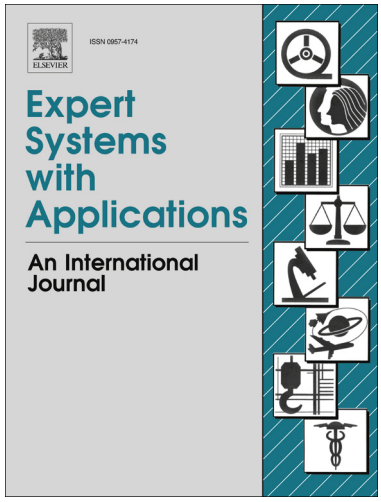

Workload modeling for virtual machine-hosted application

Weizhan Zhang, Jun Liu, Chen Liu, Qinghua Zheng, Wei Zhang

PII: S0957-4174(14)00635-6
DOI: http://dx.doi.org/10.1016/j.eswa.2014.10.012
Reference: ESWA 9613

To appear in: *Expert Systems with Applications*

Please cite this article as: Zhang, W., Liu, J., Liu, C., Zheng, Q., Zhang, W., Workload modeling for virtual machine-hosted application, *Expert Systems with Applications* (2014), doi: http://dx.doi.org/10.1016/j.eswa.2014.10.012

Workload Modeling for Virtual Machine-hosted Application

Weizhan Zhang[1,*], Jun Liu[1], Chen Liu[1], Qinghua Zheng[1], Wei Zhang[2]

SPKLSTN Lab, Department of Computer Science[1]
Xi'an Jiaotong University, Xi'an, China
{ zhangwzh@mail, liukeen@mail, gd.0818@stu, qhzheng@mail}.xjtu.edu.cn
* Corresponding author. Tel.: +86 133 2545 0992

Amazon.com, Inc.[2]
Seattle, WA 98109, USA
wzhan@amazon.com

Abstract: The workload complexity of virtual machine-hosted (VM-hosted) applications widens the gap between resource demand and allocation, which could lead to either resource underutilization or application overload. This study addresses this issue by providing a correlation analysis approach and a novel workload model for VM-hosted applications. Based on the grey system theory, the correlation analysis approach is used to determine the bottleneck resource of a given VM-hosted application, and the Workload Conversion Model (WCM) constructs a quantitative relationship between the application workload and the consumption of the bottleneck resource by modeling the resource consumption. We apply our model to a video-on-demand system as a case study. The model identifies that the VM memory is a key bottleneck resource of the system, and a bimodal relationship exists between the workload and VM memory consumption. The experimental results also show that the precision of the WCM is high, with a relative percentage error less than 5% and a variance ratio less than 25%.

Keywords: Workload modeling, VM-hosted application, Grey system theory, VOD application

1 Introduction

Server virtualization enables heterogeneous physical servers to be pooled into a shared infrastructure, and applications encapsulated in a virtual machine (VM) share these resources according to their requirements(M. Liu & Li, 2014). Server virtualization provides many benefits, such as a high resource utilization rate, considerable flexibility, and service level agreements (SLAs) about the quality of service of a VM-based application, which have contributed to the increase in VM-hosted application paradigms. Many applications, such as parallel computing, database service, and web service, use server virtualization to achieve the goal of maximizing resource utilization efficiency while adhering to the SLAs of the applications.

However, in a virtual environment, competition for system resources occurs among different applications running on the same physical server, which leads to a gap between resource demand and allocation. To illustrate, Figure 1 shows a virtualized framework for multiple VM-hosted applications, in which each node is divided into two VM instances running different applications simultaneously. Moreover, a group of VMs makes up a VM cluster that undertakes the same application workload, and each VM runs a single application. Mixed application workloads generate competition for system resources, and such competition results in excess system resource consumption. For example, *App1* and *App2* are on the same Virtual Machine Monitor (VMM) node, and the competition between *App1* and