Chapter 11
Energy–Efficiency in Cloud Data Centers

Burak Kantarci
University of Ottawa, Canada

Hussein T. Mouftah
University of Ottawa, Canada

ABSTRACT

Cloud computing aims to migrate IT services to distant data centers in order to reduce the dependency of the services on the limited local resources. Cloud computing provides access to distant computing resources via Web services while the end user is not aware of how the IT infrastructure is managed. Besides the novelties and advantages of cloud computing, deployment of a large number of servers and data centers introduces the challenge of high energy consumption. Additionally, transportation of IT services over the Internet backbone accumulates the energy consumption problem of the backbone infrastructure. In this chapter, the authors cover energy-efficient cloud computing studies in the data center involving various aspects such as: reduction of processing, storage, and data center network-related power consumption. They first provide a brief overview of the existing approaches on cool data centers that can be mainly grouped as studies on virtualization techniques, energy-efficient data center network design schemes, and studies that monitor the data center thermal activity by Wireless Sensor Networks (WSNs). The authors also present solutions that aim to reduce energy consumption in data centers by considering the communications aspects over the backbone of large-scale cloud systems.

DOI: 10.4018/978-1-4666-4522-6.ch011
INTRODUCTION

Cloud computing is a novel concept to run the Information and Communication Technology (ICT) business in a more efficient manner (Zhang et al., 2010). Many applications including e-health, scientific computation and multimedia content delivery are aimed to be provided over the cloud. Due to being data intensive and introducing data communications between high performance servers, data centers will be the main drivers in cloud computing by maximizing the computing resource utilization via virtualization technology (Sakr et al., 2011). On the other hand, according to the EPA Report on Server and Data Center Energy Efficiency (ENERGY-STAR Program, 2007) the US Data Centers contributed to 1.5% of the total electricity consumption in the country which had an annual cost of $4.5 billion. Furthermore, this ratio has been reported to almost double by the end of 2011.

Kachris and Tomkos point out power consumption as one of the most challenging issues in the design of data centers due to the doubling of the power budget to accommodate the tremendous increase in the bandwidth requirements and peak performance (2012). As mentioned in the corresponding survey, servers and storage units dominate the power consumption of the IT equipment in a data center whereas the data center network, as well as the other networking devices, contribute to around one quarter of the total IT power consumption in the data centers. Dynamic Frequency and Voltage Scaling (DFVS) is one of the promising techniques which dynamically adjusts the frequency of the CPU of a server to save power (Sarood et al., 2012) whereas energy-aware Virtual Machine (VM) consolidation is another promising technique to assure energy-efficiency in data centers. Virtualization technology enables sharing the same physical resources on a server among several applications in order to improve resource and hardware utilization and isolate the applications in terms of fault and performance. Furthermore, dynamic consolidation of VMs can significantly reduce energy consumption by offloading some physical hosts and enable them to be switched off (Belgolazov and Buyya, 2010). Besides the physical hosts, powering-off the idle network equipment can also enable significant power savings in a data center (Heller et al., 2010). Furthermore, virtualization of the data center network introduces significant savings in the energy consumption (Bari et al., in press). Use of Massive Arrays of Idle Disks offers the opportunity of nearline storage and spinning up in response to an access request while keeping the rest of the idle disks in spun down state. On the other hand, when the server is heavily loaded, redirecting the requests to other disks in a conventional RAID-based system can lead to further improvements in the energy savings (Wang et al., 2008).

Not only are the IT equipments power consuming components of a data center, non-IT equipments such as lighting, Uninterrupted Power Supply (UPS), and Heating, Ventilation and Air Conditioning (HVAC) equipment also consume significant power in a data center leading to a degradation in the Power Usage (In)Efficiency (PUE). Basically PUE denotes the ratio of the total power consumption to the power consumption of the IT equipment as shown in Equation 1. As seen in the equation, ideally, PUE is aimed to be close to one so that data center is powered mostly for running the IT equipments (Lawrence, 2006). Typical PUE value for a data center is 2 but today several data center operators report significant enhancements in PUE values such as 1.3 and even below (“Data Center Knowledge,” 2011).

\[
PUE = \frac{Total\ IT\ Power + Total\ Non - IT\ Power}{Total\ IT\ Power}\quad (1)
\]

In order to enhance data center energy-efficiency and PUE, cyber-physical systems have been proposed where workload placement is done based on cooling and/or thermal factors.