



SPEC ResearchSM Group Newsletter

CONTENTS

- 2 SPEC RG Officers and Working Groups
- 3 Welcome to the SPEC RG Newsletter
- 3 10th Anniversary of the SPEC RG
- 3 Kaivalya Dixit Distinguished Dissertation Award
- 4 New Textbook on “Systems Benchmarking”
- 5 Plan for Predictive Data Analytics Group
- 5 ICPE 2021: Statistics
- 6 ICPE 2022 in Beijing
- 6 Reports of the Working Groups
- 12 Announcements
- 12 Selected Abstracts

10TH ANNIVERSARY OF THE SPEC RESEARCH GROUP

In March 2011, the SPEC Research Working Group was established with the goal to contribute to the computing community and extend the state of the art in performance evaluation.

Read more on page 3

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2020

The award selection committee for 2020 chose the dissertation of Maher Alharby of Newcastle University. The selection committee was impressed by the potential impact of the work, the rigor of the quantitative evaluations, the balance struck between design, and development & evaluation.

Read more on page 3

ICPE 2022 WILL BE HELD IN BEIJING

Nikolas Herbst and Philipp Leitner, the PC Co-Chairs of the next ACM/SPEC International Conference on Performance Engineering (ICPE 2022), invite interesting high-quality submissions. The conference will take place April 9-13, 2022 in Beijing, China.

Read more on page 6

SPEC RESEARCH WORKING GROUPS REPORT ON THEIR PROGRESS

The five SPEC Research Working Groups Security, Cloud, DevOps Performance, Quality of Experience, and Power report on their progress, articles, benchmarks, and technical reports published in 2020. The Working Groups are always open for new members. Feel invited to join us!

Read more on pages 6-12



CONTACT

Standard Performance Evaluation Corporation (SPEC)
7001 Heritage Village Plaza, Suite 225
Gainesville, VA 20155, USA

SPEC Research Group

Chair: Samuel Kounev (rgchair@spec.org)

Web: <http://research.spec.org>

SPEC RESEARCH GROUP OFFICERS

Chair:

Samuel Kounev, University of Würzburg, Germany

Vice-Chair:

André van Hoorn, University of Stuttgart, Germany

Secretary:

Cor-Paul Bezemer, University of Alberta, Canada

Steering Committee:

J. Nelson Amaral, University of Alberta, Canada

Cor-Paul Bezemer, University of Alberta, Canada

Alexandru Iosup, VU Amsterdam, NL

Samuel Kounev, University of Würzburg, Germany

Klaus-Dieter Lange, HPE, USA

Meikel Poess, Oracle Corporation, USA

Petr Tůma, Charles Univ. of Prague, Czech Republic

Weiyi Shang, Concordia University, Canada

André van Hoorn, University of Stuttgart, Germany

Boris Zibitsker, BEZNext, USA

Publicity Officer:

Weiyi Shang, Concordia University, Canada

Release Manager:

Vojtěch Horký, Charles Univ. Prague, Czech Republic

Newsletter Editor:

André Bauer, University of Würzburg, Germany

<http://research.spec.org>

SPEC RESEARCH WORKING GROUPS

DevOps Performance Working Group

Chair:

André van Hoorn, University of Stuttgart, Germany

Vice-Chair:

Cor-Paul Bezemer, University of Alberta, Canada

Robert Heinrich, KIT, Germany

Secretary:

Simon Eismann, University of Würzburg, Germany

<https://research.spec.org/en/working-groups/rg-devops-performance.html>

Quality of Experience Working Group

Chair:

Florian Wamser, University of Würzburg, Germany

Vice-Chair:

Tobias Hoßfeld, University of Würzburg, Germany

<https://research.spec.org/working-groups/rg-quality-of-experience.html>

Cloud Working Group

Chair:

Alexandru Iosup, VU Amsterdam, The Netherlands

Vice-Chair & Secretary:

Nikolas Herbst, University of Würzburg, Germany

Secretary:

Sacheendra Talluri, VU Amsterdam, The Netherlands

Release Manager:

André Bauer, University of Würzburg, Germany

<https://research.spec.org/en/working-groups/rg-cloud.html>

Power Working Group

Chair:

Norbert Schmitt, University of Würzburg, Germany

Vice-Chair:

Klaus-Dieter Lange, HPE, USA

Secretary:

Mike Tricker, Microsoft, USA

<https://research.spec.org/working-groups/rg-power.html>

Security Working Group

Chair:

Aleksandar Milenkoski, ERNW & Univ. Würzburg, Germany

Vice-Chair:

Nuno Antunes, University of Coimbra, Portugal

Secretary and Release Manager:

José Flora, University of Coimbra, Portugal

<https://research.spec.org/working-groups/rg-ids-benchmarking.html>

WELCOME TO THE SPEC RESEARCH GROUP NEWSLETTER

With 125 members in 22 countries and nearly two dozen benchmarks spanning highly diverse aspects of computing performance and energy efficiency, SPEC has become known as a beacon of truth for computing researchers, vendors, users and analysts worldwide. These professionals rely on SPEC to ensure that the marketplace has a fair and useful set of metrics to differentiate computing systems. Founded in 2011, the SPEC Research Group is proud of being part of the recent years of this remarkable history.

We are delighted to present to you the next issue of the SPEC Research Group Newsletter. This regular publication provides information on latest developments, news, and announcements relevant to the benchmarking and quantitative system evaluation communities. Our newsletter is part of our mission to foster the exchange of knowledge and experiences between industry and academia in the field of quantitative system evaluation and analysis.

Some highlights from the last year include:

- 11th (and first virtual) ACM/SPEC **ICPE 2020** in Edmonton, Canada
- 1st IEEE International Conference on Autonomic Computing and Self-Organizing Systems **ACSOS 2020** as a merger of ICAC and SASO
- 6th International Workshop on Quality-aware Dev-Ops **QUDOS 2020** at ECSA 2020
- 3rd Workshop on Hot Topics in Cloud Computing Performance **HotCloudPerf 2020** at ICPE 2020
- New tool accepted: **Mowgli: evaluation framework for cloud-hosted DBMS**

We have been actively working on preparation, planning and organization of ICPE 2021, this year as a virtual conference. We hope that a vivid exchange of ideas will be a great motivation for the next year of scientific and engineering work.

We hope that you will enjoy reading the newsletter. We welcome and encourage your contributions for articles and suggestions for future coverage.

Samuel Kounev
(SPEC Research Chair, University of Würzburg).
André Bauer
(Newsletter Editor, University of Würzburg).

SPEC, the SPEC logo and the names SERT, SPEC SFS, SPECjbb, SPECvirt.sc, Chauffeur WDK, and SPEC PTDaemon are registered trademarks of the Standard Performance Evaluation Corporation. The SPEC Research Logo and the name SPEC Research are service marks of SPEC. Additional company, product and service names mentioned herein may be the trademarks or service marks of their respective owners. Copyright ©1988-2021 Standard Performance Evaluation Corporation (SPEC). Reprinted with permission. All rights reserved.

10TH ANNIVERSARY OF THE SPEC RESEARCH GROUP

This year both the SPEC newsletter and the SPEC Research Group (SPEC RG) celebrate their 10th anniversary: In March 2011, SPEC RG was founded (see press release <https://www.spec.org/news/rgpressrelease.html>). The then and current chair Samuel Kounev recognized a high potential in this group: “We see this as an incubator for exciting new research and a forum for exchanging ideas, methodologies and new developments in benchmarking”, further he stated: “It’s another way for SPEC to contribute to the computing community and extend the state of the art in performance evaluation.”

These statements have turned out to be true: Back then, RG started with its own Cloud Working Group only; Today, RG comprises five own working groups, which actively contribute to standardization in the community. Also, when RG was founded, the SPEC RG newsletter was established, which is published regularly each year before ICPE. During the last 10 years the newsletter was edited by the following editors:

1. Piotr Rygielski (2011–2016)
2. Nikolas Herbst (2016–2020)
3. André Bauer (since 2016)

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2020 WINNER

The SPEC Kaivalya Dixit Distinguished Dissertation Award is an annual award that aims to recognize outstanding doctoral dissertations within the scope of the SPEC Research Group in terms of scientific originality, scientific significance, practical relevance, impact, and presentation. The award selection committee for 2020 was chaired by Prof. William Knottenbelt of Imperial College London and consisted of eight members from academia.

The winning dissertation was authored by Maher Alharby of Newcastle University under the supervision of Prof. Aad van Moorsel. The selection committee was especially impressed by the potential impact of the work, the rigor of the quantitative evaluations, the balance struck between design, development and evaluation, and the variety of publications associated with Dr. Alharby’s dissertation.

The award is to be presented at the International Conference on Performance Engineering (ICPE 2021).

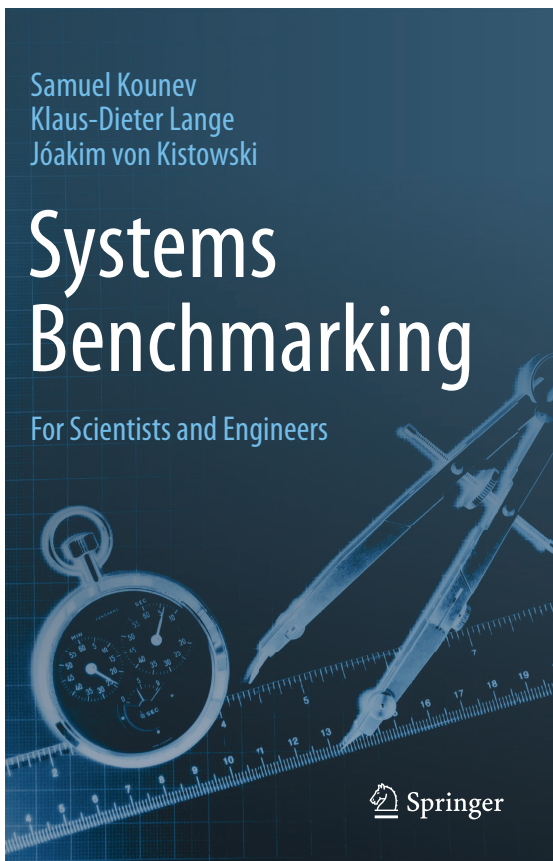
Given the high quality of dissertations nominated for this award, the committee decided to publicly recognize another dissertation as Runner Up, entitled “Performance Modelling of Distributed Stream Processing Topologies” and authored by Thomas Cooper, also of Newcastle University, under the supervision of Dr. Paul Ezhilchelvan.

Members of the committee were impressed by the industrial relevance of the application scenario, the sophistication of the modeling employed and the “exquisite, well-articulated” writing in the dissertation.

<https://research.spec.org/nc/en/news/single-view/article/winner-of-spec-kaivalya-dixit-distinguished-dissertation-award-2020.html>

NEW TEXTBOOK ON “SYSTEMS BENCHMARKING”

A new textbook on the topic of systems benchmarking written by SPEC RG Members was published recently by Springer International Publishing. The book serves as both a textbook and handbook on the benchmarking of systems and components thereof. It provides theoretical and practical foundations as well as an in-depth exploration of modern benchmarks and benchmark development. Teaching materials (slides and exercises) will be provided later for those interested in using the book for teaching a course on the topic.



“This book should be required reading for anyone interested in making good benchmarks.” – from the Foreword by David Patterson, 2017 ACM A.M. Turing Award Laureate.

More information can be found on the book web site at: <http://benchmarking-book.com/>

Acknowledgements

The authors of the book - Samuel Kounev, Klaus-Dieter Lange, and JÓakim von Kistowski - would like to thank the following people who supported this effort by contributing to chapters and reviewing content.

The following people (in alphabetical order) contributed as co-authors of chapters in the second part of the book: Jeremy A. Arnold, André Bauer, John Beckett, James Bucek, Ken Cantrell, Don Capps, Alexander Carlton, Simon Eismann, Sorin Faibish, Johannes Grohmann, Nikolas Herbst, Karl Huppler, Rouven Krebs, Mary Marquez, Aleksandar Milenkoski, David Morse, Meikel Poess, Nick Principe, David Schmidt, Norbert Schmitt, Simon Spinner, and Sitsofe Wheeler.

The following colleagues from industry (in alphabetical order) reviewed individual chapters and provided valuable feedback: Walter Bays, Hansfried Block, Karla Orozco Bucek, John Henning, Scott Hinchley, Supriya Kamthania, Mukund Kumar, Kris Langenfeld, Pranay Mahendra, John R. Mashey, Luis Mendoza, Sriranga Nadiger, Daniel Pol, Jesse Rangel, Nishant Rawtani, Jeff Reilly, David Reiner, Sanjay Sharma, and Rajesh Tadakamadla.

Book Overview

The book is divided into two parts: foundations and applications. The first part introduces the foundations of benchmarking as a discipline, covering the three fundamental elements of each benchmarking approach: metrics, workloads, and measurement methodology. The second part focuses on different application areas, presenting contributions in specific fields of benchmark development. These contributions address the unique challenges that arise in the conception and development of benchmarks for specific systems or subsystems, and demonstrate how the foundations and concepts in the first part of the book are being used in existing benchmarks. Further, the book presents a number of concrete applications and case studies based on input from leading benchmark developers from consortia such as the Standard Performance Evaluation Corporation (SPEC) and the Transaction Processing Performance Council (TPC).

Providing both practical and theoretical foundations, as well as a detailed discussion of modern benchmarks and their development, the book is intended as a handbook for professionals and researchers working in areas related to benchmarking. It offers an up-to-date point of reference for existing work as well as latest results, research challenges, and future research directions. It also can be used as a textbook for graduate and postgraduate students studying any of the many subjects related to

benchmarking. While readers are assumed to be familiar with the principles and practices of computer science, as well as software and systems engineering, no specific expertise in any subfield of these disciplines is required.

WE WANT YOU FOR OUR PLANNED NEW WORKING GROUP

The University of Würzburg and Ulm are planning to initiate a new working group in the field of **Predictive Data Analytics**. We are currently looking for potential members who are interested in working in this area. Two publications [1,2] towards this group will be presented at ICPE 2021 and the abstracts can be found at the end of the newsletter. The mission of the new group will be as follows:

The discipline of data analytics/science has grown significantly in recent years as a means to make sense of the vast amounts of data available. It has pervaded every side/aspect of computer science and engineering and is heavily used in business decision making. Despite this, there is a lack of integration and automation of data flows between data scientists/analysts and data analysis consumers.

The overarching goal of the planned group is to bridge the missing links between all the facets involved in data analytics, namely big data storage and provisioning, data versioning and performance evaluation. To this end, we envision a standardized, coherent pipeline ranging from big data storage, pre-processing through to analytics and evaluation.

The interests of the group lies in but is not limited to:

1. Performance Forecasting and Anomaly Detection
2. Time Series
 - (a) Clustering/Segmentation
 - (b) Feature Engineering
 - (c) Forecasting / Prediction
 - (d) Synthesis / Imputation
 - (e) Data Quality Assessment
 - (f) Algorithm Performance Benchmarking
3. Streamlining the data science process (DataOps)
 - (a) Standardization of data versioning
 - (b) Standardization and automation of deployment
4. Benchmarking of Big Data Infrastructure (e.g. Distributed Database Management Systems)

If you are interested in joining, please contact André Bauer (andre.bauer@uni-wuerzburg.de).

[1] André Bauer, Marwin Züfle, Simon Eismann, Johannes Grohmann, Nikolas Herbst, and Samuel Kounev. *Libra: A Benchmark for Time Series Forecasting Method*. In Proceedings of the 12th ACM/SPEC International Conference on Performance Engineering (ICPE) (2021). ACM, New York, NY, USA.

[2] Mark Leznik, Patrick Michalsky, Peter Willis, Benjamin Schanzel, Per-Olov Östberg, and Jörg Domaschka. *Multivariate Time Series Synthesis Using Generative Adversarial Networks*. In Proceedings of the 12th ACM/SPEC International Conference on Performance Engineering (ICPE) (2021). ACM, New York, NY, USA.

ICPE 2021: STATISTICS

The 12th ACM/SPEC International Conference on Performance Engineering (ICPE 2021), will be held virtually from Rennes, France, from April 19 to April 23. This year the research track of ICPE attracted 43 submissions, 12 of which were selected as full articles after a rigorous review process, yielding an acceptance ratio of 28%. There were also 8 articles accepted as short articles. The Work-in-Progress and Vision Track received 11 submissions of which 4 were accepted. Of the 18 submissions to the Industry Track, 7 were accepted, of which 4 as full articles. The following workshops are planned for ICPE 2021:

- The 1st Workshop on Performance and Energy-efficiency in Concurrent Systems (PECS 2021): <https://pecs-workshop.github.io/2021/>
- The 4th International Workshop on Autonomic Solutions for Parallel and Distributed Data Stream Processing (Auto-DaSP 2021): <http://calvados.di.unipi.it/auto-dasp-21/>
- Workshop on Challenges in Performance Methods for Software Development WOSP-C 2020: <https://wosp-c-21.github.io/>
- The 3rd Workshop on Education and Practice of Performance Engineering (WEPPE): <https://esulabsolutions.godaddysites.com/sponsored-events>
- The 4th Workshop on Hot Topics in Cloud Computing Performance (HotCloudPerf-2021): <https://hotcloudperf.spec.org/>
- The 9th International Workshop on Load Testing and Benchmarking of Software Systems (LTB 2021): <http://ltb2021.eecs.yorku.ca/>

The following distinguished speakers will give keynotes at ICPE 2021:

- Sven Apel, Saarland University
Software Performance Modelling in Spacetime
Almost any practical software system today is configurable. A configurable software system provides

a set of configuration options to adjust and optimize its functional and non-functional properties. In particular, a system's performance behavior often depends in intricate ways on individual configuration options as well as interactions among them. This complex behavior renders tasks such as identifying performance-critical options, deploying performance-optimal configurations, and pinning down reasons for performance regression challenging. Clearly, any practical performance modelling technique needs to take the dimension of configurability into account (space dimension). But this is only one side of the coin. Software systems evolve and so do their (configuration-dependent) performance behaviors (time dimension).

This talk will review recent and ongoing work on modelling performance of configurable software systems. While early approaches concentrated on modelling performance across the configuration space, more recent approaches incorporate the evolution of the software system in question, effectively modelling the system's performance behavior in space-time. Besides discussing the key ideas in this area, the talk will highlight challenges that arise from the interplay between configuration sampling and performance learning, the uncertainty that is inherent in performance measurement and behavior, and the interaction of workload-dependent and configuration-dependent behavior.

- Shan Lu, University of Chicago
Tackling performance bugs in web applications and cloud systems Software developers face the stringent task of delivering new software features while keeping the software latency low. This task has become increasingly challenging as many software systems, including web applications and cloud applications, need to process ever-growing amount of user data and contain increasingly complicated configuration settings.

This talk will discuss our effort in understanding and tackling performance problems caused by inefficient data-processing and mis-configurations. I will first discuss our empirical study about how inefficient data-processing has become the performance bottleneck of modern web applications, and then present program analysis, code refactoring, and IDE tools that we have built to automatically identify and fix thousands of inefficient data-processing in popular web applications, to help developers understand the data-processing cost behind web page elements, and to help them explore different performance-enhancing web-page designs. I will also discuss our effort in automatically identifying

performance-sensitive configurations and automatically adjusting configuration settings using control theoretic techniques.

With the COVID-19 pandemic, the face-to-face ICPE in Rennes is not possible. All authors were offered the opportunity to provide video presentations and slides to be made available on the ICPE 2021 website. Virtual sessions via teleconference are scheduled for the days of the conference for the invited speakers' presentations and for paper discussions. The detailed program can be found at <https://icpe2021.spec.org/>.

Johann Bourcier (Université de Rennes 1, France),
Zhen Ming (Jack) Jiang (York University, Canada),
Vittorio Cortellessa (Università dell'Aquila, Italy),
Cor-Paul Bezemer (University of Alberta, Canada)

ICPE 2022 IN BEIJING - PRELIMINARY ANNOUNCEMENT

The ACM/SPEC International Conference on Performance Engineering (ICPE) provides a forum for the integration of theory and practice in the field of performance engineering. It brings together researchers and industry practitioners to share ideas, discuss challenges, and present results of both work-in-progress and state-of-the-art research on performance engineering of software and systems.

ICPE 2022 will be held in Beijing (China), from April 9 (Sa) to April 13 (Wed). A decision whether the conference will be hybrid (remote participation possible) or fully virtual will be announced timely.

The contact person for ICPE 2021 is Arthur Kang (Inspur, China), who will be Local Chair along with Dan Feng (Huazhong University of Science and Technology, China) and Steffen Becker (University of Stuttgart, Germany) as General Co-Chair. The PC Co-Chairs will be Nikolas Herbst (University of Würzburg, Germany) and Philipp Leitner (Chalmers University of Technology, Sweden). The industrial track chair will be Yaoqing Gao (Huawei, Canada).

REPORT: DEVOPS PERFORMANCE WORKING GROUP

DevOps is an emerging principle for engineering and operating software systems. It aims to increase the rate and velocity of releasing new software versions, which is, for instance, achieved by a high degree of automation and by integrating development and operations responsibilities. DevOps imposes immense challenges for quality assurance, e.g., concerning performance and related attributes. Key reasons are that respective activities are constrained by time and that the environment in which a software

system is running is ever-changing. On the other hand, DevOps provides great opportunities because the integration between development and operations allows for a high degree of automation as well as a streamlined collection and analytics of performance data. The RG DevOps Performance Working Group is a forum for individuals and organizations interested in the interplay of DevOps and performance engineering. The mission of the working group is to consolidate concepts and tools to better integrate these activities. Its membership body currently includes representatives of Concordia University, Imperial College London, Karlsruhe Institute of Technology, MongoDB, University of Alberta, University of L'Aquila, University of Stuttgart, and University of Würzburg. The group as a whole meets in online meetings that are held on a monthly basis. In addition to the discussion of organizational topics, these general meetings include a technical presentation by group members or by invited guests. In total, 11 group meetings were held in 2020, including the following talks as part of our monthly lecture series:

- *“What’s Wrong With My Benchmark Results? Studying Bad Practices in JMH Benchmarks”* by Diego Costa, Concordia University
- *“A Framework for Optimizing Resource Usage of Service-Based Applications Through Multi-Version Containers”* by Sara Gholami, University of Alberta
- *“COCOA: Cold Start Aware Capacity Planning for Function-as-a-Service Platforms”* by Alim Gias, Imperial College London
- *“Incremental Calibration of Architectural Performance Models with Parametric Dependencies”* by Manar Mazkatli, Karlsruhe Institute of Technology
- *“Using Black-Box Performance Models to Detect Performance Regressions under Varying Workloads: An Empirical Study”* by Lizhi Liao, Concordia University
- *“Run-time/Design-time Interactions for Performance Assessment”* by Daniele Di Pompeo, University of L'Aquila
- *“Runtime-Data Driven Interactive Specification, Analysis, and Explanation of Transient Software Behavior under Continuous Change: Vision and WiP”* by Sebastian Frank and Alireza Hakamian, University of Stuttgart
- *“Locating Performance Regression Root Causes in the Field for Web-based Systems”* by Lizhi Liao, Concordia University
- *“Detecting Regressions in Load Test Results”* by Alexander Podelko, Oracle/MongoDB

- *“Anomaly/Root Cause Determination and Cloud Platform Selection”* by Boris Zibitsker, BEZNext

In addition to the monthly meetings, the group operates in subgroups consisting of 6-8 participants who collaborate closely on concrete topics. Collaborations include jointly supervised student projects. The subgroups meet biweekly and report to the whole group once a month in the regular meeting. The current subgroups are:

1. **Performance testing of next-generation cloud applications:** This subgroup focuses on the challenges of performance testing next-generation cloud applications. The group published and presented their work on performance regression testing of microservice applications at ICPE 2020 [1] and is currently investigating the load testing of serverless applications. The group has presented its preliminary results on this topic at SSP 2020.
2. **Model extraction and refinement in continuous software engineering:** This subgroup focuses on how models can be leveraged to improve the continuous software engineering process. The group published two papers: the first one is on the incremental calibration of the architectural performance engineering with parametric dependencies during DevOps-based software development, which was published and presented at ICSEA 2020 [2]. The second paper concentrates on optimizing the parametric dependencies for the incremental performance model extraction using genetic algorithms, published and presented at QUDOS 2020 [3].
3. **Performance of continuous delivery infrastructures.** This subgroup focuses on the evaluation and improvement of continuous delivery (CD) infrastructures, which have become a critical component of software development. In addition to the previously analyzed performance data of a CD system, we recently used the APIs to dynamically fetch up-to-date data with the aim to analyze it and optimize build configurations. We already developed a prototype that reorganizes the order of builds to reduce the overall build time.

In January 2021, the group initiated a new subgroup:

- 4 **Resilience engineering for cloud-native applications:** An important property of cloud-native software systems is resilience, i.e., the ability to withstand continuous change imposed by frequent redeployments, usage behavior, failures, or attacks. Currently, resilience requirements and properties are poorly specified, analyzed, and understood. The group will investigate novel approaches, tools, and data sets for resilience engineering including interactive visualization, runtime monitoring, scenario-

based architecture evaluation, risk analysis, chaos engineering, formal methods, and other techniques.

Several members of the group are active in the organization of international events. In 2020, the group co-organized:

- The sixth edition of the International Workshop on Quality-Aware DevOps (QUDOS 2020) [4]. It was co-located with the 14th European Conference on Software Architecture (ECSA), and organized as a joint event with the workshop on Continuous Software Engineering (CSE). The workshop included a keynote, seven paper presentations, and attracted about 40 participants.
- Journal special issue on software performance [5]. Group members have initiated a special issue on “Software Performance” in the Empirical Software Engineering Journal, which is a top-ranked software engineering journal. Submissions are due in May 2021.

For more information about the DevOps Performance Working Group (including our mission, activities, meetings, presentations, and projects), please visit our web page [6]. If you are interested in following the discussions or contributing actively, please contact the working group chairs.

Cor-Paul Bezemer (University of Alberta), André v. Hoorn (University of Stuttgart), Simon Eismann (University of Würzburg)

<https://research.spec.org/en/working-groups/rg-devops-performance.html>

[1] Simon Eismann, Cor-Paul Bezemer, Weiyi Shang, Dusan Okanovic, André van Hoorn: Microservices: A Performance Tester’s Dream or Nightmare? ICPE 2020: 138-149

[2] Manar Mazkatli, David Monschein, Johannes Grohmann, Anne Koziolk: Incremental Calibration of Architectural Performance Models with Parametric Dependencies. ICSA 2020: 23-34

[3] Sonya Voneva, Manar Mazkatli, Johannes Grohmann, Anne Koziolk: Optimizing Parametric Dependencies for Incremental Performance Model Extraction. ECSA Companion 2020: 228-240

[4] <http://2020.qudos-workshop.org/>

[5] https://emsejournal.github.io/special_issues/2021_Software_Performance.html

[6] <http://research.spec.org/devopswg/>

REPORT: QUALITY OF EXPERIENCE WORKING GROUP

In the RG QoE, topics related to benchmarking and the evaluation of systems from the user’s point of view are discussed. In particular, discussions are held on the Quality of Experience (QoE) as a metric for the user-centered and subjective evaluation of systems and infrastructure. The end user is of the utmost importance, and the evaluation metric aims to reflect the needs and requirements from the perspective of the user. As a result, metrics that are

commonly grouped under the term quality have evolved depending on the type of application. For example, a video streaming system is rated based on the playout resolution and the smoothness of the streaming rather than the throughput experienced. A server infrastructure is assessed on the basis of its ability to serve the applications running on it. In communication, for example, the term quality has been largely associated with the so-called “Quality of Service” (QoS) for many years. Now providers are starting to evaluate their network specifically for services such as video streaming or surfing the Internet in order to ensure the applicability of their network for a specific usage scenario. The idea of the RG QoE is to use such user-centered assessment metrics in benchmarks.

In view of the discussions at the last RG general meeting, the new topic of **Crowdsourcing-based Performance Benchmarking** was discussed. The benchmarking of services and infrastructures (or the quality monitoring of Internet services) with active help from users is a relevant topic in some areas. Examples of this are platform providers who examine the performance of a large service by adding monitoring and testing frameworks in various consumer applications, usually in the background. Such a type of measurement is defined as the collection and processing of data measured by the crowd. As a result, you get a large amount of data from uncontrolled measurements by the crowd. The density, number, and accuracy of the measurements differ due to the uncontrolled measurement environment. There is always the question of the reliability of such measurements. The focus lies here in the description, discussion and standardization of the fundamental basics, ideas, and concepts for the reliability of crowdsourced measurements. One goal is to provide guidelines on reliability of crowdsourced measurements for IoT, cloud apps, networks, and on how measurements can be validated, or on how reliability can be quantified.

In the Quality of Experience Research Group, we try to consolidate, summarize, and categorize different definitions of Quality of Experience (QoE). The group shall be the starting point for discussions of QoE ideas, QoE approaches, QoE measurement tools, and QoE assessment paradigms. We seek to stimulate collaboration between industry and research through the exchange of ideas, and want to use the group to promote the usefulness of QoE and highlight its scope.

Florian Wamser (University of Würzburg)

<https://research.spec.org/working-groups/rg-quality-of-experience.html>

REPORT: BIG DATA WORKING GROUP

The Big Data Working Group was retired last year.

REPORT: CLOUD WORKING GROUP

In 2020, the SPEC RG Cloud Group has driven several activities aligned with its long-term mission of furthering cloud benchmarking, quantitative evaluation, and experimental analysis, in directions relevant for both academia and industry. We have focused this year on novel cloud paradigms such as Functions-as-a-Service, Serverless Computing, the Cloud Continuum extending clouds with fog and edge devices, Convergence of HPC and Big Data as cloud services.

The scope of the group is 'to develop new methodological elements for gaining deeper understanding not only of cloud performance, but also of cloud operation and behavior, through diverse quantitative evaluation tools, including benchmarks, metrics, and workload generators'. We consider properties such as elasticity, performance isolation, dependability, and other non-functional system properties, in addition to classical performance-related metrics such as response time, throughput, scalability, and efficiency. Our work towards benchmark prototypes includes designing reference architectures, standardizing use cases, observing patterns, and methods for reproducibility.

In 2020, through monthly online meetings facilitated by SPEC's Zoom and meetings focusing on furthering specific activities, and through continuous discussion via a Slack workspace, we have advanced work on the following main topics:

1. **Serverless Use Cases:**

Serverless architectures are rapidly adopted by practitioners, but many characteristics of serverless applications are either unknown or contested. Therefore, the SPEC RG Cloud Group started collecting and analyzing serverless use cases in late 2019. This activity culminated in a collection of 89 serverless applications from academic papers, open-source repositories, industrial articles, and scientific applications analyzed along 24 characteristics. In 2020, a detailed report of these results was published as a SPEC RG Techreport [1] and a practitioner-focussed summary of these results was published at IEEE Software [2]. In a next step, we are looking into a meta-analysis of studies on the characteristics of serverless applications, currently under submission to IEEE Transactions on Software Engineering

2. **Serverless Performance Benchmark:**

Existing studies on serverless performance primarily focus on simple micro-benchmarks using opaque testing strategies [3]. Motivated by the application insights from the serverless use case activity, the SPEC RG Cloud group is working on an application-driven performance benchmark using

clear box strategies and more realistic invocation patterns. Our benchmark suite is able to integrate with existing serverless applications. We plan to release it as open source software in the coming months.

3. **Performance Variability in Serverless Computing:**

Serverless application developers expect the ability to pay for exactly the resources used, and elastically scale resources up and down as required. This leads to highly variable application resource requirements and the need for resources to be rapidly provisioned, resulting in variable application performance. The SPEC RG Cloud group is investigating if the performance variability can be tied to specific causes such as scheduling delay and resource congestion. We plan to develop scheduling and resource management techniques which can reduce performance variability, or take on more variability for lower cost. This is a starting activity

4. The **Edge activity** of the SPEC Cloud Group is

working on extending a workshop paper [4] towards a full conference paper titled "Edge workload trace gathering and analysis for benchmarking". The paper defines workload classes as well as collecting and analyzing traces for selected edge applications. Also, it will present a generic methodology for workload definition and gathering. The paper "LOOPS: A Holistic Control Approach for Resource Management in Cloud Computing" discusses a multi-loop control approach to coordinate resource management in cloud computing environments. It appears as a short paper at ICPE 2021 conference. As a call for contributions: A paper tackling the "Transfer Learning" concept is planned, as well as a survey is started to cover the computation and communication models which emerged with cloud and edge computing. In both directions, the authors are open for cooperation.

5. **Cloud Experiment Methodology:**

Our paper "Methodological Principles for Reproducible Performance Evaluation in Cloud Computing" published on IEEE Transactions on Software Engineering (TSE)[6], was invited to be included in the program of the International Conference on Software Engineering (ICSE), as a Journal-First publication. A video of the presentation is available here.

Besides these focused activities, the Cloud WG has successfully been acting in the following directions: (1) The yearly workshop HotCloudPerf, (2) an accepted Dagstuhl seminar on "Serverless Computing", and (3) in setting up an Artifact Track for ACSOS 2021. More details below in order:

HotCloudPerf workshops: Since 2018, the Cloud working group is organizing a yearly workshop in connection with their face-to-face meeting: The third, virtual edition of HotCloudPerf was well-attended in conjunction with the ACM/SPEC ICPE 2020. The workshop featured 2 keynotes and 3 papers. The fourth edition is to be held in conjunction with ICPE 2021 and received a record number of submissions. You can find the workshop program here [6].

Dagstuhl Seminar “Serverless Computing”: Alexandru Iosup, Cristian Abad, Nikolas Herbst and Ian Foster successfully proposed a Dagstuhl seminar on “Serverless Computing” [7]. The seminar is scheduled for May 16-21, 2021. With a good number of (virtual) attendees the seminar organization is now in a hot preparation phase.

1st Artifact Track at ACSOS 2021: For the first time, ACSOS 2021 is introducing an artifact evaluation (AE) process. The AE is an optional evaluation process for research works that have been accepted for publication at the conference. Other conferences in the area of cloud computing are introducing such a process either as an optional process, or as part of the submission. Two members of the SPEC Cloud WG have been selected to chair this activity, i.e., Nikolas Herbst and Alessandro Papadopoulos.

To conclude, 2020 was a full and successful year for the RG Cloud Group. We are looking forward to an even more successful 2021. For this, we are actively seeking new participants and activities. You can also join ongoing activities.

Alexandru Iosup (Vrije Universiteit Amsterdam, Chair) and Nikolas Herbst (University of Würzburg, Vice-Chair)

<http://research.spec.org/working-groups/rg-cloud-working-group.html>

[1] Simon Eismann, Joel Scheuner, Erwin van Eyk, Maximilian Schwinger, Johannes Grohmann, Nikolas Herbst, Cristina L. Abad, and Alexandru Iosup. A Review of Serverless Use Cases and their Characteristics. SPEC Technical Report, May 2020

[2] Simon Eismann et al., “Serverless Applications: Why, When, and How?,” in IEEE Software, vol. 38, no. 1, pp. 32-39, Jan.-Feb. 2021, doi: 10.1109/MS.2020.3023302.

[3] Erwin van Eyk, Joel Scheuner, Simon Eismann, Cristina L. Abad, and Alexandru Iosup. 2020. Beyond Microbenchmarks: The SPEC-RG Vision for a Comprehensive Serverless Benchmark. In Companion of the ACM/SPEC International Conference on Performance Engineering (ICPE '20). Association for Computing Machinery, New York, NY, USA, 26–31. DOI: 10.1145/3375555.3384381

[4] K. Toczé, N. Schmitt, I. Brandic, A. Aral and S. Nadjim-Tehrani, “Towards Edge Benchmarking: A Methodology for Characterizing Edge Workloads,” 2019 IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS*W), Umea, Sweden, 2019, pp. 70-71, doi: 10.1109/FAS-W.2019.00030.

[5] A. V. Papadopoulos et al., “Methodological Principles for Reproducible Performance Evaluation in Cloud Computing,” in IEEE Transactions on Software Engineering, doi: 10.1109/TSE.2019.2927908. Selected as Journal-First publication presented at ICSE 2020.

[6] <http://hotcloudperf.spec.org>

[7] <https://www.dagstuhl.de/de/programm/kalender/semhp/?semnr=21201>

REPORT: POWER WORKING GROUP

The SPEC Research Power WG (AMD, Dell, HPE, Intel, IBM, Microsoft, University of Würzburg) has operated tightly coupled with the SPECpower Committee, since its inception in 2017, to research the energy and resource efficiency of computing devices and software. Our close collaboration fosters the interaction between industry and academia by contributing research that enhances and promotes methods and tools for energy and resource efficiency evaluation to address this essential concern for the industry, academia, and regulatory institutions.

Under the chairmanship of Norbert Schmitt and Klaus-Dieter Lange, the Power WG extended its scope to include resource efficiency with a focus on cloud software. One of the key goals is to define an energy-efficiency metric for software to classify software and raise awareness among developers and operators alike. With the common conception of software energy efficiency being inseparable from the executing hardware, our research faces many diverse and exciting challenges. Challenges include, but are not limited to, separation of hardware and software, and comparability in the domain of software energy efficiency. Programming languages, compilers, their optimizations, and software architecture play their part in our research to identify suitable measures on software energy efficiency and for an applicable and relevant software classification.

As a first step, the Power Working Group has researched the influence of different software characteristics, like programming language, based on the SPEC CPU2017 benchmark suite in a WiP paper at ICPE [1]. Following up on this work, the working group analyzed how the performance, power consumption, and energy efficiency changes when the benchmarks in the SPEC CPU2017 suite are compiled with highly optimized parameters compared to a real-world scenario with less tuned and less aggressive optimizations. This work was published at the IEEE/ACM International Conference on Utility and Cloud Computing [2].

With the upcoming replacement of the SPECpower_{ssj} 2008 benchmark, the working group, in close collaboration with the SPECpower Committee, did take a closer look at the *SPECpowerNext*. This includes the reimplementation of an industry standard benchmark in ChauffeurWDK, its new Redfish integration, and the two new workloads, the Auxiliary Processing Accelerator and Wiki

workload with a focus and preliminary measurements on the latter. This industry track paper will be published at the ICPE [3].

Further steps towards software energy efficiency were taken in an effort to outline an idea for a research benchmark to measure the resource efficiency (for example CPU and memory) of cloud applications. In a vision paper to be published at the ICPE, the general setup and also the most important limitations are discussed [4]. Based on this paper, the Power WG and RETIT GmbH also applied for third party funding to develop a working demonstrator.

The SPECresearch Power WG looks forward to new and exciting challenges in power, resource, and energy efficiency benchmarking and testing. The group is happy to accept new members and visions for additional research directions in the general area of energy and resource efficiency benchmarking.

[1] N. Schmitt, J. Bucek, K. D. Lange, and S. Kounev: Energy Efficiency Analysis of Compiler Optimizations on the SPEC CPU 2017 Benchmark Suite. ACM/SPEC ICPE 2019.

[2] N. Schmitt, J. Bucek, J. Beckett, A. Cragin, K. D. Lange, and S. Kounev: Performance, Power, and Energy-Efficiency Impact Analysis of Compiler Optimizations on the SPEC CPU 2017 Benchmark Suite. IEEE/ACM UCC 2020.

[3] N. Schmitt, K. D. Lange, S. Sharma, N. Rawtani, C. Ponder, and S. Kounev: The SPECpowerNext Benchmark Suite, its Implementation and New Workloads from a Developer's Perspective. ACM/SPEC ICPE 2021.

[4] N. Schmitt, R. Vobl, A. Brunnert, and S. Kounev: Towards a Benchmark for Software Resource Efficiency. ACM/SPEC ICPE 2021.

Norbert Schmitt (University of Würzburg)
Klaus-Dieter Lange (Hewlett Packard Enterprise)

<https://research.spec.org/working-groups/power-working-group.html>

REPORT: SECURITY WORKING GROUP

The SPEC RG Security Benchmarking Working Group, after establishing its long-term agenda, is working towards devising and conducting impactful work while achieving the defined goals. In 2020, the Working Group saw a member defend his PhD thesis [1] and it had a paper published at the Mediterranean Communication and Computer Networking Conference (MedComNet 2020) [2].

Lukas Iffländer, finished his PhD by submitting and defending his thesis "Attack-aware Security Function Management" [1]. His thesis was co-reviewed by group vice-chair Nuno Antunes who also took part in the defense as an examiner. With Lukas moving his focus toward industry, he handed his function as Secretary and Release Manager – that he held for four years – over to José Flora.

The paper, entitled "A Model-Based Approach to Anomaly

Detection Trading Detection Time and False Alarm Rate", proposes a methodology that addresses the practical challenges in implementing anomaly detection approaches, emphasizing the difficulties in thoroughly defining normal behavior. Based on performance signatures, the methodology takes advantage of an analytical performance model allowing to control the rate of false positives.

Following the proposal of a hypercall testing framework in 2019 [3, 4], the group completed the implementation. Since then, the framework, now named Hyringe, has been used to extensively test the Hyper-V hypervisor regarding software aging, robustness, performance, and security. The group managed to find two robustness problems, which were reported to and fixed by Microsoft. A full paper about Hyringe and its use cases is scheduled for 2021.

The SPEC RG Security Benchmarking Working Group continues to push forward in several research directions, including:

- extending the previously proposed approach and the framework for testing the robustness and performance of hypercall interfaces;
- identifying challenges in the area of evaluating robustness and performance aspects of security-relevant system components and security mechanisms;
- proposing new methodologies to evaluate security of systems based on injecting the effects of intrusions;
- evaluating the performance of different approaches in the context of detecting security intrusions based on performance and behavior profiles; and
- constructing a methodology for the evaluation of intrusion detecting systems capable of monitoring over multiple containers' activities.

Aleksandar Milenkoski (ERNW, Germany),

Nuno Antunes (University of Coimbra),

José Flora (University of Coimbra)

<https://research.spec.org/working-groups/rg-ids-benchmarking.html>

[1] L. Iffländer, "Attack-aware Security Function Management", 2021, Ph.D. Universität Würzburg

[2] C. F. Goncalves, D. S. Menasche, A. Avritzer, N. Antunes, and M. Vieira, "A Model-Based Approach to Anomaly Detection Trading Detection Time and False Alarm Rate," 2020 Mediterranean Communication and Computer Networking Conference (MedComNet), pp. 1–8, 2020.

[3] L. Beierlieb, L. Iffländer, A. Milenkoski, C. F. Goncalves, N. Antunes, S. Kounev, "Towards Testing the Software Aging Behavior of Hypervisor Hypercall Interfaces," 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)

[4] L. Beierlieb, L. Iffländer, A. Milenkoski, and S. Kounev, "Towards Testing the Performance Influence of Hypervisor Hypercall Interface Behavior," Proceedings of the 10th Symposium on Software Performance, 2019

NEW TOOL ACCEPTED: MOWGLI: EVALUATION FRAMEWORK FOR CLOUD-HOSTED DBMS

A new tool was accepted for SPEC RG's repository of peer-reviewed tools for quantitative system evaluation and analysis: **Mowgli**. This tool provides a comprehensive evaluation framework for cloud-hosted DBMS that supports the evaluation objectives performance, scalability, elasticity and availability. Mowgli fully automates the DBMS evaluation process for each evaluation objective, i.e. allocating cloud resources, deploying the DBMS on the allocated resources, executing the workloads, executing runtime adaptations (elasticity objective) and injecting cloud resource failures (availability objective), releasing the cloud resources and processing the objective-specific metrics.

Mowgli is further developed at Ulm University in the scope of the Benchmarking as a Service (BaaS <https://baas-project.de/>) project. If you are interested in the topic, want to provide feedback, add requirements, and discuss, just get in touch.

Daniel Seybold, University of Ulm

JSS PAPER PUBLISHED

"Function-as-a-Service Performance Evaluation: A Multivocal Literature Review", published in the Journal of Systems and Software (JSS), describes 112 performance studies of Function-as-a-Service (FaaS) platforms. This paper consolidates the results from 61 industrial and 51 academic performance studies and provides actionable recommendations on reproducible FaaS experimentation. It also offers an interactive online appendix to allow readers to systematically discover relevant performance studies, including pointers to source code and data sets. DOI: 10.1016/j.jss.2020.110708

Joel Scheuner, Philipp Leitner, Chalmers, Sweden

PSI: PERVASIVE SELF-OPTIMIZING COMPUTING INFRASTRUCTURES

In October 2021, Alessandro Papadopoulos was awarded the individual Swedish Research Council (VR) starting grant, to conduct research on the design of self-optimizing infrastructure in the context of cloud computing. The project is based on some preliminary work on performance evaluation, and dynamic resource provision devel-

oped also in collaboration with the SPEC Cloud WG. PSI (Pervasive Self-Optimizing Computing Infrastructures) aims to provide a fabric of software components able to dynamically optimize the behavior of the cloud and IoT infrastructures. PSI targets an improved usage of resources at the systems' edge via continuous and distributed system-wide optimizations. This project covers both theoretical and practical aspects, and it combines different research areas including self-adaptive software, control theory, optimization, distributed and real-time systems. The main goal of the project is to develop new methodologies for the efficient usage of computational resources while providing guarantees on different key performance indicators, like, for example, response time and throughput of the system.

Alessandro Papadopoulos, Mälardalen University

SELECTED ABSTRACTS

Multivariate Time Series Synthesis Using Generative Adversarial Networks

Collection and analysis of distributed (cloud) computing workloads allows for a deeper understanding of user and system behavior and is necessary for efficient operation of infrastructures and applications. The availability of such workload data is however often limited as most cloud infrastructures are commercially operated and monitoring data is considered proprietary or falls under GDPR regulations. This work investigates the generation of synthetic workloads using Generative Adversarial Networks and addresses a current need for more data and better tools for workload generation. Resource utilization measurements such as the utilization rates of Content Delivery Network (CDN) caches are generated and a comparative evaluation pipeline using descriptive statistics and time-series analysis is developed to assess the statistical similarity of generated and measured workloads. We use CDN data open sourced by us in a data generation pipeline as well as back-end ISP workload data to demonstrate the multivariate synthesis capability of our approach. The work contributes a generation method for multivariate time series workload generation that can provide arbitrary amounts of statistically similar data sets based on small subsets of real data. The presented technique shows promising results, in particular for heterogeneous workloads not too irregular in temporal behavior.

Mark Leznik, Patrick Michalsky, Peter Willis, Benjamin Schanzel, Per-Olov Östberg, and Jörg Domaschka. Multivariate Time Series Synthesis Using Generative Adversarial Networks. In Proceedings of the 12th ACM/SPEC International Conference on Performance Engineering (ICPE) (2021). ACM, New York, NY, USA.

Libra: A Benchmark for Time Series Forecasting Methods

In many areas of decision making, forecasting is an essential pillar. Consequently, there are many different forecasting methods. According to the “No-Free-Lunch Theorem”, there is no single forecasting method that performs best for all time series. In other words, each method has its advantages and disadvantages depending on the specific use case. Therefore, the choice of the forecasting method remains a mandatory expert task. However, expert knowledge cannot be fully automated. To establish a level playing field for evaluating the performance of time series forecasting methods in a broad setting, we propose Libra, a forecasting benchmark that automatically evaluates and ranks forecasting methods based on their performance in a diverse set of evaluation scenarios. The benchmark comprises four different use cases, each covering 100 heterogeneous time series taken from different domains. The data set was assembled from publicly available time series and was designed to exhibit much higher diversity than existing forecasting competitions. Based on this benchmark, we perform a comprehensive evaluation to compare different existing time series forecasting methods.

André Bauer, Marwin Züfle, Simon Eismann, Johannes Grohmann, Nikolas Herbst, and Samuel Kounev. Libra: A Benchmark for Time Series Forecasting Method. In Proceedings of the 12th ACM/SPEC International Conference on Performance Engineering (ICPE) (2021). ACM, New York, NY, USA.

Incremental Calibration of Architectural Performance Models with Parametric Dependencies

Architecture-based Performance Prediction (AbPP) allows evaluation of the performance of systems and to answer what-if questions without measurements for all alternatives. A difficulty when creating models is that Performance Model Parameters (PMPs, such as resource demands, loop iteration numbers and branch probabilities) depend on various influencing factors like input data, used hardware and the applied workload. To enable a broad range of what-if questions, Performance Models (PMs) need to have predictive power beyond what has been measured to calibrate the models. Thus, PMPs need to be parametrized over the influencing factors that may vary. Existing approaches allow for the estimation of the parametrized PMPs by measuring the complete system. Thus, they are too costly to be applied frequently, up to after each code change. Moreover, they do not keep manual changes to the model when recalibrating. In this work, we present the Continuous Integration of Performance Models (CIPM), which incrementally extracts and calibrates the performance model, including parametric dependencies. CIPM responds to source code

changes by updating the PM and adaptively instrumenting the changed parts. To allow AbPP, CIPM estimates the parametrized PMPs using the measurements (generated by performance tests or executing the system in production) and statistical analysis, e.g., regression analysis and decision trees. Additionally, our approach responds to production changes (e.g., load or deployment changes) and calibrates the usage and deployment parts of PMs accordingly. For the evaluation, we used two case studies. Evaluation results show that we were able to calibrate the PM incrementally and accurately.

Manar Mazkatli, David Monschein, Johannes Grohmann, Anne Koziolk: Incremental Calibration of Architectural Performance Models with Parametric Dependencies. ICSA 2020: 23-34.

Performance, Power, and Energy-Efficiency Impact Analysis of Compiler Optimizations on the SPEC CPU 2017 Benchmark Suite

The growth of cloud services leads to more and more data centers that are increasingly larger and consume considerable amounts of power. To increase energy efficiency, both the actual server equipment and the software must become more energy efficient. Software has a major impact on hardware utilization levels, and subsequently, the energy efficiency. While energy efficiency is often seen as identical to performance, we argue that this may not be necessarily the case. A sizable amount of energy could be saved, increasing energy efficiency by leveraging compiler optimizations but at the same time impacting performance and power consumption over time. We analyze the SPEC CPU 2017 benchmark suite with 43 benchmarks from different domains, including integer and floating-point heavy computations on a state-of-the-art server system for cloud applications. Our results show that power consumption displays more stable behavior if less compiler optimizations are used and also confirmed that performance and energy efficiency are different optimizations goals. Additionally, compiler optimizations possibly could be used to enable power capping on a software level and care must be taken when selecting such optimizations.

Norbert Schmitt, James Bucek, John Beckett, Aaron Cragin, Klaus-Dieter Lange, Samuel Kounev. Performance, Power, and Energy-Efficiency Impact Analysis of Compiler Optimizations on the SPEC CPU 2017 Benchmark Suite. In Proceedings of the 13th IEEE/ACM International Conference on Utility and Cloud Computing (UCC2020)