Optimal Resource Usage in Multi-Cloud Computing Environment

Veena Goswami, School of Computer Applications, KIIT University, Bhubaneswar, Odisha, India

Choudhury Nishkanta Sahoo, School of Computer Applications, KIIT University, Bhubaneswar, Odisha, India

ABSTRACT

Cloud computing has emerged as a new paradigm for accessing distributed computing resources such as infrastructure, hardware platform, and software applications on-demand over the internet as services. This paper presents an optimal resource management framework for multi-cloud computing environment. The authors model the behavior and performance of applications to integrate different service-providers for end-to-end-requirements. Each service model caters to specific type of requirements and there are already number of players with own customized products/services offered. Intercloud Federation and Service delegation models are part of Multi-Cloud environment where the broader target is to achieve infinite pool of resources. They propose an analytical queueing network model to improve the efficiency of the system. Numerical results indicate that the proposed provisioning technique detects changes in arrival pattern, resource demands that occur over time and allocates multiple virtualized IT resources accordingly to achieve application Quality of Service targets.

Keywords: Cloud Computing, Federation of Clouds, Multi-Cloud, Performance, Quality Of Service, Queueing

1. INTRODUCTION

Cloud computing is a general term for system architectures that involves delivering hosted services over the Internet. Cloud computing services are offered on a pay-as-you-go basis and assure considerable reduction in hardware and software investment costs, as well as energy costs. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), which includes hardware, storage, servers, and networking components are made accessible over the Internet; Platforms-as-a-Service (PaaS), which includes computing platforms — hardware with operating systems, virtualized servers, and the like; and Software-as-a-Service (SaaS), which includes software applications and other hosted services. A cloud service differs from traditional hosting in three principal aspects. First, it is provided on demand,
typically by the minute or the hour; second, it is elastic since the user can have as much or as little of a service as they want at any given time; and third, the service is fully managed by the provider (Brunette & Mogull, 2009; Mell & Grance, 2009; Vaquero et al., 2009).

Large service centers have been set up to provide comprehensive services by sharing the IT resources to clients. Companies often outsource their IT infrastructure to third party service providers to reduce the management cost. This extends to the efficient use of resources and a step-down of the operating costs. The service providers and their clients often negotiate utility based Service Level Agreements (SLAs) to manage its resources to maximize its profits. (Ardagna et al., 2005) proposed a Service level agreements (SLA) based profit optimization in multi-tier systems. Utility based optimization approaches provides, load balancing and obtain the best trade-off between job classes for Quality of Service levels.

Efficiently managing cloud resources and maintaining Service level Agreements for cloud services is a enormous challenge. Performance virtualization techniques have been employed to provide effective performance of computer service subject to QoS metrics such as response time, throughput, network utilization, have been extensively studied in the (Slothouber, 1995; Karlapudi & Martin, 2004; Lu & Wang, 2005). Web server performance model using an open queueing network was employed to model the behavior of Web servers on the Internet (Slothouber, 1995). (Karlapudi & Martin, 2004) studied a Web application tool for the performance prediction of Web applications between specified end-points. Cloud centers as the enabling platform for dynamic and flexible application provisioning is facilitated by exhibiting data center’s capabilities as a network of virtual services. Hence, users are able to access and deploy applications from any place in the Internet driven by the demand and Quality of Service (QoS) requirements (Buyya et al., 2009). IT companies are freed from the trivial task of setting up basic hardware and software infrastructures by using clouds as the application hosting platform. Thus, they can focus more on innovation and creation of business values for their application services (Armbrust et al., 2010).

This paper focuses on an analytical model through which Quality of service (QoS) is ensured by obtaining important performance indicators such as mean request response time, blocking probability, probability of immediate service and probability distribution of number of tasks in the system. This model allows cloud operators to tune the parameters such as the number of servers on one side, and the values of blocking probability and probability that a task request will obtain immediate service, on the other. Successful provision of cloud services and, consequently, widespread adoption of cloud computing necessitates accurate performance evaluation that allows service providers to dimension their resources in order to fulfill the Service Level Agreements with their customers.

The rest of this paper is organized as follows. Section 2 contains a brief description of the system model and its analysis for the multi-cloud computing environments. Analytical model and its analysis is given in Section 3. Some important performance measures are presented in Section 4. Section 5 contains numerical analysis to show the effectiveness of the model parameters. Section 6 concludes the paper.

2. SYSTEM DESIGN

We discuss different scenarios for the need of multi-cloud Environments.

1. Multi-Cloud environment consisting of multiple SaaS cloud services:- For the purpose of optimization, an organization may need to outsource a number of marginal functions to cloud services offered by different vendors. For example, it is highly likely that an enterprise may use Gmail for the email services and SalesForce.com for the HR service. This means that the many features (e.g. address book, calendar, appointment booking, etc.) in the email
Decision Making, Dashboard Displays, and Human Performance in Service Systems

Survey of the State-of-the-Art of Cloud Computing
[www.igi-global.com/article/survey-state-art-cloud-computing/60407?camid=4v1a](www.igi-global.com/article/survey-state-art-cloud-computing/60407?camid=4v1a)