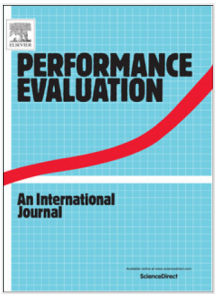
Accepted Manuscript



A quantitative model of application slow-down in multi-resource shared systems

Seung-Hwan Lim, Youngjae Kim

PII: S0166-5316(16)30230-9
DOI: http://dx.doi.org/10.1016/j.peva.2016.10.004
Reference: PEVA 1883

To appear in: *Performance Evaluation*

Received date: 20 August 2015
Revised date: 9 May 2016
Accepted date: 24 October 2016

Please cite this article as: S.-H. Lim, Y. Kim, A quantitative model of application slow-down in multi-resource shared systems, *Performance Evaluation* (2016), http://dx.doi.org/10.1016/j.peva.2016.10.004

A Quantitative Model of Application Slow-Down in Multi-Resource Shared Systems

Seung-Hwan Lim

Computational Data Analytics Group, Oak Ridge National Laboratory
Oak Ridge, TN 37831, USA

Youngjae Kim*

Department of Computer Science and Engineering, Sogang University
35 Baekbeom-ro, Mapo-gu, Seoul, Korea

Abstract

Scheduling multiple jobs onto a platform enhances system utilization by sharing resources. The benefits from higher resource utilization include reduced cost to construct, operate, and maintain a system, which often include energy consumption. Maximizing these benefits comes at a price – resource contention among jobs increases job completion time. In this paper, we analyze slow-downs of jobs due to contention for multiple resources in a system; referred to as *dilation factor*. We observe that multiple-resource contention creates non-linear dilation factors of jobs. From this observation, we establish a general quantitative model for dilation factors of jobs in multi-resource systems. A job is characterized by a vector-valued loading statistics and dilation factors of a job set are given by a quadratic function of their loading vectors. We demonstrate how to systematically characterize a job, maintain the data structure to calculate the dilation factor (loading matrix), and calculate the dilation factor of each job. We validate the accuracy of the model with multiple processes running on a native Linux server, virtualized servers, and with multiple MapReduce workloads co-scheduled in a cluster. Evaluation with measured data shows that the D-factor model has an error

*Corresponding Author
Email addresses: lims1@ornl.gov (Seung-Hwan Lim), youkim@sogang.ac.kr (Youngjae Kim)