Chapter 76
Software-Defined Networking for Scalable Cloud-Based Services to Improve System Performance of Hadoop-Based Big Data Applications

Desta Haileselassie Hagos
University of Oslo (UiO), Norway

ABSTRACT

The rapid growth of Cloud Computing has brought with it major new challenges in the automated manageability, dynamic network reconfiguration, provisioning, scalability and flexibility of virtual networks. OpenFlow-enabled Software-Defined Networking (SDN) alleviates these key challenges through the abstraction of lower level functionality that removes the complexities of the underlying hardware by separating the data and control planes. SDN has an efficient, dynamic, automated network management, higher availability and application provisioning through programmable interfaces which are very critical for flexible and scalable cloud-based services. In this study, the author explores broadly useful open technologies and methodologies for applying an OpenFlow-enabled SDN to scalable cloud-based services and a variety of diverse applications. The approach in this paper introduces new research challenges in the design and implementation of advanced techniques for bringing an SDN-enabled components and big data applications into a cloud environment in a dynamic setting. Some of these challenges become pressing concerns to cloud providers when managing virtual networks and data centers, while others complicate the development and deployment of cloud-hosted applications from the perspective of developers and end users. However, the growing demand for manageable, scalable and flexible clouds necessitates that effective solutions to these challenges be found. Hence, through real-world research validation use cases, this paper aims at exploring useful mechanisms for the role and potential of an OpenFlow-enabled SDN and its direct benefit for scalable cloud-based services. Finally, it demonstrates the impact of an OpenFlow-enabled SDN that fully embraces the opportunities and challenges of cloud infrastructures to improve the system performance of Hadoop-based big data applications by utilizing the network control capabilities of an OpenFlow to solve network congestion.

DOI: 10.4018/978-1-5225-7501-6.ch076

1. INTRODUCTION AND MOTIVATION

Cloud Computing is a model that allows users to dynamically access and configure a shared pool of computing resources (e.g., networking elements, data storage space, servers, applications and services) over the Internet stored at remote locations. The main enabling technology for *Cloud Computing* is *Virtualization*. Cloud computing has revolutionized how software applications are offered to end users, with economies of scale obtained by creating large data centers that provide vast amounts of computation, communication and storage resources. Hence, cloud-based environments are playing an increasingly important role in the context of big data analytics and storage applications. Due to the growth of cloud-based services, users expect on-demand self-service access to applications, infrastructure, and other IT resources. Cloud-based services and applications have enabled massive data centers to exploit the capabilities delivered by *Virtualization* technologies to increase automation, predictability, efficiency, self-service provisioning, continuity, manageability, quality of service, scalability and flexibility. To efficiently address this increasing demand and rapid growth of *Cloud Computing* technology, network and service providers need an effective and scalable network solution with significantly low operating costs through simplified hardware, software and management. The *Virtualization* technology has allowed cloud providers to consolidate the application workloads of a large and diverse set of end-user tenants on a smaller number of physical machines, enabling a high resource utilization while keeping management costs low. However, performance unpredictability in the cloud is a critical issue for many users and it is considered as one of the major obstacles for cloud computing (Armbrust et al., 2009). In this paper, we mainly focus on the role of SDN for scalable cloud-based services and big data applications. The main contributions of this paper are fourfold.

- Motivating the direct role of an *OpenFlow-enabled* SDN for scalable cloud-based services and performance of big data applications from the perspective of future networking.
- Exploring a finer and more effective mechanisms in the dimension of bringing SDN and cloud-based services and big data applications together.
- Introducing new research and technical challenges in the design and implementation of advanced techniques to use SDN-enabled components in a cloud setting.
- Demonstrating the impact of an *OpenFlow-enabled* SDN that fully embraces the opportunities and challenges of cloud infrastructures to improve the performance of big data applications by taking failure scenarios into account (solving network congestion).

The rest of the paper is structured as follows: In Section 2 we give a brief overview of the role of SDN to the cloud and big data applications. Section 3 provides the system architectures of our study. Section 4 presents the research validation use cases for our study. In Section 6 we discuss the state-of-the-art and related works. Section 7 presents some open research questions to be investigated in greater detail. Finally, Section 8 draws conclusions of the paper and provide an outlook on our intended future work.

2. SDN FOR SCALABLE CLOUD-BASED SERVICES AND BIG DATA APPLICATIONS

This section presents the role of SDN for scalable cloud-based services and Hadoop-based big data applications.