Chapter 12
Energy-Efficiency in a Cloud Computing Backbone

Burak Kantarci
University of Ottawa, Canada

Hussein T. Mouftah
University of Ottawa, Canada

ABSTRACT

Cloud computing combines the advantages of several computing paradigms and introduces ubiquity in the provisioning of services such as software, platform, and infrastructure. Data centers, as the main hosts of cloud computing services, accommodate thousands of high performance servers and high capacity storage units. Offloading the local resources increases the energy consumption of the transport network and the data centers although it is advantageous in terms of energy consumption of the end hosts. This chapter presents a detailed survey of the existing mechanisms that aim at designing the Internet backbone with data centers and the objective of energy-efficient delivery of the cloud services. The survey is followed by a case study where Mixed Integer Linear Programming (MILP)-based provisioning models and heuristics are used to guarantee either minimum delayed or maximum power saving cloud services where high performance data centers are assumed to be located at the core nodes of an IP-over-WDM network. The chapter is concluded by summarizing the surveyed schemes with a taxonomy including the cons and pros. The summary is followed by a discussion focusing on the research challenges and opportunities.

1. INTRODUCTION

With the advent of cloud computing, distributed computing and storage resources have become available ubiquitously via a new business model that offers dynamic provisioning and release of the service facilities with respect to the pay-as-you-go tariff (Armbrust et al., 2009; Mell & Grance, 2011). Besides several existing definitions (Vaquero et al., 2009), Q. Zhang et al. (2010) define cloud computing as a novel business model that brings several distributed system concepts together.

Cloud services are expected to dominate the future Internet to deliver several applications such as e-health (Rosenthal et al., 2010), e-learning (Sultan, 2010), scientific computation (Srirama

DOI: 10.4018/978-1-4666-6539-2.ch012

et al., 2012) and multimedia services (Shi et al., 2011). Energy management has been pointed out as one of the challenges in cloud computing (Zhang et al., 2010; Hayes, 2008). On the other hand, due to the energy bottleneck problem in the Internet, power consumption of the Information and Communication Technologies (ICTs) has become a major concern (Hinton et al., 2011). Furthermore, Greenhouse Gas (GHG) emissions of the ICTs are expected to be 8% in the next few years unless sustainability is addressed (Leisching & Pickavet, 2009).

Energy-efficiency of cloud computing has three main aspects, namely processing, storage and transport (Baliga et al., 2011). Processing and storage refers to the management policies within the data centers whereas transport of cloud services denotes transmission from/towards/between the data centers (Mouftah & Kantarci, 2013). Despite the benefits of offloading the local resources by migrating them to the data centers, widely adoption of cloud-based services introduce increased traffic intensity in the backbone network, and consequently increased power and energy consumption.

This chapter studies the energy-efficient transportation of cloud services over the Internet backbone. To this end, unicast, anycast and manycast communication modes are considered. In cloud computing, services are provisioned by a shared pool of resources; hence at the time of service request at a source node, the destination node is not known. The service request can be destined to either any eligible destination (Develder et al., 2012) or distributed among several eligible destinations (Charbonneau & Vokkarane, 2010). The former is denoted by anycast while the latter is called manycast.

Although, migration of the services to remote servers is expected to increase the network traffic, data centers are the most power hungry elements of a cloud system (Kliazovich et al., 2012). Therefore, this chapter also considers solutions which aim at energy-efficiency of both data centers and the transport network while accommodating the cloud services. To this end, this chapter provides a survey of anycast and manycast-based demand provisioning solutions, network design schemes considering the design of the backbone network with data centers and energy-efficient virtual network reconfiguration schemes. Among these surveyed schemes, a Mixed Integer Linear Programming (MILP)-based formulation and heuristics for its solution will be presented in more detail since these schemes will be used in a case study for addressing the trade-off between the provisioning delay and energy-efficiency.

There are four main models of cloud computing, namely public clouds, private clouds, hybrid clouds and community clouds. In public clouds, all resources are available to public whereas in private clouds a single organization owes and utilizes the cloud infrastructure. Community clouds enable sharing the cloud infrastructure between several organizations whereas hybrid clouds are the combinations of several cloud infrastructures. The scope of this study is limited to the public cloud deployment, and we will consider the minimalist infrastructure in Figure 1 as a basis throughout the chapter.

The chapter is organized as follows. Section 2 gives an overview of current energy efficiency solutions for the Internet of Clouds and data centers. In Section 3, we define the provisioning techniques for the Internet backbone for the delivery of cloud services based on anycast and manycast paradigms. Section 4 presents a case study where the cloud services are submitted to the data centers based on the manycast paradigm and delivered together with the non-data center demands over the same transport medium. Finally, Section 5 summarizes the chapter, presents a comparison of the studied schemes, points out the open issues in this field and gives future directions to the researchers in this field.

Related Content

A Comprehensive Survey on Techniques Based on TPM for Ensuring the Confidentiality in Cloud Data Centers
Arun Fera M., M. Saravanapriyaand J. John Shiny (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications (pp. 446-458).*
www.irma-international.org/chapter/a-comprehensive-survey-on-techniques-based-on-tpm-for-ensuring-the-confidentiality-in-cloud-data-centers/224587

An IoT-Based Framework for Health Monitoring Systems: A Case Study Approach
N. Sudhakar Yadav, K. G. Srinivasaand B. Eswara Reddy (2019). *International Journal of Fog Computing (pp. 43-60).*
www.irma-international.org/article/an-iot-based-framework-for-health-monitoring-systems/219360

Multi-Layer Token Based Authentication Through Honey Password in Fog Computing
Praveen Kumar Rayani, Bharath Bhushanand Vaishali Ravindra Thakare (2018). *International Journal of Fog Computing (pp. 50-62).*
www.irma-international.org/article/multi-layer-token-based-authentication-through-honey-password-in-fog-computing/198412

Cloud Security in E-Commerce Applications
Shah Rukh Malik, Mujahid Rafiqand Muhammad Ahmad Kahloon (2020). *Cloud Computing Applications and Techniques for E-Commerce (pp. 50-67).*
www.irma-international.org/chapter/cloud-security-in-e-commerce-applications/247594

A Novel Resource Management Framework for Fog Computing by Using Machine Learning Algorithm
Shanthi Thangam Manukumarand Vijayalakshmi Muthuswamy (2020). *Architecture and Security Issues in Fog Computing Applications (pp. 42-52).*
www.irma-international.org/chapter/a-novel-resource-management-framework-for-fog-computing-by-using-machine-learning-algorithm/236439