Digital Communications and Networks (2015) 1, 181–190

HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

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

# Modeling and performance analysis for composite network-compute service provisioning in software-defined cloud environments

Qiang Duan

*Information Sciences & Technology Department, The Pennsylvania State University Abington, PA 19001, United States*

Received 18 December 2014; received in revised form 19 February 2015; accepted 16 May 2015
Available online 1 June 2015

**KEYWORDS**
Cloud computing;
Composite service provisioning;
Software-defined Cloud environment;
Service modeling;
Performance analysis

**Abstract**
The crucial role of networking in Cloud computing calls for a holistic vision of both networking and computing systems that leads to composite network-compute service provisioning. Software-Defined Network (SDN) is a fundamental advancement in networking that enables network programmability. SDN and software-defined compute/storage systems form a Software-Defined Cloud Environment (SDCE) that may greatly facilitate composite network-compute service provisioning to Cloud users. Therefore, networking and computing systems need to be modeled and analyzed as composite service provisioning systems in order to obtain thorough understanding about service performance in SDCEs. In this paper, a novel approach for modeling composite network-compute service capabilities and a technique for evaluating composite network-compute service performance are developed. The analytic method proposed in this paper is general and agnostic to service implementation technologies; thus is applicable to a wide variety of network-compute services in SDCEs. The results obtained in this paper provide useful guidelines for federated control and management of networking and computing resources to achieve Cloud service performance guarantees.


## 1. Introduction

Cloud computing is a large scale distributed computing paradigm driven by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable computing

*E-mail address:* qduan@psu.edu
Peer review under responsibility of Chongqing University of Posts and Telecommunications.

http://dx.doi.org/10.1016/j.dcan.2015.05.003

functions are delivered on demand as services to external customers over the Internet [1]. Networking plays a crucial role in Cloud computing. From a service provisioning perspective, the services received by Cloud end users comprise not only computing functions provided by Cloud data centers but also communications functions offered by networks. Results obtained from recent study on Cloud service performance have indicated that networking has a significant impact on quality of Cloud services, and in many cases data communications become a bottleneck that limits Clouds from supporting high-performance applications [2,3]. Therefore, networks with Quality of Service (QoS) capabilities become an indispensable ingredient for high-performance Cloud service provisioning.

For example, consider a scenario in which an application utilizes the Cloud infrastructure for storing and processing a large data set and requires upper bounded response delay. This application may use the computing capability of Amazon EC2 (Elastic Compute Cloud) and the storage capacity of Amazon S3 (Simple Storage Service). In order for the user to access EC2 and S3, network services must also be provided for data transmissions from the user to S3 virtual disk, between S3 disk and the EC2 server (if EC2 and S3 are located at different sites), and from EC2 server back to the user. Therefore, the end-to-end service offered to the user is essentially a composition of both Cloud services and network services. In order to meet the service delay requirement of the application, sufficient amount of networking resources (e.g. transmission bandwidth and packet forwarding capacity) must be provided to guarantee network delay performance in addition to the computing and storage resources offered by the Cloud infrastructure for meeting data processing and storing requirements.

The significant role that networking plays in Cloud service provisioning calls for a holistic vision of the computing and networking systems involved a Cloud environment. Such a vision requires federated management, control, and optimization of computing and networking resources for composite service provisioning. Software-Defined Networking (SDN) is one of the latest revolutions in the networking field, which decouples network control and data forwarding functions; thus enabling network control to be programmable and underlying network infrastructure to be abstracted for applications [4]. The logically centralized control plane in SDN allows upper layer applications to program the underlying network platform through a standard API; therefore can provide better support for Cloud computing. Expectation of more agile Cloud services also requires a high degree of programmability for computing infrastructure, which leads to software-defined computing and storage. Software-defined network, compute, and storage systems, when integrated together, enable an environment with fully automated provisioning and orchestration of IT infrastructures for Cloud computing, which is referred to as Software-Defined Cloud Environment (SDCE) [5].

A key to realize the SDCE notion is orchestration of heterogeneous network, compute, and storage systems for composite service provisioning. The Service-Oriented Architecture (SOA) provides an effective mechanism for heterogeneous system integration through loose-coupling interactions among system components. SOA has been widely adopted in Cloud computing via the IaaS, PaaS, and SaaS paradigms. Applying SOA in the field of networking leads to the *Network-as-a-Service* (*NaaS*) paradigm that enables encapsulation and virtualization of networking resources in the form of SOA-compliant *network services*. NaaS allows network infrastructure to be virtualized, exposed, and accessed as services that can be orchestrated with computing services in a Cloud environment to provision composite network–compute services to Cloud users [6]. Recently NaaS has been proposed as a key mechanism in SDN for achieving end-to-end QoS provisioning [7]. Therefore SOA may form a basis for service provisioning in SDCEs.

As SDCE being rapidly adopted by service providers, it becomes important to obtain thorough understanding about performance of composite network–compute service provisioning, which is the service performance actually perceived by Cloud users. Since networking has a strong impact on end-to-end performance for Cloud service provisioning, Cloud service performance should be evaluated with a holistic vision of both computing and networking aspects. For example, suppose a biology lab creates 100 GB of raw data that will be processed in Amazon EC2 Cloud. Assume that the lab obtained 10 EC2 virtual machine instances and each instance can process 20 GB data per hour, then the total processing time of the Cloud service is only 30 min. However, if the lab uses a network service that offers 200 Mb/s throughput for data transmission to the EC2 server, then even the single-trip delay for data transmission from the lab to EC2 servers will be 67 min. In this example network delay for round-trip data transmission contributes more than 80% of the total service delay; thus demonstrating the significant impact of networking on Cloud service performance.

Analytical modeling and analysis provide an effective approach to obtaining insights about end-to-end service performance. Cloud performance analysis has attracted attention of the research community and many results have been reported in the literatures. However, the significant impact of network performance on Cloud service provisioning has not been sufficiently considered in these works. On the other hand, although network performance has been extensively studied, currently available techniques typically lack the ability to analyze composite systems that consist of heterogeneous networking and computing functionalities. Therefore, system modeling and performance analysis for composite network–compute service provisioning in SDCEs are still an open problem.

Network–compute service orchestration in SDCEs brings in new challenges to service modeling and performance analysis. Key features of Cloud computing and NaaS, such as virtualization and abstraction of networking and computing resources, make service provisioning independent of system implementations; thus requiring modeling and analysis methods to be general and agnostic to network and Cloud implementations. In order to tackle this challenge, a novel modeling approach is proposed in this paper for characterizing the service capabilities of composite network–compute systems. Analysis techniques are developed based on this model for evaluating delay performance of composite network–compute services. The developed modeling and analysis techniques are general and agnostic to network and Cloud implementations; thus are applicable to composite network–compute service