17) for each provider, and we assume it remains constant over the time interval $[0-T]$, while the quality can be updated throughout the game. Afterward, whenever the values of the defined parameters vary significantly enough to warrant a shift in price, the optimal price can easily be determined by solving the Stackelberg game with the new parameter values. Consequently, the new optimal price can be utilized to solve the differential competition game over the next period of time. The following definition explains the service quality factors featured in the game.

Definition 1. The quality factors of an IaaS can be any of the following elements:

- QoS: basic quality features such as response time, throughput, and availability.
- Adding new service features and innovative/customized offers to the existing service.
- Enhancing and optimizing cloud specific features such as elasticity, security, and storage space.
- Supporting customers technically or non-technically.

The next assumption establishes the initial conditions to formulate our differential competition game.

Assumption 2. Each player (IaaS provider) has perfect knowledge of:

- The function $\dot{D}_k(t)$ determining the evolution of the user demand, and the control path of $q_k(t)$ available to the two players.
- The payoff function PP_k.
- The initial demand state at time zero, $D_k(0)$.

Table 2
Notations used in the differential competition game.

Decision variable	
$q_k(t)$	Quality control path (quality improvement) of IaaS k at time t
Input parameters	
k_1	New and small IaaS provider
k_2	Existing and big IaaS provider
δ_k	Customers' defection rate to buy IaaS k
ρ	Discount rate of future IaaS provider's revenues
$\hat{\theta}_k$	Rate of IaaS k demands increase
$\check{\theta}_k$	Rate of IaaS k demands drop
η_k	Non-functional cost of IaaS quality (e.g. quality attributes achieved by preserving higher VM capacity)
f_k	Functional cost of IaaS quality (e.g. offering new features and improving technical support)
ζ_{k_2}	Rate of discounted non-functional cost due to large infrastructure for IaaS provider k_2
ϑ_{k_1}	Rate of customization value for customers of IaaS provider k_1
$[0-T]$	Time horizon of the game

However, players have no knowledge about the future states. So, they will not be able to observe the state and update their initial control path ($q_k(t)$) of quality improvement. The information structure of the game is open-loop. This means the players should make their decisions at time t only with the knowledge of the initial condition of the state at time zero. The intuition behind the selection of this information structure is that IaaS providers have to put some investment and make stable pricing strategies at the initial stage because changing these strategies bears some cost for both providers and their customers. Besides, quality improvement may result in the increase of immediate rating, but improving the average rating is a long-term strategy and is not observable in short time.

In traditional service trading models, once customers had chosen their providers, they tended to keep the relation working since the investment has been made through a long-term contract. In the subscription cloud economy, customers are much free to defect anytime from a provider and switch to another one as there is very little to no financial penalty to do so. There are several variables that affect the user's demand over time. The demand for an IaaS increases with its rating improvement. Due to the strong correlation between rating and quality as asserted in Assumption 1, we use quality instead of rating. Therefore, improvement of quality elevates the demand at a rate of $\hat{\theta}_k$. We also define a demand drop rate $\check{\theta} k$ when the other provider enhances its service quality. Moreover, customers may defect at a certain rate δ_k. Based on the predefined variables, the users' demand dynamics evolve according to the following equations:

$$\begin{cases} \dot{D}_{k_1}(t) = (\hat{\theta}_{k_1} + \vartheta_{k_1})\, q_{k_1}(t)^{\beta_n} - \check{\theta}_{k_1}\, q_{k_2}(t)^{\beta_n} - \delta_{k_1} D_{k_1}(t) \\ D_{k_1}(0) = D_{k_1 0},\, 0 < t < T \end{cases} \tag{18}$$

$$\begin{cases} \dot{D}_{k_2}(t) = \hat{\theta}_{k_2}\, q_{k_2}(t)^{\beta_m} - \check{\theta}_{k_2}\, q_{k_1}(t)^{\beta_m} - \delta_{k_2} D_{k_2}(t) \\ D_{k_2}(0) = D_{k_2 0},\, 0 < t < T \end{cases} \tag{19}$$

As discussed earlier, ϑ_{k_1} is the added-value to the quality for IaaS provider k_1. β_n and β_m denote the average users' sensitivity towards rating for k_1 and k_2 respectively. Eqs. (18) and (19) explicitly describe how the service quality of the two competitors jointly determine the dynamics of demand rate.

The marginal cost of increasing quality is considered to be quadratic in past studies [13]. We consider the same quadratic increment for the increase of quality. Thus, let $\hat{C}q_k(t)$ to be a cost associated with the efforts to increase the quality level by an amount $q_k(t)$ at time t. Two types of quality improvement are considered in our model: functional (f) and non-functional (η). The functional quality improvement is realized by adding extra functionalities to the service, such as offering new features or improving technical support. The non-functional one is related to the quality attributes that can be reached by reserving extra resources and increasing the processing capacity. Definition 2 determines the cost of quality improvement for both IaaS providers.

Definition 2. To increase the quality, k_1 and k_2 incur a quadratic cost function as follows:

$$\hat{C}q_{k_1}(t) = f_{k_1} q_{k_1}(t)^2 + \eta_{k_1} q_{k_1}(t)^2 \tag{20}$$

$$\hat{C}q_{k_2}(t) = f_{k_2} q_{k_2}(t)^2 + (\eta_{k_2} - \zeta_{k_2}) q_{k_2}(t)^2 \tag{21}$$

Concurring with the cost functions delineated by Definition 2, the instant profit of IaaS providers k_1 and k_2 at time t can be calculated according to the following formulas:

$$\begin{aligned} PP_{k_1}(D_{k_1}(t), q_{k_1}(t)) \quad &= (P^*_{k_1} - \phi^*_{k_1} C_{k_1})\, D_{k_1}(t) \\ &- (f_{k_1} q_{k_1}(t)^2 + \eta_{k_1} q_{k_1}(t)^2) - C_{0k_1} \end{aligned} \tag{22}$$

$$\begin{aligned} PP_{k_2}(D_{k_2}(t), q_{k_2}(t)) \quad &= (P^*_{k_2} - \phi^*_{k_2} C_{k_2})\, D_{k_2}(t) \\ &- (f_{k_2} q_{k_2}(t)^2 + (\eta_{k_2} - \zeta_{k_2}) q_{k_2}(t)^2) - C_{0k_2} \end{aligned} \tag{23}$$

Different from the instant profit of IaaS provider in the static game in Eq. (3), we do not consider the rating accumulation in the profit function. The reason is that we already considered the increase of future demand due to the enhanced user rating (quality) in Eqs. (18) and (19). The objective function is the total discounted IaaS provider's payoff over the planning horizon $[0 - T]$:

$$\begin{aligned} \text{maximize} \quad & \int_0^T e^{\rho t} \{PP_{k_1}(q_{k_1}(t), D_{k_1}(t))\} dt \\ \text{subject to} \quad & \dot{D}_{k_1}(t) = (\hat{\theta}_{k_1} + \vartheta_{k_1})\, q_{k_1}(t)^{\beta_n} \\ & - \check{\theta}_{k_1}\, q_{k_2}(t)^{\beta_n} - \delta_{k_1} D_{k_1}(t) \\ & D_{k_1}(0) = D_{0k_1},\, 0 < \beta_n < 1 \end{aligned} \tag{24}$$

$$\begin{aligned} \text{maximize} \quad & \int_0^T e^{\rho t} \{PP_{k_2}(q_{k_2}(t), D_{k_2}(t))\} dt \\ \text{subject to} \quad & \dot{D}_{k_2}(t) = \hat{\theta}_{k_2}\, q_{k_2}(t)^{\beta_m} - \check{\theta}_{k_2}\, q_{k_1}(t)^{\beta_m} - \delta_{k_2} D_{k_2}(t) \\ & D_{k_2}(0) = D_{0k_2},\, 0 < \beta_m < 1 \end{aligned} \tag{25}$$

ρ is a constant discount rate to rebate all the future costs and revenues' streams relative to the present. Note that Eqs. (24) and (25) formulate two optimal control problems with the service quality and the cumulative demand as control and state variables, respectively. In the following section we solve these optimal control problems.

5.3. Open-loop equilibrium solution

The optimal control theory provides appropriate techniques to analyze differential games [37]. To examine the dynamics of the payoff functions and the paths of control variables, we exploit the Hamiltonian systems. In the open-loop structures, equilibrium strategies can be discovered by computing the solution of a two-point boundary value problem for ordinary differential equations obtained from the Pontryagin maximum principle in Hamiltonian functions. The Pontryagin maximum principle allows us to derive the necessary conditions for a control path to be optimal open-loop control. The optimal control paths of quality are defined as follows.

Definition 3. For the IaaS provider k, the quality strategy $q^*_k(t)$ is optimal if the inequality $PP_k(D_k(t), q^*_k(t)) \geq PP_k(D_k(t), q_k(t))$ holds for all feasible control paths $q_k(t) \neq q^*_k(t)$.

To acquire the optimal control, we first formulate the Hamiltonian system of the IaaS providers' payoff which is quite similar to the Lagrangian method that we used in the first game.

$$\begin{aligned} & H_{k_1}(q_{k_1}(t), D_{k_1}(t), \lambda_{k_1}(t), t) = \\ & (P^*_{k_1} - \phi^*_{k_1} C_{k_1})\, D_{k_1}(t) - (f_{k_1} q_{k_1}(t)^2 + \eta_{k_1} q_{k_1}(t)^2) - C_{0k_1} \\ & + \lambda_{k_1}(t)((\hat{\theta}_{k_1} + \vartheta_{k_1})\, q_{k_1}(t)^{\beta_n} - \check{\theta}_{k_1}\, q_{k_2}(t)^{\beta_n} - \delta_{k_1} D_{k_1}(t)) \end{aligned} \tag{26}$$

$$\begin{aligned} & H_{k_2}(q_{k_2}(t), D_{k_2}(t), \lambda_{k_2}(t), t) = \\ & (P^*_{k_2} - \phi^*_{k_2} C_{k_2}) D_{k_2}(t) - (f_{k_2} q_{k_2}(t)^2 + (\eta_{k_2} - \zeta_{k_2}) q_{k_2}(t)^2) \\ & - C_{0k_2} + \lambda_{k_2}(t)(\hat{\theta}_{k_2}\, q_{k_2}(t)^{\beta_m} - \check{\theta}_{k_2}\, q_{k_1}(t)^{\beta_m} - \delta_{k_2} D_{k_2}(t)) \end{aligned} \tag{27}$$

The adjoint variable or shadow price (λ_k) related to a particular constraint reflects the change in the optimal value of the objective function per unit increase in the right-hand-side value of that constraint, under the condition that all the other problem data are unchanged. The economic interpretation of $\lambda_k(t)$ is the value of an additional unit of demand. For given $q_k(t)$, $\lambda_k(t) > 0$ implies that the IaaS provider benefits from current demands. With a zero shadow price $\lambda_k(t) = 0$, the IaaS provider does not take into account the impact of the quality on future user demands. On the other hand, when $\lambda_k(t) < 0$, the IaaS provider has no motive to sacrifice current profits for future profits, so that it will no longer elevate the service quality.

The optimal control strategy of the original problem, as outlined in the control theory, must also maximize the corresponding Hamiltonian function. Thus, based on the Pontryagin maximum principle, all candidate optimal strategies have to satisfy the following necessary conditions:

$$\begin{aligned} \frac{\partial H_{k_1}(t)}{\partial q_{k_1}(t)} = & -2(f_{k_1} q_{k_1}(t) + \eta_{k_1} q_{k_1}(t)) \\ & + \lambda_{k_1}(t)(\hat{\theta}_{k_1} + \vartheta_{k_1})\, \beta_n q_{k_1}(t)^{\beta_n - 1} = 0 \end{aligned} \tag{28}$$

$$\begin{aligned} \dot{\lambda}_{k_1}(t) = \rho \lambda_{k_1}(t) - \frac{\partial H_{k_1}(t)}{\partial D_{k_1}(t)} \\ =(\rho + \delta_{k_1})\lambda_{k_1}(t) - P^*_{k_1} + \phi^*_{k_1} C_{k_1},\, \lambda_{k_1}(T) = 0 \end{aligned} \tag{29}$$

$$\begin{aligned} \frac{\partial H_{k_2}(t)}{\partial q_{k_2}(t)} = & -2(f_{k_2} q_{k_2}(t) + (\eta_{k_2} - \zeta_{k_2}) q_{k_2}(t)) \\ & + \lambda_{k_2}(t)\hat{\theta}_{k_2}\, \beta_m q_{k_2}(t)^{\beta_m - 1} = 0 \end{aligned} \tag{30}$$

$$\begin{aligned} \dot{\lambda}_{k_2}(t) = \rho \lambda_{k_2}(t) - \frac{\partial H_{k_2}(t)}{\partial D_{k_2}(t)} \\ =(\rho + \delta_{k_2})\lambda_{k_2}(t) - P^*_{k_2} + \phi^*_{k_2} C_{k_2},\, \lambda_{k_2}(T) = 0 \end{aligned} \tag{31}$$

When only one boundary condition is specified as $D_k(0) = D_{0k}$, the free-end condition is used as $\lambda_{k_1} = \lambda_{k_2} = 0$ at $t = T$. It should be noted that the Pontryagin maximum principle is only a necessary condition, but not essentially sufficient for optimality. Consequently, the solution of the pair quality control in the above equations does not necessarily converge to the Nash equilibrium. To investigate the normality of our defined systems and to assess if Pontryagin can provide a sufficient condition for optimality in our case, we shall derive the monotonicity condition on the adjoint variables in Lemma 1. This condition is important since the adjoint variables significantly affect the payoff functions in our optimal control-based optimization.

Lemma 1. *With positive profit unit margins, we have $\lambda_{k_1}(t) > 0$ and $\lambda_{k_2}(t) > 0$ for all $t \in [0, T)$.*

Proof. Here we prove the monotonicity of $\lambda_{k_1}(t)$, and the same proof applies for $\lambda_{k_2}(t)$. As stated in Eq. (29), we have $\lambda_{k_1}(T) = 0$. Therefore, at $t = T$, $\dot{\lambda}_{k_1} = -P^*_k + \phi^*_{k_1} C_{k_1} < 0$, since $P^*_{k_1} > \phi^*_{k_1} C_{k_1}$. So, $\lambda_{k_1}(t) > 0$ as t approaches T. Now, consider $\lambda_{k_1}(t1) < 0$ for any given $t1$. Then we should have $\lambda_{k_1}(t2) = 0$ for some $t2 > t1$ and $\dot{\lambda}_{k_1}(t2) \geq 0$. Consequently, $\dot{\lambda}_{k_1}(t2) = -P^*_{k_1} + \phi^*_{k_1} C_{k_1} < 0$. This is a contradiction and $\lambda_{k_1}(t)$ is proved to be positive during the whole period of time. □

The following proposition is concluded from Lemma 1.

Proposition 1. *IaaS providers' profit optimization functions have a normal form maximum principle with positive adjoint variables* $(\lambda_{k_1}(t), \lambda_{k_2}(t))$ *associated with* $(q^*_{k_1}(t), q^*_{k_2}(t))$.

Lemma 2. *Pontryagin Maximum Principle (in Eqs. (28)–(31)) provides the necessary sufficient conditions of optimality and the control path pair of* $(q^*_{k_1}(t), q^*_{k_2}(t))$ *is optimal and unique.*

Proof. Proposition 1 asserts that the formulated profit optimization problems have a normal form. It suffices to prove that the Hamiltonian function is concave in $D_k(t)$ for both providers k_1 and k_2 in any $t \in [0 - T]$. Let $(q^*_{k_1}(t), D^*_{k_1}(t))$ be a pair that satisfies the Pontryagin condition for IaaS provider k_1, with $\lambda_{k_1}(0) = 1$, and for all admissible demand states, the limiting transversality condition holds: $\lim_{t\to T} \lambda_{k_1}(t)(D_{k_1}(t) - D^*_{k_1}(t)) \geq 0$. To prove the concavity of the dynamic function in $D_{k_1}(t)$, the following condition must hold:

$$H_{k_1}(q_{k_1}(t), D_{k_1}(t), \lambda_{k_1}(t), t) - H_{k_1}(q^*_{k_1}(t), D^*_{k_1}(t), \lambda_{k_1}(t), t) \leq \frac{\partial H_{k_1}(t)}{\partial D_{k_1}(t)}(D_{k_1}(t) - D^*_{k_1}(t)) \tag{32}$$

The left-hand-side of the inequality is negative since the Hamiltonian function in the optimal quality path is the maximum IaaS provider profit that is more than its profit at any other path in any time t. Thus, it is enough to prove that the right-hand-side of the inequality is positive. From Eq. (29), we can see that:

$$\frac{\partial H_{k_1}(t)}{\partial D_{k_1}} = \rho\lambda_{k_1}(t) - \dot{\lambda}_{k_1}(t) \tag{33}$$

Replacing Eq. (33) in Eq. (32), we get $(\rho\lambda_{k_1}(t) - \dot{\lambda}_{k_1}(t))(D_{k_1}(t) - D^*_{k_1}(t))$ in the right-hand-side. From the transversality condition, we already know that $\lambda_{k_1}(t)(D_{k_1}(t) - D^*_{k_1}(t)) \geq 0$, so it is enough to prove that $\dot{\lambda}_{k_1}(t)$ is negative. It is known in optimal control theory that the motion of shadow price is equal to the negative derivative of Hamiltonian towards the dynamic state, so that $\dot{\lambda}_{k_1}(t) = -P^*_{k_1} + \phi^*_{k_1} C_{k_1} - \delta_{k_1}\lambda_{k_1}(t) \leq 0$. The same logic applies for k_2. □

After proving the monotonicity of adjoint variables in Lemma 1 and the sufficiency of the Pontryagin maximum principle in obtaining the optimal solution in Lemma 2, we can obtain the optimal control path.

Theorem 1. *The finite horizon differential game in Eqs. (24) and (25) has a unique Nash equilibrium solution for the two IaaS providers. The optimal quality strategies are given by:*

$$q^*_{k_1}(t) = \left(\frac{(P^*_{k_1} - \phi^*_{k_1} C_{k_1})(\hat{\theta}_{k_1} + \vartheta_{k_1})\beta_n}{2(\rho + \delta_{k_1})(f_{k_1} + \eta_{k_1})}\right)^{\frac{1}{2-\beta_n}}\left(1 - e^{\frac{(\rho+\delta_{k_1})(t-T)}{2-\beta_n}}\right) \tag{34}$$

$$q^*_{k_2}(t) = \left(\frac{(P^*_{k_2} - \phi^*_{k_2} C_{k_2})\,\hat{\theta}_{k_2}\beta_m}{2(\rho + \delta_{k_2})(f_{k_2} + (\eta_{k_2} - \zeta_{k_2}))}\right)^{\frac{1}{2-\beta_m}}\left(1 - e^{\frac{(\rho+\delta_{k_2})(t-T)}{2-\beta_m}}\right) \tag{35}$$

Proof. The two formulated differential equations Eqs. (29) and (31) can lead us to the adjoint variables:

$$\lambda_{k_1}(t) = \frac{P^*_{k_1} - \phi^*_{k_1} C_{k_1}}{\rho + \delta_{k_1}}(1 - e^{(\rho+\delta_{k_1})(t-T)}) \tag{36}$$

$$\lambda_{k_2}(t) = \frac{P^*_{k_2} - \phi^*_{k_2} C_{k_2}}{\rho + \delta_{k_2}}(1 - e^{(\rho+\delta_{k_2})(t-T)}) \tag{37}$$

Replacing Eq. (36) in Eq. (28) and Eq. (37) in Eq. (30) gives us the optimal quality control paths. □

Differential games enable us to analyze the dynamic nature of competition and quality improvement. The following corollaries and propositions are inferred from Theorem 1.

Corollary 1. *Each provider's quality improvement decreases in its quality development cost.*

Proof. The decrease is straightforward from the first derivative of quality with respect to cost, $\frac{\partial q^*_{k_1}(t)}{\partial(f_{k_1}+\eta_{k_1})} < 0$ and $\frac{\partial q^*_{k_2}(t)}{\partial(f_{k_2}+(\eta_{k_2}-\zeta_{k_2}))} < 0$. However, the cost decrement slope is steeper for big providers due to serving a large number of customers. The difference is specifically reflected in the non-functional costs since functional costs are expected to be alleviated as the service becomes more mature. This corollary is an evidence of the economic benefits of continuous quality improvement for both IaaS providers to have a higher level of quality equilibrium as well as user rating. □

Corollary 2. *Higher level of customer loyalty and lower discount factor lead to a higher quality equilibrium for both providers.*

Proof. This corollary simply means the fewer IaaS providers' customer defection rate, the more incentive for the providers to improve their service quality. It can be inferred from the first order conditions for Hamiltonian systems of IaaS provider k_1 (Eq. (28)), where we have:

$$q^*_{k_1}(t) = \left(\frac{\lambda_{k_1}(t)(\hat{\theta}_{k_1} + \vartheta_{k_1})\beta_n}{2(f_{k_1} + \eta_{k_1})}\right)^{\frac{1}{2-\beta_n}}$$

This implies that the two variables of customer defection rate and discount factor are reflected through the value of the shadow price $\lambda_{k_1}(t)$. As t approaches the end of the time horizon, the negative effect of ρ and δ_{k_1} becomes more evident:

$$\lim_{t\to T} \lambda_{k_1}(t) = \frac{P^*_{k_1} - \phi^*_{k_1} C_{k_1}}{\rho + \delta_{k_1}} \quad \square$$

The same logic is applied for IaaS provider k_2. Thus, as the marginal values of customers drop, the service quality equilibrium shrinks.

Proposition 2. *The quality improvement of cloud services is higher in early stages and decreases over time.*

The service quality improvement rate is steeper at the beginning of the time horizon. As t approaches T, the improvement flattens out. The reason can be the maturity of the service, getting maximum user ratings, or adjustment of the service features and support.

Proposition 3. *Assuming that (1) both providers make the same revenue per unit service; (2)* $\delta_{k_1} = \delta_{k_2}$ *with the same user rating sensitivities; and (3)* ϑ_{k_1} *for IaaS provider* k_1 *and* ζ_{k_2} *for IaaS provider* k_2 *determine the quality level. If smaller providers do not take the advantage of customization and providing value for their target segment, then* $q^*_{k_2}(t) > q^*_{k_1}(t)$, $\forall t < T$. *The established condition outlines when quality improvement of the bigger providers always dominates the smaller ones.*

In the above propositions and corollaries, we brought a number of managerial insights into attention. We showed that in the early stage, there should be an emphasis on increasing the quality of IaaS. Also, it will be to the IaaS provider's advantage to reduce the quality cost. That will increase the optimum quality level and will give rise to a ripple effect of benefits. The dynamic differential game will be played in a time interval, so that some of the variables may change during that time. In the following section, we will analyze how these variations can affect the optimality conditions.

6. Post-optimality analysis

The input data in theoretical optimization approaches is not subject to change, however, in real life it might be found impractical. This assumption is rather valid in a static and deterministic environment, while the essence of our problem is dynamic. User demand reflects market behavior that is changing, and in some degree unpredictable. Cost and capacity estimates are sometimes prone to errors and to changes over time due to the dynamic behavior of the market. Therefore, an important question lies in the sensitivity of the obtained optimal solutions to changes in the input parameters.

We investigate the variability of VM request arrival rate due to a future increase in the number of users. Subsequently, this variability may affect the optimal capacity and price as well. As a result, two types of variations may happen in the range of: (1) objective function; and (2) constraints. The objective function's range refers to the range over which capacity and price coefficients can vary, without changing the basis associated with our optimal solution. In this case, for example, by computing the amount of change in price, we can obtain a new optimal price: $P^*_{new} = P^*_{old} + \Delta P^*$. The constraint's range refers to the user arrival range so that the values of the shadow prices in terms of the defined quality and capacity will remain unchanged.

As the number of users grows, the VM request arrival rate will expand. The value range of l and possible changes to the optimality of the VM price and capacity are investigated in Theorem 2.

Theorem 2 (*IaaS Provider Best Response Sensitivity*)**.** *The optimal solutions obtained for the IaaS provider about price P^* and capacity ϕ^* remain optimum if:*

$$\Delta l < \underbrace{(Q(\phi^* - l) - 1)(\phi^* - l)^2}_{l_{threshold}} \tag{38}$$

In that case, the optimal price and capacity vary as follows:

$$\Delta P^* = l\ \Delta l\ C(\frac{\alpha_i + \gamma_i}{\alpha_i + \gamma_i - 1}) \tag{39}$$

$$\Delta \phi^* = \Delta l \tag{40}$$

Proof. The expressions for the sensitivity derivatives can be derived based on the Kuhn–Tucker conditions. The changes in the optimum values of ϕ^* and P^* necessary to satisfy the Kuhn–Tucker conditions due to a change Δl in the user arrival rate parameter can be estimated as follows:

$$\Delta P^* = \frac{\partial P^*}{\partial l}\Delta l = l\ \Delta l\ C(\frac{\alpha_i + \gamma_i}{\alpha_i + \gamma_i - 1})$$

$$\Delta \phi^* = \frac{\partial \phi^*}{\partial l}\Delta l = \Delta l$$

Earlier, Eq. (15) proved that $\lambda_k = 0$, and the constraint is inactive in the profit maximization problem in Eq. (3). Now, Eq. (14) can be used to determine when an originally inactive constraint becomes active due to the change in VM request arrival rate, Δl. Let us consider the constraint in Eq. (14) as $g(x) = \frac{1}{\phi^* - l} - Q$. The currently inactive constraint will become critical due to Δl, if the new value of $g(x)$ converts to zero:

$$g(x) + \frac{dg(x)}{dl}\Delta l = g(x) + (\frac{\partial g(x)}{\partial \phi^*}\frac{\partial \phi^*}{\partial l} + \frac{\partial g(x)}{\partial P^*}\frac{\partial P^*}{\partial l})\Delta l = 0$$

Thus, the necessary change to Δl to make $g(x)$ active can be found as:

$$\Delta l = -\frac{g(x)}{\frac{\partial g(x)}{\partial \phi^*}\frac{\partial \phi^*}{\partial l}} = (Q(\phi^* - l) - 1)(\phi^* - l)^2 \quad \square$$

The change of the optimal price ΔP obtained from Eq. (39) shall be examined for its effect on the optimality condition of the user demand size as shown in Theorem 3.

Theorem 3 (*IaaS User Best Response Sensitivity*)**.** *The optimal solutions obtained for IaaS user i on VM request size x_i^* remain optimum if:*

$$\Delta P^* < \underbrace{\frac{P^*}{\alpha_i + \gamma_i}}_{P^*_{threshold}} \tag{41}$$

Proof. To prove Eq. (41), we should calculate how much λ_{i1} will fluctuate. Similarly, the variation in the value of Lagrange multiplier due to ΔP can be estimated as follows:

$$\Delta \lambda_{i1} = \frac{\partial \lambda_{i1}}{\partial P^*}\Delta P^*$$

The above equation can be used to determine when the originally active constraint defined for the optimization problem in Eq. (2) becomes inactive due to the change ΔP. Since the value of λ_{i1} is zero for an inactive constraint, we will have:

$$\lambda_{i1} + \Delta \lambda_{i1} = \lambda_{i1} + \frac{\partial \lambda_{i1}}{\partial P^*}\Delta P^* = 0$$

From Eq. (9) we calculate λ_{i1} as follows:

$$\lambda_{i1} = \frac{(B_i r_i^{\beta_i}\gamma_i \mu)^{\gamma_i - 1}}{P^{*(\alpha_i + \gamma_i)}}$$

Therefore, the amount of change in the optimal price to diminish its optimality is as follows:

$$\Delta P^* = \frac{-\lambda_{i1}}{\frac{\partial \lambda_{i1}}{\partial P^*}} = \frac{P^*}{\alpha_i + \gamma_i} \quad \square$$

7. Experiments and analysis

As the main purpose of our experiments is to demonstrate the effectiveness of the proposed games, we have to set meaningful data and reasonable game parameters. To do so, we obtained real-world data and previously achieved suitable values for the parameters of the Cobb–Douglas demand function [14]. Initially, we experimented with 300 IaaS users for the small provider k_1 using real customer ratings to investigate the sensitivity of pricing formula to VM request arrival rate. The data was collected from the Trust Feedback Dataset, provided by Noor et al. [38] in the CloudArmor project.[5] This dataset collects cloud service consumers' feedback from leading review websites (such as Cloud Hosting Reviews and Cloud Storage Reviews and Ratings). It includes 10,000+ feedback given by nearly 7000 consumers to 113 real-world cloud services. The feedback are based on Quality of Service attributes (e.g. availability, response time, throughput, etc.).

5 https://cs.adelaide.edu.au/~cloudarmor/ds.html.

Table 3
Assigned variables' values in simulation.

Variables	k_1	k_2
P	0.18	0.18, 0.13
C	0.000076	0.00000076
δ	0.001	0.001
ρ	0.005	0.005
$\hat{\theta} = \check{\theta}$	0.3	0.3
η	0.5	0.5
f	0.5	0.5
ζ_{k_2}	NA	0.7
ϑ	0, 0.7, 0.9	NA
l	300	900

To simulate the differential game, we assumed VM request arrival rate to be 300 tasks per hour for k_1 and 900 tasks per hour for k_2, that are realistic and commonly used values for cloud services [39]. The price for k_1 is borrowed from a local IaaS provider in Malaysia, called exabytes.[6] Due to the earlier explained reasons, k_2 price will be less or at most the same. Both cases are to be considered in our experiments. To obtain the VM process rate (capacity), which is the minimum speed of 2mbps, we referred to the IaaS promised QoS in the SLA statements. Since there is no information available about the providers' cost, we approximate the cost of k_1 to the cost of renting large cloud infrastructure from Google (per VM per hour) and the cost of k_2 is set to be 100 times less.

The value of ζ_{k_2} (the discounted non-functional cost) is approximated using the Eta-Squared statistics of the ANOVA analysis on the acquired user ratings given to non-functional quality features. The reason behind using Eta-Squared comes from the fact that the average Eta-Squared of some non-functional features (e.g., availability and response time) reflects the importance of these parameters on customers' demand. In fact, we used Eta-squared to measure the effect size of the independent variables (the non-functional attributes). On the other hand, there is no feature representing the personalization value to customers to be used for ϑ. So, we run our experiments by giving different values to ϑ. The rest of the parameters are assigned based on the past literature [13,40,41]. Table 3 depicts the utilized values of variables in the experiments. The time axis is normalized to the (0–1) interval.

As discussed previously, there is no similar work to our model or related experiments to be compared to. For this reason, only the results of our model are reported.

7.1. Significance of quality over user rating

We use statistical analysis to check the significance and confidence (i.e., reliability) of the obtained experimental results and ensure that the outcomes are statistically significant, and data are interpreted correctly, so decisions can be made with high confidence. Accordingly, one-way ANOVA was conducted to assess the effect of the provided quality on the user rating. The ANOVA test was performed over 2000 user ratings given to 78 distinct cloud services considering 8 attributes representing functional and non-functional quality features. Given that the significance value (p) is less than the α-value ($\alpha = .001$) for all quality features, as reported in Table 4, we can rest Assumption 1 and claim that quality attributes are strongly positively correlated with the overall user rating score.

The analysis of variance and Eta-squared values showed that the effect of the technical support attribute was the most significant criterion. It was followed by customer service, response time

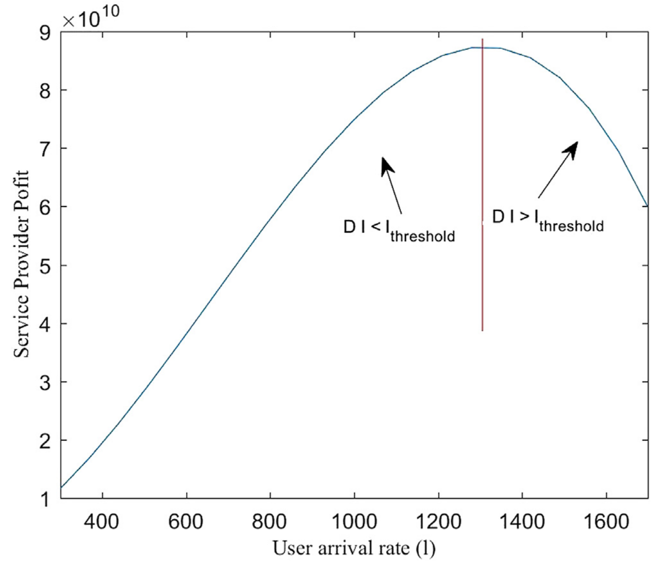


Fig. 3. Sensitivity of pricing optimality to the increase of VM request arrival rate.

and availability. Taken together, these results suggest that high levels of more tangible and measurable qualities have more effect on the user rating score.

The ANOVA results proved that user rating has a strong tie with after-sales service, and customer support has turned into a crucial tool in an organization's arsenal of sales tools. In classical business models, there is little incentive to provide excellent customer support since the majority of the revenue from a customer is already secured. In today's subscription model, however, the equation is almost reversed. Once a service is sold, the IaaS provider receives a very small fraction of the lifetime revenue at the beginning of the transaction. Afterwords, the support team is under a great pressure to keep the customer satisfied. This satisfaction is also crucial to enhance the customer loyalty that has a significant impact on quality equilibrium as asserted by Corollary 2.

7.2. Sensitivity analysis

In order to evaluate the VM request arrival rate threshold and optimality of price, we simulate an increasing number of VM request arrival with a fixed price for IaaS provider k_1. Given the speed of 2 mbps, Q (for k_1) is 900 in an hour. According to our obtained formula in Eq. (38), we have: $\Delta l < 900(0.5)(1.5)^2$, that makes a critical value of change to VM request arrival at about $\Delta l < 1000$. This means that if this provider experiences an increase of 1000 VM request arrivals, it needs to recalculate its pricing strategy since it is not making an optimized profit. This sensitivity is illustrated in Fig. 3. Once the VM request arrival rate crosses the threshold ($\Delta l = 1300 - 300 = 1000$), the profit starts sinking.

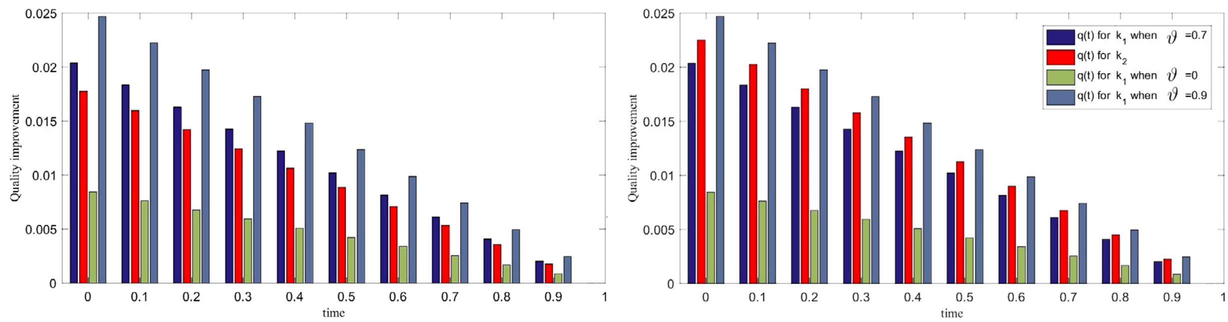
7.3. Quality improvement impact on user demand and profit

During the first game, the IaaS provider has to set the optimal price based on the predicted user demand response given the offered price. The best pricing strategy should consider users' reactions towards the given price. Nevertheless, there are two possible cases for IaaS k_1's price to be considered: (1) $P_{k_1} > P_{k_2}$; and (2) $P_{k_1} = P_{k_2}$. It was learned that the smaller provider faces higher costs, so that it cannot offer a price cheaper than the large

6 www.exabytes.my.

Table 4
ANOVA test results.

Service attributes	Sum of squares	df	Mean square	F	Sig. (p)	Eta-squared
Technical support	3215.603	1907	525.621	1701.668	.000	**.817**
Response time	768.870	698	145.287	537.118	.000	**.756**
Availability	2212.676	1414	331.846	844.832	.000	**.750**
Speed	1472.511	742	218.983	427.415	.000	**.744**
Ease of use	1541.725	1289	283.299	891.095	.000	**.735**
Accessibility	533.575	630	97.636	427.327	.000	**.732**
Operation& management features	853.372	617	149.497	358.836	.000	**.701**
Storage space	846.267	601	115.788	180.430	.000	**.547**



a) when $P_{k_1} > P_{k_2}$ b) when $P_{k_1} = P_{k_2}$

Fig. 4. The IaaS quality improvement of k_1 and k_2.

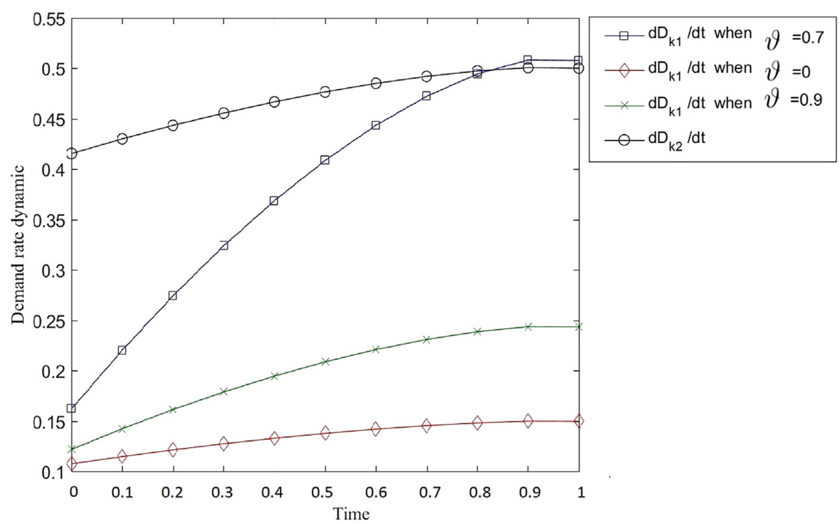


Fig. 5. The change of user demand when $P_{k_1} > P_{k_2}$.

provider. Thus, k_1 can offer either the same price as k_2 or a higher price. The following sections provide more details on these two scenarios.

7.3.1. Scenario 1: Higher pricing

To assess the role of the competitive advantage of k_1, we assign three different values to ϑ_{k_1}, each representing a specific market positioning:

1. $\vartheta_{k_1} = 0$: in this case, the IaaS provider is not using its strategic advantage and does not target any market niches nor offer customization.
2. $\vartheta_{k_1} = 0.7$: this means the IaaS provider is providing added value for customers as worthy as the non-functional quality advantages offered by the existing competitors.
3. $\vartheta_{k_1} = 0.9$: this case illustrates a situation where the IaaS provider is making extra effort to provide more value to the customers than the existing offers.

Fig. 4(a) presents the quality improvement steps over time for k_1 and k_2 when $P_{k_1} > P_{k_2}$. Agreeing with Proposition 2, the quality improvement rate is steeper at the beginning and flattens out at final stages. When k_1 is providing the same value as k_2, it needs to put more effort on quality than k_2. The difference of this extra quality is higher in the first steps, but reduces over time. Thus, k_1 can, for instance, expand the infrastructure to preserve a more capacity and networking bandwidth, or reorganize the tenant clusters. Customer support is specifically important where the small provider operates to provide a more customized solution. Examples of less costly quality improvement would be offering creative features and highly customized and targeted services. It is important to mention in this context that the numbers shown in the figure are calculated based on the obtained units from Eqs. (34) and (35), so they are not expressed as a percentage. We also used a normalized time range, but with actual time, the values of the obtained quality improvement would be higher. However, what matters in this figure is the correct strategy of the providers considering their competitive advantages, not the actual value of the quality improvement. These values do not convey any specific meaning on their own; comparing the strategies of the two providers at the equilibrium point is the key objective.

Mapping the defined quality improvement to user rating increment gives the opportunity to identify how the users feel about the trade-off between the price and quality. When the users are paying higher price and receiving lower quality, they would expect to see much higher added-value to their businesses. The strategic benefits will satisfy their expectations and would rise their ratings. This is highly significant for smaller providers to plan the right amount of investment to improve the quality in their early stages of development. Clearly, the case where k_1 is making extra effort on providing customer value requires additional improvement.

The effect of such a quality (rating) improvement on the user demand rate is quite interesting. As shown in Fig. 5, k_1 gains the highest change in user demand when it is providing the same value as k_2. The change it experiences exceeds that of k_2. This can be because of receiving a closer rating to k_2, which can reflect how valuable is providing such targeted services to customers. It is not surprising that the demand rate increases very little when the IaaS provider is not offering any special value. Remarkably, providing extra value does not necessarily lead to a higher demand rate. Overdoing that may result in limiting the range of the targeted customers. If the IaaS provider offers very specific and customized services, it may narrow its range of clients and miss the market share that it could obtain. Therefore, finding a suitable strategy is essential for the IaaS provider that wants to compete and earn its share of a profitable but competitive market.

7.3.2. Scenario 2: Equal pricing

In this scenario, the small IaaS provider sets the same price as the existing IaaS in the market. Unlike the first scenario, this time k_1 has less improvement of quality than k_2 when $\vartheta_{k_1} = 0.7$ as shown in Fig. 4(b). This is due to the fact that k_1 is lowering its pricing down to the same amount as k_2, while its cost is higher. Besides, users of k_1 have already gained the benefit of having lower prices, so the provider does not have to offer higher quality to cover the benefit of price and gain higher ratings. Meanwhile, as expected, no difference is observed between the two scenarios in the amount of the quality improvement when ϑ_{k_1} is 0 or 0.9. This is reasonable because providing zero value (resp. a very high value) demands the same quality improvement regardless of the pricing strategy.

Fig. 6 illustrates the variation in demand rate over time. When k_1 sets equal pricing as k_2, its demand increase does not reach the change of k_2's demand, which remains higher. Unexpectedly, the equal pricing strategy mainly affects k_2's demand rather than k_1. This event can be related to the impact of quality improvement on the user demand rate. When k_1 sets higher pricing, it can afford more improvement leading to enhanced rating, and its users demand gets slightly higher. Meanwhile, k_2 takes the most advantage of k_1's lower rating improvement to attract more users. It can be inferred that customers prioritize quality over price, which confirms the movement of Cloud 2.0. The trend of k_1's user demand rate variation offering the highest and lowest values does not present any significant change.

7.3.3. Provider's profit and users' loyalty

Total variations of IaaS provider's profit with both pricing strategies are presented in Fig. 7. Since the trend of profit for k_1 when $P_{k_1} = P_{k_2}$ was almost the same as when $P_{k_1} > P_{k_2}$, we provided only one plot for each different case of ϑ_{k_1}. However, as the profit of k_2 differs significantly depending on the pricing strategy ($P_{k_1} = P_{k_2}$ or $P_{k_1} > P_{k_2}$), two separate plots are depicted. This figure provides a very useful insight for small IaaS providers by showing that pricing does not alter profit optimization as long as they provide a quality level adjusted with that pricing level. Higher price demands more quality improvement to meet the user expectation and gain high rating. Consequently, the IaaS provider undergoes the burden of the cost associated with that improvement such as increasing the number of servers to reserve more capacities. Here, k_1 can rely on its identified market segment behavior through obtaining its sensitivity to price and rating to decide about setting a reasonable price.

Although the pricing strategy does not significantly affect the profit of k_1, it has enormous influence on the profit of k_2. The reason is that k_2 has already established its reputation as an IaaS leader and obtained a high rating, so that its quality improvement is saturated. In fact, when k_1 fails to attain the customers looking for high quality and customized services, k_2 attracts them. However, k_1 can recompense its profit with higher service price. In this case, k_1's customers will mainly constitute the new public IaaS adopters who could not enjoy the cloud computing benefits due to their regional, national or international barriers or because of some very customized needs. Although customization is essential for some businesses, yet the quality features of the services are very important. The majority of the users are not willing to sacrifice one for the other. Thus, if the smaller IaaS providers are ready to compete and gain more market share through their own market segment, they need to supply a reasonable amount of IaaS non-functional quality.

We further investigate the importance of customer loyalty on IaaS providers' revenue. Being loyal to an IaaS provider in today's subscription model has mutual benefit for both, the users and providers. As Fig. 8 illustrates, customer loyalty is a key motivation for quality improvement for both providers, k_1 and k_2. The effect of customer loyalty is intense when the customers exhibit a highly loyal behavior. In this case, IaaS providers commit themselves to provide high quality services. However, customers with low and even medium loyalty do not make a significant difference. Consequently, as the customer defection rate increases, the user demand and provider's profit drop. The new and small providers are slightly more vulnerable to customer defection, in particular when it comes to future demand provisioning. In fact, the new IaaS providers need to establish their credibility by increasing their users satisfaction and attracting high users' ratings. They also have a limited range of customers compared to the established providers. However, it is most likely that customers who receive customized and targeted infrastructure services would be more loyal to their providers since it is unlikely to find such services anywhere else.

In summary, IaaS quality, specially customer support, has a strong correlation with users' ratings. It offers a managerial point for IaaS providers to increase the customer loyalty through not only customized services, but also fully commitment in after-sale support. The results show that when the IaaS users are paying higher and receiving lower non-functional qualities, they would expect to see much higher added-value from the IaaS provider. The strategic benefit satisfies the user's expectation and rise the provider rating. The small IaaS providers gain the highest change in user demand when they provide the same quality as the established ones. Notably, providing extra value does not necessarily lead to a higher demand rate as it might limit the range of the targeted customers. Improving the quality and ratings of a small and new IaaS provider specifically increases its demand rate and profit in the both pricing scenarios, higher and equal. Setting equal pricing for both IaaS providers mainly favors the established provider. When the smaller provider sets equal pricing, it cannot afford more improvement that leads to a lower user satisfaction. This failure to attract high user ratings makes

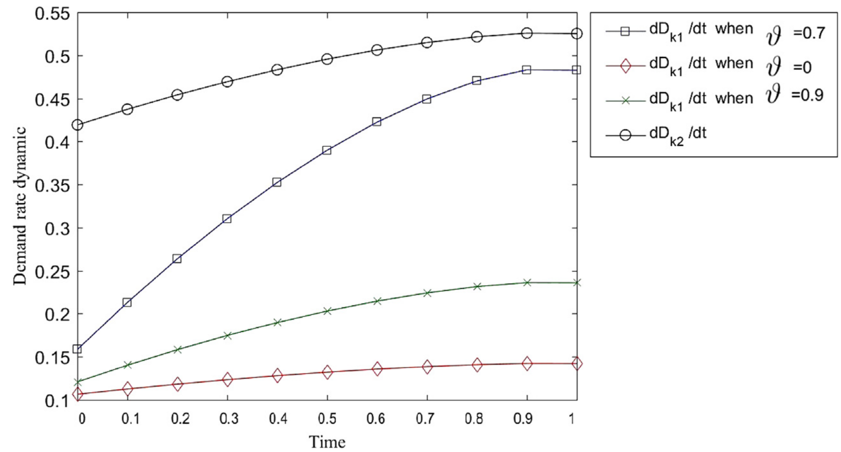


Fig. 6. The change of user demand when $P_{k_1} = P_{k_2}$.

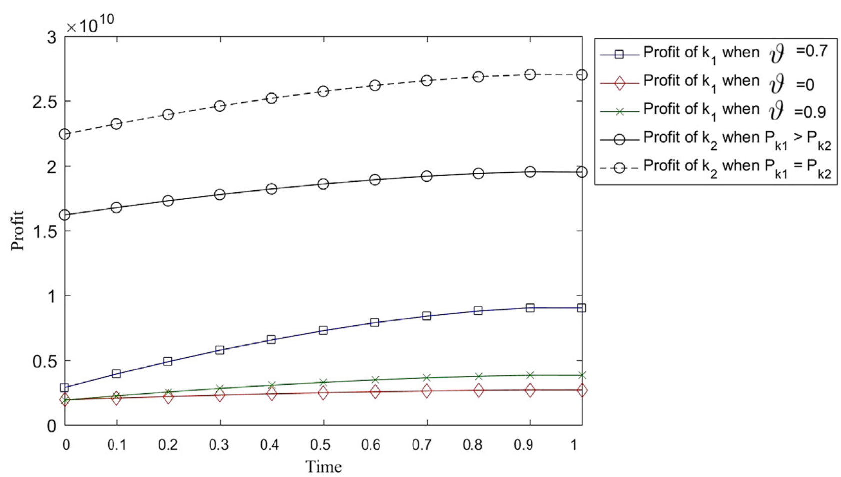


Fig. 7. IaaS provider's profit taking different quality controls and pricing strategies.

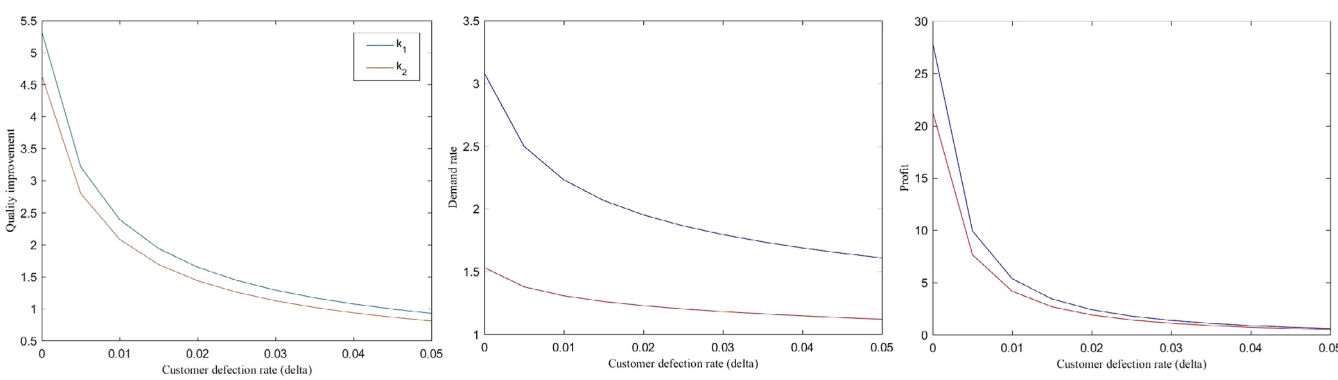


Fig. 8. Customer loyalty effect on quality and profit.

a perfect situation for its competitor to attract them, although it does not harm its own profit since it gains the difference of the demand with the difference in the price and quality, and maintains the customers who have no choice but their customized services.

8. Conclusion

This paper tackled the issue of oligopoly IaaS market that neglects the user satisfaction and threatens the grows of the cloud market industry. A conceptual game theoretical framework with two games, namely Stackelberg and differential has been introduced and designed to allow new and even small IaaS providers to obtain a market share using their own strategic advantages. The theoretically obtained results were confirmed by experiments using real-world dataset. It was found that the user demand from small IaaS providers increases the most when these providers provide added-value services equal to the value offered by the existing providers. Regardless of the pricing strategy (higher or equal), improving the quality and ratings of a small and new IaaS provider increases its demands' rate and profit. However, the best strategy for small IaaS providers is to set higher price and improve the quality of their provided added-value solutions, specifically in the early stages of development. The reason is that service customization increases the customer loyalty in today's subscription cloud economy model, where customers are free to defect anytime. Higher customer loyalty elevates the provider's profit and increases the quality equilibrium. Higher level of service quality leads to a higher user satisfaction and improves the small IaaS provider's market position through higher ratings. This research drew many technical, strategic and managerial insights to guide IaaS providers on how to utilize their strengths in deciding on future opportunities and target markets. By offering value beyond simply providing computing resources, the IaaS provider will play a strategic role in the future of Cloud 2.0.

9. Future work and research directions

Although game theory has been applied to real problems in different domains such as political science, biology, and engineering, the assumption that players are rational and have common knowledge so they aim at maximizing profit and minimizing cost is not always practical as shown by some experimental studies [42]. These experimental proposals demonstrated that in some cases, players consider in their decision-making other preferences than simply maximizing profits, for instance psychological, ideological, societal, or environmental preferences. Behavioral game theory deals with this type of problems and tries to explain decision making using experimental data [42].

The main purpose of our work was to theoretically analyze the behavior of the small cloud providers to gain a share in the market. This research is important and significant in the specific context of cloud computing where providers care mainly about profits and where small providers care predominantly about pure economical attributes to gain a market share. However, proving that our theoretical findings, supported by simulations involving real data, match experimental choices of real cloud providers in real settings is yet to be explored. The problem is highly challenging and will be investigated as future work. To fully investigate the practical implication of our proposal, we will study different cloud market players and analyze if their behavior is as expected in theory, considering different considerations, social, political, etc. This line of research is highly appealing as it has been demonstrated that real players play naturally towards the equilibrium solutions, in particular when the game is played many times so players gain experience and understand better the game [43].

We further plan to investigate learning approaches to provide cloud players with better mechanisms to learn (1) the behavior of customers in order to increase their satisfaction; and (2) better strategies to compete against different providers within our two-stage game. Supporting and deploying mobile-edge technologies are other directions for further research towards the future of Cloud 2.0. We will focus in particular on two key issues: security and computation offloading to tackle the problem of limited computational power, storage, and energy [44,45].

In this paper, we considered the IaaS market and its specific characteristics, in particular multi-tenancy. However, multi-tenancy in other layers is quite different and goes beyond the infrastructure. Investigating other layers within the new vision of value-based Cloud 2.0, such as SaaS with a high degree of multi tenancy, higher economy of scale and different pricing models, with different quality factors, is another challenging research direction. Finally, we plan to study the cooperation strategies in the context of Cloud 2.0 where the concepts of communities and incentives to cooperate [46] will be elaborated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to thank Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support including Discovery Grant for Jamal Bentahar and Vanier Canada Graduate Scholarship for Mona Taghavi. We would also like to thank the Center on Cyber-Physical Systems at Khalifa University, Abu Dhabi.

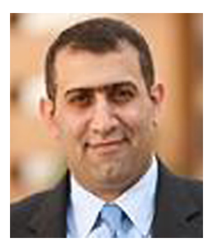
References

[1] J. Chen, C. Wang, B.B. Zhou, L. Sun, Y.C. Lee, A.Y. Zomaya, Tradeoffs between profit and customer satisfaction for service provisioning in the cloud, in: HPDC, 2011, pp. 229–238.
[2] A. Bestavros, O. Krieger, Toward an open cloud marketplace: Vision and first steps, IEEE Internet Comput. 18 (1) (2014) 72–77.
[3] D. Durkee, Why cloud computing will never be free, Commun. ACM 53 (5) (2010) 62–69.
[4] B. Yang, Z. Li, S. Jiang, K. Li, Envy-free auction mechanism for VM pricing and allocation in clouds, Future Gener. Comput. Syst. 86 (2018) 680–693.
[5] B. Xu, T. Qin, G. Qiu, T.-Y. Liu, Optimal pricing for the competitive and evolutionary cloud market., in: IJCAI, 2015, pp. 139–145.
[6] J. Nair, V.G. Subramanian, A. Wierman, Provisioning of ad-supported cloud services: The role of competition, Perform. Eval. 120 (2018) 36–48.
[7] L.P. Willcocks, W. Venters, E.A. Whitley, Moving to the Cloud Corporation - How to face the challenges and harness the potential of cloud computing, Palgrave Macmillan, 2014.
[8] P. Hoberg, J. Wollersheim, H. Krcmar, The business perspective on cloud computing - a literature review of research on cloud computing, in: 18th Americas Conference on Information Systems, AMCIS 2012, Seattle, Washington, USA, August 9-11, 2012, p. Paper 5.
[9] B. Suleiman, S. Sakr, D.R. Jeffery, A. Liu, On understanding the economics and elasticity challenges of deploying business applications on public cloud infrastructure, J. Internet Serv. Appl. 3 (2) (2012) 173–193.
[10] V. Chang, Y. Kuo, M. Ramachandran, Cloud computing adoption framework: A security framework for business clouds, Future Gener. Comput. Syst. 57 (2016) 24–41.
[11] A.S. Sohal, R. Sandhu, S.K. Sood, V. Chang, A cybersecurity framework to identify malicious edge device in fog computing and cloud-of-things environments, Comput. Secur. 74 (2018) 340–354.
[12] G. Sun, Y. Li, Y. Li, D. Liao, V. Chang, Low-latency orchestration for workflow-oriented service function chain in edge computing, Future Gener. Comput. Syst. 85 (2018) 116–128.
[13] M. Fan, S. Kumar, A.B. Whinston, Short-term and long-term competition between providers of shrink-wrap software and software as a service, European J. Oper. Res. 196 (2) (2009) 661–671.
[14] M. Taghavi, J. Bentahar, H. Otrok, O.A. Wahab, A. Mourad, On the effects of user ratings on the profitability of cloud services, in: 2017 IEEE International Conference on Web Services, ICWS 2017, Honolulu, HI, USA, June 25-30, 2017, 2017, pp. 1–8.
[15] J. Zhang, N. Xie, X. Zhang, W. Li, An online auction mechanism for cloud computing resource allocation and pricing based on user evaluation and cost, Future Gener. Comput. Syst. 89 (2018) 286–299.
[16] G. Nan, Z. Mao, M. Yu, M. Li, H. Wang, Y. Zhang, Stackelberg game for bandwidth allocation in cloud-based wireless live-streaming social networks, IEEE Syst. J. 8 (1) (2014) 256–267.
[17] Z. Diamadi, A. Dubey, D. Pleasance, A. Vora, Winning in the SMB Cloud: Charting a Path to Success, McKinsey, New York, 2014.
[18] R. Smith, Computing in the cloud, Res.-Technol. Manag. 52 (5) (2009) 65–68.
[19] Q. Yu, Cloudrec: a framework for personalized service recommendation in the cloud, Knowl. Inf. Syst. 43 (2) (2015) 417–443.
[20] R. Pal, P. Hui, Economic models for cloud service markets: Pricing and capacity planning, Theoret. Comput. Sci. 496 (2013) 113–124.
[21] O.A. Wahab, J. Bentahar, H. Otrok, A. Mourad, A stackelberg game for distributed formation of business-driven services communities, Expert Syst. Appl. 45 (2016) 359–372.
[22] G. Sun, D. Liao, D. Zhao, Z. Sun, V. Chang, Towards provisioning hybrid virtual networks in federated cloud data centers, Future Gener. Comput. Syst. 87 (2018) 457–469.

[23] O.A. Wahab, J. Bentahar, H. Otrok, A. Mourad, Towards trustworthy multi-cloud services communities: A trust-based hedonic coalitional game, IEEE Trans. Serv. Comput. 11 (1) (2018) 184–201.
[24] X. He, A. Prasad, S.P. Sethi, G.J. Gutierrez, A survey of stackelberg differential game models in supply and marketing channels, J. Syst. Sci. Syst. Eng. 16 (4) (2007) 385–413.
[25] M. Breton, R. Jarrar, G. Zaccour, A note on feedback sequential equilibria in a lanchester model with empirical application, Manage. Sci. 52 (5) (2006) 804–811.
[26] G.J. Gutierrez, X. He, Life-cycle channel coordination issues in launching an innovative durable product, Prod. Oper. Manage. 20 (2) (2011) 268–279.
[27] P.-Y. Chen, Y.-C. Chou, R.J. Kauffman, Community-based recommender systems: Analyzing business models from a systems operator's perspective, in: IEEE HICSS, 2009, pp. 1–10.
[28] M. Taghavi, J. Bentahar, K. Bakhtiyari, C. Hanachi, New insights towards developing recommender systems, Comput. J. 61 (3) (2018) 319–348.
[29] W. Duan, B. Gu, A.B. Whinston, Do online reviews matter? An empirical investigation of panel data, Decis. Support Syst. 45 (2008) 1007–1016.
[30] M. Hadji, W. Louati, D. Zeghlache, Constrained pricing for cloud resource allocation, in: IEEE NCA, 2011, pp. 359–365.
[31] B. Shen, Y. Shen, W. Ji, Profit optimization in service-oriented data market: A stackelberg game approach, Future Gener. Comput. Syst. 95 (2019) 17–25.
[32] L. Zhao, L. Lu, Z. Jin, C. Yu, Online virtual machine placement for increasing cloud provider's revenue, IEEE Trans. Serv. Comput. 10 (2) (2017) 273–285.
[33] C. Kilcioglu, J.M. Rao, Competition on price and quality in cloud computing, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016, pp. 1123–1132.
[34] M. Habibi, M. Fazli, A. Movaghar, Efficient distribution of requests in federated cloud computing environments utilizing statistical multiplexing, Future Gener. Comput. Syst. 90 (2019) 451–460.
[35] A.S. Goldberger, The interpretation and estimation of cobb-douglas functions, Econometrica (1968) 464–472.
[36] S.C. Misra, A. Mondal, Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding return on investment, Math. Comput. Modelling 53 (3) (2011) 504–521, Telecommunications Software Engineering: Emerging Methods, Models and Tools.
[37] L.M. Hocking, Optimal Control: An Introduction to the Theory with Applications, Oxford University Press, 1991.
[38] T.H. Noor, Q.Z. Sheng, L. Yao, S. Dustdar, A.H. Ngu, Cloudarmor: Supporting reputation-based trust management for cloud services, IEEE Trans. Parallel Distrib. Syst. 27 (2) (2016) 367–380.
[39] H. Khazaei, Performance Modeling of Cloud Computing Centers (Ph.D. thesis), University of Manitoba, (Canada), 2013.
[40] K. Begnum, M. Burgess, T.M. Jonassen, S. Fagernes, On the stability of adaptive service level agreements, IEEE Trans. Netw. Serv. Manag. 3 (1) (2006) 13–21.
[41] S.K. Mukhopadhyay, R. Setaputra, A dynamic model for optimal design quality and return policies, European J. Oper. Res. 180 (3) (2007) 1144–1154.
[42] C.F. Camerer, Behavioral Game Theory: Experiments in Strategic Interaction, Princeton University Press, 2003.
[43] L. Samuelson, Game theory in economics and beyond, J. Econ. Perspect. 30 (4) (2016) 107–130.
[44] F. Zhang, J. Ge, C. Wong, C. Li, X. Chen, S. Zhang, B. Luo, H. Zhang, V. Chang, Online learning offloading framework for heterogeneous mobile edge computing system, J. Parallel Distrib. Comput. 128 (2019) 167–183.
[45] B. Huang, Z. Li, P. Tang, S. Wang, J. Zhao, H. Hu, W. Li, V. Chang, Security modeling and efficient computation offloading for service workflow in mobile edge computing, Future Gener. Comput. Syst. (2019).
[46] B. Khosravifar, J. Bentahar, A. Moazin, P. Thiran, Analyzing communities of web services using incentives, Int. J. Web Serv. Res. 7 (3) (2010) 30–51.

Mona Taghavi is a Ph.D. Candidate at Concordia Institute for Information System Engineering, Concordia University, Canada. She obtained her M.Sc. in management information systems from University Kebangsaan Malaysia in 2012 and BSc. in information technology engineering from Islamic Azad University, Iran in 2008. Her research interests include services computing, game theory, blockchain, machine learning and recommender systems.

Jamal Bentahar received the bachelor's degree in software engineering from the National Institute of Statistics and Applied Economics, Morocco, in 1998, the M.Sc. degree in the same discipline from Mohamed V University, Morocco, in 2001, and the Ph.D. degree in computer science and software engineering from Laval University, Canada, in 2005. He is a Professor with Concordia Institute for Information Systems Engineering, Faculty of Engineering and Computer Science, Concordia University, Canada. From 2005 to 2006, he was a Postdoctoral Fellow with Laval University, and then NSERC Postdoctoral Fellow at Simon Fraser University, Canada. He is an NSERC Co-Chair for Discovery Grant for Computer Science (2016–2018). His research interests include the areas of cloud and services computing, game theory, computational logics, model checking, multi-agent systems, and software engineering.



Hadi Otrok received his Ph.D. in ECE from Concordia University. He is an Associate Professor in the Department of ECE at Khalifa University. He is an associate editor of the Ad-hoc networks, IEEE communications letters and a co-chair of several committees at various IEEE conferences. His research interests include services computing, network security, game theory and mechanism design.