February 12, 2020
Grascomp/EuroSys joint workshop

“Achieving high impact in experimental computer science and systems.”

UCLouvain Brussels campus, Auditoire 42 A
Jardin Martin V, 1200 Woluwe-Saint-Lambert
Directions: https://goo.gl/maps/o79JDoXCS3PpSmWM6

The location is just a short walk from Metro station "Alma" (line 1) directly reachable from Brussels central or Brussels Schumann.

Organizer/contact: Prof. Etienne Rivière, UCLouvain -- etienne.riviere@uclouvain.be

Introduction

This workshop is organized in cooperation between Grascomp, the Belgian doctoral school in computer science and EuroSys, the European chapter of the ACM Special Interest Group in Computer Systems (SIGOPS).

The goal of this one-day workshop is to discuss how to reach a high impact in computer systems (broadly construed) research. It will feature two keynotes, one panel on experimental reproducibility, and 12 research presentations featuring recent research work in systems with a potential for high impact.

Important: Participation to the workshop for Grascomp PhD students is free, but **registration is mandatory!** Grascomp will cover your travel (2nd class train + metro) and the food during the workshop day. Post-docs and academics are also welcome to join, and must also register, but their costs will be reimbursed only if funding allows.

Registration for Grascomp participants is mandatory with deadline Sunday 9 2020, 23:59 UTC: https://forms.gle/prpoGoZB4Kdf6trT9

The abstracts and titles are listed after the program. All research talks and panels will be 15 minutes followed by 5 minutes of Q&A.

PROGRAM

8:30-9:00 Welcome and distribution of badges

9:00-10:00 **Keynote 1** -- Prof. Ivan Beschastnikh, University of British Columbia (Canada)
Successfully publishing your research (link to slides)

10:00-10:30 Coffee break

10:30-12:30 Research session 1 -- Data centers, clouds and storage
Session chair: Andrea Rosa, USI, Switzerland

- **Henri Maxime (Max) Demoulin, University of Pennsylvania, USA:**
  - Building data center runtimes in the age of acceleration
- **Marios Kogias, EPFL, Switzerland:**
  - R2P2: Making RPCs first-class datacenter citizens
- ~~**Djob Mvondo, Université Grenoble Alpes, France:**~~
  - ~~FaasCache - Speeding up Function As A Service Execution speed~~
- **Vladimir Podolskiy, Technische Universität München, Germany:**
  - Topology-awareness in distributed container-orchestration systems on the example of Kubernetes
- **Leander Jehl, University of Stavanger, Norway**
  - BBChain: Blockchain technology for transparent degree certificates
- **Vero Estrada-Galiñanes, University of Stavanger, Norway**
  - Entangled Merkle Trees for Reliable Off-chain Distributed Storage Systems

12:30-14:00 Lunch

14:00-15:00 **Keynote 2** -- **Prof. Alexandru Iosup, Vrije Universiteit Amsterdam and Young Royal Academy of Arts and Sciences (Netherlands)**
Will It Rain Today? Understanding the Weather of Computing Clouds, Before it Happens

15:00-16:00 Research session 2 -- Machine learning and data processing
Session chair: Michał Król, UCLouvain

- **Stefanos Laskaridis and Stylianos Venieris, Samsung AI center, Cambridge, UK**
  - Improving the performance of convolutional neural networks
- **Thaleia Dimitra Doudali, Georgia Institute of Technology, USA**
  - Kleio: a Hybrid Memory Page Scheduler with Machine Intelligence
- **Jean-Sebastien Legare, University of British Columbia, Canada**
  - Reproducible Cloud-Scale Genomics

16:00-16:30 Coffee break and refreshments

16:30-16:50 Feature presentation: **Andrew Quinn, University of Michigan, USA**

Artifacts evaluation, reproducible computer systems in practice.
With the contribution of: Kostis Kaffes, Amoghavarsha Suresh, Anjo Vahldiek-Oberwagner, and Abhinav Jangda

> *In this presentation, we will report on our experience in participating in the SOSP 2019 artifact evaluation committee, discuss how artifact evaluation can help computer systems research and improve the impact of our contributions. We will also advertise the organization of the upcoming OSDI 2020 artifact evaluation and encourage workshop participants to volunteer.*

16:50-17:50 Research session 3 -- Dependability and reliability
Session chair: Guillaume Rosinosky, UCLouvain

- **Nikos Vasilakis, MIT, USA**
  - Multi-library Program Dialysis with Lya
- **Poonam Yadav, University of York, UK**
  - Network service dependencies in commodity internet-of-things devices
- **Guthemberg Silvestre, ENAC, France**
  - Reliable distributed systems for dynamic environments

17:50 Closing

The workshop will be followed by an **informal drink offered by the organization**.

Note for EuroSys shadow PC participants: dinner is on your own.

Keynotes

Keynote 1 -- Prof. **Ivan Beschastnikh**, University of British Columbia (Canada)

Title: Successfully publishing your research

Abstract: Writing a research paper that will be accepted by a top computer science conference (or journal) is a skill that takes years of training to develop. To a casual observer it may seem that doing good research necessarily leads to a published paper. So, investing time into doing good research is part of the training to write an acceptable paper. My experience indicates that this is not entirely the case. Academic publishing is a social enterprise that hinges on peer review, a process that is evolving, and is often an enigmatic barrier to graduate students around the world. In this talk I will discuss the paper writing process, focusing on peer review and its implications for publishing research in experimental computer science and systems.

Keynote 2 -- Prof. **Alexandru Iosup**, Vrije Universiteit Amsterdam and Young Royal Academy of Arts and Sciences (Netherlands)

Title: Will It Rain Today? Understanding the Weather of Computing Clouds, Before it Happens

Slides and abstract available online.

Abstract: Cloud computing services play an important role in today's modern society. They enable daily operation and advances in key application domains, from banking to e-commerce, from science to gaming, from governance to education. Combining technology developed since the 1960s (e.g., modes of resource sharing) with new paradigms that could only have emerged in the 2010s (e.g., FaaS), they promise to enable unprecedented efficiency and seamless access to services for many. However successful, we cannot take the cloud for granted: its core does not yet rely on sound principles of science and design, its engineering is often based on hacking, and there have already been worrying signs of unstable operation. In this talk, we posit that we can address the current challenges by focusing on the relatively large complex of systems (that is, systems of systems or even ecosystems), and by increasing and focusing the effort put into performance experiments, load testing, and benchmarking. We contrast this to the current focus on single or relatively small systems, and on experimentation that is not always principled. We show examples of how our approach could work in practice, presenting (i) results related to performance variability, (ii) discovery methods that feed into the engineering of future load testing and benchmarking frameworks, and (iii) processes that could improve the reproducibility and credibility of experimental results in this field. This leads us to formulate the vision of a community-wide effort to create the Distributed Systems Memex, to share and preserve operational and especially performance traces collected from the distributed systems

that currently underpin our society. Part of this work has been conducted in the international collaboration provided by the SPEC RG Cloud Group.

Biography: Prof.dr.ir. Alexandru Iosup is full tenured professor and University Research Chair at Vrije Universiteit Amsterdam, and a member of the Young Royal Academy of Arts and Sciences of the Netherlands. He received his PhD in computer science from TU Delft, the Netherlands. He is the chair of the Massivizing Computer Systems research group at the VU and of the SPEC-RG Cloud group. His work in distributed systems and ecosystems has received prestigious recognition, including the 2016 Netherlands ICT Researcher of the Year, the 2015 Netherlands Higher-Education Teacher of the Year, and several SPEC community awards SPECtacular (last in 2017). He can be contacted at A.Iosup@vu.nl or @AIosup.

Research presentations (12)

Research session 1 -- Data centers, clouds and storage

Henri Maxime (Max) Demoulin, University of Pennsylvania, USA

Title: Building data center runtimes in the age of acceleration

Abstract: Datacenters are increasingly equipped with specialized hardware, such as smart NICs. To allow programmers to make use of those, datacenter operators want fast dataplane stacks that can not only provide high performance to applications, but also high resource utilization across clusters. Such stacks typically come in the form of a library OS linked to the application or a microkernel responsible for interfacing with specialized hardware and dispatching requests to application threads or processes.
Each combination of application and workload benefit from specific request dispatch policies, that usually require application-level information. However, current state-of-the-art dataplane solutions are designed ad-hoc for a set of application and expected workload. Those systems typically embed domain-specific knowledge and thus force a strong tie between the libOS and the application. Such design is detrimental to the adoption of fast dataplanes: separation of concerns between actors with different constraints (high utilization for datacenter operators, high performance for application owners) is desirable to satisfy both objectives.
To fill this gap, we design Persephone, a fast dataplane platform where datacenter operators provide the libOS and programmers provide an annotated DAG of their application. This DAG embeds domain specific information used by the libOS to perform cross-layer optimizations. Persephone provides the flexibility required to achieve optimal performances for a variety of application and workloads, while allowing datacenter operators to arbitrate the resource consumption of contending applications.

We demonstrate the need for Persephone by detailing how request dispatching for three representative datacenter applications (REST API, key-value store, and network function) benefits from cross-layer optimization.

Marios Kogias, EPFL, Switzerland

Title: R2P2: Making RPCs first-class datacenter citizens

Abstract: Remote Procedure Calls are widely used to connect datacenter applications with strict tail-latency service level objectives in the scale of µs. Existing solutions utilize streaming or datagram-based transport protocols for RPCs that impose overheads and limit the design flexibility. Our work exposes the RPC abstraction to the endpoints and the network, making RPCs first-class datacenter citizens and allowing for in-network RPC scheduling. We propose R2P2, a UDP-based transport protocol specifically designed for RPCs inside a datacenter. R2P2 exposes pairs of requests and responses and allows efficient and scalable RPC routing by separating the RPC target selection from request and reply streaming. Leveraging R2P2, we implement a novel join-bounded-shortest-queue (JBSQ) RPC load balancing policy, which lowers tail latency by centralizing pending RPCs in the router and ensures that requests are only routed to servers with a bounded number of outstanding requests. The R2P2 router logic can be implemented either in a software middlebox or within a P4 switch ASIC pipeline. Our evaluation, using a range of microbenchmarks, shows that the protocol is suitable for µs-scale RPCs and that its tail latency outperforms both random selection and classic HTTP reverse proxies. The P4-based implementation of R2P2 on a Tofino ASIC adds less than 1µs of latency whereas the software middlebox implementation adds 5µs latency and requires only two CPU cores to route RPCs at 10 Gbps line-rate. R2P2 improves the tail latency of web index searching on a cluster of 16 workers operating at 50% of capacity by 5.7× over NGINX. R2P2 improves the throughput of the Redis key-value store on a 4-node cluster with master/slave replication for a tail-latency service-level objective of 200µs by more than 4.8× vs. vanilla Redis.

Djob Mvondo, Université Grenoble Alpes, France

Title: FaasCache - Speeding up Function As A Service Execution speed
Unfortunately cancelled due to french train workers strike :(

Abstract: FaaS which stands for Function As A Service is an emerging cloud model that rapidly gains market shares and attracts new consumers due to its pricing models. With FaaS, the customer writes his code in the form of functions and pushes to the cloud FaaS platform which abstracts the underlying infrastructure and manages resource provisioning. The customer pays for the execution time and memory consumed by the function. Additionally, these functions can be registered as a response to events or feeds which can be external or internal to the cloud platform. These functions being stateless, to keep track of processed data between different

invocations/calls of functions, they usually persist the latter to external databases such as Google Cloud FireStore or AWS Dynamo DB. As a result, the overall execution time lengthens as the interaction with these storage platforms is usually synchronous. This talk presents an ongoing work that aims at introducing a distributed cache system whose main goal is to shorten the interaction between FaaS functions and the external databases. Designing such a cache system can be more tricky than it seems due to underlying constraints and optimization.

Vladimir Podolskiy, Technische Universität München, Germany

Title: Topology-awareness in distributed container-orchestration systems on the example of Kubernetes

Abstract: This talk will be about a work-in-progress on the hardware topology-aware container-orchestration platforms such as Kubernetes. In particular, the talk will be focused on NUMA-aware scheduling of applications in Kubernetes.

Leander Jehl, University of Stavanger, Norway

Title: BBChain: Blockchain technology for transparent degree certificates

Abstract: In this talk, I will present ongoing work from the BBChain project, which aims to build a decentralized database of academic degree certificates. Until today, such certificates are issued on paper and are subject to human error and fraud. On the other hand, degree certificates are a common, if not standard example for digital credentials. Especially Blockchain technology promises to provide highly available, publicly verifiable, and tamperproof credentials. Several early adopters already use Blockchain technology to store emission proofs of digital certificates. I will discuss the usefulness of blockchain technology for digital degree certificates. In particular, how blockchain-based solutions may prevent fraud within the certification system and increase transparency of the certification process. Last year's reports of bribery in the admissions processes at US academic institutions show the importance of addressing these issues. In the BBChain project, we rely on three mechanisms which are all enabled by blockchain technology and which help to prevent or detect such fraud: secure timestamping, causal event composition, and hierarchy modeling.

Vero Estrada-Galiñanes, University of Stavanger, Norway

Title: Entangled Merkle Trees for Reliable Off-chain Distributed Storage Systems

Abstract: In the near future, digitization will transform daily operations and paper-based documents may disappear. Thus, our society needs systems that can guarantee long-term storage with an affordable cost. In particular, our ongoing project BBChain (bbchain.no) brings into focus the opportunities that result from storing and computing over degree certificates in a global decentralized system.
We investigate how the BBChain system can be realized in a decentralized ecosystem and particularly how to protect data against arbitrary failures. In principle, a decentralized, distributed, and immutable database can be built atop a blockchain. But to assure the economic viability and the scalability of the solution, BBChain needs to minimize the information stored on blockchains. Only diploma integrity information will be written on the blockchain (on-chain data), whereas the diploma documents will be stored off-chain. Novel off-chain solutions that are built on peer-to-peer systems have many challenges to overcome while working towards the provision of a highly reliable service. For example, poor maintenance is an insidious problem that is more difficult to solve in p2p systems.
In this talk I will share insights on a reliable service based on entanglement codes and provide the big picture of our proof-of-concept built on top of Swarm.
Swarm is a native storage layer of the Ethereum blockchain, which aims at providing redundant (off-chain) storage for dapp code, data, as well as, blockchain and state data. Our proof-of-concept improves Swarm's resilience to failures and seeks to balance repairs and storage costs.

Research session 2 -- Machine learning and data processing

Stefanos Laskaridis and Stylianos Venieris, Samsung AI center, Cambridge, UK

Title: Improving the performance of convolutional neural networks

Abstract: Convolutional neural networks (CNNs) have recently become the state-of-the-art in a diversity of AI tasks, ranging from image classification to speech processing. Despite their popularity, CNN inference comes at a high computational cost. One way of alleviating this and accelerating CNN inference is to exploit the difference in the classification difficulty among samples and early-exit at different stages of the network. Existing studies on early exiting focus on the training scheme, without taking into account the application-level requirements or the deployment platform capabilities. This work presents a novel methodology for generating high-performance early-exit networks by co-optimising the placement of intermediate exits together with the early-exit strategy at inference time. Furthermore, an efficient design space exploration algorithm is proposed which enables the traversal of a large number of alternative architectures and generates the highest-performing design, tailored to the specific use-case requirements and target hardware. Quantitative evaluation shows that our system outperforms state-of-the-art early-exit schemes and pushes further the performance of highly-optimised

hand-crafted early-exit architectures, while delivering significant speedup over hand-tuned lightweight models on imposed latency-driven SLAs for embedded devices.

Thaleia Dimitra Doudali, Georgia Institute of Technology, USA

Title: Kleio: a Hybrid Memory Page Scheduler with Machine Intelligence

Abstract: The increasing demand of big data analytics for more main memory capacity in datacenters and exascale computing environments is driving the integration of heterogeneous memory technologies. The new technologies exhibit vastly greater differences in access latencies, bandwidth and capacity compared to the traditional NUMA systems. Leveraging this heterogeneity while also delivering application performance enhancements requires intelligent data placement. The focus of this talk will be Kleio, a page scheduler with machine intelligence for applications that execute across hybrid memory components. Kleio is a hybrid page scheduler that combines existing, lightweight, history-based data tiering methods for hybrid memory, with novel intelligent placement decisions based on deep neural networks. Performance evaluation indicates that Kleio reduces on average 80% of the performance gap between the existing solutions and an oracle with knowledge of future access pattern. Finally, the talk will conclude with future research directions focusing on the impact of the data movement cost to application performance and its effect on efficient data placement decisions.

Jean-Sebastien Legare, University of British Columbia, Canada

Title: Reproducible Cloud-Scale Genomics

Abstract: Answering genetic queries with common bioinformatics software requires core-years of compute time. Many bioinformatics studies are conducted from the same datasets, so the potential gains for reuse of final and partial results of past experiments are high. Yet, common data flow frameworks do not easily track provenance of results, which leads to ad hoc reuse of results, low fidelity in drawing conclusions, and low reproducibility of the experiments. We present a framework, Bunnies, designed to improve reproducibility of experiments in bioinformatics, and reduce compute costs by taking advantage of provenance information.

Research session 3 -- Dependability and reliability

Nikos Vasilakis , MIT, USA

Title: Multi-library Program Dialysis with Lya

Abstract: Applications today rely on hundreds of libraries, to the point where code written by their nominal developers is only a small fraction of their total line count. Despite its benefits, this over-reliance creates many challenges, precisely due to the lack of knowledge and visibility into library internals---for example, in the presence of deeply-nested third-party code, security auditing and performance profiling, already challenging in monolithic applications, become extremely difficult.
To address these challenges, we present program dialysis, a dynamic analysis technique specifically tailored to applications with many third-party libraries. Combining name shadowing, context wrapping, and transformation of the underlying dependency graph, program dialysis automates dynamic fragmentation, analysis, and reassembly of programs at the level of individual libraries during program execution. It bolts onto existing languages, enabling analysis expressions in the source language, with only a few lines of analysis-specific code. Our dialysis prototype, Lya, targets the JavaScript ecosystem counting over 1M libraries for the web, server, and mobile. We develop a series of case-studies that motivate Lya's design, and demonstrate how Lya allows the analysis of both individual libraries as well as multi-library programs with low developer effort and performance overhead: insightful analyses can be expressed in a few lines of code, add a minimal increase in the load latency of individual libraries, and scale to large programs with hundreds of libraries.

Poonam Yadav, University of York, UK

Title: Network service dependencies in commodity internet-of-things devices

Abstract: We continue to see increasingly widespread deployment of IoT devices, with apparent intent to embed them in our built environment likely to accelerate if smart city and related programmes succeed. We are concerned with the ways in which current generation IoT devices are being designed in terms of their ill-considered dependencies on network connectivity and services. Through our research work, our hope is to provide evidence that such dependencies need to be better thought through in design, and better documented in implementation so that those responsible for deploying these devices can be properly informed as to the impact of device deployment (at scale) on infrastructure resilience. We believe this will be particularly relevant as we feel that commodity IoT devices are likely to be commonly used to retrofit ""smart"" capabilities to existing buildings, particularly domestic buildings.
To the existing body of work on network-level behaviour of IoT devices, we add (i) a protocol-level breakdown and analysis of periodicity, (ii) an exploration of the service and

infrastructure dependencies that will implicitly be taken in ""smart"" environments when IoT devices are deployed, and (iii) examination of the robustness of device operation when connectivity is disrupted. We find that many devices make use of services distributed across the planet and thus appear dependent on the global network infrastructure even when carrying out purely local actions. Some devices cease to operate properly without network connectivity (even where their behaviour appears, on the face of it, to require only local information, e.g., the Hive thermostat). Further, they exhibit quite different network behaviours, typically involving significantly more traffic and possibly use of otherwise unobserved protocols, when connectivity is recovered after some disruption.

Guthemberg Silvestre, ENAC, France

Title: Reliable distributed systems for dynamic environments

Abstract: In this talk, I will briefly present the ongoing doctoral research projects on reliable distributed systems for dynamic environments conducted in my team. The talk highlights some of our research's guidelines on mobile distributed systems and applications, such as the study of new distributed algorithms for emerging services running on a swarm of drones.