Issue 14
May 2025

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 2024

The award selection committee for 2024 selected the dissertation of Jingzhi Gong from Loughborough University, UK. The selection committee particularly appreciated this thesis for its innovative approach to optimizing software performance using deep learning. Its contributions offer valuable solutions for improving software maintenance and performance optimization in complex modern systems, demonstrating a substantial advancement in the field.

Read more on page 3

ICPE 2026 WILL BE HELD IN FLORENCE

The ICPE organizing committee and local general chairs Roberto Verdecchia and Enrico Vicario invite interesting high-quality submissions for the next ACM/SPEC International Conference on Performance Engineering (ICPE 2026). The conference will take place on May 4-8, 2026 in Florence, Italy.

Read more on page 5

SPEC RESEARCH WORKING GROUPS REPORT ON THEIR PROGRESS

The SPEC Research Working Groups DevOps Performance, Cloud, Security, Power, and Predictive Data Analytics report on their progress, articles, benchmarks, and technical reports published in 2024. The Working Groups are always open for new members. Feel invited to join us!

Read more on pages 5-11

NEW SPEC RG TOOLS AND ARTIFACTS

The newest SPEC RG tools and artifacts let users benchmark heterogeneous multi-kernel systems and trace and measure the performance and resource utilization of software systems.

Read more on pages 13-14

SPEC RESEARCH GROUP OFFICERS

Chair:

Samuel Kounev, University of Würzburg, Germany

Vice-Chair:

Cristina Abad, ESPOL University, Ecuador

Secretary:

Diego Elias Costa, Concordia University, Canada

Steering Committee:

Cristina Abad, ESPOL University, Ecuador
Diego Elias Costa, Concordia University, 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, University of Waterloo, Canada
Michele Tucci, University of L'Aquila, Italy
Petr Tůma, Charles Univ. of Prague, Czech Republic

Publicity Officer:

Weiyi Shang, University of Waterloo, Canada

Release Manager:

Michele Tucci, University of L'Aquila, Italy

Newsletter Editor:

Matthijs Jansen, VU Amsterdam, The Netherlands
Jesse Donkervliet, VU Amsterdam, The Netherlands

https://research.spec.org

SPEC RESEARCH WORKING GROUPS

DevOps Performance Working Group

Chair:

Heng Li, Polytechnique Montréal, Canada

Vice-Chair:

Diego Elias Costa, Concordia University, Canada
Michele Tucci, University of L'Aquila, Italy

Secretary:

Lizhi Liao, Memorial University of Newfoundland, Canada

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

Thomas Prantl, University of Würzburg, Germany

Vice-Chair:

Aleksandar Milenkoski, SentinelLabs/Univ. Würzburg, Germany

Secretary:

Lukas Horn, University of Würzburg, Germany

Release Manager:

Simon Engel, University of Würzburg, Germany

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

Predictive Data Analytics Working Group

Chair:

André Bauer, Illinois Institute of Technology, USA

Vice-Chair:

Robert Leppich, University of Würzburg, Germany

Secretary:

Veronika Lesch, DHBW Mosbach, Germany

Release Manager:

Michael Stenger, University of Würzburg, Germany
Daniel Grillmeier, University of Würzburg, Germany

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

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:

- 15th ACM/SPEC ICPE 2024 in London, UK
- 5th IEEE International Conference on Autonomic Computing and Self-Organizing Systems ACSOS 2024 in Aarhus, Denmark
- 7th Workshop on Hot Topics in Cloud Computing Performance HotCloudPerf 2024 at ICPE 2024

We have been actively working on the preparation, planning, and organization of ICPE 2025, which will be held in Toronto, Canada. 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 2024 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 *Pushing the Boundary: Specialising Deep Configuration Performance Learning* was authored by Jingzhi Gong from Loughborough University (UK) under the supervision of Dr. Tao Chen. The selection committee particularly appreciated this thesis for its innovative approach to optimizing software performance using deep learning. The research introduces robust frameworks that significantly enhance prediction accuracy even in dynamic environments by identifying critical challenges such as selecting effective encoding methods and managing sparse data. These contributions offer valuable solutions for improving software maintenance and performance optimization in complex modern systems, demonstrating a substantial advancement in the field. The award will be presented at the 16th ACM/SPEC International Conference on Performance Engineering (ICPE), scheduled in Toronto (Canada) in May 2025.

Given the high quality of the dissertations nominated for this award, the committee decided to publicly recognize two other dissertations as Runner-Ups: *Serverless Control Planes for Orchestration of Cloud Resources* by Dr. Alexander Joseph Fuerst of Indiana University Bloomington (USA), under the supervision of Prof. Prateek Sharma, and *Robust System Software for Quantum Computing* by Dr. Tirthak Patel of Northeastern University (USA), under the supervision of Prof. Devesh Tiwari.

The award selection committee for 2024 was chaired by Prof. Alessandro V. Papadopoulos of Mälardalen University (Sweden).

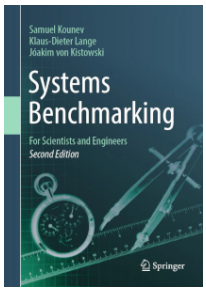
The SPEC Kaivalya Dixit Distinguished Dissertation Award was established in 2011 to recognize outstanding dissertations within the scope of the SPEC Research Group. Contributions of interest span the design of metrics for system evaluation and the development of methodologies, techniques, and tools for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

Dissertations defended **between October 2023 and September 2025** will be eligible to be nominated for the 2025 award (a thesis can be nominated only once).

https://research.spec.org/news/2025-01-31-00-00-winner-of-spec-kaivalya-dixit-distinguished-dissertation-award-2024/

2ND EDITION OF "SYSTEMS BENCHMARKING" TEXTBOOK

The second edition of a textbook on the topic of systems benchmarking written by SPEC RG Members was published by Springer.



The second edition of the book *"Systems Benchmarking - For Scientists and Engineers"* by Sam Kounev, Klaus-Dieter Lange, and Jóakim von Kistowski was just published by Springer. The first edition published in 2020 was received very well by the community and has now been established as both a textbook and a practical guide covering the state-of-the-art and future directions in the field of systems benchmarking. Together with the supplementary materials, the book is being adopted for teaching benchmarking in a growing number of graduate and postgraduate courses at universities around the world. It provides theoretical and practical foundations as well as an in-depth exploration of modern benchmarks and benchmark development.

Besides a number of updates in almost all chapters, for this new edition three chapters are added in Part II of the book: (1) "Machine Learning and Artificial Intelligence" to cater the growing need to evaluate and benchmark ML and AI systems, (2) "Scalability of Networks and Systems" focusing on novel metrics and techniques to evaluate scalability, and (3) "PC, Workstation, Graphics, and Network Benchmarks" covering popular benchmarks like SYSmark, PCMark, Phoronix Test Suite, 3DMark, the Blender benchmark, and end-to-end network performance tools.

For more information about the new edition, including a foreword by Ian T. Foster (Argonne National Laboratory and University of Chicago), you can check the book web site at: https://www.benchmarking-book.com/

The authors would like to thank the following SPEC members who were involved as chapter co-authors in Part II ("Applications") of the book: Jeremy Arnold (AMD), André Bauer (IIT Chicago), John Beckett (AMD), James Bucek (HP Inc.), Ken Cantrell (NetApp, Inc.), Don Capps (Netapp), Alexander Carlton (AMD), Simon Eismann (MongoDB), Sorin Faibish (DellEMC), Johannes Grohmann (Meta), Nikolas Herbst (Uni-Würzburg), Tobias Hossfeld (Uni-Würzburg), Rouven Krebs (SAP SE), Yannik Lubas (Uni-Würzburg), Mary Anne Marquez (IBM), Maximilian Meissner (Uni-Würzburg), Aleksandar Milenkoski (SentinelOne), David Morse (Broadcom), Nicholas A. Principe (iXsystems, Inc.), David Schmidt (Red Hat Inc.), Norbert Schmitt (IBM), Van Smith (AMD), Simon Spinner (IBM), Sitsofe Wheeler (Ocean Blue Software),and Marwin Züfle (Uni-Würzburg).

ICPE 2025: STATISTICS

The 16th ACM/SPEC International Conference on Performance Engineering (ICPE 2025) will be held in Toronto, Canada, from May 5 to May 9, 2025. This year, the research track of ICPE attracted 49 submissions, of which 17 were selected as full articles and 7 as short articles after a rigorous review process, yielding an acceptance rate of 34.69%. Of the 11 submissions to the Industry Track, 8 were accepted as full articles. Of the 6 submissions to the Industry Presentation Track, 5 were accepted as presentations. In the emerging research track, 3 out of 5 submissions were accepted as emerging research papers. There were 18 submissions to the Artifact Track, where 4 of these submissions were standalone artifacts not linked to an ICPE paper; 17 were awarded the functional badge. In addition, 13 were awarded the functional badge and 6 with the results reproduced badge. Even 2 artifact reusable badges were awarded and 2 artifacts invited to the SPEC Research Tool Repository. For the fifth time, ICPE features a Data Challenge with 2 accepted short papers out of 2 submissions. 5 tutorials were proposed, out of which 4 were accepted, and ultimately, 2 were included in the program due to travel issues. Furthermore, 2 posters will be presented in interactive sessions. For the second 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 2025:

- The Third International Workshop on AI Performance and Optimization in the LLM World (AIPerf 2025) https://sites.google.com/view/aipreflm2025
- The Sixth Workshop on Benchmarking in the Data Center: Expanding to the Cloud (BID 2025) https://parallel.computer/
- Fifth Workshop on Education and Practice of Performance Engineering (WEPPE 2025) https://esulabsolutions.godaddysites.com/sponsored-events
- The Eight Workshop on Hot Topics in Cloud Computing Performance (HotCloudPerf 2025) https://hotcloudperf.spec.org/

- The Thirteenth International Workshop on Load Testing and Benchmarking of Software Systems (LTB 2025)
  https://ltb2025.github.io/
- The Tenth Workshop on Challenges in Performance Methods for Software Development (WOSP-C 2025)
  https://wosp-c.github.io/wosp-c-25/

The following distinguished speakers will give keynotes at ICPE 2025:

- Ahmed E. Hassan (Queen's University, Canada)
  Software Performance Engineering for Foundation Model-Powered Software (FMware)
- Marc Brooker (AWS, USA)
  Great Performance for Bad Days
- Lizy Kurian John (UT Austin, USA)
  AI for Performance Engineering and Performance Engineering for AI

Marin Litoiu (York University, Canada),
Evgenia Smirni (College of William and Mary, USA),
Katinka Wolter (Freie Universitat Berlin, Germany),
Alessandro Papadopoulos (Mälardalen University, Sweden)

ICPE 2026 IN FLORENCE; ICPE 2027 IN VIENNA — 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 2026 will be held in Florence, May 4-8, 2026. Florence is a relatively small city in central Italy (around 400,000 inhabitants) with an extraordinarily rich artistic, architectural, and cultural heritage, attracting about 10 million tourists every year. Renowned as the cradle of the Renaissance, it is home to iconic landmarks such as the Cathedral of Santa Maria del Fiore (the Duomo), the Uffizi Gallery, Ponte Vecchio, and Palazzo Vecchio. The city is surrounded by the gentle hills of the Tuscan countryside, known for its vineyards, olive groves, and picturesque medieval towns.

The contact person for ICPE 2026 is Roberto Verdecchia, from the University of Florence, Florence, Italy. Enrico Vicario has accepted the role of general co-chair. The PC chairs and organizing committee will be announced later.

ICPE 2027 will take place in Vienna, Austria, with Ivona Brandic serving as local general-chair.

REPORT: DEVOPS PERFORMANCE WORKING GROUP

DevOps has become one of the main principles 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 runs 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 and 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, University of Würzburg, Memorial University of Newfoundland, and Wuhan University.

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 invited guests. In total, eight group meetings were held in 2024, including the following ten talks as part of our monthly lecture series:

- "Bridging Local Performance Testing and Architectural Models: A Performance Regression Detection Approach" by Lizhi Liao (Memorial University of Newfoundland) in January.
- "Introducing Interactions in Multi-Objective Optimization of Software Architectures" by Michele Tucci (University of L'Aquila) in January.
- "Resilience engineering for cloud-native applications" by Alireza Hakamian (University of Hamburg) in February.
- "Performance Time Series: Characterization and Efficiency of Change Point Detection Methods" by Diego Elias Costa (Concordia University) in February.
- "Mitigating Operating System Noise for Enhanced Application Performance" by Naser Ezzati-Jivan (Brock University) in March.

- "Time Series Forecasting of Runtime Software Metrics: An Empirical Study" by Federico Di Menna (University of L'Aquila) in April.
- "Reducing the Length of Field-Replay Based Load Testing" by Yuanjie Xia (University of Waterloo) in June.
- "Innovative Approaches for Anomaly Detection and Root Cause Analysis in AIOps" by Mahsa Panahandeh (University of Alberta) in September.
- "Understanding Web Application Workloads and Their Applications: Systematic Literature Review and Characterization" by Roozbeh Aghili (Polytechnique Montréal) in October.
- "BFQ, Multiqueue-Deadline, or Kyber? Performance Characterization of Linux Storage Schedulers in the NVMe Era" by Zebin Ren (Vrije Universiteit Amsterdam) 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. Current work [1] aims to solve the challenges of detecting system-level performance regressions as early as during the development phase. 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 subgroup 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 working on a paper expected to be submitted to ICSE 2026 (due March 2025).
3. **Search-based software performance engineering**. This subgroup focuses on developing novel approaches and research directions on the multi-criteria optimization of performance-related quality attributes in software architectures. The current work aims to enable the designer to interact with search-based heuristics in order to steer the search process towards desirable areas of the solution space [2].
4. **Resilience engineering for cloud-native applications**. This subgroup investigates novel approaches, tools, and data sets for resilience engineering, including chaos engineering, architecture extraction, runtime monitoring, interactive resilience scenario improvement, and other techniques. In 2024, the group published a tool named "DiSpel Cockpit" that supports continuous and iterative specification, verification, and refinement of resilience scenarios [3].
5. **Configuration-aware performance modeling**. This is a new subgroup aiming to investigate performance modeling methods for highly configurable systems. The current work is on studying the relationship between software workloads and configurations in order to develop a novel technique for comprehensively modeling the performance of highly configurable systems.
6. **Energy-aware DevOps**. This is another new subgroup aiming to investigate challenges in energy-aware DevOps practices. The work will focus on code-level energy measurement and energy awareness in the SDLC.

For more information about the DevOps Performance Working Group (including our mission, activities, meetings, presentations, and projects), please visit our web page at https://research.spec.org/working-groups/rg-devops/.

If you are interested in following the discussions or contributing actively, please contact the working group chairs.

Heng Li (Polytechnique Montreal),
Diego Elias Costa (Concordia University),
Michele Tucci (University of L'Aquila),
Lizhi Liao (Memorial University of Newfoundland)

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

[1] L. Liao, S. Eismann, H. Li, *et al.*, "Early detection of performance regressions by bridging local performance data and architectural models," *CoRR*, vol. abs/2408.08148, 2024. DOI: 10.48550/ARXIV.2408.08148. arXiv: 2408.08148. [Online]. Available: https://doi.org/10.48550/arXiv.2408.08148.

[2] V. Cortellessa, J. A. Diaz-Pace, D. D. Pompeo, *et al.*, "Introducing interactions in multi-objective optimization of software architectures," *CoRR*, vol. abs/2308.15084, 2023. DOI: 10.48550/ARXIV.2308.15084. arXiv: 2308.15084. [Online]. Available: https://doi.org/10.48550/arXiv.2308.15084.

[3] S. Frank, A. E. Tefur, M. A. Hakamian, and A. van Hoorn, "Dispel cockpit: Specification, verification, and refinement of resilience scenarios," in *Software Architecture. ECSA 2024 Tracks and Workshops - Luxembourg City, Luxembourg, September 3-6, 2024, Proceedings*, A. Ampatzoglou, J. Pérez, B. Buhnova, *et al.*, Eds., ser. Lecture Notes in Computer Science, vol. 14937, Springer, 2024, pp. 3–11. DOI: 10.1007/978-3-031-71246-3_1. [Online]. Available: https://doi.org/10.1007/978-3-031-71246-3%5C_1.

REPORT: CLOUD WORKING GROUP

In 2024, 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 to both academia and industry. This year, we have focused on **surveying** the computing continuum from the cloud through the edge to IoT, **characterizing** the overhead of distributed tracing systems, and **implementing** systems for large-scale geospatial applications.

The group's **scope** is to develop new methodological elements for gaining a deeper understanding of performance and operations across the compute continuum, from the cloud to the user. We develop quantitative benchmarks and performance measurement tools, enabling novel large-scale applications to be developed. 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 2024, 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. **Compute Continuum**: We published a survey titled "The computing continuum: From IoT to the cloud" in Elsevier Internet of Things [1]. The survey discusses computing models beyond the cloud that enable new use cases. The survey summarizes the development of computing models from the grid through the cloud to the plethora of current models, such as edge and fog. The survey presents two reference architectures for the new computing models. The survey analyzes use cases of these new architectures in the science, industrial, healthcare, and cloud gaming domains.
2. **Distributed Tracing**: Cloud applications are composed of multiple services. They use architectures such as microservices and serverless computing. This results in complex ecosystems that are hard to debug when functional and performance problems arise. Distributed tracing helps engineers make sense of this complexity and diagnose problems. However, distributed tracing is limited in what information it can collect. Engineers have to choose in advance which information they want to collect. We are investigating the limits of distributed tracing and the challenges ahead if we want to collect all the information all the time. Towards this, we published preliminary work characterizing the overhead of distributed tracing systems such as OpenTelemetry [2].
3. **Earth Observation Workflow Scalability**: Satellites record images of the earth every day, and we have 100s of satellites in orbit doing this. This generates petabytes of data. This data is highly specialized, and it is difficult to analyze it to gain insight. To investigate the challenges in modelling and scaling an earth observation workflow, we conducted a case study using a complex workflow for delineation urban areas and analyzing the spatial relationships between the settlement clusters [3]. It uses a graph data structure to model the relationships between settlements, which are aggregated to urban areas by contracting the graph for edges, which meet a user-specified condition. The multiple stages of the workflow are modelled with the Common Workflow Language (CWL). The workflow execution measurements for different input regions and various configurations, revealed the significant impact of parametric dependencies on the time to result. Especially, implicit parameters, like the settlement area and its distribution in the region of interest, influenced the workflow's completion time the most.

Besides these focused activities, the Cloud WG has been successfully organizing and growing the yearly workshop HotCloudPerf 2025 [4]. On May 6, the workshop will be held in Toronto, co-located with ICPE 2025, and feature an invited keynote and eight accepted paper presentations.

Keynote presentations:

- Padma Apparao (Intel Corporation, United States).

Paper presentations:

- Sharod Roychoudhury et. al., GenAI for bottleneck Detection in Cloud Architecture
- Anders Nõu et. al., Investigating Performance Overhead of Distributed Tracing in Microservices and Serverless Systems
- Zhiqi Li et. al., IrisBench: An Open-Source Benchmark Suite for Video Processing Systems in Cloud
- James McMahon et. al., Remote Memory Prefetching: Is Coarse-grained Fine?
- Vinita Pawar et. al., ObjecTier: Non-Invasively Boosting Memory Tiering Performance
- Sándor Battaglini-Fischer et. al., FAILS: A Framework for Automated Collection and Analysis of LLM Service Incidents
- Klervie Toczé, Introducing Resource Awareness Levels in Edge Computing Resource Management
- Yannik Lubas et. al., Microservice Applications and Their Workloads on GitHub

The Cloud working group congratulates Klervie Toczé, a multi-year active group member, on successfully defend-

ing her doctoral studies [5]. Klervie is now at the Vrije Universiteit Amsterdam. For more insights on her work, please check the Graduated PhDs section.

To conclude, this was a successful year for the RG Cloud Group. We started two new activities in 2024. We look forward to them succeeding in 2025. For this, we are actively seeking new participants. You can also propose new activities that would be of interest to members.

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] A. Al-Dulaimy, M. Jansen, B. Johansson, *et al.*, "The computing continuum: From IoT to the cloud," *Internet Things*, vol. 27, p. 101 272, 2024. DOI: 10.1016/J.IOT.2024.101272. [Online]. Available: https://doi.org/10.1016/j.iot.2024.101272.

[2] A. Nõu, S. Talluri, A. Iosup, and D. Bonetta, "Investigating performance overhead of distributed tracing in microservices and serverless systems," in *Companion of the 16th ACM/SPEC International Conference on Performance Engineering, ICPE 2025, Toronto, Canada, May 5-9, 2025*, ACM, 2025.

[3] L. Gruber, N. Herbst, S. Kounev, T. Esch, and T. Nguyen, "Modelling and optimizing a graph-based settlement delineation and analysis workflow," *Softwaretechnik-Trends*, vol. 44, no. 4, pp. 24–26, 2024. [Online]. Available: https://fb-swt.gi.de/fileadmin/FB/SWT/Softwaretechnik-Trends/Verzeichnis/Band%5C_44%5C_Heft%5C_4/SSP24%5C_07%5C_camera-ready%5C_4458.pdf.

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

[5] M. Beckman, *New doctor in computer science: Klervie toczé*, https://liu.se/en/news-item/nybliven-doktor-i-datalogi-klervie-tocze, [Online; accessed 6. Mar. 2025], Nov. 2024.

REPORT: SECURITY WORKING GROUP

The SPEC RG Security Benchmarking Working Group had personnel changes in the last year. Alexander Milenkoski is stepping down from his role as Chair, to become the new Vice-Chair with Thomas Prantl from University of Würzburg taking over. Furthermore, Simon Engel becomes the new Release Manager and Lukas Horn, the new secretary, both from Würzburg as well. The group would like to thank Alexander Milenkoski and José Flora for their commitment to the group over the past few years and are delighted that they will continue to be part of the group in new roles.

The Working Group congratulates their longstanding member José Flora on successfully defending his PhD thesis thesis in January 2025. He is now working at Dashlane.

With the new personnel in place, the group now also focuses on Privacy-Preserving Machine Learning (PPML), mainly through homomorphic encryption. Another new topic is IoT security. In 2024, the group had papers published at ARES [1], EDCC [2], ICSOFT [3] and in Springer International Journal on Information Security [4], Discover Data [5], [6] and Journal of Systems and Software [7].

The ARES paper uncovers an attack vector in an attribute-based-encryption scheme that enables malicious users to decrypt content without the required permissions and attributes. We also provided a solution to mitigate it.

In the EDCC publication, three techniques are proposed that allow to obtain the core behavior of a microservice into intrusion detection models. An evaluation with generated datasets shows that obtaining the core helps reusing detection models across several releases.

At ICSOFT a hypercall logger for the Hyper-V hypercall interface was presented that logs the inputs, outputs and sequence of hypercalls. These logs can be leveraged to create test cases for hypercall handlers. Evaluations showed that the execution is slowed down by 100,000 to 200,000 and the logs are correct.

The IJIS paper is the basis for the future work of the group on homomorphic encryption. It presents novel solutions to approximate typical mathematical functions (square root, division, exponential, logarithm) in such cryptosystems. Additionally, it shows the feasibility of these principles by applying them to the known Box-Cox transformation that therefore can be computed homomorphically.

The two contributions in Discover Data are focused on IoT Benchmarking. One of them defines the business problems of group encryption schemes and provides a benchmark around it. The second one analyzes the network impact of such schemes and demonstrates that the execution time of different group operations depends more on the current network situation than on the group sizes.

The publication in the Journal of Systems and Software is a benchmark to evaluate and compare different intrusion detection approaches. It also presents a benchmarking methodology that allows to standardize the process of the evaluation. Three case studies show the usefulness and wide application range of this benchmark.

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

- Propose a representative benchmark approach to evaluate mathematical functions in homomorphic cryptosystems.
- Develop new metrics to classify the performance of function approximation in homomorphic encryption.
- Improve accuracy of function approximation in homomorphic encryption by solving optimization problems.
- Identify challenges and applications in the area of boolean and integer homomorphic cryptosystems.
- Explore the foundation of homomorphic training of neural networks.
- Work towards a comparing benchmark of current

PPML techniques like homomorphic encryption, secure multiparty computation and functional encryption.
- Identify robustness of neural networks by exploiting their mathematical properties.

The SPEC RG Security WG is looking forward to its biggest goal for the coming year, establishing a new benchmark for homomorphic encryption. The group is happy to accept new members that like to contribute but also have new visions in the general area of security.

Thomas Prantl (University of Würzburg),
Aleksandar Milenkoski (SentinelLabs),
Lukas Horn (University of Würzburg),
Simon Engel (University of Würzburg)

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

[1] T. Prantl, M. Lauer, L. Horn, *et al.*, "Security analysis of a decentralized, revocable and verifiable attribute-based encryption scheme," in *Proceedings of the 19th International Conference on Availability, Reliability and Security, ARES 2024, Vienna, Austria, 30 July 2024 - 2 August 2024*, ACM, 2024, 24:1–24:11. DOI: 10.1145/3664476.3664487. [Online]. Available: https://doi.org/10.1145/3664476.3664487.

[2] J. Flora and N. Antunes, "Doing more with less? A study on models for intrusion detection in microservices," in *19th European Dependable Computing Conference, EDCC 2024, Leuven, Belgium, April 8-11, 2024*, IEEE, 2024, pp. 49–56. DOI: 10.1109/EDCC61798.2024.00021. [Online]. Available: https://doi.org/10.1109/EDCC61798.2024.00021.

[3] L. Beierlieb, N. Bellmann, L. Iffländer, and S. Kounev, "Logging hypercalls to learn about the behavior of hyper-v," in *Proceedings of the 19th International Conference on Software Technologies, ICSOFT 2024, Dijon, France, July 8-10, 2024*, H. Fill, F. J. D. Mayo, M. van Sinderen, and L. A. Maciaszek, Eds., SCITEPRESS, 2024, pp. 411–418. DOI: 10.5220/0012768100003753. [Online]. Available: https://doi.org/10.5220/0012768100003753.

[4] T. Prantl, L. Horn, S. Engel, *et al.*, "De bello homomorphico: Investigation of the extensibility of the openfhe library with basic mathematical functions by means of common approaches using the example of the CKKS cryptosystem," *Int. J. Inf. Sec.*, vol. 23, no. 2, pp. 1149–1169, 2024. DOI: 10.1007/S10207-023-00781-0. [Online]. Available: https://doi.org/10.1007/s10207-023-00781-0.

[5] T. Prantl, A. Bauer, S. Engel, *et al.*, "Benchmarking of secure group communication schemes with focus on iot," *Discover Data*, vol. 2, no. 1, May 2024, ISSN: 2731-6955. DOI: 10.1007/s44248-024-00010-6. [Online]. Available: http://dx.doi.org/10.1007/s44248-024-00010-6.

[6] T. Prantl, P. Amann, C. Krupitzer, S. Engel, A. Bauer, and S. Kounev, "Network impact analysis on the performance of secure group communication schemes with focus on iot," *Discover Data*, vol. 2, no. 1, Sep. 2024, ISSN: 2731-6955. DOI: 10.1007/s44248-024-00015-1. [Online]. Available: http://dx.doi.org/10.1007/s44248-024-00015-1.

[7] J. Flora and N. Antunes, "Evaluating intrusion detection for microservice applications: Benchmark, dataset, and case studies," *J. Syst. Softw.*, vol. 216, p. 112 142, 2024. DOI: 10.1016/J.JSS.2024.112142. [Online]. Available: https://doi.org/10.1016/j.jss.2024.112142.

REPORT: PREDICTIVE DATA ANALYTICS WORKING GROUP

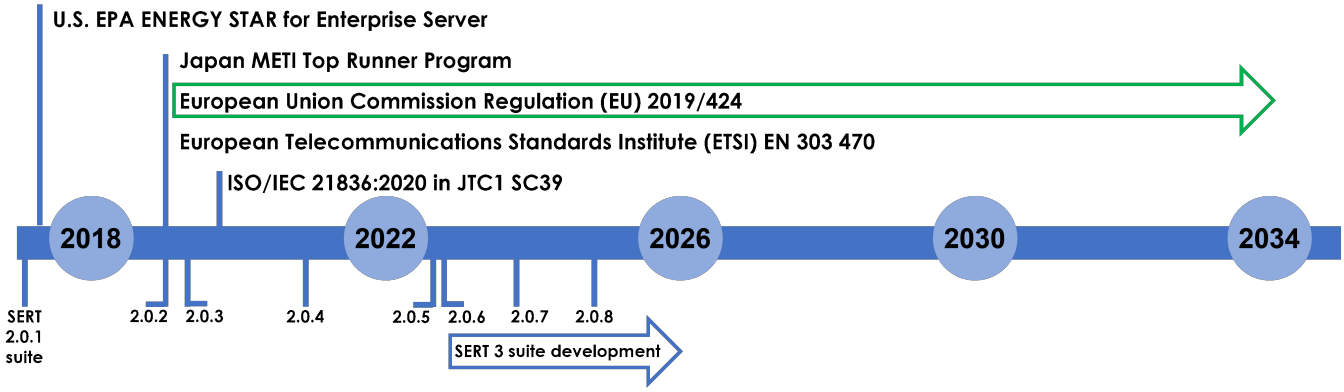
The mission of the group is to bridge the missing links between the facets involved in data analytics, namely big data storage and provisioning, data versioning, and performance evaluation. The group's ambition is to standardize and benchmark the entire data lifecycle, i.e., the analytics/prediction 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 2024, our group was working on standardizing the evaluation of time series synthesis. To this end, we have published two journal articles. "Thinking in Categories: A Survey on Assessing the Quality for Time Series Synthesis" [1] in the ACM Journal of Data and Information Quality and "Evaluation is key: a survey on evaluation measures for synthetic time series" [2] in the Springer Journal of Big Data. In the first journal, we address the challenge of evaluating synthetic time series quality by identifying key aspects of synthesis and proposing a systematic evaluation procedure. By establishing a common language and criteria, we aim to promote rigorous and reproducible research in time series synthesis. In the second journal, we surveyed evaluation measures for synthetic time series generation, addressing the lack of consensus on defining and quantifying quality. By organizing measures into a taxonomy and providing a concise selection guide, we aim to standardize evaluation practices and prevent inconsistent approaches in the field.

At ICPE 2025, we are presenting the paper "Quantifying Data Leakage in Failure Prediction Tasks" [3]. In this paper, we introduce a novel data leakage measure to quantify data leakage induced by the non-independence between train and test samples. We evaluate the impact of different data splitting techniques on the extent of leakage on an HDD failure prediction dataset. The paper highlights how data leakage leads to over-optimistic results and demonstrates the effectiveness of the proposed measure.

Besides the above-mentioned activities, the interests of the group lie in but are not limited to:

1. Performance modeling, analysis, testing, and pre-



A timeline of the SPEC Server Efficiency Rating Tool (SERT) development and adoption.

diction
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. Federated learning
8. Streamlining the data science process (DataOps)
9. Benchmarking of big data infrastructure

André Bauer (Illinois Institute of Technology),
Michael Stenger (University of Würzburg)

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

[1] M. Stenger, A. Bauer, T. Prantl, *et al.*, "Thinking in categories: A survey on assessing the quality for time series synthesis," *ACM J. Data Inf. Qual.*, vol. 16, no. 2, 14:1–14:32, 2024. DOI: 10.1145/3666006. [Online]. Available: https://doi.org/10.1145/3666006.

[2] M. Stenger, R. Leppich, I. T. Foster, S. Kounev, and A. Bauer, "Evaluation is key: A survey on evaluation measures for synthetic time series," *J. Big Data*, vol. 11, no. 1, p. 66, 2024. DOI: 10.1186/S40537-024-00924-7. [Online]. Available: https://doi.org/10.1186/s40537-024-00924-7.

[3] D. Grillmeyer, M. Hadry, V. Lesch, *et al.*, "Quantifying data leakage in failure prediction tasks," in *Proceedings of the 16th ACM/SPEC International Conference on Performance Engineering (ICPE '25), May 5–9, 2025, Toronto, Canada*, ACM, 2025.

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 been closely aligned with the SPECpower Committee since its launch in 2017.

Our mission is to research the energy and resource efficiency of computing devices and software. Our close collaboration strengthens the interaction between industry and academia by contributing research that enhances and promotes methods and tools for evaluating energy and resource efficiency, addressing this essential concern for both sectors as well as regulatory institutions.

Closely linked to the SPEC RG Power Working Group and the SPECpower Committee is the International Standards Group (ISG), which oversees the establishment of standardized benchmarks primarily developed for use in government regulations and programs. It collaborates with national and international standard development organizations to enhance global standards. Within the ISG, the ISG Server Efficiency Committee (Chair: Klaus-Dieter Lange) develops and establishes standardized server benchmarks primarily for regulatory use. Current members and associates of the ISG are AMD, AWS, Ampere, Apple, Bull, Dell, HPE, IBM, IEIT, Intel, Microsoft, NVIDIA, and Quanta.

With the release of ISO/IEC 21836:2020 in August 2020 [1], the founding members of ISG have already 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 U.S. EPA has already adopted the SPEC Server Efficiency Rating Tool (SERT®) 2 suite in their version 4.0 of their Server ENERGY STAR program. In the meantime, the ISG has already started the development of the SERT 3 suite [2]. 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.

At ICPE 2025, the RG Power WG [3], the SPECpower Committee [4], and the SPEC ISG [5] present a tutorial and tutorial-paper on the topic of efficiency benchmarking. It offers a guide to conducting energy-efficiency-related experiments and discusses current challenges in efficiency benchmarking. Power and efficiency measurements are highly sensitive to subtle changes in experiment setup, which can easily lead to misleading or

inconclusive results. Our goal is to share our expertise with computer scientists across disciplines, helping them improve efficiency in their fields through reliable and reproducible experiments while avoiding common pitfalls.

The tutorial aligns with the group's focus on expanding the SPEC Power and Performance Benchmark Methodology [6]. As general-purpose servers adopt new technologies—such as liquid cooling, direct current power supplies, accelerators (GPUs, FPGAs, ASICs), and advanced memory features—we examine key factors influencing efficiency and their relevancy for the next generation of efficiency benchmarks and standards. Further, 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] https://www.iso.org/standard/71926.html.
[2] https://www.spec.org/sert3/.
[3] https://research.spec.org/working-groups/rg-power/.
[4] https://www.spec.org/power/.
[5] https://www.spec.org/isg/.
[6] https://www.spec.org/power_ssj2008/docs/SPECpower-Methodology.pdf.

GRADUATED LONG-TERM MEMBERS OF THE SPEC RG

In the past year, two long-term members and well-known researchers in SPEC RG, Klervie Toczé and José Flora, received their PhD. Congratulations! You can find a summary of their work in the following paragraphs.

Klervie Toczé

Managing the edge computing infrastructure so that the appropriate resources are available at the required time and location (i.e. orchestrating) is especially challenging in case of sudden load spikes and when the orchestration impact itself has to be limited.

This thesis [1] enables edge computing orchestration with increased resource-awareness. First, it proposes two methods to better understand the edge resource demand. One for categorizing use cases, and one for characterizing, modeling and gathering traces from applications. The important insight is that application descriptions or models that are not based on a real application may not be giving an accurate picture of the load, thus driving incorrect supply-side decisions.

Second, it provides solutions on the supply side for orchestrating edge resources. Two frameworks are proposed for handling load spikes while avoiding over-provisioning. The first one utilizes mobile edge devices while the second leverages the concept of spare devices. The thesis also formalizes microservice request placement and studies the influence of two proposed energy metrics on placement decisions.

Finally, the thesis includes a critical perspective on edge computing. Sustainability challenges that should be highlighted more are collected. As a way forward, tools for adopting the strategy of sufficiency are proposed. Sufficiency involves aiming at only using the needed resources (no more, no less) with a goal of reducing resource usage.

Klervie now works as a postdoctoral researcher at the Vrije Universiteit Amsterdam. She is part of the interdisciplinary FEED4FOOD EU project [2], and belongs to the Software and Sustainability (S2) [3] and Development Economics groups. Have a look at her personal website [4] for updated information.

[1] K. Toczé, *Orchestrating a Resource-aware Edge*. Linkopings Universitet (Sweden), 2024. [Online]. Available: https://doi.org/10.3384/9789180757485.
[2] http://www.feed4food.eu/.
[3] https://s2group.cs.vu.nl/.
[4] http://www.klervietocze.eu/.

José Flora

The adoption of microservices is growing rapidly due to the agility of development and deployment. However, studies that focus on the effectiveness of such approaches in dealing with failures or attacks are missing. As a result, there are concerns about their dependability and security, which studies show is a slowing factor in adoption.

Intrusion detection is a technique to identify malicious activity on a host or network that has been used to increase the security of many systems. The success of past efforts makes it a promising approach for enhancing the security of microservices, but it currently faces significant challenges. Namely, effectively processing all the information produced by the services, keeping up with constant changes and releases, and maintaining effective operation even in such dynamic environments. Therefore, it is extremely important to develop techniques that can be deployed in real-world scenarios to keep applications secure.

We present two studies that demonstrate: i) the inability of state-of-the-art fault-tolerance mechanisms from popular container orchestrator platform in detecting failures caused by a diverse set of software faults, which compromises the reliability and dependability of the applications that use them, and ii) the applicability and effectiveness of host intrusion detection based on the system calls a

container uses during its execution.

We propose components to improve the detection effectiveness of detection models used in scalable, elastic, and continuously changing environments. We devise a set of techniques that allow detection models to monitor all active service replicas and techniques that allow to obtain the core behavior of a microservice into intrusion detection models.

This work also presents a benchmarking approach that considers the challenges and characteristics of microservices and provides a representative evaluation and comparison of intrusion detection models with configurable options. The benchmark user can utilize a 900-hour dataset with three types of data (training, testing, validation) that we generated following a well-defined methodology that ensures a representative and comprehensive evaluation base.

The work converges on the µSherlock framework, which provides continuous and adaptive intrusion detection that aims to evolve with microservice applications and provide a constant level of security regardless of changes and updates. We implement a tool for end-to-end continuous monitoring of microservices in Kubernetes.

José now works as a Software Engineer at Dashlane [1]. Find regular updates on his personal website [2].

[1] https://www.dashlane.com/.
[2] https://jeflora.github.io/.

TENURED LONG-TERM MEMBERS OF THE SPEC RG

André Bauer becomes Professor at the Illinois Institute of Technology

André Bauer, Chair of the Predictive Data Analytics Working Group, was appointed assistant professor in computer science at the Illinois Institute of Technology in August 2024. Prior to his current position at Illinois Tech, he held a postdoctoral research role at the University of Chicago, after earning his PhD from the University of Würzburg. André Bauer's research bridges the gap between performance engineering and data science, aiming to unlock the full potential of data science within scientific computing. He is actively recruiting PhD candidates to conduct cutting-edge research on microservices and serverless computing. Candidates with a strong background or interest in benchmarking, modeling, and performance engineering are particularly encouraged to apply.

IN MEMORIAM: ANDRÉ VAN HOORN

We regret to announce the passing of André van Hoorn on June 8th, 2024 following a serious illness.

SPEC Research Group Vice Chair André van Hoorn, a longtime member of our community, passed away on Saturday, June 8, 2024. André is survived by his wife, Merle, and their three children aged 11, 7, and 3. André has been an invaluable member of our community and contributor to both SPEC RG and ICPE since their inception back in 2010. Over the years, he has served in numerous roles including RG Vice Chair (since 2015), Co-Founder and Chair of the RG DevOps Performance Working Group (since 2014), RG Publicity Officer (2013-2015), and RG Steering Committee Member (since 2014).

His service was recognized among others with a SPECtacular Contributor Award from David Reiner, SPEC President, presented in Austin at the SPEC 2015 Annual Meeting. Despite his condition, André decided to attend ICPE 2024 in London, which speaks to his courage and strong dedication to the SPEC and ICPE community until the last moment.

In addition to SPEC RG, André has been a key member of the ICPE community, contributing high-quality papers on a regular basis, being a regular Program Committee (PC) member, and serving in numerous roles, including RG Steering Committee member, Web and Publicity Chair (2015), Tutorials Chair (2016), Finance Chair (2017), PC Chair (2018), Awards Chair (2019), and Artifacts Chair (2020). He has also been involved in co-organizing RG-supported events (e.g., GI-Dagstuhl seminar, QUDOS workshop series, journal special issues) as well as many additional activities behind the scenes.

His precious friendship, dedication, and excellent work will be immensely missed! André's brother Boris asks everyone to have a beer to toast him, that is what André would have wanted!

NEW SPEC RG TOOL: HETEROBENCH: MULTI-KERNEL BENCHMARKS FOR HETEROGENEOUS SYSTEMS

HeteroBench is a comprehensive, vendor-agnostic benchmark suite designed to evaluate the performance of heterogeneous computing systems that combine CPUs, GPUs, and FPGAs. Its primary motivation is to address the challenges associated with comparing these complex systems, which are often plagued by varying architectures, programming models, and toolchains. By providing a standardized framework for evaluating heterogeneous systems, HeteroBench aims to accelerate the development of innovative software solutions that can fully leverage the unique capabilities of these systems.

HeteroBench offers consistent implementations across multiple programming environments to cater to diverse developer needs. The suite features several multi-kernel benchmark applications in various versions, allowing users to choose from standard Python, Numba-accelerated Python, and a range of C++ implementations, including OpenMP-enhanced and OpenACC-enhanced variants for both CPUs and GPUs, as well as CUDA for NVIDIA GPUs and Vitis HLS-enhanced C++ for FPGAs. HeteroBench also supports an extensive range of accelerator options, encompassing CPUs, GPUs (NVIDIA, AMD, and Intel), as well as FPGAs (AMD Xilinx). This comprehensive support enables users to perform thorough performance evaluations on heterogeneous systems.

HeteroBench has the potential to play a role in advancing the development of more efficient software solutions that can harness the power of heterogeneous systems. The suite's user-friendly high-level parallel programming features, including Python and OpenMP/OpenACC directives, provide enhanced accessibility and adaptability for users. Furthermore, the multi-kernel design guides users to customize the placement of kernels across different hardware platforms, facilitating performance evaluation and optimization for heterogeneous systems. Extensive testing has validated the effectiveness and versatility of HeteroBench, showcasing the performance characteristics of various hardware configurations and providing a valuable resource for researchers and practitioners in heterogeneous computing systems. The suite is publicly available at [1].

[1] H. Tian, A. Mishra, Z. Chen, *et al.*, *HeteroBench*, https://github.com/HewlettPackard/HeteroBench, [Online; accessed 8. Apr. 2025], Apr. 2025.

NEW SPEC RG TOOL: KIEKER OBSERVABILITY FRAMEWORK VERSION 2

The Kieker Observability Framework Version 2, the successor of the Kieker Monitoring Framework, allows for tracing performance behavior and the measurement of system resources that are critical for the performance engineering of software systems. The new Kieker framework incorporates "observability", which lets you understand the internal states of a software system with its external outputs. Observability Engineering is about designing, building, and maintaining systems that enable us to learn the internal states of the target system with monitored data, in which, observability and monitoring are complementary.

With the new version of the Kieker framework, we will focus on its interoperability with the existing cloud-native observability technologies, such as the OpenTelemetry API, the Prometheus backend, and the Grafana dashboard frontend. With this approach, we will expand the Kieker traceability beyond the native Kieker monitoring tools, providing a library of Kieker analysis tools to the software systems that are not aware of Kieker.

Our initial effort on Kieker's interoperability with OpenTelemetry is demonstrated through our ICPE 2025 tool artifact,[1] a visualization of the TeaStore reference microservices using Kieker traces exported as OpenTelemetry traces, received and visualized by the ExplorViz visual analytics framework. The off-the-shelf Docker compose scripts are available on GitHub [1], and as online services.[2,3] The latest Kieker release is available for download on GitHub.[4]

[1] S. Yang, D. G. Reichelt, R. Jung, M. Hansson, and W. Hasselbring, *Kieker Observability Framework Version 2*, https://github.com/kieker-monitoring/tool-artifact, [Online; accessed 11. Apr. 2025], Apr. 2025.

NEW SPEC RG ARTIFACT: SPECIFICATION, VERIFICATION, AND REFINEMENT OF RESILIENCE SCENARIOS WITH DISPEL COCKPIT

Chaos Engineering is an established method to assess the resilience of software systems by injecting failures and learning from experiments in production. Existing Chaos Engineering tools, such as Chaos Toolkit, facilitate

[1] https://doi.org/10.1145/3680256.3721972
[2] https://teastore.sustainkieker.kieker-monitoring.net
[3] https://explorviz.sustainkieker.kieker-monitoring.net
[4] https://github.com/kieker-monitoring/kieker

creating and executing various failures but lack support for the entire process of resilience scenario elicitation, specification, execution, and refinement. Sebastian Frank et al. introduce DiSpel Cockpit [1] for software architects to continuously and iteratively specify (testable) resilience scenarios, collect data from different sources (simulation, monitoring), verify the resilience scenarios against results, and refine the response specifications. The functionalities of the DiSpel Cockpit include [2]:

1. Scenario Creation: Users can create scenarios by providing specifications of the stimuli and responses using the Property Specification Patterns. Additional information must be provided to allow for quantitative analyses, e.g., architecture and experiment files in JSON format.
2. Data Collection: Scenarios can be simulated and monitoring data can be searched for existing scenario occurrences based on the stimuli specifications.
3. Verification: After data collection, the output can be verified. The response text color indicates whether the response passes validation and statistics are computed.
4. Refinement: Finally, users can refine the responses using the built-in TQPropRefiner interface by investigating and adjusting parameters and timing specifications.
5. Configuration: The Cockpit provides an overview of all required services, indicating their availability.

[1] S. Frank, A. E. Tefur, M. A. Hakamian, and A. van Hoorn, "Dispel cockpit: Specification, verification, and refinement of resilience scenarios," in *Software Architecture. ECSA 2024 Tracks and Workshops - Luxembourg City, Luxembourg, September 3-6, 2024, Proceedings*, A. Ampatzoglou, J. Pérez, B. Buhnova, *et al.*, Eds., ser. Lecture Notes in Computer Science, vol. 14937, Springer, 2024, pp. 3–11. DOI: 10.1007/978-3-031-71246-3_1. [Online]. Available: https://doi.org/10.1007/978-3-031-71246-3%5C_1.

[2] S. Frank, A. E. Tefur, M. A. Hakamian, and A. van Hoorn, *Dispel cockpit artifact*, https://github.com/Cambio-Project/DiSpel-Cockpit, [Online; accessed 7. Mar. 2025], 2024.

NEW SPEC RG ARTIFACT: DATASET FOR INTRUSION DETECTION IN MICROSERVICE APPLICATIONS

Microservices are predominant for cloud-based applications, which serve millions of customers daily, that commonly run business-critical systems on software containers and multi-tenant environments; so, it is of utmost importance to secure these systems. Intrusion detection is a widely applied technique that is now being used in microservices to build behavior detection models and report possible attacks during runtime. However, it is cumbersome to evaluate and compare the effectiveness of different approaches. Standardized frameworks are non-existent and without fairly comparing new techniques to the state-of-the-art, it is difficult to understand their pros and cons. We present a comprehensive approach to evaluate and compare different intrusion detection approaches for microservice applications. A benchmarking methodology is proposed to allow users to standardize the process for a representative and reproducible evaluation. José Flora et al. also present a dataset that applies representative workloads and technologies based on state-of-the-art microservice applications. A sample of the dataset is available at https://doi.org/10.5281/zenodo.10655126. We use the benchmark and dataset in three case studies, characterized by dynamicity, scalability, and continuous delivery, to evaluate and compare state-of-the-art algorithms to tackle intrusion detection in microservices. Experiments in our peer-reviewed article [1] show the usefulness and wide application range of the benchmark while showing the capacity of intrusion detection algorithms in different applications and deployments.

[1] J. Flora and N. Antunes, "Doing more with less? A study on models for intrusion detection in microservices," in *19th European Dependable Computing Conference, EDCC 2024, Leuven, Belgium, April 8-11, 2024*, IEEE, 2024, pp. 49–56. DOI: 10.1109/EDCC61798.2024.00021. [Online]. Available: https://doi.org/10.1109/EDCC61798.2024.00021.

SELECTED ABSTRACTS

Early Detection of Performance Regressions by Bridging Local Performance Data and Architectural Models

During software development, developers often make numerous modifications to the software to address existing issues or implement new features. However, certain changes may inadvertently have a detrimental impact on the overall system performance. To ensure that the performance of new software releases does not degrade, existing practices rely on system-level performance testing, such as load testing, or component-level performance testing to detect performance regressions. However, performance testing for the entire system is often expensive and time-consuming, posing challenges to adapting to the rapid release cycles common in modern DevOps practices. System-level performance testing cannot be conducted until the system is fully built and deployed. On the other hand, component-level testing focuses on isolated components, neglecting overall system performance and the impact of system workloads. In this paper, we propose a novel approach to early detection of performance regressions by bridging the local performance data generated by component-level testing and the system-level architectural models. Our approach uses local performance data to identify deviations at the component level, and then propagate these deviations to the architectural model. We then use the architectural model to predict regressions in the performance of the overall system. We evaluate

our approach on two open-source benchmark systems and show that it can effectively detect end-to-end system performance regressions from local performance deviations with different intensities and under various system workloads. More importantly, our approach can detect regressions as early as in the development phase, in contrast to existing approaches that require the system to be fully built and deployed. Our approach is lightweight and can complement traditional system performance testing when testing resources are scarce.

L. Liao, S. Eismann, H. Li, *et al.*, "Early detection of performance regressions by bridging local performance data and architectural models," in *Proceedings of the 47th IEEE/ACM International Conference on Software Engineering, ICSE 2025, Ottawa, Ontario, Canada, 27 April - 3 May, 2025*, ACM, 2025. [Online]. Available: https://arxiv.org/abs/2408.08148.

Introducing Interactions in Multi-Objective Optimization of Software Architectures

Software architecture optimization aims to enhance non-functional attributes like performance and reliability while meeting functional requirements. Multi-objective optimization employs metaheuristic search techniques, such as genetic algorithms, to explore feasible architectural changes and propose alternatives to designers. However, this resource-intensive process may not always align with practical constraints. This study investigates the impact of designer interactions on multi-objective software architecture optimization. Designers can intervene at intermediate points in the fully automated optimization process, making choices that guide exploration towards more desirable solutions. Through several controlled experiments as well as an initial user study (14 subjects), we compare this interactive approach with a fully automated optimization process, which serves as a baseline. The findings demonstrate that designer interactions lead to a more focused solution space, resulting in improved architectural quality. By directing the search towards regions of interest, the interaction uncovers architectures that remain unexplored in the fully automated process. In the user study, participants found that our interactive approach provides a better trade-off between sufficient exploration of the solution space and the required computation time.

V. Cortellessa, J. A. Diaz-Pace, D. Di Pompeo, *et al.*, "Introducing interactions in multi-objective optimization of software architectures," *ACM Trans. Softw. Eng. Methodol.*, Jan. 2025, Just Accepted, ISSN: 1049-331X. DOI: 10.1145/3712185. [Online]. Available: https://doi.org/10.1145/3712185.

Columbo: A Reasoning Framework for Kubernetes' Configuration Space

Resource managers such as Kubernetes are rapidly evolving to support low-latency and scalable computing paradigms such as serverless and granular computing. As a result, Kubernetes supports dozens of workload deployment models and exposes roughly 1,600 configuration parameters. Previous work has shown that parameter tuning can significantly improve Kubernetes' performance, but identifying which parameters impact performance and should be tuned remains challenging. To help users optimize their Kubernetes deployments, we present Columbo, an offline reasoning framework to detect and resolve performance bottlenecks using configuration parameters. We study Kubernetes and define its workload deployment pipeline of 6 stages and 26 steps. To detect bottlenecks, Columbo uses an analytical model to predict the best-case deployment time of a workload per pipeline stage and compares it to empirical data from a novel benchmark suite. Columbo then uses a rule-based methodology to recommend parameter updates based on the detected bottleneck, deployed workload, and mapping of configurations to pipeline stages. We demonstrate that Columbo reduces workload deployment time across its benchmark suite by 28% on average and 79% at most. We report a total execution time decrease of 17% for data processing with Spark and up to 20% for serverless workflows with OpenWhisk. Columbo is open-source and available at https://github.com/atlarge-research/continuum/tree/columbo.

M. Jansen, S. Talluri, K. Doekemeijer, N. Tehrany, A. Iosup, and A. Trivedi, "Columbo: A Reasoning Framework for Kubernetes' Configuration Space," in *Proceedings of the 16th ACM/SPEC International Conference on Performance Engineering (ICPE)*, ACM, May 2025.

Quantifying Data Leakage in Failure Prediction Tasks

With the ever increasing importance of cloud computing and a strong focus on realiable data centers, a high amount of research has been done on failure prediction for hard disk drives. The collection of monitoring data of SMART statistics from operational HDDs enables operators to obtain predictions about the expected remaining useful life. Numerous methods for HDD failure prediction have been published in recent years, and their evaluation has shown decent results. However, a naive splitting into training and test sets can lead to data leakage and, thus, over-optimistic results that cannot be achieved on the data of scientific interest. In this paper, we propose a novel data leakage measure for quantifying the amount of data leakage in training and test data sets. Further, we define four splitting techniques and evaluate our measure in terms of performance optimism of classification models using different splitting strategies. Our results consistently show that splitting techniques prone to data leakage induce an overestimation of predictive performance. Overall, we were able to show the usefulness of the defined data leakage measure as well as its connection with different splitting techniques and performance optimism of prediction models.

D. Grillmeyer, M. Hadry, V. Lesch, *et al.*, "Quantifying Data Leakage in Failure Prediction Tasks," in *Proceedings of the 16th ACM/SPEC International Conference on Performance Engineering (ICPE)*, ACM, May 2025.

Bridging Clusters: A Comparative Look at Multicluster Networking Performance in Kubernetes

Microservices and containers have transformed the way applications are developed, tested, deployed, scaled, and managed. Several container orchestration platforms, like Kubernetes, have emerged, streamlining container management at scale and providing enterprise-grade support for application modernization. Driven by application, compliance, and end-user requirements, companies opt to deploy multiple Kubernetes clusters across public and private clouds. However, deploying applications in multi-cluster environments presents distinct challenges, especially managing the communication between the microservices spread across clusters. Traditionally, custom configurations, like VPNs or firewall rules, were required to connect such complex setups of clusters spanning the public cloud and on-premise infrastructure. This industry paper presents a comprehensive analysis of network performance characteristics for three popular open-source multi-cluster networking solutions (namely, Skupper, Submariner, and Istio), addressing the challenges of microservices connectivity across clusters. We evaluate key factors such as latency, throughput, and resource utilization using established tools and benchmarks, offering valuable insights for organizations aiming to optimize the network performance of their multi-cluster deployments. Our experiments revealed that each solution involves unique trade-offs in performance and resource efficiency: Submariner offers low latency and consistency, Istio excels in throughput with moderate resource consumption, and Skupper stands out for its ease of configuration while maintaining balanced performance.

S. S. Malleni, R. Sevilla, J. C. Lema, and A. Bauer, "Bridging Clusters: A Comparative Look at Multicluster Networking Performance in Kubernetes," in *Proceedings of the 16th ACM/SPEC International Conference on Performance Engineering (ICPE)*, ACM, May 2025.

Security Analysis of a Decentralized, Revocable and Verifiable Attribute-Based Encryption Scheme

In recent years, digital services have experienced significant growth, exemplified by platforms like Netflix achieving unprecedented revenue levels. Some of these services employ subscription models, with certain content requiring additional payments or offering third-party products. To ensure the widespread availability of diverse digital services anytime and anywhere, providers must have control over content accessibility. To address the multifaceted challenges in this domain, one promising solution is the adoption of attribute-based encryption (ABE). Over the years, various approaches have been proposed in the literature, offering a wide range of features. In a prior study, we assessed the security of one of these proposed approaches and identified one that did not meet its promised security standards. In this research we focuses on conducting a security analysis for another ABE scheme to pinpoint its shortcomings and emphasize the critical importance of evaluating the safety and effectiveness of newly proposed schemes. Specifically, we uncover an attack vector within this ABE scheme, which enables malicious users to decrypt content without the required permissions or attributes. Furthermore, we propose a solution to rectify this identified vulnerability.

T. Prantl, M. Lauer, L. Horn, *et al.*, "Security analysis of a decentralized, revocable and verifiable attribute-based encryption scheme," in *Proceedings of the 19th International Conference on Availability, Reliability and Security*, ser. ARES '24, Vienna, Austria: Association for Computing Machinery, 2024, ISBN: 9798400717185. DOI: 10.1145/3664476.3664487. [Online]. Available: https://doi.org/10.1145/3664476.3664487.

Evaluating intrusion detection for microservice applications: Benchmark, dataset, and case studies

Microservices are predominant for cloud-based applications, which serve millions of customers daily, that commonly run business-critical systems on software containers and multi-tenant environments; so, it is of utmost importance to secure these systems. Intrusion detection is a widely applied technique that is now being used in microservices to build behavior detection models and report possible attacks during runtime. However, it is cumbersome to evaluate and compare the effectiveness of different approaches. Standardized frameworks are non-existent and without fairly comparing new techniques to the state-of-the-art, it is difficult to understand their pros and cons. This paper presents a comprehensive approach to evaluate and compare different intrusion detection approaches for microservice applications. A benchmarking methodology is proposed to allow users to standardize the process for a representative and reproducible evaluation. We also present a dataset that applies representative workloads and technologies based on microservice applications state-of-the-art. The benchmark and dataset are used in three case studies, characterized by dynamicity, scalability, and continuous delivery, to evaluate and compare state-of-the-art algorithms with the objective of tackling intrusion detection in microservices. Experiments show the usefulness and wide application range of the benchmark while showing the capacity of intrusion detection algorithms in different applications and deployments.

J. Flora and N. Antunes, "Evaluating intrusion detection for microservice applications: Benchmark, dataset, and case studies," *Journal of Systems and Software*, vol. 216, p. 112 142, 2024, ISSN: 0164-1212. DOI: https://doi.org/10.1016/j.jss.2024.112142. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164121224001870.