This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPDS.2015.2473850, IEEE Transactions on Parallel and Distributed Systems

Cloud Customer's Historical Record Based Resource Pricing

Mohammad Aazam, Eui-Nam Huh, *Members*, IEEE, Marc St-Hilaire, *Senior Member*, IEEE, Chung-Horng Lung, Ioannis Lambadaris, *Members*, IEEE

Abstract — Media content in its digital form has been rapidly scaling up, resulting in popularity gain of cloud computing. Cloud computing makes it easy to manage the vastly increasing digital content. Moreover, additional features like, omnipresent access, further service creation, discovery of services, and resource management also play an important role in this regard. The forthcoming era is interoperability of multiple clouds, known as cloud federation or inter-cloud computing. With cloud federation, services would be provided through two or more clouds. Once matured and standardized, inter-cloud computing is supposed to provide services which would be more scalable, better managed, and efficient. Such tasks are provided through a middleware entity called cloud broker. A broker is responsible for reserving resources, managing them, discovering services according to customer's demands, Service Level Agreements (SLAs) negotiation, and match-making between the involved service provider and the customer. So far existing studies discuss brokerage in a narrow focused way. In the research outcome presented in this paper, we provide a holistic brokerage model to manage on-demand and advance service reservation, pricing, and reimbursement. A unique feature of this study is that we have considered dynamic management of customer's characteristics and historical record in evaluating the economics related factors. Additionally, a mechanism of incentive and penalties is provided, which helps in trust build-up for the customers and service providers, prevention of resource underutilization, and profit gain for the involved entities. For practical implications, the framework is modeled on Amazon Elastic Compute Cloud (EC2) On-Demand and Reserved Instances service pricing. For certain features required in the model, data was gathered from Google Cluster trace.

Index Terms — **c**loud broker, resource management, pricing, Inter-cloud computing, cloud federation.

1 INTRODUCTION

CLOUD computing has become worldwide because of its increasing demands and requires more heterogeneous infrastructure which results in making interoperability an area of interest. This increasing demand is challenging for cloud customers as far as selection of an appropriate Cloud Service Provider (CSP) is concerned and it connects them to a particular CSP [1]. Inter-cloud computing is needed at this place. This paradigm of inter-cloud computing is in its initial stages but still it permits uninterrupted interoperability between clouds, no matter whatever their underlying infrastructure is. This enables users to shift their workloads across clouds in an easy way. Moreover, resources can be handled effectively through cloud brokerage which is an advantageous aspect of inter-cloud computing [1], [2].

On the clouds, most of the data-intensive applications are now installed. These applications, storage, and data resource are located in so many different manners that they have to reach even cross-continental networks. Due to this issue, the performance of cloud systems and user requests is affected by performance degradation in networks. The need to ensure service quality, specially for bulk-data transfer, makes resource reservation and utilization a serious issue [3].

Mohammad Aazam is with Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada and Computer Engineering Department, Kyung Hee University, Korea (e-mail: aazam@ieee.org).

Eui-Nam Huh is with Computer Engineering Department, Kyung Hee University, Korea (e-mail: johnhuh@khu.ac.kr).

Marc St-Hilaire (email: marc_st_hilaire@carleton.ca), Chung-Horng Lung (email: chlung@sce.carleton.ca), and Ioannis Lambadaris (email: ioannis@sce.carleton.ca) are with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada.

Jayant Baliga et al. [4] stresses upon the fact that more resources are utilized with the increasing digital content which causes more energy consumption. The performance and the overall cost of the services provided, both are affected by greater energy consumption. This intensifies the need for efficient resource management and dynamic pricing.

Ewa Deelman et al. [5] reinforces the fact in their study that significant amount of cost can be reduced by right amount of resource allocation, without having any effect on the performance. If we take costumer's perspective in consideration, honesty is a major concern in resource allocation and pricing. The customers are usually charged on hourly basis in current pay-as-you-go billing mechanisms which is subject to unfairness [6]. Being an economically-oriented idea, cloud computing focuses on fairness as its key feature in a pricing scheme. Personal fairness and social fairness are two types of pricing fairness in the terminology of economics. Personal pricing fairness is meant to be subjective and reasonable for consumers. On the other hand, social fairness refers to overall fairness maintained among the users, using the same service. Charging the same cost for the same service being consumed is called social fairness. If pricing is unfair, it becomes a reason of dissatisfaction for the customers and as a result, service providers would fail to gain the loyalty of its customers [6]. In addition to that, service underutilization also depends upon the loyalty of the customers. It is estimated that many datacenters experience 5% to 20% of utilization of their total resources [42], which shows that the datacenters are significantly underutilized. Underutilization can be countered by having a dynamic and customer's historical record based resource management and