A Novel and Efficacious Iterative Algorithm for Cloud Optimization

Pasam Venkateshwar Rao[1], Jyotsna Ravula[2]

[1,2]*Assistant Professor, Department of CSE*
[1,2]*Malla Reddy Engineering College (A), Hyderabad, Telangana, India.*

Abstract

To design a service mechanism for profit optimizations of both a cloud provider and its multiple users. We consider the problem from a game theoretic perspective and characterize the relationship between the cloud provider and its multiple users as a Stackelberg game, in which the strategies of all users are subject to that of the cloud provider. The cloud provider tries to select and provision appropriate servers and configure a proper request allocation strategy to reduce energy cost while satisfying its cloud users at the same time. We approximate its server's selection space by adding a controlling parameter and configure an optimal request allocation strategy. For each user, we design a utility function which combines the net profit with time efficiency and try to maximize its value under the strategy of the cloud provider. We formulate the competitions among all users as a generalized Nash equilibrium problem (GNEP). We solve the problem by employing variational inequality (VI) theory and prove that there exists a generalized Nash equilibrium solution set for the formulated GNEP. Finally, we propose an iterative algorithm (IA), which characterizes the whole process of our proposed service mechanism. We conduct some numerical calculations to verify our theoretical analyses. The experimental results show that our IA algorithm can benefit both of a cloud provider and its multiple users by configuring proper strategies.

Keywords: *Cloud computing, Generalized Nash equilibrium, Non-cooperative game theory, Profit optimization, Resource allocation, variation inequality theory*

1. Introduction

Over recent years, the growth of the IoT (Internet of Things) is expected to create a pervasive connection of things such as embedded devices, sensors, and actuators. This will inevitably result in the generation of enormous amount of data, which have to be autonomously stored, processed, accessed, and managed. Cloud computing has been recognized as a paradigm for the big data problem [1]. Cloud computing allows the sensing data to be stored and used intelligently for smart applications.

One of challenges which arose when IoT meets cloud is resource allocation by which a cloud provider efficiently allocates its resources to cloud users with SLA constraints. The dominating performance factors in resource allocation include provider's and user's profit, resource utilization, and QoS (Quality of Service) [2]. For resource allocation in cloud, auction-based model is a popular approach in resource allocation and pricing [3]. In particular, combinatorial auction is preferred in cloud computing because it allows a user to buy a package of resources rather than an individual resource. Since the auction is done on group of resources, it provides efficient resource allocation and helps to improve the profit to both a provider and a user. The other issue of resource allocation in cloud computing is meeting SLA (Service Level Agreement) established with a user. Before a provider provisions a service to a user, a provider and a user need to establish SLA contract. The SLA is an agreement that specifies QoS between a provider and a user [4].

We define two-levels of SLA to represent different objectives: class-based SLA and job-based SLA [5]. In class-based SLA, for each job class, QoS is measured based on performance metrics. In job-based SLA, QoS is measured using the metrics of individual jobs. Any single job with a poor service quality immediately affects the measured QoS and incurs some SLA penalty cost. We believe the job-based SLA is a more robust type of SLAs from the users' perspective than providers' perspective and we focus on this. In general, SLA is defined in terms of various performance metrics, such as service latency, throughput, consistency, and security. In our paper, SLA is defined in terms of deadline along with execution time of each job. The deadline violation is important from QoS guarantees perspective. It causes loss of profit for a provider due to incompletion to execute the certain jobs and SLA penalty cost [6]. Considering SLA guarantee, we propose a resource allocation mechanism in combinatorial auction system for cloud computing. For efficient resource allocation, winner determination problem in the auction should be solved. The conventional winner determination mechanisms are proposed to maximize resource utilization and a provider's profit for each time interval. However, the profit can be maximized by reducing the penalty cost for SLA violation. To reduce SLA violations, the deadline constraints need to be considered when winner is determined in the auction. Thus, we propose a winner determination mechanism with consideration for the deadline constraints to maximize the provider's profit.

2. Existing System

Here hardly any previous works investigate multiple users′ profit optimizations, let alone optimizing the profits of a cloud provider and its users at the same time. In existing system, based on DVFS technique and the concept of slack sharing among processors, the authors also proposed two novel energy-aware scheduling algorithms.

1. Since multiple users will try to access the data application performance depends on the user’s data requests.
2. The existing system unable to avoid the server energy cost.

3. Proposed System

In this paper, we try to design a new service mechanism for profit optimizations of both a cloud provider and its multiple users. We consider the problem from a game theoretic perspective and characterize the relationship between the cloud provider and its users as a Stackelberg game, in which the strategies of all users are subject to that of the cloud provider. In our mechanism, the cloud provider tries to select appropriate servers and configure a proper request allocation strategy to reduce energy cost while satisfying its users at the same time.

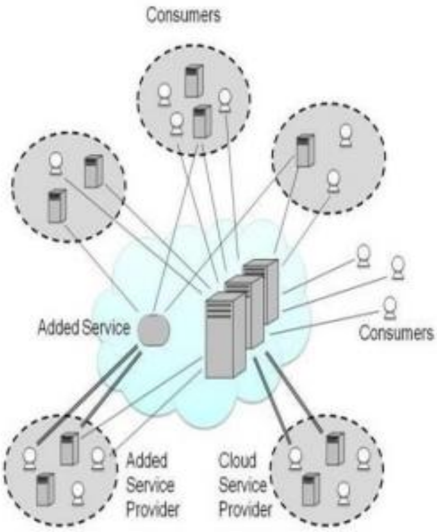

Figure 1. Architecture of proposed model

In this subsection, we model the architecture of our proposed service mechanism, in which the cloud provider can select an appropriate servers subset S from M (i.e.,S ⊆M) to provide services for the H future time slots, and configure a proper strategy pS =(p1S, . . . , pHS) with ph S =(phj)j∈ S (h ∈ H) to allocate the aggregated requests to the selected servers, such that the average response time over all cloud users is minimized, while its multiple users can make an appropriate request decision according to the selected servers and allocation strategy. As shown in Fig. each user i (i ∈ N) is equipped with a utility function (Ui) and a request configuration strategy i.e., the request strategy over H future time slots. All requests enter a queue to be processed by the cloud center. The cloud provider tries to select an appropriate servers subset S, configure an appropriate allocation strategy pS, and publishes some information (e.g., per request charge r, server subset S, and the corresponding allocation strategy pS, current aggregated requests) on the information exchange model. When multiple users try to configure appropriate request strategies, they first get information from the exchange module, then compute proper request strategies such that their own utilities are maximized and send the newly strategies to the cloud provider.

Advantages

- Cost effectiveness will be provided.
- Application performance will be improved.
- In this work, we first try to optimize multiple users′ since multiple cloud users compete for using the resources of a cloud provider, and the utility of each user is affected by the decisions (service request strategies) of other users, it is natural to analyze the behaviors of such systems as strategic games.
- We characterize the relationship between the cloud provider and its users as a Stackelberg game and try to optimize the profits of both a cloud provider and its users at the same time.
- We formulate the competitions among all users as a generalized Nash equilibrium problem (GNEP) and prove that there exists a generalized Nash equilibrium solution set for the formulated GNEP.
- We solve the GNEP by employing vocational inequality (VI) theory and propose an iterative algorithm (IA) to characterize the whole process of our proposed service mechanism.

4. Implementation

4.1. Customer

In this module, we develop the Customer entity functionalities. Customer is important for a cloud provider to select appropriate servers to provide services, such that it reduces cost as much as possible while satisfying its users at the same time. The reason behind lies in that both of them are not just for the profit of a cloud provider, but for the appeals to more cloud users in the market to use cloud service and thus also impact the profit.

4.2. Cloud provider

The experimental results show that our IA algorithm can benefit both of a cloud provider and its multiple users by configuring proper strategies. The reason behind lies in that both of them are not just for the profit of a cloud provider, but for the appeals to more cloud users in the market to use cloud service and thus also impact the profit.

4.3. Admin

Administrator has the responsibility of ensuring that the administrative activities within an organization run efficiently, by providing structure to other employees throughout the organization. These activities can range from being responsible for the management of human resources, budgets and records, to undertaking the role of supervising other customer. These responsibilities can vary depending on the customer and level of education.

5. Conclusion

Due to numerous advantages, the popularity of cloud computing and its cost effectiveness, flexibility, and scalability, more applications are transmitted from the local computing environment in the center of the cloud. In this paper, we try to develop a new process for the service provider of cloud optimization and gain more profit. We consider the problem from a game theory perspective and characterize the relationship between the cloud provider and its multiple users as a Stackelberg game, in which the strategies of all users are subject to the cloud provider. The cloud provider tries to select appropriate servers and set up a proper request allocation strategy to reduce energy costs while at the same time satisfying its cloud users. We approach the server's server space by adding a control parameter and setting up an optimal delivery assignment strategy. For each user, we design a utility that combines the net profit with time efficiency and tries to maximize its value under the cloud provider's strategy. We formulate the competitions among all users as a common Nash equilibrium problem (GNEP)

References

[1] Chubo Liu, Kenli Li, Senior Member, IEEE, Keqin Li, Fellow, IEEE, and Rajkumar Buyya, Fellow, IEEE, “A New Service Mechanism for Profit Optimizations of a Cloud Provider and Its Users”, IEEE Transactions on Cloud Computing, 2017.

[2] A. Prasad and S. Rao, “A mechanism design approach to resource procurement in cloud computing,”Computers, IEEE Transactions on, vol. 63, no. 1, pp. 17–30, Jan 2014.

[3] R. Pal and P. Hui, “Economic models for cloud service markets: Pricing and capacity planning,”Theoretical Computer Science, vol.496, no. 0, pp. 113 – 124, 2013.

[4] P. D. Kaur and I. Chana, “A resource elasticity framework for qos-aware execution of cloud applications,” Future Generation Computer Systems, vol. 37, no. 0, pp. 14 – 25, 2014.

[5] L. Duan, D. Zhan, and J. Hohnerlein, “Optimizing cloud data center energy efficiency via dynamic prediction of cpu idle inter-vals,” in 2015 IEEE 8th International Conference on Cloud Computing. IEEE, 2015, pp. 985–988.

[6] Z. Li, J. Ge, H. Hu, W. Song, H. Hu, and B. Luo, "Cost and energy aware scheduling algorithm for scientific workflows with dead-line constraint in clouds," IEEE Transactions on Services Computing,2015, doi: 10.1109/TSC.2015.2466545.

[7] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks,"ACM SIGCOMM computer communication review, vol. 39, no. 1, pp. 68–73, 2008.

[8] J. Cao, K. Hwang, K. Li, and A. Zomaya, "Optimal multiserver configuration for profit maximization in cloud computing,"Parallel and Distributed Systems, IEEE Transactions on , vol. 24, no. 6,pp. 1087–1096, June 2013.

[9] Y. Feng, B. Li, and B. Li, "Price competition in an oligopoly market with multiple iaas cloud providers,"Computers, IEEE Transactions on, vol. 63, no. 1, pp. 59–73, Jan 2014.

[10] J. Cao, K. Li, and I. Stojmenovic, "Optimal power allocation and load distribution for multiple heterogeneous multicore server processors across clouds and data centers,"Computers, IEEE Transactions on, vol. 63, no. 1, pp. 45–58, Jan 2014.

