Information Sciences 319 (2015) 132–151

Contents lists available at ScienceDirect

Information Sciences

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

Prediction-based energy policy for mobile virtual desktop infrastructure in a cloud environment

Tien-Dung Nguyen, Pham Phuoc Hung, Tran Hoang Dai, Nguyen Huu Quoc, Cong-Thinh Huynh, Eui-Nam Huh *

Department of Computer Engineering, Kyung Hee University, Republic of Korea

A R T I C L E I N F O

Article history:
Received 15 June 2014
Received in revised form 9 February 2015
Accepted 14 February 2015
Available online 20 February 2015

Keywords:
Energy policy
VM state management
mVDI
Remote desktop service

A B S T R A C T

Using cloud services from mobile devices has become a growing trend because of its mobility and convenience. However, mobile devices join and leave cloud services more frequently than traditional computers, which causes energy inefficiency in a cloud data center. Waste, in the form of energy and cooling requirements, particularly occurs when a mobile device disconnects from a service, but the cloud servers, known as virtual machines (VMs), continue running. The VMs should transition to lower-power states instead remaining active. However, transition to a lower-power state causes a service delay when users reconnect to the service because VMs in a lower-power state are not ready to serve. Therefore, an efficient energy policy must not only maximize energy savings but also minimize service delays. In this paper, we propose two approaches to energy efficiency: an Instant Energy Policy (IEP) that can quickly find an appropriate low-power state based on a predicted disconnection time and a Prediction-based Energy Policy (PrEP) that determines when to transition VMs to a low-power state and when to return them to the active state based on each users activity history. IEP predicts the unknown disconnection time using the multisize sliding windows workload estimation technique, which supports a non-stationary environment. This method can quickly obtain an energy policy, but it is limited when disconnection time fluctuates widely. PrEP presents an improved approach to achieve an optimal global result with respect to both energy consumption and service delay. Through simulations with a real-world dataset collected by the MIT Human Dynamics Lab, we show that PrEP provides approximately 20% power saving over the benchmark policies while guaranteeing minimal service delay.

1. Introduction

The mobile computing revolution is changing many aspects of modern life in positive ways. By using mobile devices such as smartphones or tablets, people can easily play and work anytime, anywhere. However, mobile devices have their own limitations in CPU, memory, storage, battery, etc. These limitations cause cloud service performance degradation (including on-demand movies, 3D games, graphical design tools, etc.). Many promising approaches have been proposed to overcome

* Corresponding author.
E-mail addresses: ntiendung@khu.ac.kr (T.-D. Nguyen), hungpham@khu.ac.kr (P.P. Hung), dai.tran@khu.ac.kr (T.H. Dai), quoc@khu.ac.kr (N.H. Quoc), thinhhc@gmail.com (C.-T. Huynh), johnhuh@khu.ac.kr (E.-N. Huh).

http://dx.doi.org/10.1016/j.ins.2015.02.022
0020-0255/

these issues. The Mobile Virtual Desktop Infrastructure (mVDI) [8] is the evolution of virtualization technology for mobile devices. mVDI has the same look and feel as a traditional mobile device, but all its actions happen remotely in a cloud server. On the server side, audio/video output is not delivered to locally attached devices (such as the monitor and speakers) but redirected over the network to the client for display. Therefore, mVDI not only enhances the performance of mobile services but also reduces the CPU and energy demand of mobile devices.

In an mVDI scenario, each customer remotely controls one or more virtual machines (VMs) in a data center. Every VM consumes a specific amount of energy depending on the services running inside and the type of infrastructure (such as Openstack, Xen, and HyperV). If the number of users increases, the data center sees the effect in rising energy consumption and energy costs as the number of running VMs increases. [18] reports that the energy consumption of data centers rose by 56% from 2005 to 2010, and in 2010 accounted for between 1.1% and 1.5% of global electricity use. Apart from high operation costs, this results in substantial carbon dioxide emissions, about 2% of global emissions [13]. In this paper, we focus on reducing data center energy consumption by implementing efficient energy policies [5].

Regardless of the type of services or infrastructure, VMs consume different amounts of energy depending on their operational state (Fig. 1). In the following table, we present the power consumption of each VM state as measured in an Openstack system.

The active state consumes more energy than the paused or suspended state, which are also called lower-power states. We can reduce energy consumption by efficiently switching a VM between the active and lower-power states instead of keeping it in the active state. Herein, we will continue to discover when and how a VM is best transitioned to lower-power states.

Practically, as customers use VDI-based cloud services, they can face unexpected events, such as receiving phone/Kakao/Viber calls or messages, being interrupted by friends or a low battery, experiencing a weak wifi/3G/4G signal or a sudden OS system reboot, etc. Whenever such a thing happens, customers temporarily disconnect their devices from their VMs for a length of time that depends on the interrupting event. On the data center side, the VMs continue running without any connection to mobile devices, which wastes energy. Thus, disconnected VMs should be transitioned to lower-power states to reduce energy consumption.

However, the transition of VM states causes service delay. For example, imagine a person using cloud services who receives an incoming phone call. The user temporarily disconnects the services and receives the incoming call. The network monitor module detects that the VM has become disconnected and can be switched to a lower-power state. A service delay occurs when the user finishes the call and tries to reconnect the service because the VMs are in a lower-power state and not ready to serve. The delay time is indicated by the time needed to transition from a lower-power state to an active state. Fig. 1 shows that the delay for the paused and suspended states is 2 s and 6 s, respectively. Intuitively, VMs consuming less power have to spend more time to return to the active state. Maximizing energy savings and minimizing service delays are the challenges that will be addressed in this paper.

The most common power management policy is a timeout policy [23] which manages VM states based on time. VMs automatically transition to lower-power states after one given time and return to an active state after another given time. This solution is not optimal because the service time and the disconnection time are not fixed values. A better idea is known

VM State	Power (W)	Time to return active state (s)
Active	55.7	-
Paused	44.5	2
Suspended	43.6	6

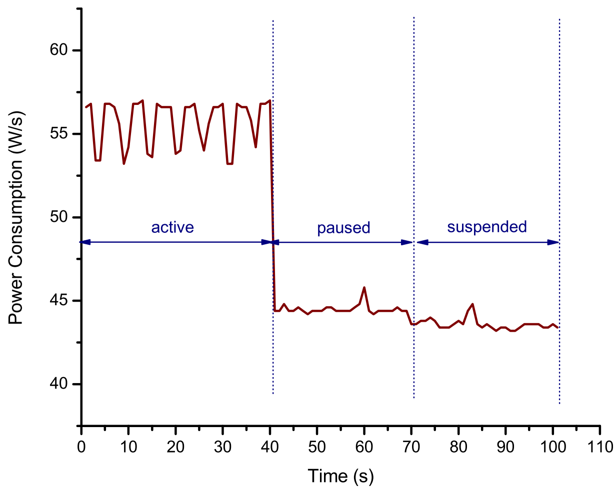

Fig. 1. Power consumption of server with one VM.