Lecture Notes in Computer Science **11332**

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7407

Dalibor Klusáček · Walfredo Cirne
Narayan Desai (Eds.)

Job Scheduling Strategies for Parallel Processing

22nd International Workshop, JSSPP 2018
Vancouver, BC, Canada, May 25, 2018
Revised Selected Papers

Springer

Editors
Dalibor Klusáček
CESNET
Prague, Czech Republic

Narayan Desai
Google
Seattle, WA, USA

Walfredo Cirne
Google
Mountain View, CA, USA

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-10631-7 ISBN 978-3-030-10632-4 (eBook)
https://doi.org/10.1007/978-3-030-10632-4

Library of Congress Control Number: 2018965504

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains the papers presented at the 22nd workshop on Job Scheduling Strategies for Parallel Processing that was held in Vancouver, Canada, on May 25, 2018, in conjunction with the 32nd IEEE International Parallel and Distributed Processing Symposium (IPDPS 2018). The proceedings of previous workshops are also available from Springer as LNCS volumes 949, 1162, 1291, 1459, 1659, 1911, 2221, 2537, 2862, 3277, 3834, 4376, 4942, 5798, 6253, 7698, 8429, 8828, 10353, and 10773.

This year 12 papers were submitted to the workshop, of which we accepted seven. All submitted papers went through a complete review process, with the full version being read and evaluated by an average of four reviewers. We would like to especially thank to our Program Committee members and additional reviewers for their willingness to participate in this effort and their excellent, detailed, thoughtful reviews.

From its very beginning, JSSPP has strived to balance practice and theory in its program while encouraging vivid discussions with the audience. This combination was repeatedly shown to provide a rich environment for technical debate about scheduling approaches. This year, the workshop opened with a keynote delivered by John Wilkes. Principal Engineer at Google, John motivated and described Google's *Flex*, the key piece of Google's resource management system. The main goal of *Flex* is to assure that internal users have access to enough resources to meet their business needs. Doing it efficiently, reliably, and scalably (i.e., with little human intervention) is very challenging. John described how techniques like controlled over-subscription, risk management, and leveraging different service-level objectives are used to meet this challenge. The presentation is available at: http://jsspp.org/papers18/Google-Flex-JSSPP.pdf.

Papers accepted for this year's JSSPP focused on several interesting problems in resource management and scheduling domain. The first two papers discuss the issues related to imprecise job walltimes estimates. Job walltimes estimates, usually specified by users, are known to be very imprecise, which causes problems both to the users and to the scheduling policies. Soysal et al. present a novel approach to use job metadata for job classification and improved walltime prediction. Klusáček et al. present an experimental analysis that discusses how the use of walltime predictors impacts the actual performance of a job scheduler.

Azevedo and Suter present an experience report from a real infrastructure, describing their efforts to reduce the need for a "human expert" when scheduling large HTC workloads in a system that is subject to many operational constraints that may impede the optimization efforts of the scheduler.

Merzkyet et al. describe a new pilot-based scheduling system called RADICAL-Pilot. Unlike classic HPC scheduling systems that schedule jobs on a job-per-job basis, pilot-based systems decouple workload specification, resource selection, and task execution via job placeholders and late-binding, helping to satisfy the resource requirements of workloads comprising multiple tasks. In their paper,

Merzkyet et al. describe RADICAL-Pilot's design, architecture, and implementation, and characterize the good performance of RADICAL-Pilot when executing multiple concurrent tasks.

Bashizade et al. propose a dynamic mechanism for sharing GPUs among multiple tenants, i.e., users. This adaptive simultaneous multi-tenancy allows the GPU to be shared among multiple kernels, as opposed to single-kernel multi-tenancy that only runs one kernel on the GPU at any given time and static simultaneous multi-tenancy that does not adapt to events in the system. By dynamically adjusting the kernels' parameters at run-time — when a new kernel arrives or a running kernel ends — Bashizade et al. show that system throughput is improved by an average of 9.8%, compared with sequentially executed kernels.

Bhuiyan et al. present a stochastic optimization-based framework for robust decision-making in the selection of distributed resources for scientific workflows with uncertain demands over a given planning horizon. Using their novel two-stage stochastic programming model for resource selection, they demonstrate up to 30% and 54% cost reductions relative to solutions lacking explicit considerations of demand uncertainties for 24-month and 36-month planning horizons, respectively.

Last but not least, Abdelmoamen et al. present an approach to control resource usage among multiple tenants in a distributed system. In their approach they built upon the concept of actors, which are autonomous concurrently executing active objects. In this paper, authors compare two different ways of supporting resource control for actor systems built using the Scala's Akka library. Abdelmoamen et al. then experimentally establish the performance cost of using these approaches, as well as their impact on resource utilization.

We hope you can join us at the next JSSPP workshop, this time in Rio de Janeiro, Brazil, on May 24, 2019. Enjoy your reading!

September 2018

Walfredo Cirne
Narayan Desai
Dalibor Klusáček

Organization

Workshop Organizers

Walfredo Cirne	Google, USA
Narayan Desai	Google, USA
Dalibor Klusáček	CESNET, Czech Republic

Program Committee

Henri Casanova	University of Hawaii at Manoa, USA
Julita Corbalan	Barcelona Supercomputing Center, Spain
Hyeonsang Eom	Seoul National University, South Korea
Dror Feitelson	Hebrew University, Israel
Liana Fong	IBM T. J. Watson Research Center, USA
Eitan Frachtenberg	Facebook, USA
Alfredo Goldman	University of Sao Paulo, USA
Allan Gottlieb	New York University, USA
Virajith Jalaparti	Microsoft, USA
Kostantinos Karanasos	Microsoft, USA
Zhiling Lan	Illinois Institute of Technology, USA
Bill Nitzberg	Altair, USA
P-O. Östberg	Umeå University, Sweden
Larry Rudolph	Two Sigma, USA
Gonzalo Rodrigo	Berkeley Lab, USA
Uwe Schwiegelshohn	TU Dortmund University, Germany
Yingchong Situ	Google, USA
Leonel Sousa	Universidade de Lisboa, Portugal
Mark Squillante	IBM, USA
Wei Tang	Google, USA
Ramin Yahyapour	University of Göttingen, Germany

Additional Reviewers

Emilio Francesquini
Pedro Bruel
Sergio Santander-Jiménez

Contents

Analysis of Job Metadata for Enhanced Wall Time Prediction 1
Mehmet Soysal, Marco Berghoff, and Achim Streit

Evaluating the Impact of Soft Walltimes on Job Scheduling Performance. . . . 15
Dalibor Klusáček and Václav Chlumský

Reducing the Human-in-the-Loop Component of the Scheduling of Large HTC Workloads. 39
Frédéric Azevedo, Luc Gombert, and Frédéric Suter

Using Pilot Systems to Execute Many Task Workloads on Supercomputers 61
Andre Merzky, Matteo Turilli, Manuel Maldonado, Mark Santcroos, and Shantenu Jha

Adaptive Simultaneous Multi-tenancy for GPUs 83
Ramin Bashizade, Yuxuan Li, and Alvin R. Lebeck

Stochastic Programming Approach for Resource Selection Under Demand Uncertainty 107
Tanveer Hossain Bhuiyan, Mahantesh Halappanavar, Ryan D. Friese, Hugh Medal, Luis de la Torre, Arun Sathanur, and Nathan R. Tallent

Approaching Actor-Level Resource Control for Akka 127
Ahmed Abdelmoamen, Dezhong Wang, and Nadeem Jamali

Author Index 147