

CONTENTS

- 2 SPEC RG Officers and Working Groups
- 3 Welcome to the SPEC RG Newsletter
- 3 Kaivalya Dixit Distinguished Dissertation Award
- 4 ICPE 2023: Statistics
- 4 ICPE 2024 in London
- 5 Working Group Reports
- 11 Graduated SPEC RG Members
- 11 New SPEC RG Tools
- 12 Selected Abstracts

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2022

The award selection committee for 2022 selected the dissertation of Lishan Yang from the College of William & Mary, USA. The selection committee was impressed by the high quality of scientific work conducted by Dr. Yang and particularly by her ability to bridge multiple academic disciplines spanning software engineering, performance, and computer systems.

Read more on page 3

SPEC RESEARCH WORKING GROUPS REPORT ON THEIR PROGRESS

The SPEC Research Working Groups Security, Cloud, DevOps Performance, Predictive Data Analytics, and Power report on their progress, articles, benchmarks, and technical reports published in 2022. The Working Groups are always open for new members. Feel invited to join us!

Read more on pages 5-11

GRADUATED LONG-TERM MEMBERS OF SPEC RG

In the past year, two long-term members and well-known researchers from SPEC RG, Joel Scheuner and Simon Eismann, received their PhDs. SPEC RG thanks Joel and Simon for their work in the last years. Find summaries of their work in this newsletter.

Read more on page 11

CONTACT



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

SPEC Research Group (SPEC RG)

Chair: Samuel Kounev (rgchair@spec.org)

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

NEW SPEC RG TOOLS

Theodolite and Libra are the newest SPEC RG tools. Theodolite is a framework for benchmarking the scalability of cloud-native applications, while Libra is a benchmark for time series forecasting.

Read more on page 11

SPEC RESEARCH GROUP OFFICERS

Chair:

Samuel Kounev, University of Würzburg, Germany

Vice-Chair:

André van Hoorn, University of Hamburg, Germany

Secretary:

Cristina Abad, ESPOL University, Ecuador

Steering Committee:

Cristina Abad, ESPOL University, Ecuador

Cor-Paul Bezemer, University of Alberta, Canada

Nikolas Herbst, University of Würzburg, Germany

Alexandru Iosup, VU Amsterdam, NL

Samuel Kounev, University of Würzburg, Germany

Klaus-Dieter Lange, HPE, USA

Alex Podelko, Amazon, USA

Weiyi Shang, Concordia University, Canada

Petr Tůma, Charles Univ. of Prague, Czech Republic

André van Hoorn, University of Hamburg, Germany

Publicity Officer:

Weiyi Shang, Concordia University, Canada

Release Manager:

Michele Tucci, Charles Univ. Prague, Czech Republic

Newsletter Editors:

Martin Straesser, University of Würzburg, Germany

André Bauer, University of Chicago, USA

<https://research.spec.org>

SPEC RESEARCH WORKING GROUPS

DevOps Performance Working Group

Chair:

André van Hoorn, University of Hamburg, Germany

Vice-Chair:

Cor-Paul Bezemer, University of Alberta, Canada

Weiyi Shang, Concordia University, Canada

Secretary:

Heng Li, Polytechnique Montréal, Canada

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

Cloud Working Group

Chair:

Alexandru Iosup, VU Amsterdam, The Netherlands

Vice-Chair:

Nikolas Herbst, University of Würzburg, Germany

Secretary:

Sacheendra Talluri, VU Amsterdam, The Netherlands

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

Power Working Group

Chair:

Maximilian Meissner, University of Würzburg, Germany

Vice-Chair:

Klaus-Dieter Lange, HPE, USA

Secretary:

Aaron Cragin, Microsoft, USA

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

Security Working Group

Chair:

Aleksandar Milenkoski, SentinelLabs/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>

Predictive Data Analytics Working Group

Chair:

André Bauer, University of Chicago, USA

Vice-Chair:

Pooyan Jamshidi, University of South Carolina, USA

Secretary:

Mark Leznik, University Ulm, Germany

Release Manager:

Daniel Seybold, University Ulm, Germany

<https://research.spec.org/working-groups/rg-predictive-data-analytics/>

WELCOME TO THE SPEC RESEARCH GROUP NEWSLETTER

With over 120 members in 30 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 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:

- 13th ACM/SPEC **ICPE 2022** (virtual event)
- 3rd IEEE International Conference on Autonomous Computing and Self-Organizing Systems **ACSOS 2022** (virtual event)
- 5th Workshop on Hot Topics in Cloud Computing Performance **HotCloudPerf 2022** at ICPE 2022

We have been actively working on the preparation, planning, and organization of ICPE 2023, this year (the first time after the pandemic) as an on-site 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 this newsletter. We welcome and encourage your contributions for articles and suggestions for future coverage.

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

SPEC, the SPEC logo and the names SERT, SPEC SFS, SPECjbb, SPECvirt, Chauffeur, and 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 ©1988-2023 Standard Performance Evaluation Corporation (SPEC). Reprinted with permission. All rights reserved.

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2022 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 winning dissertation *Practical GPGPU Application Resilience Estimation and Fortification* was authored by Lishan Yang at the College of William & Mary (US) under the supervision of Prof. Evgenia Smirni. The selection committee was impressed by the high quality of scientific work conducted by Dr. Yang and particularly by her ability to bridge multiple academic disciplines spanning software engineering, performance, and computer systems. The committee also appreciates the high industrial relevance of this dissertation as well as the ability and willingness of Dr. Yang to work with industry. The award is to be presented at the International Conference on Performance Engineering (ICPE 2023).

Given the high quality of dissertations nominated for this award, the committee decided to publicly recognize another dissertation as runner-up, "Model Learning for Performance Prediction of Cloud-native Microservice Applications" authored by Dr. Johannes Grohmann at the University of Würzburg (Germany) under the supervision of Prof. Samuel Kounev.

The award selection committee for 2022 was chaired by Dr. Philipp Leitner (Chalmers University of Technology, Sweden).

The SPEC Kaivalya Dixit Distinguished Dissertation Award was established in 2011 to recognize outstanding dissertations within the scope of the SPEC Research Group. Contributions of interest span the design of metrics for system evaluation as well as the development of methodologies, techniques and tools for measurement, load testing, profiling, workload characterization, dependability and efficiency evaluation of computing systems.

Dissertations defended **between Oct 2021 and Sep 2023** will be eligible to be nominated for the 2023 award (a thesis can be nominated only once, which means that a 2022 thesis nominated in 2023 cannot be nominated again in 2024).

<https://research.spec.org/news/2023-01-16-00-00-winner-of-spec-kaivalya-dixit-distinguished-dissertation-award-2022/>

ICPE 2023: STATISTICS

The 14th ACM/SPEC International Conference on Performance Engineering (ICPE 2023) will be held in Coimbra, Portugal from April 15 to April 19, 2023. This year, the research track of ICPE attracted 46 submissions, 15 of which were selected as full articles after a rigorous review process, yielding an acceptance rate of 32.6%. Of the 8 submissions to the Industry Track, 4 were accepted as full articles. In the emerging research track, 9 out of 13 submissions were accepted as emerging research papers. 4 papers of the research and industry tracks were able to achieve ACM artifact badges. In addition, one standalone artifact was accepted and achieved ACM artifact badges. For the second time, ICPE features a Data Challenge with 7 accepted short papers out of 9 submissions. 6 tutorials have been proposed out of which 3 were selected to complement the program. Furthermore, 7 posters, 4 demonstrations, and 3 tutorials will be presented in interactive sessions.

The following workshops are planned for ICPE 2023:

- The First Workshop on Artificial Intelligence for Performance Modeling, Prediction, and Control (AIPerf 2023)
<https://aiperf.github.io/aiperf2023/>
- The First FastContinuum Workshop (FastContinuum 2023)
<https://sites.google.com/view/fastcontinuum-2023>
- The First Workshop on Serverless, Extreme-Scale, and Sustainable Graph Processing Systems (GraphSys 2023)
<https://sites.google.com/view/graphsys23/home>
- The Sixth Workshop on Hot Topics in Cloud Computing Performance (HotCloudPerf 2023)
<https://hotcloudperf.spec.org>
- The Eleventh International Workshop on Load Testing and Benchmarking of Software Systems (LTB 2023)
<https://ltb2023.github.io>
- The First Practically FAIR 2023 (PFAIR 2023)
<https://sites.google.com/view/pfair23/home>
- The Fourth Workshop on Education and Practice of Performance Engineering (WEPPE 2023)
<https://esulabsolutions.godaddysites.com/sponsored-events>
- The Eighth Workshop on Challenges in Performance Methods for Software Development (WOSP-C 2023)
<https://wosp-c.github.io/wosp-c-23/>

The following distinguished speakers will give keynotes at ICPE 2023:

- Mathieu Nayrolles (Ubisoft, Canada)
Pushing the Limits of Video Game Performance: A Performance Engineering Perspective
- Federica Sarro (University College London, UK)
Automated Optimisation of Modern Software System Properties
- Georg Hager (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany)
Application Knowledge Required: Performance Modeling for Fun and Profit

Marco Vieira (University of Coimbra, Portugal),
Valeria Cardellini (University of Rome Tor Vergata, Italy),
Antinisa Di Marco (University of L'Aquila, Italy),
Petr Tuma (Charles University, Czechia)

ICPE 2024 IN LONDON — 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 2024 will be held in London, UK, in April 2024. London is the capital city of England and the United Kingdom, known for its rich history, diverse culture, and iconic landmarks. The city is home to some of the world's most famous attractions, such as Big Ben, Tower Bridge, London Eye, and Buckingham Palace, which make it a popular destination for tourists from around the globe. Apart from its historical significance, London is also a hub for arts, entertainment, and fashion. The city boasts numerous theaters, galleries, museums, and music venues that attract millions of visitors each year. London's diverse population also makes it a foodie's paradise with an array of culinary delights from different cultures and cuisines. From traditional fish and chips to contemporary fusion dishes, London has it all.

The contact person for ICPE 2024 is Prof. William Knottenbelt, from Imperial College London. The General Chairs, Program Chairs, and Organization Committee will be announced soon.

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 Charles University, Concordia University, Imperial College London, Karlsruhe Institute of Technology, MongoDB, Polytechnique Montreal, University of Alberta, University of L'Aquila, University of Hamburg, University of South Carolina, University of Stuttgart and University of Würzburg.

Simon Eismann, who was the group's secretary for several years, indicated that he would take a step back from his SPEC RG DevOps activities after his PhD graduation. The group would like to thank Simon for his contributions and effort in the past few years. Heng Li (Polytechnique Montreal) has agreed to take over the secretary role.

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, 7 group meetings were held in 2022, including the following talks as part of our monthly lecture series:

- *Improving Inter-Team Coordination and Architectural Consistency in DevOps Contexts* by Robert Heinrich and Rebecca Wohlrab (Karlsruhe Institute of Technology)
- *FOCloud: Feature Model Guided Performance Prediction and Explanation for Deployment Configurable Cloud Applications* by Inka Weerasingha Dewage (Tilburg University)
- *Studying the Performance Risks of Upgrading Docker Hub Images: A Case Study of WordPress* by Mikael Sabuhi (University of Alberta)

- *Simulation and Benchmarking of Container Orchestration Frameworks* by Martin Straesser (University of Würzburg)
- *Breaking Type Safety in Go: An Empirical Study on the Usage of the unsafe Package* by Diego Elias Costa (Université du Québec à Montréal)
- *Automated Generation and Evaluation of JMH Microbenchmark Suites from Unit Test* by Mostafa Jangali (Concordia University)
- *Scalability Benchmarking of Cloud-Native Applications with Theodolite* by Sören Henning (Kiel University)

In addition to the monthly meetings, the group operates in subgroups consisting of 6-8 participants who collaborate closely on concrete topics. Collaborations include joint research papers and 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. In 2022, the subgroup published an empirical study on load testing of serverless applications in the *Journal of Systems and Software* [5]. The current work focuses on solving the challenges of detecting system-level performance regressions as early as during the development phase. Current efforts include developing novel approaches to constructing and analyzing system-level analytic performance models incorporating the insights of performance deviations in local components.
2. **Performance change point detection:** This newly established group focuses on the challenges of using performance detection methods in real large-systems performance data. In collaboration with three major tech companies, SAP, MongoDB, and Oracle, this subgroup has established the first curated dataset of real performance changes and is currently assessing the quality of current change detection methods. The group is now working on writing the first paper, which is expected to be submitted in the beginning of 2023.
3. **Search-based software performance engineering:** The focus of this new subgroup (started in Jan 2022) is on developing novel approaches and research directions on the multi-criteria optimization of performance-related quality attributes in software architectures. The group is now working on introducing the interaction of the designer in the architectural optimization process.

4. **Performance of continuous delivery infrastructures:** This subgroup focuses on the evaluation and improvement of continuous delivery (CD) infrastructures. Building on previously analyzed performance data of a CD system, the group has a current focus on simulating and optimizing CD pipelines. Based on the StalkCD CI/CD metamodel for pipeline specifications, the group continued its efforts to make the metamodel compatible with more specification languages. Particularly by evaluating the compatibility with and extending the metamodel, to improve the compatibility with the GitHub Actions CI/CD specification language.
5. **Resilience engineering for cloud-native applications:** The group investigates novel approaches, tools, and data sets for resilience engineering, including chaos engineering, architecture extraction, runtime monitoring, interactive resilience scenario improvement, and other techniques. Current joint efforts include the evolution of a resilience simulator and its application to scenario-based analysis of resilience properties and container orchestration strategies. In 2022, the group has published two papers [1,2] and given three presentations [P1-P3], which include contributions by three jointly supervised student projects in 2022.

We are happy that a member of the group was awarded at the SPEC Annual Meeting 2023: Michele Tucci (University of L'Aquila; currently abroad at Charles University Prague) with a SPEC Impact Award for his contributions to SPEC RG, SPEC RG DevOps, and the International Conference on Performance Engineering (ICPE). In the context of our working group, Michele has made a significant impact in the aforementioned subgroups on performance change point detection and search-based software performance engineering. Congratulations and thank you, Michele!

For more information about the DevOps Performance Working Group (including our mission, activities, meetings, presentations, and projects), please visit our web page [4]. 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 Hamburg), Weiyi Shang (Concordia University), Heng Li (Polytechnique Montreal)

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

[1] Sebastian Frank, Lion Wagner, Alireza Hakamian, Martin Straesser, and André van Hoorn: MiSim: A Simulator for Resilience Assessment of Microservice-based Architectures. IEEE 22nd International Conference on Software Quality, Reliability, and Security (QRS2022), 2022 (in print). Pre-print available at: <https://dqualizer.github.io/files/QRS22-FrankWagnerHakamianStraesserVanHoorn2022MiSim-preprint.pdf>

[2] Sebastian Frank, Alireza Hakamian, Lion Wagner, Dominik Kesim, Christoph Zorn, Jóakim von Kistowski, and André van Hoorn: Interactive Elicitation of Resilience Scenarios Based on Hazard Analysis Techniques. Selected and revised paper from 15th European Conference on Software Architecture (ECSA). Springer. 2022. Pre-print available at: <https://dqualizer.github.io/files/ECSA2022-InteractiveElicitation.pdf>

[3] https://emsejournal.github.io/special-issues/2021_Software_Performance.html

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

[5] Simon Eismann, Diego Elias Costa, Lizhi Liao, Cor-Paul Bezemer, Weiyi Shang, André van Hoorn, Samuel Kounev: A case study on the stability of performance tests for serverless applications, Journal of Systems and Software, Volume 189, 2022.

[P1] Tim Thüring, Gabriel Glaser, Marcel Hafner, Abel Gitzing, Sebastian Frank, Alireza Hakamian and André van Hoorn: Extracting Software Architectures from Traces for the Simulation of Microservice-based Architectures, 13th Symposium on Software Performance (<https://www.performance-symposium.org/2022/>), 2022.

[P2] Denis Zahariev, Sebastian Frank, Alireza Hakamian and André van Hoorn: Supporting and Verifying Transient Behavior Specifications in Chaos Engineering, 13th Symposium on Software Performance (<https://www.performance-symposium.org/2022/>), 2022.

[P3] Martin Straesser, Patrick Haas and Samuel Kounev: Leveraging Kubernetes Source Code for Performance Simulation, 13th Symposium on Software Performance (<https://www.performance-symposium.org/2022/>), 2022.

REPORT: CLOUD WORKING GROUP

In 2022, 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 **measuring** serverless application performance, **adapting** stateful interactive applications like gaming to use serverless computing, **defining** the compute continuum through surveys and a reference architecture.

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 2022, 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 tracing:** Making serverless computing widely applicable requires detailed performance understanding. We design and implement ServiBench [1], a serverless benchmarking suite. ServiBench (i) leverages synchronous and asynchronous serverless applications representative of production usage, (ii) extrapolates cloud-provider data to generate realistic workloads, (iii) conducts comprehensive, end-to-end experiments to capture application-level performance, (iv) analyzes results using a novel approach based on (distributed) serverless tracing, and (v) supports comprehensively serverless performance analysis. With ServiBench, we conduct comprehensive experiments on AWS. We find that the median end-to-end latency of serverless applications is often dominated not by function computation but by external service calls, orchestration, or trigger-based coordination. We release collected experimental data under FAIR principles and ServiBench as a tested, extensible open-source tool.
 2. **Serverless gaming:** Gaming is a large industry with an increasing social, cultural, and economic impact. The discussion around this topic is converging to a reenvisioning of a metaverse: a digital ecosystem of virtual worlds that further integrates human society in a digital space. However, the development of such a metaverse requires addressing challenges related to realism, ubiquity, interoperability, and scalability. We are investigating how to address the latter class of challenges, scalability, through the use of serverless computing. Serverless computing is an emerging cloud-computing technology that promises simple and efficient scalability improvements across application domains. However, serverless computing is not designed for real-time multi-user systems, and it is unclear how serverless computing can be applied, or must evolve, for this use case. To this end, we explored serverless computational offloading across several projects. Currently, we are exploring novel architectures for large-scale virtual worlds with serverless computing at the core, with a focus on consistency and scheduling.
 3. **Serverless storage:** Data-driven interactive computation is widely used for business analytics, search-based decision-making, and log mining. These applications' short duration and bursty nature makes them a natural fit for serverless computing. Data processing serverless applications are composed of many small tasks. Application tasks that use remote storage encounter bottlenecks in the form of high latency, performance variability, and throttling. The use of caching for input data, albeit widely used in industry, has yet to be studied. We present the first performance study of scaling, a key feature of serverless computing, on serverless clusters with input data caches. We compare 8 task placement algorithms and quantify their impact on task slowdown and resource usage before and after scaling. We find up to a 400% increase in task slowdown after scaling without work stealing and a 15% slowdown with work stealing. We also find that cache misses after scaling can lead to an additional 18% resource usage.
 4. **Edge Reference Architecture:** The first paper in this activity [2] (Titled: The SPEC-RG Reference Architecture for the Compute Continuum) analyzes commonalities and differences between cloud and edge computing models, makes a case for a single unified computing model for the compute continuum, and presents the design of the SPEC-RG reference architecture for the compute continuum. This paper will be published at the 23rd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing.
 5. **Edge Workload Characterization:** Following our work titled "Edge Workload Trace Gathering and Analysis for Benchmarking [3], published at the 6th IEEE International Conference on Fog and Edge Computing 2022, we are working on defining a trace format for edge workload with the aim of creating an open trace repository. As a part of the definition, the group is also looking into what fine-grained characteristics are of interest for edge workloads and how these can be used to (semi-) automatically categorize and differentiate various edge computing workloads.
 6. **An overview of the Computing Continuum:** We are working on providing a comprehensive and unified view of the computing continuum, from IoT to the cloud. We are aiming at discussing the computing models in general with a focus on cloud computing, the computation models that emerged beyond the cloud, and the communication technologies that enable computing in the continuum. The work will include two novel reference architectures: one for the edge-cloud computing models, and the other for edge-cloud communication technologies. Several real use cases from different sectors to validate the proposed reference architectures will be demonstrated. The SPEC RG Cloud Group members are planning to highlight some points that express their vision about how to efficiently enable and utilize the computing continuum in the future.
 7. **Non-von Neumann Architectures Survey:** We are working on a survey titled "Beyond von Neumann
-

Architectures in the Computing Continuum”. The survey discusses the recent trends in artificial intelligence that sparked the research for developing more efficient and sustainable distributed computing architectures. To begin with, the paper provides a detailed taxonomy of the currently available computer architectures based on their memory characteristics concerning the classical von Neumann model. Furthermore, it explores a large set of applications and analyzes their suitability for different kinds of non-von Neumann architectures. Finally, the paper discusses the main barriers related to adopting modern non-von Neumann systems, such as Quantum and Neuromorphic systems, in the computing continuum.

Besides these focused activities, the Cloud WG has been acting successfully in organizing and growing the yearly workshop HotCloudPerf 2023 [4]. On April 16, the workshop will be held in Coimbra co-located with ICPE2023 and feature two invited keynotes, six accepted paper presentations and a joint panel discussion with FastContinuum workshop.

Keynote speakers:

- Tania Lorido Botran (Roblox, USA)
- Cristian Klein (Umeå University, Sweden)
- Radu Prodan (Klagenfurt University, Austria)

Paper presentations:

- Xiaoyu Chu et al, VUA, NL: How Do ML Jobs Fail in Datacenters? Analysis of a Long-Term Dataset from a HPC Cluster
- Francisc-Josep Lordan Gomis et.al, BSC, ES & TUD, DE: Securing the Execution of ML Workflows across the Compute Continuum
- Floriment Klinaku et al., U. Stuttgart, DE: Hitchhiker’s Guide for Explainability in Autoscaling
- Jessica Leone et al., U. l’Aquila, IT: Enhancing Trace Visualizations for Microservices Performance Analysis
- Matthijs Jansen et al., VUD, NL: Can My WiFi Handle the Metaverse? A Performance Evaluation Of Meta’s Flagship Virtual Reality Hardware
- George Kousiouris et al., HU Athens, GR: Performance experiences from running an E-health inference process as FaaS across diverse clusters

Two group members, namely Joel Scheuner (Chalmers, SE) and Simon Eismann (JMU Wü), successfully defended their PhD theses. You can find a thesis summary and links in the respective sections below. We would

like to thank Simon and Joel for the multi-year fruitful collaboration within the working group.

To conclude, 2022 was a full and successful year for the RG Cloud Group. We are looking forward to an even more successful 2023. 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] Joel Scheuner, Simon Eismann, Sacheendra Talluri, Erwin van Eyk, Cristina Abad, Philipp Leitner, Alexandru Iosup: Let’s Trace It: Fine-Grained Serverless Benchmarking using Synchronous and Asynchronous Orchestrated Applications. arXiv preprint arXiv:2205.07696, 2022.

[2] Matthijs Jansen, Auday Al-Dulaimy, Alessandro V. Papadopoulos, Animesh Trivedi, Alexandru Iosup: The SPEC-RG Reference Architecture for the Compute Continuum. IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID), 2023.

[3] Klervie Toczé, Norbert Schmitt, Ulf Kargén, Atakan Aral, Ivona Brandić: Edge Workload Trace Gathering and Analysis for Benchmarking. IEEE International Conference on Fog and Edge Computing (ICFEC), 2022.

[4] <https://hotcloudperf.spec.org>

REPORT: SECURITY WORKING GROUP

The SPEC RG Security Benchmarking Working Group, after establishing its long-term agenda, continues to work towards devising and conducting impactful results while achieving the defined goals. In 2022, the Working Group had papers published at the Virus Bulletin Conference 2022 [1] and the IEEE Access Journal [2].

The paper entitled (*Encryption*) *time flies when you’re having fun: the case of the exotic BlackCat ransomware* [1] where the BlackCat ransomware was reverse engineered to provide a first look into the inner workings of the encryption modes implemented. The analysis provides a unique insight into the design decisions that ransomware developers make to achieve an optimal balance between encryption speed and encryption completeness.

Following the paper entitled *My Services Got Old! Can Kubernetes Handle the Aging of Microservices?* [3] published in 2021, we extended the work and published an article entitled *A Study on the Aging and Fault Tolerance of Microservices in Kubernetes* [2] on the IEEE Access Journal. The work studies the effectiveness of Kubernetes in dealing with faults and aging in microservices, and on the possibility of using faults to accelerate aging effects for testing purposes. For this, we conducted an analysis of the implementation and tuning of Kubernetes probes, followed by experiments with varying load and fault in-

jection into two distinct and representative microservice testbeds to analyze the capacity of probes in detecting issues in applications.

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
- collaboration between Würzburg and Coimbra on injecting intrusions into hypervisors
- evaluating the performance of different approaches in the context of detecting security intrusions based on performance and behavior profiles
- methodologies to allow understanding hypercall interfaces
- design and develop a methodology to profile ransomware encryption processes

Aleksandar Milenkoski (SentinelLabs/Univ. Würzburg, Germany),
 Nuno Antunes(University of Coimbra),
 José Flora (University of Coimbra)

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

[1] Aleksandar Milenkoski: (Encryption) time flies when you're having fun: the case of the exotic BlackCat ransomware, presented at the 2022 Virus Bulletin (VB2022), 2022.

[2] José Flora, Paulo Gonçalves, Miguel Teixeira, Nuno Antunes: A Study on the Aging and Fault Tolerance of Microservices in Kubernetes. IEEE Access, vol. 10, 2022.

[3] José Flora, Paulo Gonçalves, Miguel Teixeira, Nuno Antunes: My Services Got Old! Can Kubernetes Handle the Aging of Microservices? IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), 2021.

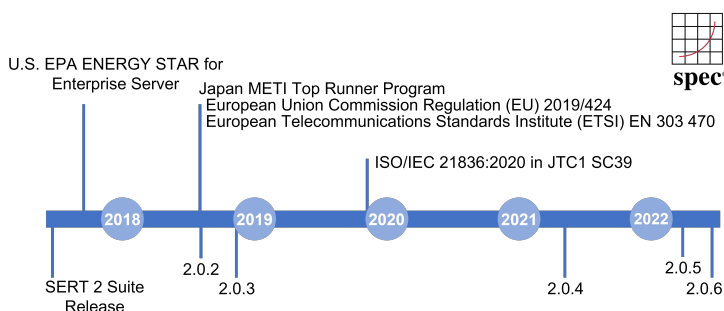
REPORT: POWER WORKING GROUP

The SPEC Research Power Working Group (*AMD, Ampere, Dell, HPE, IBM, Inspur, Intel, Microsoft, University of Würzburg*) has operated tightly coupled with the SPECpower Committee since its inception in 2017.

Our mission is 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.

Closely connected to the SPEC RG Power WG and the SPECpower Committee is the International Standards Group (ISG), which we introduced to the public at ICPE 2021 [1]. The ISG oversees the establishment of standardized benchmarks primarily developed for the use in government regulations and programs, and collaborates with national and international standard development organizations to enhance global standards. Within the ISG, the ISG Server Efficiency Committee (Chair: Sanjay Sharma) and the ISG Client Efficiency Committee (Chair: Roger Tiple) have been established. The goal of the former is to oversee the establishment and development of standardized server benchmarks primarily for the use in government regulations and programs. The latter's mission is similar, but with a focus on client devices. Current members and associates of the ISG are *AMD, Ampere, Apple, Bull, Dell, HPI, HPE, IBM, Inspur, Intel, Microsoft, Quanta, and University of Würzburg*.

With the release of ISO/IEC 21836:2020 in August 2020 [2], the founding members of ISG already have achieved the standardization of the server energy effectiveness metric (SEEM). ISG is striving towards global harmonization, where one internationally standardized benchmark is used in every national regulation for server energy efficiency. The Server Efficiency Rating Tool (SERT) suite, which the U.S. EPA already has adopted with Version 3.0 of their Server Energy Star program, will be enhanced further under the responsibility of the ISG (since the release of the SERT 2.0.5 suite).



Apart from our contributions to government regulations, one of our goals is to raise practitioners' awareness with respect to the influence different designs of the same functionality in their software can have on energy efficiency. To this end, we investigated the energy efficiency of common sorting algorithms in different implementation variants and for different problem sizes [3][4]. The results demonstrate that simple design choices can have significant impact on energy efficiency, a fact often neglected by developers.

At ICPE 2022, members of the RG Power WG [5], the SPECpower Committee [6], the SPEC ISG [7], and The Green Grid [8] jointly offered a tutorial titled *SPEC Server Efficiency Benchmark Development - How to Contribute to the Future of Energy Conservation* [9].

The group currently focuses on expanding the SPEC Power and Performance Benchmark Methodology [10] to include technologies such as liquid cooling, Direct-Current PSUs, and servers with Auxiliary Processing Accelerators (APA) such as GPUs, FPGAs, and ASICs. Furthermore, for the upcoming SPECpowerNext benchmark and SERT 3 suite, the evaluation of novel workloads will be a major focus. At ICPE 2023, we present an overview outlining our current research focus [11].

The SPEC RG Power WG is looking 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.

Maximilian Meissner (University of Würzburg),
Klaus-Dieter Lange (Hewlett Packard Enterprise)
<https://research.spec.org/working-groups/rg-power/>

[1] Norbert Schmitt, Klaus-Dieter Lange, Sanjay Sharma, Aaron Cragin, David Reiner, and Samuel Kounev: SPEC — Spotlight on the International Standards Group (ISG). Companion of the ACM/SPEC International Conference on Performance Engineering, 2021.

[2] <https://www.iso.org/standard/71926.html>

[3] Norbert Schmitt, Supriya Kamthania, Nishant Rawtani, Luis Mendoza, Klaus-Dieter Lange, Samuel Kounev: Energy-Efficiency Comparison of Common Sorting Algorithms. International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), 2021.

[4] Maximilian Meissner, Supriya Kamthania, Nishant Rawtani, James Bucek, Klaus-Dieter Lange, and Samuel Kounev: Experience and Guidelines for Sorting Algorithm Choices and Their Energy Efficiency. Companion of the ACM/SPEC International Conference on Performance Engineering, 2022.

[5] <https://research.spec.org/working-groups/rg-power/>

[6] <https://www.spec.org/power/>

[7] <https://www.spec.org/isg/>

[8] <https://www.thegreengrid.org/en/about-us>

[9] Maximilian Meissner, Klaus-Dieter Lange, Jeremy Arnold, Sanjay Sharma, Roger Tiple, Nishant Rawtani, David Reiner, Mike Petrich, and Aaron Cragin: SPEC Efficiency Benchmark Development: How to Contribute to the Future of Energy Conservation. Companion of the ACM/SPEC International Conference on Performance Engineering, 2022.

[10] https://www.spec.org/power_ss2008/docs/SPECpower-Methodology.pdf

[11] Maximilian Meissner, Klaus-Dieter Lange, Sanjay Sharma, Jeremy Arnold, Aaron Cragin, Patrick Galizia, Mike Petrich, Bin Zhang, and Samuel Kounev: Challenges and Future Directions in Efficiency Benchmarking (Vision Paper). Companion of the ACM/SPEC International Conference on Performance Engineering, 2023.

REPORT: PREDICTIVE DATA ANALYTICS WORKING GROUP

To bridge the missing links between the facets involved in data analytics, namely big data storage and provisioning, data versioning, and performance evaluation, the SPEC predictive data analytics working group was established in June 2021. The group's ambition is to standardize and benchmark the entire data lifecycle, i.e., the analytics/prediction methods and especially pipelines for data analytics ranging from big data storage and preprocessing to analytics and assessment, as well as to provide heuristics for the selection of tools, patterns, and infrastructure. Furthermore, the group members are interested in promoting the interaction between industry and academia by contributing research towards standardization and benchmarking of the different aspects of data analytics. For this, the group members investigate data analytics-related methodologies, systems, and metrics. Another important goal is to support open data and promote the reproducibility of experiments and benchmarking of data analytics methods. The interests of the group lie in but are not limited to:

1. Performance modeling, analysis, testing, and prediction
2. Performance analysis of ML Systems
3. Resource autoscaling and reconfigurable systems
4. Performance behavior in resource-constrained environments
5. Change point and anomaly detection
6. Time series analysis and forecasting
7. Streamlining the data science process (DataOps)
8. Benchmarking of big data infrastructure

In 2022, we discussed how to standardize and guide modeling in the context of cloud computing. Another topic of interest was evaluating the quality of synthetic data and their generators. In addition, we will present a Data

Challenge paper [1] at ICPE 2023. Moreover, our chair, André Bauer, won the SPECTacular Award 2023 for his dedicated service for the SPEC RG.

In general, the group is meeting online every first Tuesday in a month at 9 am CST; you are welcome to attend. If you are interested in joining the group, please contact André Bauer (andrebauer@uchicago.edu).

André Bauer, University of Chicago, USA

<https://research.spec.org/working-groups/rg-predictive-data-analytics/>

[1] André Bauer, Martin Straesser, Mark Leznik, Lukas Beierlieb, Marius Hadry, Nathaniel Hudson, Kyle Chard, Samuel Kounev, and Ian Foster: Searching for the Ground Truth: Assessing the Similarity of Benchmarking Runs. Proceedings of the 14th ACM/SPEC International Conference on Performance Engineering, 2023.

GRADUATED LONG-TERM MEMBERS OF THE SPEC RG

In the past year, two long-term members and well-known researchers in SPEC RG, Joel Scheuner and Simon Eismann, received their PhDs. Congratulations! You can find summaries of their work in the following paragraphs.

Joel Scheuner

The growth of established cloud services, such as Infrastructure-as-a-Service (IaaS), and the emergence of new serverless services, such as Function-as-a-Service (FaaS), has led to an unprecedented diversity of cloud services with different performance characteristics. Measuring these characteristics is difficult in dynamic cloud environments due to performance variability in large-scale distributed systems with limited observability. This thesis aims to enable reproducible performance evaluation of serverless applications and their underlying cloud infrastructure.

His thesis has several components each with unique findings. The review of 112 FaaS performance studies from academic and industrial sources found a strong focus on a single cloud platform using artificial micro-benchmarks and discovered that most studies do not follow reproducibility principles on cloud experimentation. Characterizing 89 serverless applications revealed that they are most commonly used for short-running tasks with low data volume and bursty workloads. A novel trace-based serverless application benchmark shows that external service calls often dominate the median end-to-end latency and cause long tail latency. The evaluation of different cloud benchmarking methodologies has shown that only selected micro-benchmarks are suitable for estimating application performance, performance variability depends on the resource type, and batch testing on the same instance with repetitions should be used for reliable performance testing.

Joel now works as a Backend Engineer at the cloud emulation company LocalStack [1]. Find regular updates on his personal website [2].

Simon Eismann

Due to the rapid adoption of serverless computing, much information surrounding serverless computing is inconsistent and often outdated as the serverless paradigm evolves. This makes the performance engineering of serverless applications and platforms challenging. In this thesis, we address the lack of performance knowledge surrounding serverless applications and platforms from multiple angles. We conduct empirical studies to characterize serverless applications and platforms, and identify a community consensus on eight characteristics of serverless applications. We introduce automated optimization methods that simplify the operation of serverless applications, and achieve an average speedup of 39.7% while simultaneously reducing average costs by 2.6%. We enable the analysis of design tradeoffs of serverless platforms by speeding up white-box performance models up to 94.8%.

Simon now works as a Performance Engineer at the database company MongoDB [3]. Have a look at his PhD Website [4] for his full record and achievements.

[1] <https://localstack.cloud/>

[2] <https://joelscheuner.com/>

[3] <https://www.mongodb.com/>

[4] <https://se.informatik.uni-wuerzburg.de/software-engineering-group/staff/simon-eismann/>

NEW SPEC RG TOOL: THEODOLITE: BENCHMARKING FRAMEWORK FOR SCALABILITY OF CLOUD-NATIVE APPLICATIONS

Theodolite is a framework for benchmarking the scalability of cloud-native applications, running in Kubernetes. It automates the benchmarking process by deploying the system under test (SUT) to a Kubernetes cluster, generating load on the SUT, and collecting performance metrics during load generation. Theodolite comes as a Kubernetes Operator being installed inside the Kubernetes cluster.

Theodolite allows both designing new benchmarks and running existing ones. Designing a benchmark involves specifying the SUT, a load generator, a dimension along load is increased, a dimension along resources are scaled, and SLOs which are evaluated by Theodolite. All of this is defined in declarative files, which can be written and managed using established Kubernetes tooling. Likewise, running a benchmark only involves defining the experimental setup in declarative files and deploying them to Kubernetes.

Theodolite comes with a set of benchmarks for distributed stream processing engines. These benchmarks represent typical use cases for analyzing Industrial Internet of Things sensor data such as writing measurements to a database or performing different types of aggregations on streaming data. Along with a configurable load generator, Theodolite provides benchmark implementations for the stream processing engines Apache Flink, Kafka Streams, Hazelcast Jet as well as other engines, which are supported by the Apache Beam SDK.

Sören Henning, Kiel University

NEW SPEC RG TOOL: LIBRA: A BENCHMARK FOR TIME SERIES FORECASTING METHODS

Time series forecasting is essential in various disciplines for decision-making. Therefore, it is also an active research area where many methods are proposed. According to the "No-Free-Lunch Theorem", there is no forecasting method that performs best for all use cases. That is, every method has its own advantages and disadvantages. Although there are forecasting competitions such as the M-Competitions which can be considered as benchmarks, these are barely applied in scientific works. They typically consider only small sets of methods and do not provide any information on the time-to-result of the studied methods. Consequently, methods until now fail to provide a reliable approach to guide the choice of an appropriate forecasting method for a particular use case. The question of how to solve this problem remains.

To tackle this question, the forecasting benchmark Libra is presented. Libra automatically evaluates a forecasting method based on the choices of the user. The user uploads a code artifact of the forecasting method to be benchmarked. Then, the user specifies one of four use cases and selects one out of three evaluation types for the benchmarking process. Based on that, the method in question is evaluated and compared to other state-of-the-art methods, so that it allows the user to compare the respective forecasting method to other forecasting methods. In addition to that, different forecast error measures are included for a more detailed insight. A highly diverse data set was gathered comprising 400 publicly available time series taken from different domains, providing a higher heterogeneity compared to prior forecasting competitions. To sum it up, Libra offers a broad data set exhibiting, a high degree of diversity, different measures, three types of evaluation approaches, and is publicly available at [1].

André Bauer, University of Chicago

[1] <https://github.com/DescartesResearch/ForecastBenchmark>

SPEC LAUNCHES NEW BLOG SECTION

SPEC (Standard Performance Evaluation Corporation) has launched a new blog on its website [1]. The blog features the latest developments and updates from all SPEC groups and members. Members of the Research Group are invited and encouraged to submit content proposals. The RG steering committee is in talks about a formal process outlining how contents can be proposed, and will inform RG members about the outcome in the near future.

Samuel Kounev, SPEC Research Chair
André van Hoorn, SPEC Research Vice-Chair

[1] <https://www.spec.org/blog/>

SELECTED ABSTRACTS

MiSim: A Simulator for Resilience Assessment of Microservice-based Architectures

Increased resilience compared to monolithic architectures is both one of the key promises of microservice-based architectures and a big challenge, e.g., due to the systems' distributed nature. Resilience assessment through simulation requires fewer resources than the measurement-based techniques used in practice. However, there is no existing simulation approach that is suitable for a holistic resilience assessment of microservices comprised of (i) representative fault injections, (ii) common resilience mechanisms, and (iii) time-varying workloads. This paper presents MiSim — an extensible simulator for resilience assessment of microservice-based architectures. It overcomes the stated limitations of related work. MiSim fits resilience engineering practices by supporting scenario-based experiments and requiring only lightweight input models. We demonstrate how MiSim simulates (1) common resilience mechanisms — i.e., circuit breaker, connection limiter, retry, load balancer, and autoscaler — and (2) fault injections — i.e., instance/service killing and latency injections. In addition, we use TeaStore, a reference microservice-based architecture, aiming to reproduce scaling behavior from an experiment by using simulation. Our results show that MiSim allows for quantitative insights into microservice-based systems' complex transient behavior by providing up to 25 metrics.

Sebastian Frank, Lion Wagner, Alireza Hakamian, Martin Straesser, and André van Hoorn: MiSim: A Simulator for Resilience Assessment of Microservice-based Architectures. Proceedings of the IEEE 22nd International Conference on Software Quality, Reliability, and Security (QRS2022), 2022 (in print). Pre-print available at: <https://dqualizer.github.io/files/QRS22-FrankWagnerHakamianStraesserVanHoorn2022MiSim-preprint-nodoi.pdf>

Interactive Elicitation of Resilience Scenarios Based on Hazard Analysis Techniques

Context. Microservice-based architectures are expected to be resilient. **Problem.** In practice, the elicitation of resilience requirements and the quantitative evaluation of whether the system meets these requirements is not systematic or not even conducted. **Objective.** We explore (1) the usage of the scenario-based Architecture Trade-Off Analysis Method (ATAM) and established hazard analysis techniques, i.e., Fault Trees and Control Hazard and Operability Study (HAZOP), for interactive resilience requirement elicitation and (2) resilience testing through chaos experiments for architecture assessment and improvement. **Method.** In an industrial setting, we design a structured ATAM-based workshop, including the system's stakeholders, to elicit resilience requirements. To complement the workshop, we develop RESIRIO—a semi-automated, chatbot-assisted, and HAZOP-based approach—for elicitation. We evaluate RESIRIO through a user study. The requirements from both sources are specified using the ATAM scenario template. We use and extend Chaos Toolkit to transform and automate two scenarios. We quantitatively evaluate these scenarios and suggest resilience improvements based on resilience patterns. **Result.** We identify 12 resilience scenarios in the workshop. We share lessons learned from the study. In particular, our work provides evidence that an ATAM-based workshop is intuitive to stakeholders in an industrial setting and that stakeholders can quickly learn to use RESIRIO in order to successfully obtain new scenarios. **Conclusion.** Our approach helps requirements and quality engineers in interactive resilience requirements elicitation.

Sebastian Frank, Alireza Hakamian, Lion Wagner, Dominik Kesim, Christoph Zorn, J akim von Kistowski, and Andr  van Hoorn: Interactive Elicitation of Resilience Scenarios Based on Hazard Analysis Techniques. Selected and revised paper from 15th European Conference on Software Architecture (ECSA). Springer. 2022. Pre-print available at: <https://dqualizer.github.io/files/ECSA2022-InteractiveElicitation.pdf>

A Case Study on the Stability of Performance Tests for Serverless Applications

Context: While in serverless computing, application resource management and operational concerns are generally delegated to the cloud provider, ensuring that serverless applications meet their performance requirements is still a responsibility of the developers. Performance testing is a commonly used performance assessment practice; however, it traditionally requires visibility of the resource environment. **Objective:** In this study, we investigate whether performance tests of serverless applications are stable, that is, if their results are reproducible, and what implications the serverless paradigm has for perfor-

mance tests. **Method:** We conduct a case study where we collect two datasets of performance test results: (a) repetitions of performance tests for varying memory size and load intensities and (b) three repetitions of the same performance test every day for ten months. **Results:** We find that performance tests of serverless applications are comparatively stable if conducted on the same day. However, we also observe short-term performance variations and frequent long-term performance changes. **Conclusion:** Performance tests for serverless applications can be stable; however, the serverless model impacts the planning, execution, and analysis of performance tests.

Simon Eismann, Diego Elias Costa, Lizhi Liao, Cor-Paul Bezemer, Weiyi Shang, Andr  van Hoorn, Samuel Kounev: A case study on the stability of performance tests for serverless applications, *Journal of Systems and Software*, Volume 189, 2022.

The SPEC-RG Reference Architecture for the Compute Continuum

As the next generation of diverse workloads like augmented/virtual reality and autonomous driving evolves, computation is shifting from cloud-based services to the edge, leading to the emergence of a cloud-edge compute continuum. This continuum promises a wide spectrum of deployment opportunities for workloads that can leverage the strengths of cloud (scalable infrastructure, high reliability) and edge (energy efficient, low latencies). Despite its promises, the continuum has only been studied in silos of various computing models, thus lacking strong end-to-end theoretical and engineering foundations for computing and resource management across the continuum. Consequently, developers resort to ad hoc approaches to reason about performance and resource utilization of workloads in the continuum. In this work, we conduct a first-of-its-kind systematic study of various computing models, identify salient properties, and make a case to unify them under a compute continuum reference architecture. This architecture provides an end-to-end analysis framework for developers to reason about resource management, workload distribution, and performance analysis. We demonstrate the utility of the reference architecture by analyzing two popular continuum workloads, deep learning and industrial IoT. We have developed an accompanying open-source benchmarking framework and first-order analytical model for quantitative reasoning of continuum workloads.

Matthijs Jansen, Auday Al-Dulaimy, Alessandro V. Papadopoulos, Animesh Trivedi, Alexandru Iosup: The SPEC-RG Reference Architecture for the Compute Continuum. *IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, 2023.

The State of Serverless Applications: Collection, Characterization, and Community Consensus

Over the last five years, all major cloud platform providers have increased their serverless offerings. Many early adopters report significant benefits for serverless-based over traditional applications, and many companies are considering moving to serverless themselves. However, currently there exist only few, scattered, and sometimes even conflicting reports on when serverless applications are well suited and what the best practices for their implementation are. We address this problem in the present study about the state of serverless applications. We collect descriptions of 89 serverless applications from open-source projects, academic literature, industrial literature, and domain-specific feedback. We analyze 16 characteristics that describe why and when successful adopters are using serverless applications, and how they are building them. We further compare the results of our characterization study to 10 existing, mostly industrial, studies and datasets; this allows us to identify points of consensus across multiple studies, investigate points of disagreement, and overall confirm the validity of our results. The results of this study can help managers to decide if they should adopt serverless technology, engineers to learn about current practices of building serverless applications, and researchers and platform providers to better understand the current landscape of serverless applications.

Simon Eismann, Joel Scheuner, Erwin van Eyk, Maximilian Schwinger, Johannes Grohmann, Nikolas Herbst, Cristina Abad, Alexandru Iosup: The State of Serverless Applications: Collection, Characterization, and Community Consensus. *Transactions on Software Engineering*, 2021.

Meterstick: Benchmarking Performance Variability in Cloud and Self-hosted Minecraft-like Games

Due to increasing popularity and strict performance requirements, online games have become a cloud-based and self-hosted workload of interest for the performance engineering community. One of the most popular types of online games is the Minecraft-like Game (MLG), in which players can terraform the environment. The most popular MLG, Minecraft, provides not only entertainment, but also educational support and social interaction, to over 130 million people world-wide. MLGs currently support their many players by replicating isolated instances that support each only up to a few hundred players under favorable conditions. In practice, as we show here, the real upper limit of supported players can be much lower. In this work, we posit that performance variability is a key cause for the lack of scalability in MLGs, investigate experimentally causes of performance variability, and derive actionable insights. We propose an operational model for MLGs, which extends the state-of-the-art with essential aspects, e.g., through the consideration of environment-based workloads, which are sizable workload components that

do not depend on player input (once set in action). Starting from this model, we design the first bench-mark that focuses on MLG performance variability, defining specialized workloads, metrics, and processes. We conduct real-world benchmarking of MLGs, both cloud-based and self-hosted. We find environment-based workloads and cloud deployment are significant sources of performance variability: peak-latency degrades sharply to 20.7 times the arithmetic mean, and exceeds by a factor of 7.4 the performance requirements. We derive actionable insights for game-developers, game-operators, and other stakeholders to tame performance variability.

Jerrit Eickhoff, Jesse Donkervliet, Alexandru Iosup: Benchmarking Performance Variability in Cloud and Self-hosted Minecraft-like Games. *ACM/SPEC International Conference on Performance Engineering*, 2023.

(Encryption) time flies when you're having fun: the case of the exotic BlackCat ransomware

Time is critical for ransomware operators – the faster they encrypt the victim's files, the less likely they are to be detected in the process. Encryption can be a time-consuming process, and ransomware developers know this. That is why they get creative when programming encryption routines – the goal is to minimize the time spent on encryption and maximize the amount of encrypted file content. In this way, the greatest possible irretrievable damage is done in the shortest possible time. BlackCat is a new and very high-profile player in the current ransomware scene. The ALPHV threat group, which is behind the ransomware, provides the malware to affiliates in exchange for a share in the ransom payments. The way BlackCat performs encryption is highly customizable and ALPHV uses this as an advertising tool to attract affiliates. BlackCat operators can choose between six encryption modes and two encryption algorithms. Ransomware operators can further configure each encryption mode with mode-specific settings. Each encryption mode and algorithm occupies a specific position on the trade-off scale between encryption speed and completeness. We reverse-engineered the BlackCat ransomware to provide a first look into the inner workings of the encryption modes that BlackCat implements. Our analysis provides a unique insight into the design decisions that ransomware developers make to achieve an optimal balance between encryption speed and encryption completeness. This work also tests the encryption modes and encryption algorithms that BlackCat implements. We conducted a series of experiments to measure in numbers the trade-off between encryption speed and completeness that the different modes achieve. We examine metrics such as encryption speed, time spent on encryption, and amount of file content encrypted. For example, we observed differences in the time spent encrypting a file in the order

of minutes for different ransomware configuration points. Our measurements provide a hard look at the numbers – they show how much response time is available once a carefully configured BlackCat has started encrypting files.

Aleksandar Milenkoski: (Encryption) time flies when you're having fun: the case of the exotic BlackCat ransomware, presented at the 2022 Virus Bulletin (VB2022), 2022.

A Study on the Aging and Fault Tolerance of Microservices in Kubernetes

Microservice-based applications are increasingly being adopted along with cloud service models, and nowadays serve millions of customers daily. They are supported by container-based architectures which are managed by orchestration platforms, such as Kubernetes, that monitor, manage, and automate most of the tasks. Although these tools provide failover capabilities, it is not yet studied how effective they are in dealing with diverse types of faults. Fault injection is an effective methodology for validating components that are supposed to detect the malfunctions and report/correct them. This paper studies the effectiveness of Kubernetes in dealing with faults and aging in microservices, and on the possibility of using faults to accelerate aging effects for testing purposes. For this, we conducted an analysis of the implementation and tuning of Kubernetes probes, followed by experiments with varying load and fault injection into two distinct and representative microservice testbeds to analyze the capacity of probes in detecting issues in applications. The goal is to improve the knowledge of researchers and developers on whether Kubernetes can detect different faults and aging issues. Also, even though some services tend to accumulate aging effects, with increasing resource consumption, Kubernetes does not detect them nor acts on them, indicating that probes may be insufficient for aging scenarios. Results also showed that fault injection is useful to accelerate aging effects for the testing and evaluation purposes.

José Flora, Paulo Gonçalves, Miguel Teixeira, Nuno Antunes: A Study on the Aging and Fault Tolerance of Microservices in Kubernetes. *IEEE Access*, vol. 10, 2022.

A Systematic Approach for Benchmarking of Container Orchestration Frameworks

Container orchestration frameworks play a critical role in modern cloud computing paradigms such as cloud-native or serverless computing. They significantly impact the quality and cost of service deployment as they manage many performance-critical tasks such as container provisioning, scheduling, scaling, and networking. Consequently, a comprehensive performance assessment of container orchestration frameworks is essential. However, until now, there is no benchmarking approach that covers the many different tasks implemented in such platforms

and supports evaluating different technology stacks. In this paper, we present a systematic approach that enables benchmarking of container orchestrators. Based on a definition of container orchestration, we define the core requirements and benchmarking scope for such platforms. Each requirement is then linked to metrics and measurement methods, and a benchmark architecture is proposed. With COFFEE, we introduce a benchmarking tool supporting the definition of complex test campaigns for container orchestration frameworks. We demonstrate the potential of our approach with case studies of the frameworks Kubernetes and Nomad in a self-hosted environment and on the Google Cloud Platform. The presented case studies focus on container startup times, crash recovery, rolling updates, and more.

Martin Straesser, Jonas Mathiasch, André Bauer, Samuel Kounev: A Systematic Approach for Benchmarking of Container Orchestration Frameworks. *ACM/SPEC International Conference on Performance Engineering*, 2023.

An Empirical Study of Container Image Configurations and Their Impact on Start Times

A core selling point of application containers is their fast start times compared to other virtualization approaches like virtual machines. Predictable and fast container start times are crucial for improving and guaranteeing the performance of containerized cloud, serverless, and edge applications. While previous work has investigated container starts, there remains a lack of understanding of how start times may vary across container configurations. We address this shortcoming by presenting and analyzing a dataset of approximately 200,000 open-source Docker Hub images featuring different image configurations (e.g., image size and exposed ports). Leveraging this dataset, we investigate the start times of containers in two environments and identify the most influential features. Our experiments show that container start times can vary between hundreds of milliseconds and tens of seconds in the same environment. Moreover, we conclude that no single dominant configuration feature determines a container's start time, and hardware and software parameters must be considered together for an accurate assessment.

Martin Straesser, André Bauer, Robert Leppich, Nikolas Herbst, Kyle Chard, Ian Foster, Samuel Kounev: An Empirical Study on Container Image Configurations and Their Impact on Start Times. *IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, 2023.

Searching for the Ground Truth: Assessing the Similarity of Benchmarking Runs

Stable and repeatable measurements are essential for comparing the performance of different systems or applications, and benchmarks are used to ensure accuracy and replication. However, if the corresponding measure-

ments are not stable and repeatable, wrong conclusions can be drawn. To facilitate the task of determining whether the measurements are similar, we used a data set of 586 micro-benchmarks to (i) analyze the data set itself, (ii) examine our previous approach, and (iii) propose and evaluate a heuristic. To evaluate the different approaches, we perform a peer review to assess the dissimilarity of the benchmark runs. Our results show that this task is challenging even for humans and that our heuristic exhibits a sensitivity of 92%.

André Bauer, Martin Straesser, Mark Leznik, Lukas Beierlieb, Marius Hadry, Nathaniel Hudson, Kyle Chard, Samuel Kounev, and Ian Foster: Searching for the Ground Truth: Assessing the Similarity of Benchmarking Runs. Proceedings of the 14th ACM/SPEC International Conference on Performance Engineering, 2023.

Streaming vs. Functions: A Cost Perspective on Cloud Event Processing

In cloud event processing, data generated at the edge is processed in real-time by cloud resources. Both distributed stream processing (DSP) and Function-as-a-Service (FaaS) have been proposed to implement such event processing applications. FaaS emphasizes fast development and easy operation, while DSP emphasizes efficient handling of large data volumes. Despite their architectural differences, both can be used to model and implement loosely-coupled job graphs. In this paper, we consider the selection of FaaS and DSP from a cost perspective. We implement stateless and stateful workflows from the Theodolite benchmarking suite using cloud FaaS and DSP. In an extensive evaluation, we show how application type, cloud service provider, and runtime environment can influence the cost of application deployments and derive decision guidelines for cloud engineers. We find that in terms of pure costs for the cloud services, FaaS is superior for applications that are subject to small event rates and require small state. Once event rates increase and utilize at least a single DSP instance, DSP engines can be operated at lower costs. This observation holds independently of the cloud provider and implementation technology. However, when choosing among FaaS and streaming also “hidden” costs should be taken into account.

Tobias Pfandzelter, Sören Henning, Trever Schirmer, Wilhelm Hasselbring, David Bermbach: Streaming vs. Functions: A Cost Perspective on Cloud Event Processing. Proceedings of the IEEE International Conference on Cloud Engineering (IC2E), 2022.

Efficient Data Processing: Assessing the Performance of Different Programming Languages

This paper compares the performance of R, Python, and Rust in the context of data processing tasks. A real-world data processing task in form of an aggregation of benchmark measurement results were implemented in each language, and their execution times were measured. The results indicate that while all languages are capable of performing the tasks effectively, there are significant differences in performance. Even the same code showed significant runtime differences depending on the interpreter used for execution. Rust and Python were found to be the most efficient, with R requiring much longer execution times. Additionally, the paper discusses the potential implications of these findings for data scientists and developers when choosing a language for data processing projects.

Lukas Beierlieb, André Bauer, Robert Leppich, Lukas Iffländer, Samuel Kounev: Efficient Data Processing: Assessing the Performance of Different Programming Languages. Proceedings of the 14th ACM/SPEC International Conference on Performance Engineering, 2023.

Power to the Applications: The Vision of Continuous Decentralized Autoscaling

Autoscaling has been one of the most active research areas since the beginning of the cloud computing era. Nearly all previously proposed approaches focus on decision-making based on averaged monitoring values of many service instances at fixed points in time. This limits responsiveness and can lead to service level objective (SLO) violations when the load suddenly increases. Our vision of continuous decentralized autoscaling avoids these issues by giving individual service instances the power to make scaling decisions in a distributed fashion. Each instance performs self-monitoring and evaluates its state. The service instance initiates upscaling if it detects an overload or downscaling if its load is below a specified threshold. By randomly determining scaling timing, we achieve quasi-continuous scaling behavior when multiple service instances are deployed. We discuss challenges regarding analytical modeling, simulation, and real-world evaluation of this approach.

Martin Straesser, Stefan Geißler, Tobias Hoßfeld, Samuel Kounev: Power to the Applications: The Vision of Continuous Decentralized Autoscaling. Proceedings of the IEEE/ACM International Conference on Cluster, Cloud and Internet Computing Workshops (CCGRIDW), 2023.