FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Thesis Advisor Position

Thesis: Reproducible Partial-Load Experiments in Workload Colocation Analysis

Author: Andrej Podzimek

Advisor: Lubomír Bulej

Submitted: August 2016

Performance evaluation of computer systems is an important area of computer science research with a broad audience—understanding the performance of computer systems is of interest to data-center operators, software architects, and researchers alike. Performance evaluation traditionally focuses on peak performance of a system at certain task, measured using real-world or synthetic workloads executing on dedicated hardware. However, recent advances in virtualization technology have turned dedicated hardware into a mere design option, and increasing amount of computation takes place in virtual machines deployed on public or private cloud platforms.

In data centers, multiple virtual machines often share a single physical machine to increase the utilization of the hardware, because a server generally tends to be more energy-efficient at higher utilization. This is possible because very often virtual machines do not utilize all the hardware resources all the time. While the impact of interference between the colocated workloads on peak performance has been previously studied, the existing performance evaluation methodologies generally fail to take into account that the workloads still operate at partial load most of the time. In this context, Andrej Podzimek proposes a novel methodology for evaluating performance at partial load scenarios and demonstrates its relevance in an extensive experimental evaluation—showing that a common practice related to pinning of workloads to disjunctive subsets of processors to improve performance on NUMA hardware or to avoid performance interference through shared hardware resources does not lead to the best results when it comes to energy efficiency.

Even without dealing with partial-load scenarios, performance evaluation is a technically difficult discipline, complicated by various power management features of modern hardware platforms that researchers tend to disable to obtain more stable results. While technically non-trivial, the solution proposed by Andrej Podzimek does not require throwing away the existing approaches—instead it provides an additional experimental dimension that supports systematic exploration of partial-load scenarios. In addition, the solution requires no special power management settings, and thus enables experimental evaluation of realistic hardware configurations found in data centers.

Department of Distributed and Dependable Systems
Malostranské nám. 2/25, 118 00 Praha 1, Czech Republic
phone +420-951554245, fax +420-951554323
http://d3s.mff.cuni.cz

The work of Andrej Podzimek on this topic has been published at highly ranked peer-reviewed international conferences and in scientific journals. In this domain, peer-reviewed conference publications are the primary form of dissemination of research results, and the publication at the CCGRID 2015 conference deserves a special mention for having received the Best Paper Runner-Up Award, among five contenders for the Best Paper Award. Another highlight is a summary article which extends the CCGRID conference publication, published in the Journal on Future Generation Computer Systems. The FGCS journal is highly ranked (Q1) in the areas of Computer Networks and Communications; Hardware and Architecture; and Software. The versatility of Andrej Podzimek as a researcher is underscored by publications in other systems-related areas, such as non-blocking synchronization mechanisms, workload characterization, and performance awareness in software development.

I would also like to emphasize that the contributions of the thesis were achieved within the framework of an international research collaboration between the Department of Distributed and Dependable Systems at Charles University, Università della svizzera italiana in Lugano, Switzerland, and IBM Research in Zurich, Switzerland. All collaborators, and especially professor Petr Tůma, professor Walter Binder, and Lydia Chen, deserve credit for their role in the research this thesis contributes to. Having participated in this collaboration myself, I can attest that the work of Andrej Podzimek represents a major part of the research outlined in the thesis.

In conclusion, I believe that Andrej Podzimek has made a substantial contribution to the performance evaluation research, and am honored to recommend that Andrej Podzimek be awarded the doctoral degree.

Prague, September 5, 2016

Lubomír Bulej