Volume 2, Issue 4
April 2019

SPEC Research℠ Group Newsletter

CONTENTS

ICPE 2020 WILL BE HELD IN EDMONTON, CANADA

J. Nelson Amaral and Anne Koziolek, the General Chairs of the next ACM/SPEC International Conference on Performance Engineering (ICPE 2020), invite interesting high-quality submissions. The conference will take place on April 20-24, 2020 in Edmonton, Alberta, Canada.

Read more on page 5

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2018

The Kaivalya Dixit Distinguished Dissertation Selection committee has chosen this year to select two winners based on the high quality of both winning submissions:
• Matteo Nardelli of the University of Rome under the supervision of Professor Valeria Cardellini, and
• Nikolas Herbst of the University of Würzburg, under the supervision of Professor Samuel Kounev.

Read more on page 4

SIX SPEC RESEARCH WORKING GROUPS REPORT ON THEIR PROGRESS

The SPEC Research Working Groups Cloud, Big Data, DevOps Performance, Security and Power report on their progress, articles, benchmarks, and technical reports published in the year 2018. The new SPEC Research Working Group Quality of Experience introduces itself. The working groups are always open for new members, feel invited to join us!

Read more on pages 5-9

HOT TOPICS AND SELECTED ABSTRACTS

You find a selection of recently published article abstracts as part of working group activities at the end.

Read more on page 10-12

CONTACT

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

SPEC Research Group
Chair: Samuel Kounev (rgchair@spec.org)
Web: http://research.spec.org

SPEC RESEARCH GROUP OFFICERS

Chair:
Samuel Kounev, University of Würzburg, Germany

Vice-Chair:
André van Hoorn, University of Stuttgart, Germany

Secretary:
Cor-Paul Bezemer, University of Alberta, Canada

Steering Committee:
J. Nelson Amaral, University of Alberta, Canada
Cor-Paul Bezemer, University of Alberta, Canada
Alexandru Iosup, VU Amsterdam, NL
Samuel Kounev, University of Würzburg, Germany
Klaus-Dieter Lange, HPE, USA
Meikel Poess, Oracle Corporation, USA
Petr Tůma, Charles Univ. of Prague, Czech Republic
Weiyi Shang, Concordia University, Canada
André van Hoorn, University of Stuttgart, Germany
Boris Zibitsker, BEZNext, USA

Publicity Officer:
Henning Schulz, NovaTec Consulting GmbH, Germany

Release Manager:
Vojtěch Horký, Charles Univ. Prague, Czech Republic

Newsletter Editors:
André Bauer, University of Würzburg, Germany
Nikolas Herbst, University of Würzburg, Germany

http://research.spec.org

SPEC RESEARCH WORKING GROUPS

Cloud Working Group

Chair:
Alexandru Iosup, VU Amsterdam, The Netherlands

Vice-Chair & Secretary:
Nikolas Herbst, University of Würzburg, Germany

Release Manager:
André Bauer, University of Würzburg, Germany

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

Quality of Experience Working Group

Chair:
Florian Wamser, University of Würzburg, Germany

Vice-Chair:
Tobias Hoßfeld, University of Würzburg, Germany

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

Security Working Group

Chair:
Aleksandar Milenkoski, ERNW & Univ. Würzburg, Germany

Vice-Chair:
Nuno Antunes, University of Coimbra, Portugal

Secretary and Release Manager:
Lukas Iffländer, University of Würzburg, Germany

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

Big Data Working Group

Chair:
Tilmann Rabl, bankmark, Germany

Vice-Chair:
Todor Ivanov, Goethe University Frankfurt, Germany

Secretary:
Meikel Poess, Oracle Corporation, USA

Release Manager:
John Poelman, IBM, USA

https://research.spec.org/en/working-groups/rg-big-data.html

DevOps Performance Working Group

Chair:
André van Hoorn, University of Stuttgart, Germany

Vice-Chair:
Cor-Paul Bezemer, Queen's University, Canada
Robert Heinrich, KIT, Germany

Secretary:
Simon Eismann, University of Würzburg, Germany

Release Manager:
Vincenzo Ferme, Università della Svizzera Italiana, Switzerland

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

Power Working Group

Chair:
Jóakim von Kistowski, University of Würzburg, Germany

Vice-Chair:
Klaus-Dieter Lange, HPE, USA

Secretary:
Mike Tricker, Microsoft, USA

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

WELCOME TO THE SPEC RESEARCH GROUP NEWSLETTER

November 2018 marks the 30th anniversary of SPEC. With 125 members in 22 countries and nearly two dozen benchmarks spanning highly diverse aspects of computing performance and energy efficiency, SPEC has become known as a beacon of truth for computing researchers, vendors, users and analysts worldwide. These professionals rely on SPEC to ensure that the marketplace has a fair and useful set of metrics to differentiate computing systems. Founded in 2011, the SPEC Research Group is proud being part of the recent years of this remarkable history.

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

Some highlights from the last year include:

- 9^{th} ACM/SPEC ICPE 2018 in Berlin, Germany
- 15^{th} IEEE International Conference on Autonomic Computing ICAC 2018 in Trento, ITaly
- 4^{nd} International Workshop on Quality-aware DevOps QUDOS 2018 in Berlin, Germany
- 1^{st} Workshop on Hot Topics in Cloud Computing Performance HotCloudPerf 2018 at ICPE 2018
- New Technical Report published: Methodological Principles for Reproducible Performance Evaluation in Cloud Computing
- New artifact accepted: Alberta Workloads

We have been actively working on preparation, planning and organization of ICPE 2019. We hope that the vivid exchange of ideas during the upcoming ICPE 2019 will be a great motivation for the next year of scientific and engineering work.

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

Samuel Kounev
(SPEC Research Chair, University of Würzburg).
André Bauer, Nikolas Herbst
(Newsletter Editors, University of Würzburg).

ANNOUNCEMENTS

https://www.spec.org/news/

SPEC CLOUD® IAAS 2018 RELEASED

The SPEC Cloud® IaaS 2018 benchmark is SPEC's second benchmark suite to measure cloud performance. The benchmark suite's use is targeted at cloud providers, cloud consumers, hardware vendors, virtualization software vendors, application software vendors, and academic researchers.

The current version of the benchmark is version 1.0, released on December 18, 2018. The SPEC Cloud IaaS 2018 benchmark builds on the original 2016 release with a variety of enhancements and new primary metrics. Please note that due to the workload and the methodology changes for metric calculations, the results of the SPEC Cloud IaaS 2018 benchmark are not comparable to those from the SPEC Cloud IaaS 2016 benchmark.

The benchmark is designed to stress provisioning as well as runtime aspects of a private or public cloud using I/O and CPU intensive cloud computing workloads. SPEC selected the social media NoSQL database transaction and K-Means clustering using map/reduce as two significant and representative workload types within cloud computing. Each workload runs in multiple instances, referred to as an application instance. The benchmark instantiates multiple application instances during a run. The application instances and the load they generate stress the provisioning as well as run-time aspects of a cloud. The run-time aspects include CPU, memory, disk I/O, and network I/O of these instances running in a cloud. The benchmark runs the workloads until quality of service (QoS) conditions are reached. The tester can also limit the maximum number of application instances that are instantiated during a run.

U.S. EPA ADOPTS THE SPEC SERT SUITE

A significant milestone was achieved on September 17, 2018, as the United States Department of Energy's (DOE) Environmental Protection Agency (EPA) released the final version of the "ENERGY STAR Version 3.0 Computer Servers Program Requirements" which defines the Active State Efficiency Thresholds that will determine ENERGY STAR eligibility effective June 17, 2019. The new thresholds have been determined using data collected by running the SPEC Server Efficiency Rating Tool (SPEC SERT™). Server vendors seeking ENERGY STAR certification under the version 2.0 of the EPA requirements were mandated to submit SPEC SERT data to the U.S. EPA.

In the United States, the government purchases only computer servers that have a U.S. ENERGY STAR certification. Hence it is crucial for all vendors to strive to improve the efficiency of their servers to be able to sell to the government, and as a result, improve the energy efficiency of the entire computer server industry.

Background Story: Environmental agencies around the world have a goal to promote practices that progressively reduce the environmental impact caused by various consumer products. Computer systems are evaluated for their energy efficiency, and only a certain percentage get a seal of approval. In the US, the DOE's EPA grants the ENERGY STAR certification to the computer systems that meet its prescribed criteria. This criteria is tightened periodically to ensure the industry continues to produce more efficient systems.

Beginning June 17, 2019, the EPA will not accept any other form of certification collateral besides SPEC SERT™. All computer server vendors who wish to acquire ENERGY STAR certification must run the latest version of SPEC SERT™ 2.0.1 and submit their results.

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2018 WINNERS

The Kaivalya Dixit Distinguished Dissertation Selection committee has chosen this year to select two winners based on the high quality of both winning submissions.

The first winning dissertation is titled QoS-aware Deployment and Adaptation of Data Stream Processing Applications in Geo-Distributed Environments, authored by Matteo Nardelli of the University of Rome under the supervision of Professor Valeria Cardellini.

Matteo Nardelli addresses the problem of QoS-aware deployment and adaptation of data stream processing (DSP) applications in geo-distributed environments. The contributions are the formulation of the DSP operator placement problem, formulation of the run-time problem, design of a framework, optimizer and heuristics, and implementation into an open source framework.

The second winning dissertation is titled Methods and Benchmarks for Auto-Scaling Mechanisms in Elastic Cloud Environments by Nikolas Herbst of the Julius-Maximilians-Universität Würzburg, under the supervision of Professor Samuel Kounev.

Nikolas Herbst proposes several methods and related tools for automatically provisioning and scaling elastic resources for cloud systems. The thesis is also exceptional in the effort put into validation using plausible workloads.

The award selection committee for 2018 was chaired by Evgenia Smirni (College of William and Mary) and consisted of the following members: André van Hoorn (Universität Stuttgart, Germany), Diwakar Krishnamurthy (University of Calgary, Canada), Arif Merchant (Google, USA), Erich Nahum (IBM Research, USA), Vittoria de Nitto Personè (University of Rome, Italy), Xipeng Shen (NCSU, USA), Xiaoyun Zhu (Hyperpilot, USA).

The SPEC Kaivalya Dixit Distinguished Dissertation Award aims to recognize outstanding doctoral dissertations in the field of computer benchmarking, performance evaluation, and experimental system analysis in general. Nominated dissertations will be evaluated in terms of scientific originality, scientific significance, practical relevance, impact, and quality of the presentation.

Contributions of interest span the design of metrics for system evaluation as well as the development of methodologies, techniques and tools for measurement, load testing, profiling, workload characterization, dependability and efficiency evaluation of computing systems. Dissertations defended between October 2017 and September 2018, were eligible to be nominated.

ICPE 2019: STATISTICS

The 10th ACM/SPEC International Conference on Performance Engineering (ICPE 2019), being held in Mumbai, India from April 7 to 11, 2019, attracted a record number of high-quality submissions. Through a rigorous review process, we selected the most meritorious ones to create a diverse and interesting conference program. In the research track, 13 out of 71 papers were accepted as full papers. Five full papers received an ACM artifact badge after the subsequent review process in the artifact evaluation track. Seven submissions were accepted as short research papers. In the industry/experience track, 5 full papers and 8 short papers have been accepted out of 21 submissions. In the work-in-progress/vision track, 10 out of 23 papers were accepted. In addition to those scientific papers, the technical program features three keynotes, two tutorials, 7 posters and demonstrations, the presentation of the SPEC Kaivalya Dixit Distinguished Dissertation Award and one workshop. A number of co-located events will be held alongside ICPE2019, including a "Simulation Modeling Hackathon", a first hackathon on performance topics, the "India Performance Workshop", which will feature invited speakers from academia and industry, and the "TCS Quantum Computing Workshop". Details about the conference and its program are provided at http://icpe2019.spec.org/.

Antinisca Di Marco (University of L'Aquila),
Varsha Apte (IIT Bombay),
Marin Litoiu (York University), and
José Merseguer (Universidad de Zaragoza)

ICPE 2020 IN EDMONTON, ALBERTA, CANADA – 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 2020 will be held in Edmonton, Alberta (Canada), from April 20 to April 24. Edmonton is the capital of Alberta and is located in the Western Canadian prairies. It can be easily reached from several airline hubs in the USA, Canada, and Europe. This modern Winter city has a vibrant cultural downtown with many coffee shops, restaurants, galleries, and an extensive shopping area. The conference will be located downtown in the Sutton Place Hotel with direct access to public transportation, including public transit from the airport. Within walking distance are the Edmonton Gallery, the extensive new Royal Alberta Museum, and the extensive river valley. The campus of the University of Alberta is a few stations away on the metro.

The contact person for ICPE 2020 is J. Nelson Amaral, who will be General Co-Chair along with Anne Koziolek from the Karlsruhe Institute of Technology (KIT) in Germany. The PC Co-Chairs will be Alexandru Iosup from Vrije Universiteit Amsterdam in the Netherlands and Catia Trubiani from the Gran Sasso Science Institute (GSSI) in Italy. The industrial track chair will be Andreas Brunnert, from RETiT, Germany.

J. Nelson Amaral (University of Alberta)

NEW WORKING GROUP: RG QUALITY OF EXPERIENCE

In recent years, Quality of Experience (QoE) has become an important means of determining customer satisfaction for all types of services on the Internet. Reason enough to bring QoE into the benchmarking domain!

Since the late 90s, the term ”Quality of Experience” (QoE) has gained importance in various contexts. For example, when evaluating applications and services, and hardware and systems, their real needs and requirements are often taken into account. They are rated according to their specific purpose and field of application, as this reflects the usefulness for the user. As a result, metrics such as quality have evolved depending on the type of application. For example, a video streaming system is rated based on the playout resolution and smoothness of the streaming. A server infrastructure is rated based on the ability to satisfy the applications running on it. In communication, for example, the term quality has for many years been largely associated with the so-called ”Quality of Service” (QoS) but now the providers start to evaluate their network especially for services such as video streaming or web browsing to ensure the applicability of their network for a specific usage scenario.

All these evaluations with the objective of focusing on the end user perception and evaluating the infrastructure, system, or hardware according to application needs can largely be summarized under the term and metric ”Quality of Experience (QoE)”. QoE goes beyond the simple determination of objective evaluations and incorporates subjective user views and contextual factors. With the proliferation of Internet services, telecommunications networks and specialized software in daily environments, QoE provides a tool to assess and compare applications and their associated ecosystem.

Accurate and representative measurement and characterization of the QoE is needed to correlate QoE values with deployment options or optimizations for software and networks. Furthermore, benchmarking a system requires an exact model of the QoE for each class of application. As a result, there is currently a variety of different definitions in various research areas, including telecommunications, hardware development and, in particular, application and software development. Consequently, there is widespread interest in the term „Quality of Experience”.

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

The vision of the RG QoE is to promote Quality of Experience (QoE) as a new evaluation metric. In a nutshell QoE assesses how the end-user ultimately perceives a service or system. QoE can be seen as a comprehensive, universal metric for comparison that allows ratings and evaluations with respect to users and services. Current topics include currently

- QoE Modelling for gaming, **video streaming**, virtual reality, immersive applications, **multimedia applications**
- QoE within technical systems **cloud** systems, **mobile context**, trade-off between **energy efficiency**

- QoE **modeling and metrics**, crowdsourcing for QoE evaluation, **benchmarking** scenarios for different metrics
- QoE resource management: Direct application of QoE as a control metric

We seek for collaboration opportunities and offer help and discussion in (1) the definition of QoE-related metrics, (2) application workloads, (3) user studies, and (4) benchmarking with respect to QoE, which can provide another dimension for benchmarks out there. Many benchmarks to date already use a variety of application and QoE-related metrics, which can be specified and expanded by the members of the RG QoE. Talks are currently underway with OSG groups and HPGW, which started at the annual SPEC meeting in Houston this year.

We are looking for partners to jointly explore QoE in a specific context. We offer help with the use of QoE in modeling, measurement, and application.

Contact: Florian Wamser, rgqoe@spec.org

Florian Wamser (Uni. of Würzburg)
https://research.spec.org/working-groups/rg-quality-of-experience.html

REPORT: SECURITY WORKING GROUP

RG Security held elections in December. Aleksandar Milenkoski, who is with ERNW GmbH (Germany), has been elected as the group's Chair. Nuno Antunes, who is with the University of Coimbra (Portugal), has been elected as the group's Vice Chair. Lukas Iffländer, who is with the University of Würzburg (Germany), continues to serve as the group's Secretary and Publication Manager.

The group has recently broadened its focus, now focusing not only on evaluating intrusion detection systems, but also on security benchmarking in general. The group has therefore been renamed to RG Security. The group has expanded its membership with several new members. RG Security is currently working on two publications, which define the scope of the group's short- and long-term research activities. The group works on a technical report that aims at identifying and discussing relevant challenges in developing effective security benchmarks. It also works on a position paper with a focus on the industry requirements for security benchmarking. In addition, Lukas Iffländer, the group's Secretary, has recently published two papers [1],[2] in collaboration with the working group.

Aleksandar Milenkoski, (ERNW GmbH),
Nuno Antunes (University of Coimbra), and
Lukas Iffländer (University of Würzburg)
https://research.spec.org/working-groups/rg-ids-benchmarking.html

[1] Lukas Iffländer, Jonathan Stoll, Nishant Rawtani, Veronika Lesch, Klaus-Dieter Lange, and Samuel Kounev. 2019. Performance Oriented Dynamic Bypassing for Intrusion Detection Systems. ACM/SPEC ICPE 2019, April 7–11, 2019, Mumbai, India. ACM, New York, NY, USA, 8 pages.

[2] Lukas Iffländer and Nicolas Fella. 2019. Performance Influence of Security Function Chain Ordering. ACM/SPEC ICPE 2019, April 7–11, 2019, Mumbai, India. ACM, New York, NY, USA, 2 pages.

REPORT: POWER RESEARCH AND SPEC OSG POWER COMMITTEE

Since its inception in 2017, the Power Research working group has operated tightly coupled and in close collaboration with the OSG SPECpower Subcommittee.

This year, the RG Power group has focussed on contributing to the upcoming SPECpower benchmark. It has extended SPEC's existing power methodology so that it can be used to measure transactional load on accelerator devices in servers. RG Power implemented an initial FFT workload to be used in evaluating and testing the methodology. Internal development kits of the SPECpower benchmark utilizing the new methodology and the initial FFT workload are already available. With these kits, OSG and RG Power hope to test and implement a plethora of novel and exiting power benchmarking workloads for accelerators. The methodology itself was published at the ICPE 2019 [1]. In addition, RG Power has contributed a cryptography workload, using the BCrypt algorithm, to enhance the main body of workloads for the upcoming benchmark.

With the previous years' work on the Server Efficiency Rating Tool (SERT) metric reaching its conclusion, the RG Power working group is looking forward to new and exciting challenges in power and energy efficiency benchmarking and testing. Potential future topics might include: Power behavior under varying loads, energy efficiency of new accelerator devices, such as dedicated ML hardware, and energy efficiency benchmarking in new server application areas, such as NFV. The group is happy to accept new members and visions for new research directions in the general area of energy efficiency benchmarking.

Jóakim von Kistowski (Uni. of Würzburg),
Klaus-Dieter Lange (HPE)
https://research.spec.org/working-groups/rg-power.html

[1] Jóakim von Kistowski, Johann Pais, Tobias Wahl, Klaus-Dieter Lange, Hansfried Block, John Beckett, and Samuel Kounev. Measuring the Energy Efficiency of Transactional Loads on GPGPU. ACM/SPEC ICPE 2019, Mumbai, India, 2019. ACM, New York, NY, USA. 2019.

REPORT: CLOUD WORKING GROUP

In 2018, the SPEC RG Cloud Group has pursued through a diverse set of activities its long-term mission of furthering cloud benchmarking, quantitative evaluation, and experimental analysis, in directions relevant for both academia and industry.

Through our mission, we focus on novel cloud paradigms such as Functions-as-a-Service, Serverless Computing, clouds integrating fog and edge devices, convergence of HPC and Big Data as cloud services.

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

Among the keywords most discussed in the group, over 2018, are serverless computing and Function-as-a-Service, edge, convergence, metrics, experiment methodology, reproducibility, elasticity, and auto-scaling.

We are proud to announce that the PhD thesis by Nikolas Herbst, serving as RG Cloud working group's co-chair, receives the SPEC Kaivalya Dixit Distinguished Dissertation Award 2018. His work entitled "Methods and Benchmarks for Auto-Scaling Mechanisms in Elastic Cloud Environments" is highly related to the groups scope and activities.

In 2018, through monthly online meetings facilitated by WebEx and SPEC as well a numerous activity meetings, we have advanced work on the following main topics:

1. **Serverless/FaaS platforms**: This sub-group aims at standardizing, understanding and improving the emerging technologies for serverless computing and FaaS platforms. We developed a reference architecture for FaaS platforms and mapped to it 50 open-source and closed-source implementations of FaaS platforms. We submitted an article on these results to IEEE Internet Computing. Our next goal is the development of a comprehensive benchmark for FaaS platforms.

2. **Auto-Scaler Benchmarking**: This activity conducts experimental auto-scaler competitions based on earlier developed metrics and measurement methodology for a level-playing field. In a published IEEE TPDS article [1], we conduct such a broad competition for HTTP applications that is won by the hybrid auto-scaler Chameleon. Follow up work includes the assessment of coordinated auto-scaling for applications consisting of multiple distributed services with an article accepted for IEEE ICDCS. We plan to consider optimized integrated auto-scaling in the context of serverless computing platforms and in the context of complex workflows.

3. **Cloud Metrics**: This successful activity is closing with the publication of an ACM ToMPECS article [3] entitled "Quantifying Cloud Performance and Dependability: Taxonomy, Metric Design, and Emerging Challenges". The article covers elasticity, performance isolation, availability and operational risk - each accompanied with a set of metrics and examples. The other activities – especially (2) auto-scaler benchmarking and (4) cloud experiment methodology – build upon the results and can be seen as logical continuations.

4. **Cloud Experiment Methodology**: This activity is devoted to the identification of the main principles that should be used for a sound performance evaluation in cloud systems. The activity started in spring 2017. Since then, the involved researchers analyzed what are the current guidelines for reproducibility proposed by ACM, and principles proposed in other fields of science, focusing mostly on the computer science domain. The current state-of-the-art was reviewed by adopting a systematic literature review approach, analyzing some of the main venues for the cloud computing community. A paper was submitted to IEEE Transactions on Software Engineering, and is currently under revision. A SPEC technical report [4] on this is now publicly available. You can read the abstract at the end of this newsletter.

5. **Edge computing activity**: The long term goal of this starting activity is to create a benchmark for edge technologies. At the moment, the group is discussing different use cases and their characteristics, as well as ways forward to gather data sets from edge applications. The next planned steps include defining common quality attributes for the edge use cases working towards an Edge benchmark prototype.

Since 2018, the Cloud working group is organizing a yearly workshop in connection with their face-to-face meeting: The first edition of the HotCloudPerf workshop was well attended in Berlin co-located with ICPE 2018 featuring 4 full workshop papers, 2 short paper and 4 additional talks plus a joint panel discussion. The second HotCloudPerf 2019 is co-located with the FAS* confer-

ences IEEE ICAC, and SASO in Umea (Sweden) on June 17 supported by a high-profile technical program committee. The third edition of HotCloudPerf 2020 is planned to co-locate again with the ACM/SPEC ICPE 2020 in Edmonton, Alberta (Canada).

To conclude, 2018 was a full and successful year for the RG Cloud Group. We are looking forward to an even more successful 2019. For this, we are actively seeking new participants and activities.

Nikolas Herbst (University of Würzburg),
Alexandru Iosup (VU Amsterdam)
http://research.spec.org/working-groups/rg-cloud.html

[1] André Bauer, Nikolas Herbst, Simon Spinner, Ahmed Ali-Eldin, and Samuel Kounev. Chameleon: A Hybrid, Proactive Auto-Scaling Mechanism on a Level-Playing Field. IEEE Transactions on Parallel and Distributed Systems, 30(4):800 – 813, April 2019, IEEE.

[2] André Bauer, Veronika Lesch, Laurens Versluis, Alexey S. Ilyushkin, Nikolas Herbst and Samuel Kounev Chamulteon: Coordinated Auto-Scaling of Micro-Services. IEEE International Conference on Distributed Computing Systems (IEEE ICDCS 2019). To appear.

[3] Nikolas Herbst, André Bauer, Samuel Kounev, Giorgos Oikonomou, Erwin van Eyk, George Kousiouris, Athanasia Evangelinou, Rouven Krebs, Tim Brecht, Cristina L. Abad, and Alexandru Iosup. Quantifying Cloud Performance and Dependability: Taxonomy, Metric Design, and Emerging Challenges. ACM Transactions on Modeling and Performance Evaluation of Computing Systems (ToMPECS), 3(4):19:1–19:36, August 2018, ACM, New York, NY, USA.

[4] Alessandro Vittorio Papadopoulos, Laurens Versluis, André Bauer, Nikolas Herbst, Jóakim von Kistowski, Ahmed Ali-Eldin, Cristina Abad, J. Nelson Amaral, Petr Tuma, and Alexandru Iosup. Methodological Principles for Reproducible Performance Evaluation in Cloud Computing - A SPEC Research Technical Report. April 2019.

REPORT: DEVOPS PERFORMANCE WORKING GROUP

The DevOps Performance Working Group focuses on how to address performance concerns in DevOps. DevOps aims to increase the velocity and frequency of bringing software changes into stable production. These goals are achieved by a tighter integration of development (Dev) and operations (Ops), as well as a high degree of automation (e.g., via continuous delivery/deployment).

The DevOps Performance Working Group fosters and facilitates research in combining model-based and measurement-based APM (application performance management) and SPE (software performance engineering) activities for DevOps, e.g., by experience sharing, agreement on definitions, specification of metrics, and dissemination of novel methods, techniques, and tools for quantitative evaluation. We aim to identify cross-community collaboration, and to set the path for long-lasting collaborations towards performance-aware DevOps.

The major joint activity of 2017 and early 2018 was our survey on the current state of performance in DevOps. Our study reveals that automatic performance evaluations are usually not integrated into automatic delivery pipelines and not performed regularly. In addition, performance modeling is not applied in most companies. The results will be in the industry/experience track of ICPE 2019 [1]. The abstract of the paper is included at the end of this newsletter.

In mid 2018, we restructured the group to operate mostly in subgroups consisting of 6-8 participants who collaborate closely on concrete topics. The motivation for working in subgroups was that the group as a whole was becoming too large to work in a productive manner on a single joint project. The subgroups meet biweekly, and report to the whole group once a month.

The following subgroups were created:

1. **Performance regression testing of microservices**: This subgroup focuses on the challenges of performance regression testing microservices. The subgroup is currently working on a full research paper in which several of these challenges are demonstrated using a case study on the Tea Store application. This paper will soon be submitted to a major software engineering venue.

2. **Model extraction and refinement in continuous software engineering**: This subgroup focuses on how models can be leveraged to improve the continuous software engineering process.

3. **Performance engineering for big data**: This subgroup focuses on the challenges and opportunities of performance engineering for big data.

4. **Performance of continuous delivery infrastructures**: This subgroup focuses on the evaluation and improvement of continuous delivery (CD) infrastructures, which have become a critical component of software development. The subgroup is currently analyzing empirical performance data of a CD system. Next steps include the performance modeling of these kinds of systems.

Moreover, collaborations included jointly supervised student projects.

In April 2018, the group met for its fourth face-to-face meeting, co-located with the 9th ACM/SPEC International Conference on Performance Engineering (ICPE 2018) in Berlin, Germany. The next meeting is scheduled for the 10th ACM/SPEC International Conference on Performance Engineering (ICPE 2019) in Mumbai, India.

There were several events organized by group members in 2018. The working group co-organized the fourth edition of the International Workshop on Quality-Aware DevOps (QUDOS 2018 [2]). QUDOS 2018, technically supported by the SPEC RG, was co-located with ICPE 2018. This year, the afternoon program was joint with the Seventh International Workshop on Load Testing and Benchmarking of Software Systems (LTB 2018 [3]), also organized by two working group members. Both workshops were well attended with 25-30 participants. In March 2019, the fifth fifth QUDOS edition [2] was held in Hamburg, Germany, co-located with the International Conference on Software Architectures (ICSA 2019). The workshop included a keynote, 9 paper presentations, and attracted more than 30 participants. Finally, a group member co-organized the Dagstuhl seminar on Visualizing Systems and Software Performance [4] in 2018. This seminar brought together experts from the software visualization, performance engineering and high performance computing communities. This seminar was attended by several working group members.

In January 2019, a new research project has started in the context of the group. The RADON project, funded by the European Commission in the Horizon 2020 program, aims at creating a DevOps framework to create and manage microservices-based applications that can optimally exploit serverless computing technologies. The involved group members focus particularly on the aspect of non-functional properties such as performance.

In total, 14 group meetings were held in 2018 (not including the subgroup meetings), including talks about “Pet Supply Store: A Micro-Service Application for Benchmarking, Modeling and Resource Management Research”, “Continuous Performance Model Extraction”, “SPECjEnterprise2018”, “Autoscaling Microservices with Layered Queueing Network Models” and “DICE: a Quality-Aware DevOps Suite For Big Data Applications”.

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

Cor-Paul Bezemer (University of Alberta), André v. Hoorn (University of Stuttgart, Robert Heinrich (Karlsruhe Institute of Technology)

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

[1] C. Bezemer, S. Eismann, V. Ferme, J. Grohmann, R. Heinrich, P. Jamshidi, W. Shang, A. van Hoorn, M. Villaviencio, J. Walter, and F. Willnecker. How is performance addressed in DevOps? A survey on industrial practices. In Proceedings of the ACM/SPEC International Conference on Performance Engineering (ICPE 2019). ACM, 2019.

[2] International Workshop on Quality-aware DevOps (QUDOS). http://qudos-workshop.org/.

[3] The Seventh International Workshop on Load Testing and Benchmarking of Software Systems (LTB). http://ltb2018.eecs.yorku.ca/.

[4] F. Beck, A. Bergel, C. Bezemer, K.E. Isaacs. Visualizing systems and software performance - Report on the GI-Dagstuhl seminar for young researchers, July 9-13, 2018. PeerJ Preprints e27253v1, 2018.

REPORT: BIG DATA WORKING GROUP

Distributed big data processing and analytics applications demand a comprehensive end-to-end architecture stack consisting of big data technologies. However, there are many possible architecture patterns (e.g., Lambda, Kappa, or Pipeline architectures) to choose from when implementing the application requirements. A big data technology in isolation may be best performing for a particular application, but its performance in connection with other technologies depends on the connectors and the environment. Similarly, existing big data benchmarks evaluate the performance of different technologies in isolation, but no work has been done on benchmarking big data architecture stacks as a whole.

The Big Data Working Group was established to research this gap in current benchmarks. A focus is set on big data, NoSQL, and AI systems. The group holds weekly meetings, with internal calls and open, public calls alternating. For the open calls, presentations on big data benchmarking, big data systems, performance tuning, and related research are invited. These calls are well attended and received. Among the presented topics in 2017/2018 were PolyBench, SLAB, MLPerf, SparkBench, BigBench V2, and OpenML. Presentations can be found on the Big Data working group’s website.

A recent ICPE 2018 vision paper, presented by members of the Big Data Working Group, proposes a new category of benchmark, called ABench, to fill this gap and discusses key aspects necessary for the performance evaluation of different big data architecture stacks. At the moment the Big Data group is working actively on the proof of concept implementation of the benchmark. ABench uses the flexibility of Docker container technology and Kubernetes, which helps to deploy and manage the different benchmark components and systems under test on both cloud and on-premise environments. Additional tools like Helm and Kubernetes Operator are utilized to automate the deployment process. Using these technologies help us build a standardized benchmark infrastructure that is independent of the system under test and do not require any specific driver modifications in order to be compatible with emerging processing and storage technologies.

Tilmann Rabl (bankmark), Meikel Poess (Oracle),
Rekha Singhal (Tata Consultancy Services)

https://research.spec.org/en/working-groups/rg-big-data.html

WORKFLOW TRACE ARCHIVE

With the diversity the cloud offers, so do the use cases of its users. Workflows are a common way to represent programs in cloud environments. Workflows are composed of tasks with dependency constraints between them. This makes scheduling such workflows a non-trivial challenge. To be able to validate a new approach in different scenarios or quickly develop a prototype, many researchers use simulation. These simulations are often driven by realistic traces, representing a realistic workload.

Recent work by Amvrosiadis et al. [1] in ATC 2018 showed that a lot of research is biased towards the popular Google traces. This is troublesome as perceived state-of-the-art may perform poorly when encountering situation not covered by such traces. This has also been observed in prior work from SPEC [2]. It is therefore no surprise that the adoption of new scheduling systems by industry is poor [3]. Even though Amvrosiadis et al. introduce new traces, the SPEC RG believes the diversity of open-source traces is too little. Currently, there are too few traces from existing and emerging fields such as IoT, Edge, and Serverless. Further, the amount of information in the available traces is often limited, reducing credibility of experiments using these.

To address the situation, the SPEC RG Cloud group is actively working on gathering traces from different domains and fields. By working together with current and new partners, we attempt to convince all parties that sharing data and problems will aid all involved parties in better understanding the challenges modern cloud computing faces. SPEC believes in open-access; traces obtained will be made public for everyone to benefit from.

Laurens Versluis (Vrije Universiteit Amsterdam)

[1] G. Amvrosiadis, J. W. Park, G. R. Ganger, G. A. Gibson, E. Baseman, N. DeBardeleben: On the diversity of cluster workloads and its impact on research results. USENIX Annual Technical Conference 2018: 533-546 https://www.usenix.org/conference/atc18/presentation/amvrosiadis.

[2] L. Versluis, M. Neacsu and A. Iosup: A Trace-Based Performance Study of Autoscaling Workloads of Workflows in Datacenters. IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), Washington, DC, 2018, pp. 223-232. doi: 10.1109/CCGRID.2018.00037 https://ieeexplore.ieee.org/document/8411026 .

[3] G. Andreadis, L. Versluis, F. Mastenbroek, and A. Iosup. 2018. A reference architecture for datacenter scheduling: design, validation, and experiments. International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '18). IEEE Press, Piscataway, NJ, USA, Article 37, 15 pages. https://atlarge-research.com/Presentations/2018/sc18-reference-architecture.pdf.

COME TO THE SERVERLESS SIDE!

Microservices, containers, and serverless computing are part of a larger trend toward applications integrating many small, self-contained, and easy to manage components [1]. Embodying serverless computing in the cloud, Function-as-a-Service (FaaS) platforms employ existing, state-of-the-art container technology and microservices-based architectures, to give users the benefits of running complex applications, without the need for operational and distributed systems expertise.

Tens of FaaS platforms have emerged in the past couple of years, proposing and implementing widely varying designs. Understanding their key architectural and operational properties is important to selecting and tuning designs, and to identifying open research challenges and good design practices for the future. Yet, the community has limited understanding of FaaS platforms – due to the amount, complexity, and novelty of such platforms, and to their often closed-source nature.

Addressing this problem, we – the serverless activity within the Cloud working group – design a reference architecture as a layered superset of the components found through a comprehensive investigation of over 50 relevant serverless and related platforms. We illustrate how this reference architecture captures real-world architectures by mapping all platforms to its components and layers – we illustrate this process through a representative selection of popular platforms, including the serverless-specific platforms AWS Lambda, IBM OpenWhisk, and Platform9 Fission, and the generic platforms Kubernetes and Apache AirFlow.

Based on these mappings, we show common patterns of utilizing the components in the reference architecture—from the deployment and execution of FaaS functions to the orchestration of complex function workflows. We further encourage more in-depth research, by presenting both academic and engineering challenges derived from our findings. Our next steps include the development of a benchmark prototype for serverless plattforms [2].

Feel free to contact us at rgcloud@spec.org, if you are interested to contribute.

Erwin van Eyk (Platform9 & Delft University of Technology)

[1] Erwin van Eyk, Lucian Toader, Sacheendra Talluri, Laurens Versluis, Alexandru Uta, Alexandru Iosup. 2018. Serverless is more: From paas to present cloud computing. IEEE Internet Computing, 22(5), 8-17.

[2] Erwin van Eyk, Alexandru Iosup, Cristina L. Abad, Johannes Grohmann, and Simon Eismann. 2018. A SPEC RG Cloud Group's Vision on the Performance Challenges of FaaS Cloud Architectures. Companion of ACM/SPEC ICPE 2018.

SELECTED ABSTRACTS

How is performance addressed in DevOps? A survey on industrial practices

DevOps is a modern software engineering paradigm that is gaining widespread adoption in industry. The goal of DevOps is to bring software changes into production with a high frequency and fast feedback cycles. This conflicts with software quality assurance activities, particularly with respect to performance. For instance, performance evaluation activities, — such as load testing —, require a considerable amount of time to get statistically significant results.

We conducted an industrial survey to get insights into how performance is addressed in industrial DevOps settings. In particular, we were interested in the frequency of executing performance evaluations, the tools being used, the granularity of the obtained performance data, and the use of model-based techniques. The survey responses, which come from a wide variety of participants from different industry sectors, indicate that the complexity of performance engineering approaches and tools is a barrier for widespread adoption of performance analysis in DevOps. The implication of our results is that performance analysis tools need to have a short learning curve, and should be easy to integrate into the DevOps pipeline.

Cor-Paul Bezemer, Simon Eismann, Vincenzo Ferme, Johannes Grohmann, Robert Heinrich, Pooyan Jamshidi, Weiyi Shang, André van Hoorn, Monica Villaviencio, Jürgen Walter, and Felix Willnecker. How is performance addressed in DevOps? A survey on industrial practices. In Proceedings of the ACM/SPEC International Conference on Performance Engineering (ICPE 2019). ACM, 2019.

Performance Oriented Dynamic Bypassing for Intrusion Detection Systems

Attacks on software systems are becoming more and more frequent, aggressive, and sophisticated. In 2018, with the changing threat landscape, organizations are looking at 'when' they will be attacked, not 'if'. An Intrusion Detection System (IDS) can help in defending against these attacks. The systems that host IDS require extensive computing resources as IDS tend to detect attacks under overloaded conditions wrongfully. With the end of Moore's law and the growing adoption of the Internet of Things, designers of security systems can no longer expect processing power to keep up the pace. This limitation requires ways to increase the performance of these systems without additional computation power. In this work, we present two dynamic and a static approach to bypass IDS for traffic deemed benign.

We provide a prototype implementation and evaluate our solution. Our evaluation shows promising results. Performance is increased up to the level of a system without an IDS. Attack detection is within the margin of error from the 100% rate. However, our findings show that dynamic approaches perform best when using software switches. The use of a hardware switch reduces the detection rate and performance significantly.

Lukas Iffländer, Jonathan Stoll, Nishant Rawtani, Veronika Lesch, Klaus-Dieter Lange, and Samuel Kounev. 2019. Performance Oriented Dynamic Bypassing for Intrusion Detection Systems. ACM/SPEC ICPE '19, April 7–11, 2019, Mumbai, India. ACM, New York, NY, USA, 8 pages.

SELECTED ABSTRACTS (CONT.)

Methodological Principles for Reproducible Performance Evaluation in Cloud Computing - A SPEC Research Technical Report

The rapid adoption and the diversification of cloud computing technology exacerbate the importance of a sound experimental methodology for this domain. This work investigates how to measure and report performance in the cloud, and how well the cloud research community is already doing it. We propose a set of eight important methodological principles that combine best-practices from nearby fields with concepts applicable only to clouds, and with new ideas about the time-accuracy trade-off.

We show how these principles are applicable using a practical use-case experiment. To this end, we analyze the ability of the newly released SPEC Cloud IaaS benchmark to follow the principles, and showcase real-world experimental studies in common cloud environments that meet the principles.

Last, we report on a systematic literature review including top conferences and journals in the field, from 2012 to 2017, analyzing if the practice of reporting cloud performance measurements follows the proposed eight principles. Worryingly, this systematic survey and the subsequent two-round human reviews, reveal that few of the published studies follow the eight experimental principles.

We conclude that, although these important principles are simple and basic, the cloud community is yet to adopt them broadly to deliver sound measurement of cloud environments.

Alessandro Vittorio Papadopoulos, Laurens Versluis, André Bauer, Nikolas Herbst, Jóakim von Kistowski, Ahmed Ali-Eldin, Cristina Abad, J. Nelson Amaral, Petr Tuma, and Alexandru Iosup. Methodological Principles for Reproducible Performance Evaluation in Cloud Computing - A SPEC Research Technical Report. April 2019. https://research.spec.org/publications.html

Chameleon: A Hybrid, Proactive Auto-Scaling Mechanism on a Level-Playing Field

Auto-scalers for clouds promise stable service quality at low costs when facing changing workload intensity. The major public cloud providers provide trigger-based auto-scalers based on thresholds. However, trigger-based auto-scaling has reaction times in the order of minutes. Novel auto-scalers from literature try to overcome the limitations of reactive mechanisms by employing proactive prediction methods. However, the adoption of proactive auto-scalers in production is still very low due to the high risk of relying on a single proactive method.

This paper tackles the challenge of reducing this risk by proposing a new hybrid auto-scaling mechanism, called Chameleon, combining multiple different proactive methods coupled with a reactive fallback mechanism. Chameleon employs on-demand, automated time series-based forecasting methods to predict the arriving load intensity in combination with run-time service demand estimation to calculate the required resource consumption per work unit without the need for application instrumentation.

We benchmark Chameleon against five different state-of-the-art proactive and reactive auto-scalers in three different private and public cloud environments. We generate five different representative workloads each taken from different real-world system traces. Overall, Chameleon achieves the best scaling behavior based on user and elasticity performance metrics, analyzing the results from 400 hours aggregated experiment time.

André Bauer, Nikolas Herbst, Simon Spinner, Ahmed Ali-Eldin, and Samuel Kounev. Chameleon: A Hybrid, Proactive Auto-Scaling Mechanism on a Level-Playing Field. IEEE Transactions on Parallel and Distributed Systems, 30(4):800 – 813, April 2019, IEEE.