

Chapter 2

Resource Scheduling for Energy–Aware Reconfigurable Internet Data Centers

Mohammad Shojafar

Sapienza University of Roma, Italy

Nicola Cordeschi

Sapienza University of Rome, Italy

Enzo Baccarelli

Sapienza University of Rome, Italy

ABSTRACT

The pervasive use of cloud computing and the resulting growing number of Internet data centers have brought forth many concerns, including electrical energy cost, energy dissipation, cooling and carbon emission. Therefore, the need for efficient workload schedulers which are capable of minimizing the consumed energy becomes increasingly important. Green computing, a new trend for high-end computing, attempts to approach this problem by delivering both high performance and reduced energy consumption. Motivated by these considerations, in this chapter, we propose a joint computation-and-communication adaptive resource-provisioning scheduler for virtualized data centers, e.g., the Internet Data Center (IDC) scheduler, which exploits the DVFS-enabled reconfiguration capability of the underlying virtualized computing/communication platform. Specifically, we present and test a dynamic resource provisioning scheduler, which adaptively controls the execution time and bandwidth usage of each input job, as well as the internal and external switching costs on per-Virtual Machine (VM) basis.

BACKGROUND AND MOTIVATIONS

Green Cloud Computing (GCC) refers to the environmental benefits that Internet-based Information Technology (IT) services may offer. The term combines the words Green – meaning environmentally friendly – and Cloud. GCC aims at providing various models and techniques to seamlessly integrate

DOI: 10.4018/978-1-5225-0287-6.ch002

the management of computing-communication virtualized platforms, which provide Quality of Service (QoS), robustness and reduced energy utilization (Cugola & Margara, 2012; Shamshirband, Petković, Čojbašić, Nikolić, Anuar, Shuib, & Akib, 2014). The main challenge in GCC is to minimize the energy usage, while still meeting the QoS requirements of the supported applications. For this purpose, an energy-aware scheduler is needed that jointly accounts for the networking and computing resource allocation in the Cloud. In this respect, several works focused on the energy efficiency of Cloud infrastructures, by exploiting (Mishra, Jain, & Durresi, 2012; Baliga, Ayre, Hinton, & Tucker, 2011), and/or virtualization of the cloud resources (Azodolmolky, Wieder, & Yahyapour, 2013). DVFS is applied in most of the modern computing units, such as cluster computing and supercomputing, in order to reduce power consumption and achieve high reliability and availability. Resource virtualization refers to instantiating several VMs on the same physical server, in order to reduce the number of physical CPUs, while improving the resource utilization. Furthermore, one of the main requirements for Cloud computing environments is the provision of reliable Service Level Agreements (SLAs). SLAs can be managed *globally* or *locally* by the Cloud providers by relying on suitable scheduling policies (Shamshirband, Shojafar, Hosseinabadi, Kardgar, Nasir, & Ahmad, 2015). Another issue of growing concern in Cloud environments is to evenly distribute huge amount of workloads over various servers, which is referred to as *load balancing*. Load balancing algorithms seek to distribute workloads across a number of servers, so that the average execution times are minimized (Warneke & Kao, 2011). Load balancing schemes may be *static* or *dynamic* ones. In static schemes, the current state of the servers is not considered when dispatching the workloads. Examples of such schemes include Random Selection of servers and Round Robin policies. However, dynamic schemes involve direct notification or indirect inference of the servers' states by the load balancer (Warneke & Kao, 2011).

In this chapter, we present and test the performance of a new adaptive resource scheduler. It aims at minimizing the energy consumption induced by the computing, communication and reconfiguration costs in Internet-based virtualized data centers. In such environments, work is performed under hard limits on the execution times of the offered jobs. Our scheduler accounts for the dynamic load balancing, by using online job decomposition. For this purpose, we have developed a framework for the Cloud resource management that considers the optimum joint allocation of VM processing frequencies and link bandwidths. Consider that the resulting energy model is non-convex, so that, we have developed a mathematical approach to turn non-convexity into convexity.

RELATED WORK

Updated surveys of current technologies and open communication challenges about energy-efficient data centers have been recently presented in (Hirzel, Soule, Schneider, Gedik, & Grimm, 2014; Cordeschi, Shojafar, & Baccarelli, 2013; Cordeschi, Amendola, & Baccarelli, 2014; Cordeschi, Shojafar, Amendola & Baccarelli, 2015; Shojafar, Cordeschi, Amendola & Baccarelli, 2015a; Shojafar, Javanmardi, Abolfazli & Cordeschi, 2015b; Baccarelli & Biagi, 2003a, 2003b). Specifically, power management schemes that exploit DVFS techniques for performing resource provisioning are the focus of (Baliga, Ayre, Hinton, & Tucker, 2011; Azodolmolky, Wieder, & Yahyapour, 2013; Qian, He, Su, Wu, Zhu, Zhang, & Zhang, 2013; Javanmardi, Shojafar, Amendola, Cordeschi, Liu, & Abraham, 2014). Although these contributions consider hard deadline constraints, they do not consider, indeed, the performance penalty and the energy-vs.-delay tradeoff stemming from the finite transmission rate of the utilized network infra-

24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/resource-scheduling-for-energy-aware-reconfigurable-internet-data-centers/159038

Related Content

A Grid Aware Large Scale Agent-based Simulation System

Dan Chen (2009). *Quantitative Quality of Service for Grid Computing: Applications for Heterogeneity, Large-Scale Distribution, and Dynamic Environments* (pp. 299-319).

www.irma-international.org/chapter/grid-aware-large-scale-agent/28283

A Parallel Fuzzy Load Balancing Algorithm for Distributed Nodes Over a Cloud System

Mostefa Hamdani, Youcef Akloufand Hadj Ahmed Bouarara (2022). *International Journal of Grid and High Performance Computing* (pp. 1-22).

www.irma-international.org/article/a-parallel-fuzzy-load-balancing-algorithm-for-distributed-nodes-over-a-cloud-system/301576

Probabilistic Indices of Quality of Approximation

Annibal Parracho Sant'Anna (2008). *Rough Computing: Theories, Technologies and Applications* (pp. 162-174).

www.irma-international.org/chapter/probabilistic-indices-quality-approximation/28472

Applications of Supercomputers in Population Genetics

Gerard G. Dumancas (2015). *Research and Applications in Global Supercomputing* (pp. 176-200).

www.irma-international.org/chapter/applications-of-supercomputers-in-population-genetics/124342

A Waste Management Robot System: Its Implementation and Experimental Results

Keita Matsuo, Yi Liu, Donald Elmaziand Leonard Barolli (2015). *International Journal of Distributed Systems and Technologies* (pp. 1-12).

www.irma-international.org/article/a-waste-management-robot-system/126174