Computers and Electrical Engineering 39 (2013) 15–23

Contents lists available at SciVerse ScienceDirect

Computers and Electrical Engineering

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

High availability and efficient energy consumption for cloud computing service with grid infrastructure ☆

Young-Sik Jeong [a], Jong Hyuk Park [b,*]

[a] Department of Computer Engineering, Wonkwang University, Iksan, Republic of Korea
[b] Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Available online 7 April 2012

ABSTRACT

The main services in cloud computing are SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service). Among these services, server virtualization falls into IaaS which is a service for lowering server maintenance costs. Generally, the primary aim of server virtualization has been to improve system performance by replacing multiple servers with a single server with better performance. But it carries the problem of increased cost, needed for the installation of new servers for server virtualization.

In this paper, grid infrastructure is used for server virtualization in which existing servers are used rather than bringing in new servers. Server virtualization service is provided using scheduling algorithms for distributed servers or resources in grid computing. To overcome potential performance limitations that come from using existing servers, mathematical models of Meta and Sleep Servers under the grid infrastructure environment are used to provide server virtualization service with high availability.

1. Introduction

In knowledge information society, increasingly large amounts of data are driving the advancement of systems that process them [1–3]. The additional servers required for this advancement is rapidly increasing the maintenance costs for the information technology infrastructure. Furthermore, there are additional costs incurred, because more servers have to be installed, more space needs to be secured to house the servers, manpower has to be increased to maintain the servers, and more energy costs are needed to power the servers. In order to lower these costs needed for maintaining the IT infrastructure, many are turning to cloud computing for a solution [2,4,5].

Server virtualization is an integrated technology comprised of physical and logical aspects, whose aim is to increase availability of multiple servers. Server virtualization systems such as EMC (VMware) [4], Microsoft (Virtual Server 2005), SWsoft (Virtuozzo), Sun (Solaris Container) and XenSource (Xen Enterprise3) [7] are based on integration after switching to high-performance servers from existing servers. But to take advantage of server virtualization, existing servers have to be replaced for new high-performance servers because that's how server virtualization works: providing multiple guest OS environments using a single high-performance server. Thus a lot of costs are needed for the upgrade [4,6–9].

In this paper, to overcome the problem of increased costs needed for the server upgrade, grid infrastructure is used. In particular, server virtualization is implemented using grid infrastructure in order to use existing servers. It also presents a scheme for increasing availability of existing servers and for reducing energy costs. The issue of lowered performance that would result from using existing servers for server virtualization is handled using Sleep Server (sub server) and Grid

☆ Reviews processed and proposed for publication to Editor-in-Chief by Associate Editor Dr. Taeshik Shon.

* Corresponding author.
E-mail addresses: ysjeong@wku.ac.kr (Y.-S. Jeong), parkjonghyuk1@hotmail.com, jhpark1@seoultech.ac.kr (J.H. Park).

0045-7906/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compeleceng.2012.03.005

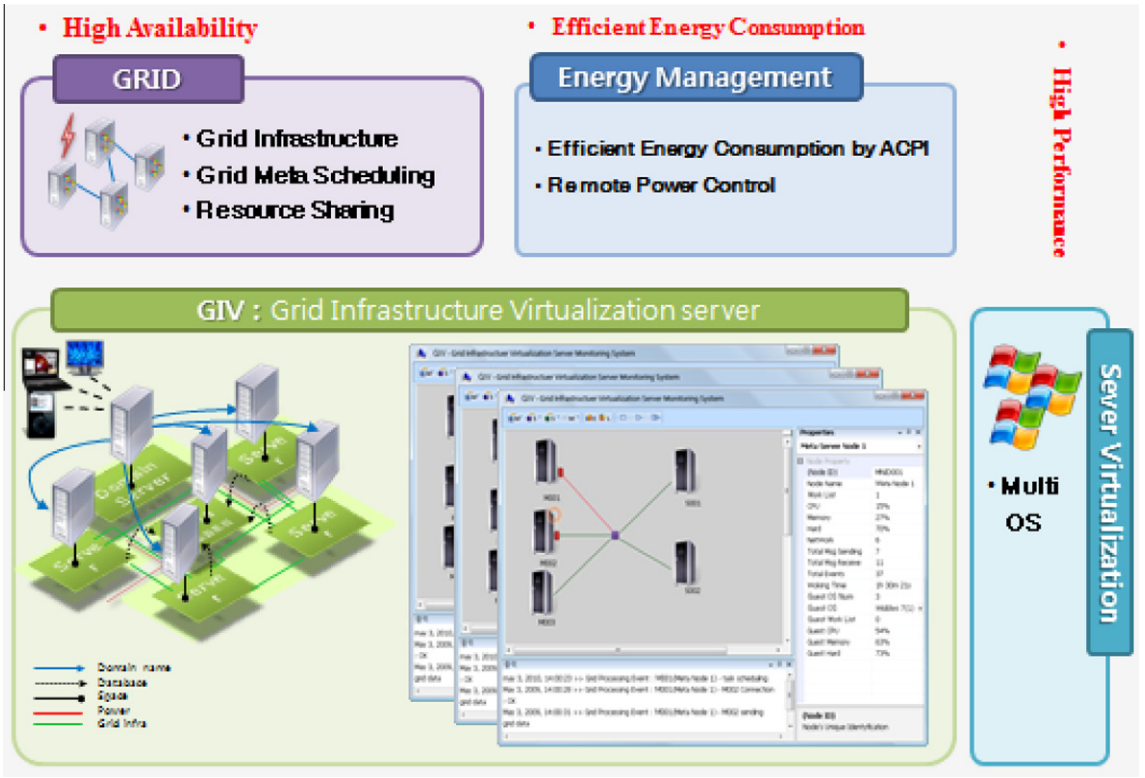

Fig. 1. Efficient server virtualization based on grid infrastructure.

infrastructure. This system is also designed to reduce energy consumption. In the system, the existing servers are divided into the Meta Server group and the Sleep Server group based on the grid infrastructure (as shown Fig. 1). For Meta Server, existing servers are turned into Virtual Nodes or Virtual Servers for the grouping while for Sleep Server, existing servers are mapped to Sleep Nodes for the grouping. The Grid Infrastructure Virtualization server (GIV) is designed and implemented, which performs scheduling of jobs and resources by controlling the virtualized server system. It also performs monitoring of job list and statistical processing.

2. Related works

In this section, two key technologies for this paper are discussed, namely server virtualization and grid infrastructure. Server virtualization is subcategorized into H/W partitioning, logical partitioning and SW partitioning. H/W partitioning refers to a method of partitioning in which entities are completely separated into hardware components. In logical partitioning, partitioning is done in the unit of firmware and resources can be used more efficiently compared to H/W partitioning because more partitions can be created. In S/W partitioning, partitioning is done in the unit of software codes. It is the most flexible of all the partitioning methods and many partitions and resources, such as processes, memory, I/O and disks are shared [1,6,7,10,11].

In this paper, S/W partitioning with a hypervisor is used to manage resources efficiently, as it has high flexibility and partitioning ability. The hypervisor offers abstraction and isolation features for virtual machines through CPU interrupt of the host system and state management. It is also an abstract software layer in the physical hardware that efficiently shares and distributes physical server resources.

Second, grid computing refers to performing computation-intensive jobs and/or data-intensive jobs by connecting every computing device such as PCs, servers, PDAs, etc. to a single network [1]. Generally, since it is rare to use 100% of a CPU while performing computing tasks, it is always idle to a certain extent. Grid computing is a technique to share the idle resources within a Grid Service Infrastructure and increase the performance speed focusing on a certain task [10,12,13].

Scheduling techniques for grid computing are broadly divided into centralized scheduling and de-centralized scheduling according to the way in which the tasks are treated [14]. This paper presents a system in which de-centralized scheduling is used to achieve flexible job scheduling by independently scheduling each Meta Node, which lowers the likelihood of occurrence of bottlenecks.

In this paper, we attempt to provide an efficient virtual computing environment by applying resource/task scheduling techniques based on grid infrastructure and by utilizing efficient server power maintenance techniques.