Issue 13
May 2024

SPEC Research℠ Group Newsletter

CONTENTS

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

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2023

The award selection committee for 2023 selected the dissertation of Shreshth Tuli from the Imperial College London, UK. The selection committee was impressed by the high quality of scientific work conducted by Dr. Tuli that combined deep neural networks with digital twins for resource management in the Fog.

Read more on page 3

ICPE 2025 WILL BE HELD IN TORONTO

The ICPE organizing committee and local general chair Marin Litoiu invite interesting high-quality submissions for the next ACM/SPEC International Conference on Performance Engineering (ICPE 2025). The conference will take place on May 5-9, 2025 in Toronto, Canada.

Read more on page 4

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 2023. The Working Groups are always open for new members. Feel invited to join us!

Read more on pages 5-10

NEW SPEC RG ARTIFACTS

The newest SPEC RG artifacts let users (1) benchmark Linux I/O scheduler performance when using NVME devices and (2) evaluate the performance of data parsing libraries.

Read more on page 10

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, University of L'Aquila, Italy

Newsletter Editors:

Matthijs Jansen, VU Amsterdam, NL
Jesse Donkervliet, VU Amsterdam, NL

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:

- 14th ACM/SPEC ICPE 2023 in Coimbra, Portugal
- 4th IEEE International Conference on Autonomic Computing and Self-Organizing Systems ACSOS 2023 in Toronto, Canada
- 6th Workshop on Hot Topics in Cloud Computing Performance HotCloudPerf 2023 at ICPE 2023

We have been actively working on the preparation, planning, and organization of ICPE 2024, which will be held in London, UK. 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).
Matthijs Jansen (Newsletter Editor, VU Amsterdam).
Jesse Donkervliet (Newsletter Editor, VU Amsterdam).

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2023 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 *AI and Co-Simulation Driven Resource Management in Fog Computing Environments* was authored by Shreshth Tuli from Imperial College London, under the supervision of Professor Nicholas R. Jennings and Dr. Giuliano Casale. The selection committee was impressed with the proposal to deliver QoS management algorithms for Fog computing by coupling deep neural networks (DNN) models at runtime with a live digital twin (co-simulator) of the Fog infrastructure that acts as a data generator to fine-tune the DNNs whenever the predictions are observed to lower in confidence. The combination of DNN and co-simulator was shown in the thesis to be capable of tackling hard scheduling and fault-tolerance problems that arise in the Fog. The award is to be presented at the 15th ACM/SPEC International Conference on Performance Engineering (ICPE) scheduled to be held in London in May 2024.

Given the high quality of dissertations nominated for this award, the committee decided to publicly recognize another dissertation as Runner-Up, *Performance Engineering of Serverless Applications and Platforms* authored by Dr. Simon Eismann, of Julius-Maximilians-Universität Würzburg, under the supervision of Prof. Samuel Kounev.

The award selection committee for 2023 was chaired by Prof. Cristina L. Abad of ESPOL University in Ecuador.

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 that are defended **between October 2022 and September 2024** will be eligible to be nominated for the 2024 award (a thesis can be nominated only once).

https://research.spec.org/news/2024-01-19-00-00-winner-of-spec-kaivalya-dixit-distinguished-dissertation-award-2023/

ICPE 2024: STATISTICS

The 15th ACM/SPEC International Conference on Performance Engineering (ICPE 2024) will be held in South Kensington, London, from May 7 to May 11, 2024. This year, the research track of ICPE attracted 52 submissions, 18 of which were selected as full articles after a rigorous review process, yielding an acceptance rate of 34.6%. Of the 13 submissions to the Industry Track, 8 were accepted as full articles. In the emerging research track, 7 out of 13 submissions were accepted as emerging research papers. There were 16 submissions to the Artifact Track, where 3 of these submissions were standalone artifacts not linked to an ICPE paper; 5 were awarded all badges, 7 were awarded the available and functional badges, and 4 received the available badge. For the third time, ICPE features a Data Challenge with 5 accepted short papers out of 5 submissions. 5 tutorials have been proposed out of which 3 were selected to complement the program. Furthermore, 4 posters and 3 demonstrations will be presented in interactive sessions. For the first time, ICPE features a Journal First track, with 3 accepted papers that have already been published in selected journals and will be presented at the conference.

The following workshops are planned for ICPE 2024:

- The Second Workshop on Artificial Intelligence for Performance Modeling, Prediction, and Control (AIPerf 2024)
  https://ai-perf.github.io/AIPerf2024/
- The Fifth Workshop on Benchmarking in the Data Center: Expanding to the Cloud (BID 2024)
  https://parallel.computer/index.html
- The Second Workshop on Serverless, Extreme-Scale, and Sustainable Graph Processing Systems (GraphSys 2024)
  https://sites.google.com/view/graphsys24/home
- The Seventh Workshop on Hot Topics in Cloud Computing Performance (HotCloudPerf 2024)
  https://hotcloudperf.spec.org
- The Twelfth International Workshop on Load Testing and Benchmarking of Software Systems (LTB 2024)
  https://ltb2024.github.io/
- The First Workshop on Performance Optimization in the LLM World (POLLMW 2024)
  https://sites.google.com/view/pollmw
- The Ninth Workshop on Challenges in Performance Methods for Software Development (WOSP-C 2024)
  https://wosp-c.github.io/wosp-c-24/

The following distinguished speakers will give keynotes at ICPE 2024:

- Manzoor Mohammed (Capitas, UK)
  How the Cloud Made Performance a Priority on Board Agendas
- Jane Hillston (University of Edinburgh, UK)
  What Does Performance Mean for Large Language Models?
- Giuliano Casale (Imperial College London, UK)
  Optimizing Edge AI: Performance Engineering in Resource-Constrained Environments

Simonetta Balsamo (Ca' Foscari University of Venice, Italy),
William Knottenbelt (Imperial College London, UK),
Cristina L. Abad (Escuela Superior Politecnica del Litoral, Ecuador),
Weiyi Shang (University of Waterloo, Canada)

ICPE 2025 IN TORONTO — 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 2025 will be held in Canada, May 5-9, 2025. Toronto, the "unofficial capital of Canada," is the largest city of the country and a vivid and multicultural cosmopolitan center. Toronto is visited on average by 27.5 million people every year, who have many a good reason to do so. Its diverse districts reflect its unique character including such places as the Kensington Market, Chinatown, Koreatown, Little Italy, Greektown, Little Portugal, the Entertainment District, the Distillery District, or the CN Tower. As well as being an iconic hub of culture and tourism, Toronto sits in a prime location for adventures beyond its city limits. Accessible by both train and car, Ontario's Niagara Falls is a must see! Comprised of three waterfalls, the Horseshoe Falls, American Falls, and the Bridal Veil Fall, a trip to Niagara Falls is quintessential for Toronto visit.

The contact person for ICPE 2025 is Prof. Marin Litoiu, from York University, Toronto, Canada. Evgenia Smirni has accepted the role of general co-chair. The PC chairs and the organizing committee will be announced later.

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 Amazon Web Services, Charles University, Concordia University, Imperial College London, MongoDB, Polytechnique Montreal, University of Alberta, University of L'Aquila, University of Hamburg, University of South Carolina, University of Stuttgart, University of Waterloo, and University of Würzburg.

The whole group meets in online meetings that are held on a monthly basis. In addition to discussing organizational topics, these general meetings include a technical presentation by group members or invited guests. In total, nine group meetings were held in 2023, including the following talks as part of our monthly lecture series:

- "How Software Refactoring Impacts Execution Time" by Luca Traini (University of L'Aquila) in January.
- "Enhancing Performance Modeling of Serverless Functions via Static Analysis" by Runan Wang (Imperial College London) in February.
- "The Impact of Hyperparameter Tuning and Model Optimization on the Performance Properties of Deep Neural Networks" by Lizhi Liao (Concordia University) in April
- "Towards effective assessment of steady state performance in Java software: are we there yet?" by Luca Traini (University of L'Aquila) in May.
- "Views From the Trenches: Current Trends in Performance Engineering" by Alex Podelko (Amazon AWS) in June.
- "Impact of Architectural Smells on Software Performance: an Exploratory Study" by Catia Trubiani (Gran Sasso Science Institute) in July.
- "Adapting Performance Analytic Techniques in a Real-World Database-Centric System: An Industrial Experience Report" by Lizhi Liao (University of Waterloo) in September.
- "BlockCompass: A benchmarking platform for blockchain performance" by Mohammadreza Rasolroveicy (York University) in October.
- "End-To-End Performance Testing, Profiling, and Analysis at Redis" by Filipe Oliveira (Redis) in November.

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. The current work focuses on solving the challenges of detecting system-level performance regressions as early as during development. 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 research 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 papers, which are expected to be submitted in the first trimester of 2024.
3. Search-based software performance engineering: This subgroup focuses on developing new approaches and research directions on the multi-criteria optimization of performance-related quality attributes in software architectures. The group is now working on techniques to let the designer interact with search-based heuristics in order to steer the search process towards desirable areas of the solution space (preprint [1]; paper under revision).
4. 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. Joint efforts in 2023 included the extension of a resilience simulator to include Kubernetes code [2] and tool-supported approaches to comprehend, refine, and verify transient behav-

ior [3], [4]. All mentioned activities included contributions by jointly supervised student projects. In addition to the publications, the group gave one presentation [P1].

We are happy that a member of the group was awarded at the SPEC Annual Meeting 2023: Michele Tucci (University of L'Aquila) has received a SPEC Impact Award for his contributions to SPEC RG, SPEC RG DevOps, and the International Conference on Performance Engineering (ICPE). Regarding 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!

Congratulations also to our long-term member Simon Eismann (meanwhile at MongoDB), having been selected the Runner-Up for the 2023 Kaivalya Dixit Distinguished Dissertation Award for his work, "Performance Engineering of Serverless Applications and Platforms". Simon has worked on some of the dissertation contributions in the context of our group.

For more information about the DevOps Performance Working Group (including our mission, activities, meetings, and projects), please visit our web page [5]. 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 (University of Waterloo),
Heng Li (Polytechnique Montreal)

https://research.spec.org/working-groups/rg-devops/

[1] V. Cortellessa, J. Andres Diaz-Pace, D. Di Pompeo, *et al.*, "Introducing interactions in multi-objective optimization of software architectures," *CoRR*, vol. abs/2308.15084, 2023. DOI: 10.48550/ARXIV.2308.15084.

[2] M. Straesser, P. Haas, S. Frank, M. Alireza Hakamian, A. van Hoorn, and S. Kounev, "Kubernetes-in-the-loop: Enriching microservice simulation through authentic container orchestration," in *Performance Evaluation Methodologies and Tools - 16th EAI International Conference, VALUETOOLS 2023, Crete, Greece, September 6-7, 2023, Proceedings*, vol. 539, 2023, pp. 82–98. DOI: 10.1007/978-3-031-48885-6_6.

[3] S. Frank, M. A. Hakamian, D. Zahariev, and A. van Hoorn, "Verifying transient behavior specifications in chaos engineering using metric temporal logic and property specification patterns," in *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering, ICPE 2023, Coimbra, Portugal, April 15-19, 2023*, 2023, pp. 319–326. DOI: 10.1145/3578245.3584314.

[4] S. Frank, M. A. Hakamian, J. Brott, and A. van Hoorn, "TQPropRefiner: Interactive comprehension and refinement of specifications on transient software quality properties," in *QUALIFIER @ ECSA 2023*, 2023.

[5] https://research.spec.org/working-groups/rg-devops/.

[P1] S. Frank, J. Brott, A. Hakamian, and A. van Hoorn, "TQPropRefiner: How to interactively comprehend and refine specifications on transient software quality properties?" In *Symposium on Software Performance (SSP 2023)*.

REPORT: CLOUD WORKING GROUP

In 2023, 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 **investigating** the future of computing beyond non-von Neumann architectures, **characterizing** the performance of schedulers for serverless computing, and **implementing** energy-aware serverless graph-processing systems.

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 2023, 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. **Non-von Neumann Architectures Survey**: We published a survey titled "Beyond von Neumann in the computing continuum: Architectures, applications, and future directions" in IEEE Internet Computing [1]. 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.
2. **Serverless Scheduling**: Serverless data-processing systems have shown promising results for processing highly-varying workloads. Such systems can benefit from architectural features such as node-local caches, distributed schedulers, and work-stealing mechanisms. The performance-impact and interplay of these features is understudied. We implement all these features in the OpenDC simulator.

We then use the simulator to characterize the performance of these features [2]. We observe that workstealing improves performance in all cases and all decentralized task-placement algorithms have similar performance.

3. **Fine-grained Energy-usage Characterization of Scientific Computing**: Multiple scientific data centers have publicly released application resource-usage and energy-usage data at high-granularity (sampled every 30 seconds). We are collecting and using such data to determine the resource-usage difference between machine learning and traditional scientific computing applications. We are investigating if thermal performance of scientific computing clusters has an impact on application performance. We are also investigating if there are any links between application reliability, energy-usage, and thermal performance.
4. **Serverless Graph-processing**: Graph-processing is a challenging workload with irregular-memory access patterns. Graph-processing workload's resource bottleneck varies across storage, memory, networking, and compute based on the user query, the algorithm used, and the analyzed dataset. Such diverse resource requirements make serverless computing a good fit for this workload. We are investigating Basic Graph Operations (BGOs) as substrates for modeling graph applications. We are also investigating their use in simulation and their inclusion in serverless graph-processing systems.

Besides these focused activities, the Cloud WG has been acting successfully in organizing and growing the yearly workshop HotCloudPerf 2024 [3]. On May 11, the workshop will be held in London co-located with ICPE2024 and feature three invited keynotes, nine accepted paper presentations.

Keynote presentations:

- Josef Spillner (ZHAW Zurich, Switzerland): Upscaling messaging and stateful computation
- Robert Chatley (Imperial College London, UK): Continuous Developer Feedback for Cloud Native Systems
- Sashko Ristov (University of Innsbruck, Austria): Engineering serverless application life-cycles in federated serverless infrastructures

Paper presentations:

- Zebin Ren et. al., A Systematic Configuration Space Exploration of the Linux Kyber I/O Scheduler
- Dante Niewenhuis et. al., FootPrinter: Quantifying Data Center Carbon Footprint
- Dragi Kimovski, Hypergraphs: Facilitating High-Order Modeling of the Computing Continuum
- Runyu Jin et.al., Baking Disaster-Proof Kubernetes Applications with Efficient Recipes
- Chetan Phalak et. al., Towards Geo-Distributed Training of ML Models in a Multi-Cloud Environment
- Radu Apşan et. al., Towards a Workload Trace Archive for Metaverse Systems
- Ivo Rohwer et. al., Resource Demand Profiling of Monolithic Workflows
- Debajyoti Halder et. al., Empirical Evaluation of ML Models for Per-Job Power Prediction
- Simon Triendl and Sashko Ristov et. al., Peeking behind the serverless implementations and deployments of the Montage Workflow

To conclude, 2023 was a successful year for the RG Cloud Group. We concluded a lot of activities. We are looking forward to starting new ones and an even more successful 2024. For this, we are actively seeking new participants and activities. You can also join ongoing activities.

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

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

[1] S. Kounev, N. Herbst, C. L. Abad, *et al.*, "Serverless computing: What it is, and what it is not?" *Commun. ACM*, vol. 66, no. 9, pp. 80–92, 2023. DOI: 10.1145/3587249.
[2] S. Talluri, N. Herbst, C. Abad, T. De Matteis, and A. Iosup, "Exde: Design space exploration of scheduler architectures and mechanisms for serverless data-processing," *Future Generation Computer Systems*, vol. 153, pp. 84–96, 2024. DOI: https://doi.org/10.1016/j.future.2023.11.013.
[3] 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 2023, the Working Group had papers published at the ICPE [1], DSN [2], the IEEE Access Journal [3], and PRDC [4].

The paper entitled *Efficient Data Processing: Assessing the Performance of Different Programming Languages* [1] compared the performance of R, Python, and Rust in data processing tasks. The paper discusses the potential implications of these findings for data scientists and developers based on the findings from the benchmark.

The DSN paper proposes a novel concept and approach of intrusion injection for virtualized environments, which consists of directly driving the system into the erroneous states that mimic the ones resulting from actual intrusions (in the same way errors are injected to mimic the effects of residual faults) [2]. The work published in IEEE proposes a methodology that uses performance modeling for the detection of anomalies in virtualized environments that can be caused, for instance, by cyberattacks [3].

The paper published at PRDC introduced three data processing techniques that enable intrusion detection in scalable and elastic microservice applications utilizing CI/CD approaches [4]. The techniques manipulate data collected from active microservice replicas and feed it to algorithms, resulting in reliable intrusion detection even after scaling operations.

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

- identifying challenges in the area of evaluating robustness and performance aspects of security-relevant system components and security mechanisms;
- explore the methodology for injecting the effects of intrusions in the evaluation of virtualized environments;
- collaboration between Würzburg and Coimbra on injecting intrusions into hypervisors;
- propose a representative benchmarking approach to evaluate the effectiveness of intrusion detection models for the security microservice-based systems;
- propose methodologies and tools that increase the security of CI/CD pipelines for microservice applications;
- methodologies to allow understanding hypercall interfaces; and
- 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] L. Beierlieb, A. Bauer, R. Leppich, L. Iffländer, and S. Kounev, "Efficient data processing: Assessing the performance of different programming languages," in *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering, ICPE 2023, Coimbra, Portugal, April 15-19, 2023*, ACM, 2023, pp. 83–87. DOI: 10.1145/3578245.3584691.

[2] C. F. Gonçalves, N. Antunes, and M. Vieira, "Intrusion injection for virtualized systems: Concepts and approach," in *53rd Annual IEEE/IFIP International Conference on Dependable Systems and Network, DSN 2023, Porto, Portugal, June 27-30, 2023*, IEEE, 2023, pp. 417–430. DOI: 10.1109/DSN58367.2023.00047.

[3] C. F. Gonçalves, D. S. Menasché, A. Avritzer, N. Antunes, and M. Vieira, "Detecting anomalies through sequential performance analysis in virtualized environments," *IEEE Access*, vol. 11, pp. 70 716–70 740, 2023. DOI: 10.1109/ACCESS.2023.3293643.

[4] J. Flora, P. Gonçalves, and N. Antunes, "Intrusion detection for scalable and elastic microservice applications," in *28th IEEE Pacific Rim International Symposium on Dependable Computing, PRDC 2023, Singapore, October 24-27, 2023*, IEEE, 2023, pp. 39–45. DOI: 10.1109/PRDC59308.2023.00014.

REPORT: POWER WORKING GROUP

The SPEC Research Power Working Group (AMD, Ampere, Dell, HPE, IBM, IEIT, Intel, Microsoft, University of Würzburg; Chair: Maximilian Meissner, Vice-Chair: Klaus-Dieter Lange) 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 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: Klaus-Dieter Lange) is responsible for the establishment and development of standardized server benchmarks primarily for use in government regulations and programs. Current members and associates of the ISG are AMD, AWS, Ampere, Apple, Bull, Dell, Google, HPE, IBM, IEIT, Intel, Microsoft, NVIDIA, 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. The next major version, the SERT 3 suite, is currently under development. At the same time, the SPECpower Committee is eagerly working on the successor to the SPECpower_ssj® 2008 benchmark. Further advancing these industry-standard efficiency benchmarks is critical to account for new trends in hardware- and software-development, and to include feedback from government organizations and industry.

At ICPE 2023, members of the RG Power WG [3], the SPECpower Committee [4], and the SPEC ISG [5] presented a vision paper titled "*Challenges and Future Directions in Efficiency Benchmarking*" [6]. In this paper, we discuss these challenges and the work our group currently focuses on. For instance, the group currently works on expanding the SPEC Power and Performance Benchmark

Methodology [7] 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 generation of SPEC efficiency benchmarks, the evaluation of novel workloads is a major focus.

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] N. Schmitt, K. Lange, S. Sharma, A. Cragin, D. Reiner, and S. Kounev, "SPEC - spotlight on the international standards group (ISG)," in *ICPE '21: ACM/SPEC International Conference on Performance Engineering, Virtual Event, France, April 19-21, 2021, Companion Volume*, 2021, pp. 167–168. DOI: 10.1145/3447545.3451171.
[2] https://www.iso.org/standard/71926.html.
[3] https://research.spec.org/working-groups/rg-power/.
[4] https://www.spec.org/power/.
[5] https://www.spec.org/isg/.
[6] M. Meissner, K. Lange, S. Sharma, *et al.*, "Challenges and future directions in efficiency benchmarking (vision paper)," in *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering, ICPE 2023, Coimbra, Portugal, April 15-19, 2023*, 2023, pp. 51–55. DOI: 10.1145/3578245.3585034.
[7] https://www.spec.org/power_ssj2008/docs/SPECpower-Methodology.pdf.

REPORT: PREDICTIVE DATA ANALYTICS WORKING GROUP

The mission of the group is to bridge the missing links between the facets involved in data analytics, namely big data storage and provisioning, data versioning, and performance evaluation. 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.

In 2023, our group conducted a survey titled "Comprehensive Exploration of Synthetic Data Generation: A Survey" [1], aiming to serve as a guide for selecting synthetic data generation models and identifying areas for future exploration. We examined 417 synthetic data generation models from the past decade, offering a detailed overview of model types, functionalities, and advancements. Through this investigation, we identified common attributes, enabling a classification and trend analysis. Additionally, we have two journals currently under review, focusing on the categorization and evaluation of synthetic time series generation, respectively.

At ICPE 2024, we are set to present a paper titled "Unveiling Temporal Performance Deviation: Leveraging Clustering in Microservices Performance Analysis" [2] in the Data Challenge track. This paper introduces a methodology that reveals temporal performance variations in microservices by clustering containers based on their performance characteristics at different time intervals. By applying our methodology to the Alibaba dataset, we uncovered both stable and dynamically changing performance patterns.

Besides the above mentioned activities, 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, forecasting, and synthesis
7. Streamlining the data science process (DataOps)
8. Benchmarking of big data infrastructure

André Bauer, University of Chicago, USA,
Michael Stenger (University of Würzburg, Germany)
https://research.spec.org/working-groups/rg-predictive-data-analytics/

[1] A. Bauer, S. Trapp, M. Stenger, *et al.*, "Comprehensive exploration of synthetic data generation: A survey," *arXiv preprint arXiv:2401.02524*, Jan. 2024. DOI: https://doi.org/10.48550/arXiv.2401.02524.
[2] A. Bauer, T. Dittus, M. Straesser, *et al.*, "Unveiling Temporal Performance Deviation: Leveraging Clustering in Microservices Performance Analysis," in *Companion of the 15th ACM/SPEC International Conference on Performance Engineering (ICPE)*, ACM, May 2024.

MEMBER NEWS

André van Hoorn becomes Professor at the University of Hamburg

SPEC DevOps Performance Working Group Chair André van Hoorn was appointed professor in software engineering and construction methods at the University of Hamburg in December 2023.

André received his doctorate in trustworthy software systems from Kiel University. He was a postdoctoral researcher, senior researcher, and interim professor in the reliable software systems research group at the University of Stuttgart and a senior researcher in the software engineering and construction methods group at the University of Hamburg.

Lukas Iffländer becomes Professor at the University of Applied Sciences Dresden

SPEC Security Working Group member Lukas Iffländer was appointed professor in information security at the University of Applied Sciences Dresden (HTW) in February 2024. He teaches students cybersecurity and data protection, as well as the basics of computer science. Lukas received his doctorate from the Julius Maximilian University of Würzburg, where he wrote and successfully defended his doctoral thesis in adaptive safety mechanisms. Afterwards, he worked as a research consultant for the German Center for Rail Transport Research (DZSF).

NEW SPEC RG ARTIFACT: PERFORMANCE CHARACTERIZATION OF LINUX STORAGE SCHEDULERS IN THE NVME ERA

Flash SSDs have become the de-facto choice to deliver high I/O performance to modern data-intensive workloads. These workloads are often deployed in the cloud, where multiple tenants share access to flash-based SSDs. Cloud providers use various techniques, including I/O schedulers available in the Linux kernel, such as *BFQ*, *Multiqueue-Deadline* (*MQ-Deadline*), and *Kyber*, to ensure certain performance qualities (i.e., service-level agreements, SLAs).

We have shown that the Linux I/O schedulers lead to performance degradation when the CPU is the bottleneck [1]. This performance degradation varies with the relative performance of the CPUs, the storage devices, and workload characterizations, making it non-trivial to select the best-fit I/O scheduler. Thus, we release the benchmarking scripts used in our performance characterization [2].

This artifact identifies when and what becomes the bottleneck for a specific hardware configuration. Specifically, the artifact can be used to (1) find what is the bottleneck and when the bottleneck occurs with an increasing number of latency-sensitive applications or throughput-bound applications, (2) measure the lock contention introduced by the Linux I/O scheduler with a specified hardware and workload setting, and (3) identify the performance guarantees that the I/O schedulers provide when a foreground workload is running with background-interfering workloads.

Overall, this artifact helps users select the best-fit I/O scheduler for a hardware and workload configuration.

[1] Z. Ren, K. Doekemeijer, N. Tehrany, and A. Trivedi, "Bfq, multiqueue-deadline, or kyber? performance characterization of linux storage schedulers in the nvme era," 2024.

[2] Z. Ren, *icpe24_io_scheduler_study_artifact*, https://github.com/stonet-research/icpe24_io_scheduler_study_artifact, [Online; accessed 17. Apr. 2024], Apr. 2024.

NEW SPEC RG ARTIFACT: FRAMEWORK FOR ASSESSING DATA PROCESSING PERFORMANCE

Data handling (reading, aggregating, etc.) is crucial in any domain. Consequently, inefficient processing can waste considerable time, which is especially problematic when lots of data have to be handled or when it is desired that data streams have to be processed continuously. The choice of the programming language as well as the compiler or interpreter can have considerable impact on processing runtimes, and our artifact aims at quantifying these differences.

The aggregation of the dataset of the 2023 ICPE Data Challenge has been selected as a representative task for data processing. There, 65GB of JSON files containing raw benchmarking measurements are aggregated by calculating averages of measurement batches. Current implementations of this task exist in Rust, Python, and R. For control over the execution environment and reproducibility, a docker container is built for each variant. With this approach, the framework has been successfully used to not only compare Rust, Python, and R variants but also the performance differences between different CPython and PyPy interpreters as well as the effect of using more threads for parallel processing with the Rust implementation [1]. The project is available at [2].

[1] L. Beierlieb, A. Bauer, R. Leppich, L. Iffländer, and S. Kounev, "Efficient data processing: Assessing the performance of different programming languages," in *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering, ICPE 2023, Coimbra, Portugal, April 15-19, 2023*, ACM, 2023, pp. 83–87. DOI: 10.1145/3578245.3584691.

[2] L. Beierlieb, *icpe23data_challenge*, https://github.com/lbeierlieb/icpe23data_challenge, [Online; accessed 17. Apr. 2024], Apr. 2024.

SELECTED ABSTRACTS

Kubernetes-in-the-Loop: Enriching Microservice Simulation Through Authentic Container Orchestration

Microservices deployed and managed by container orchestration frameworks like Kubernetes are the bases of modern cloud applications. In microservice performance modeling and prediction, simulations provide a lightweight alternative to experimental analysis, which requires dedicated infrastructure and a laborious setup. However, existing simulators cannot run realistic scenarios, as performance-critical orchestration mechanisms

(like scheduling or autoscaling) are manually modeled and can consequently not be represented in their full complexity and configuration space. This work combines a state-of-the-art simulation for microservice performance with Kubernetes container orchestration. Hereby, we include the original implementation of Kubernetes artifacts enabling realistic scenarios and testing of orchestration policies with low overhead. In two experiments with Kubernetes' kube-scheduler and cluster-autoscaler, we demonstrate that our framework can correctly handle different configurations of these orchestration mechanisms boosting both the simulation's use cases and authenticity.

M. Straesser, P. Haas, S. Frank, M. Alireza Hakamian, A. van Hoorn, and S. Kounev, "Kubernetes-in-the-loop: Enriching microservice simulation through authentic container orchestration," in *Performance Evaluation Methodologies and Tools - 16th EAI International Conference, VALUETOOLS 2023, Crete, Greece, September 6-7, 2023, Proceedings*, vol. 539, 2023, pp. 82–98. DOI: 10.1007/978-3-031-48885-6_6.

Serverless Computing: What It Is, and What It Is Not?

Full automation of IT infrastructure and the delivery of efficient IT operations as billed services have been long-standing goals of the computing industry since at least the 1960s. A newcomer—serverless computing—emerged in the late 2010s with characteristics claimed to be different from those of established IT services, including Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) clouds. Even though serverless computing has gained significant attention in industry and academia over the past five years, there is still no consensus about its unique distinguishing characteristics and precise understanding of how these characteristics differ from classical cloud computing. What is serverless computing, and what are its implications?

S. Kounev, N. Herbst, C. L. Abad, *et al.*, "Serverless computing: What it is, and what it is not?" *Commun. ACM*, vol. 66, no. 9, pp. 80–92, 2023. DOI: 10.1145/3587249.

ExDe: Design space exploration of scheduler architectures and mechanisms for serverless data-processing

Serverless computing is increasingly used for data-processing applications in both science and business domains. At the core of serverless data-processing systems is the scheduler, which ensures dynamic decisions about task and data placement. Due to the variety of user, cluster, and workload properties, the design space for high-performance and cost-effective scheduling architectures and mechanisms is vast. The large design space is difficult to explore and characterize. To help the system designer disentangle this complexity, we present ExDe, a framework to systematically explore the design space of scheduling architectures and mechanisms. The framework includes a conceptual model and a simulator to assist in design space exploration. We use the framework, and real-world workloads, to characterize the performance of three scheduling architectures and two mechanisms. Our framework is open-source software available on Zenodo.

S. Talluri, N. Herbst, C. Abad, T. De Matteis, and A. Iosup, "Exde: Design space exploration of scheduler architectures and mechanisms for serverless data-processing," *Future Generation Computer Systems*, vol. 153, pp. 84–96, 2024. DOI: https://doi.org/10.1016/j.future.2023.11.013.

BFQ, Multiqueue-Deadline, or Kyber? Performance Characterization of Linux Storage Schedulers in the NVMe Era

Flash SSDs have become the de-facto choice to deliver high I/O performance to modern data-intensive workloads. These workloads are often deployed in the cloud, where multiple tenants share access to flash-based SSDs. Cloud providers use various techniques, including I/O schedulers available in the Linux kernel, such as BFQ, Multiqueue-Deadline (MQ-Deadline), and Kyber, to ensure certain performance qualities (i.e., service-level agreements, SLAs). Though designed for fast NVMe SSDs, there has not been a systematic study of these schedulers for modern, high-performance SSDs with their unique challenges. In this paper. we systematically characterize the performance, overheads, and scalability properties of Linux storage schedulers on NVMe SSDs with millions of I/O operations/s. We report 23 observations and 5 key findings that indicate that (i) CPU performance is the primary bottleneck with the Linux storage stack with high-performance NVMe SSDs; (ii) Linux I/O schedulers can introduce 63.4% performance overheads with NVMe SSDs; (iii) Kyber and BFQ can deliver 99.3% lower P99 latency than None or MQ-Deadline schedulers in the presence of multiple interfering workloads. We open-source the scripts and datasets of this work at https://zenodo.org/records/10599514.

Z. Ren, K. Doekemeijer, N. Tehrany, and A. Trivedi, "Bfq, multiqueue-deadline, or kyber? performance characterization of linux storage schedulers in the nvme era," 2024.

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 the form of an aggregation of benchmark measurement results was implemented in each language, and their execution times were measured. The results indicate that while all languages can perform 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 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.

L. Beierlieb, A. Bauer, R. Leppich, L. Iffländer, and S. Kounev, "Efficient data processing: Assessing the performance of different programming languages," in *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering, ICPE 2023, Coimbra, Portugal, April 15-19, 2023*, ACM, 2023, pp. 83–87. DOI: 10.1145/3578245.3584691.

Intrusion Injection for Virtualized Systems: Concepts and Approach

Virtualization is drawing attention due to countless benefits, leaving Hypervisors with the paramount responsibility for performance, dependability, and security. However, while there are consolidated approaches to assessing the performance and dependability of virtualized systems, solutions to assess security are very limited. Key difficulties are evaluating the system in the presence of unknown attacks and vulnerabilities and comparing the security attributes of different systems and configurations when an intrusion occurs. In this paper, we propose a novel concept and approach of intrusion injection for virtualized environments, which consists of directly driving the system into the erroneous states that mimic the ones resulting from actual intrusions (in the same way errors are injected to mimic the effects of residual faults). We present a prototype capable of injecting erroneous states related to memory-corruption in the Xen Hypervisor to show that the concept and approach proposed here are feasible. The prototype is evaluated using publicly disclosed exploits across three different versions of Xen. Results show that our tool can inject erroneous states equivalent to those resulting from attacks that exploit existing vulnerabilities, even on versions where those vulnerabilities do not exist.

C. F. Gonçalves, N. Antunes, and M. Vieira, "Intrusion injection for virtualized systems: Concepts and approach," in *53rd Annual IEEE/IFIP International Conference on Dependable Systems and Network, DSN 2023, Porto, Portugal, June 27-30, 2023*, IEEE, 2023, pp. 417–430. DOI: 10.1109/DSN58367.2023.00047.

Intrusion Detection for Scalable and Elastic Microservice Applications

The growing complexity and dynamicity of microservices, combined with their ability to scale, present significant challenges to security monitoring tools. Integrating these tools into a DevSecOps pipeline is currently impractical, necessitating research into adaptive intrusion detection approaches. This paper introduces three data processing techniques that enable intrusion detection in scalable and elastic microservice applications utilizing CI/CD approaches. These techniques manipulate data collected from active microservice replicas and feed it to algorithms, resulting in reliable intrusion detection even after scaling operations. To evaluate these techniques, we integrate them into a state-of-the-art intrusion detection tool developed for microservice environments. Their effectiveness is evaluated using two lightweight algorithms (STIDE and BoSC) with representative workloads, attacks, and a microservice-based application, demonstrating their ability to detect most attacks, even in scenarios involving multiple replicas.

J. Flora, P. Gonçalves, and N. Antunes, "Intrusion detection for scalable and elastic microservice applications," in *28th IEEE Pacific Rim International Symposium on Dependable Computing, Singapore, October 24-27, 2023*, 2023, pp. 39–45. DOI: 10.1109/PRDC59308.2023.00014.

Challenges and Future Directions in Efficiency Benchmarking (Vision Paper)

SPEC benchmarks are crucial contributors behind the improvement of server efficiency since 2007, given their role in making the power consumption and efficiency of servers transparent for government regulators, customers, and the manufacturers themselves. As the IT landscape experiences radical transformations, efficiency benchmarks need to be updated accordingly to generate results relevant to government regulators, manufacturers, and customers. In this paper, we outline current challenges efficiency benchmark developers are tackling and highlight recent technological developments the next generation of efficiency benchmarks should take into account.

M. Meissner, K. Lange, S. Sharma, *et al.*, "Challenges and future directions in efficiency benchmarking (vision paper)," in *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering, ICPE 2023, Coimbra, Portugal, April 15-19, 2023*, 2023, pp. 51–55. DOI: 10.1145/3578245.3585034.

Comprehensive Exploration of Synthetic Data Generation: A Survey

Recent years have witnessed a surge in the popularity of Machine Learning (ML), applied across diverse domains. However, progress is impeded by the scarcity of training data due to expensive acquisition and privacy legislation. Synthetic data emerges as a solution, but the abundance of released models and limited overview literature pose challenges for decision-making. This work surveys 417 Synthetic Data Generation (SDG) models over the last decade, providing a comprehensive overview of model types, functionality, and improvements. Common attributes are identified, leading to a classification and trend analysis. The findings reveal increased model performance and complexity, with neural network-based approaches prevailing, except for privacy-preserving data generation. Computer vision dominates, with GANs as primary generative models, while diffusion models, transformers, and RNNs compete. Implications from our performance evaluation highlight the scarcity of common metrics and datasets, making comparisons challenging. Additionally, the neglect of training and computational costs in literature necessitates attention in future research. This work serves as a guide for SDG model selection and identifies crucial areas for future exploration.

A. Bauer, S. Trapp, M. Stenger, *et al.*, "Comprehensive exploration of synthetic data generation: A survey," *arXiv preprint arXiv:2401.02524*, Jan. 2024. DOI: https://doi.org/10.48550/arXiv.2401.02524.