



# HPI Research Symposium 2021

Hasso Plattner Institute  
Potsdam | April 19 - 23, 2021

Design IT. Create Knowledge.



# Agenda

## Monday, April 19, 2021 Master's Program Day

Room: Zoom

13:00 - 13:30 **Welcome**

Prof. Dr. Tilmann Rabl, Head of Data Engineering Systems Group, Hasso Plattner Institute, Potsdam  
**Successful Systems Research**

13:30 - 14:30 **Session I**

Master Project Renard: Julian Hugo, Spoorthi Kashyap, Nataniel Müller, Justus Zeinert,  
**Improving Network Integration Algorithms for Drug Predictions**

Master Project Weske: Finn Klessascheck, Tom Lichtenstein, Simon Siegert,  
**Process Mining in Personalized Medicine**

Henrik Wenck,  
**A Bayesian Analysis of the Effectiveness of Non-Pharmaceutical COVID19 Interventions**

Master Project Baudisch: Oliver Adameck, Lukas Fritzsche, Jonas Noack,  
**Super Fast Fabrication of 3D Models—by Folding Laser-Cut Foam Core**

14:30 - 15:00 **Coffee Break**

15:00 - 16:00 **Session II**

Lars Jonas Bollmeier, Björn Daase,  
**Maximizing Persistent Memory Bandwidth Utilization for OLAP Workloads**

Christian Flach,  
**Call Graphs for Live Programming - Implementing Call Tracing in Babylonian/S based on a Survey of Property Extraction Techniques for Dynamic Analysis**

Master Project Friedrich: Nicolas Klodt, Lars Seifert, Arthur Zahn,  
**Chromatic Correlation Clustering: Approximation and Heuristics**

Master Project Naumann: Tobias Bredow, Jona Otholt, Emanuel Metzenthin,  
**Multimodal Analysis for Cultural Data**

## Tuesday, April 20, 2021 Future SOC Lab and Data Lab Day

Room: Zoom

- 9:15 - 9:20     **Opening Words**  
Prof. Dr. Andreas Polze, Head of Operating Systems and Middleware Group and  
Speaker of the HPI Research School  
Welcome to the 22nd Future SOC Lab Day
- 9:20 - 10:20    **Session I**  
Jeroen Mackenbach, Dell Technologies  
Why the edge is losing its... edge (and how to get it back)  
Dr. Nitinder Mohan, Technical University of Munich  
Reality Check on Edge Computing
- 10:20 - 10:30   **Coffee Break**
- 10:30 - 12:00   **Session II**  
Stefan Halfpap, Hasso Plattner Institute, Potsdam  
Query-Driven Partial Database Replication  
Dario Bertazioli, University of Milan-Bicocca  
Fast and non invasive diagnosis of SARS-CoV-2 via raman spectroscopy and deep  
learning  
Jossekin Beilharz, Hasso Plattner Institute, Potsdam  
Edge computing for railway systems: Dependable V2X-communication at level  
crossings
- 12:30            **End of Future SOC Lab Day**

## Wednesday, April 21, 2021 International Doctoral Symposium

Room: Zoom

9:00 - 10:00 **Podium with Research School Alumni**

Hosts:

Prof. Dr. Melissa Densmore, Human-Computer Interaction, University of Cape Town

Prof. Dr. Felix Naumann, Head of Information Systems Group and Speaker of the HPI Research School

Alumni on the panel: Dr. Julian Risch (HPI DSE), Dr. Shikoh Gitau (UCT), Dr. Michael Perscheid (HPI SSE), Dr. Naama Kraus (Technion), Dr. Zhan Qianyi (Nanjing University)

10:00 - 10:15 **Coffee Break**

10:15 - 11:00 **Session I**

Christian Adriano, Ph.D. Student, Hasso Plattner Institute, Potsdam  
Towards Causal Bandits to Explore, Exploit, and Expire Questions for Quickly Understanding Software Faults

Yuan Zhi, Ph.D. Student, Nanjing University  
An Explainable Sampling Strategy for Video Action Recognition

Wang Yue, Ph.D. Student, Hasso Plattner Institute, Potsdam  
Desis: General Distributed Window Aggregation

11:00 - 11:15 **Coffee Break**

11:15 - 12:00 **Session II**

Muhammad Abdullah, Ph.D. Student, Hasso Plattner Institute, Potsdam  
Roadkill: Nesting Laser-Cut Objects for Fast Assembly

Enock Mbewe, Ph.D. Student, University of Cape Town  
Cost-aware Internet Security Decision Model

Shir Cohen, Ph.D. Student, Technion  
Not a COINcidence: Sub-Quadratic Asynchronous Byzantine Agreement WHP

12:00 - 12:30 **Lunch Break**

12:30 - 13:30 **Poster Session**

Virtual Meeting in Gather Town  
Students of research schools introduce their work

# Thursday, April 22, 2021 Research Symposium

Room: Zoom

- 15:00 - 15:15 **Opening Words**  
Prof. Dr. Christoph Meinel, President and CEO of HPI for Digital Engineering  
Prof. Dr. Andreas Polze, Head of Operating Systems and Middleware Group and Speaker of the HPI Research School  
Prof. Dr. Felix Naumann, Head of Information Systems Group and Speaker of the HPI Research School  
[Opening of the HPI Research Symposium](#)
- 15:15 - 15:45 **Keynote**  
Prof. Dr. Viktor Leis, Professor at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)  
[LeanStore: A High-Performance Storage Engine for Modern Hardware](#)
- 15:45 - 16:00 **Coffee Break**
- 16:00 - 16:45 **Elevator Pitches**  
Ph.D. students introduce their work
- 16:45 - 17:00 **Break**
- 17:00 - 17:30 **Invited Talk I**  
Prof. Dr. Boris Glavic, Professor of Computer Science, Illinois Institute of Technology  
[Efficient Uncertainty Tracking for Complex Queries with Attribute-level Bounds](#)
- 17:30 - 18:00 **Invited Talk II**  
Prof. Dr. Padhraic Smyth, Chancellor's Professor, Department of Computer Science, UC Irvine  
[Active and Bayesian Learning of Classifier Performance](#)
- 18:00 - 19:00 **Closing**

## Friday, April 23, 2021 Research Symposium

Room: Zoom

- 9:00 - 9:30 **Talk I**  
Prof. Dr. Yoav Etsion, Professor of Computer Science and Electrical Engineering, Speedata / Technion  
**Accelerating Data Analytics with the Speedata Analytics Processing Unit (APU): From Academia to a Startup**
- 9:30 - 10:00 **Talk II**  
Felix Eberhardt, Ph.D. Student, Hasso Plattner Institute, Potsdam  
**Memory Disaggregation - Impact on Operating System and Application**
- 10:00 - 10:30 **Talk III**  
Prof. Dr. Christian Dörr, Head of Cybersecurity and Enterprise Security Group, Hasso Plattner Institute, Potsdam  
**A Quantification of Cryptojacking at Internet Scale**
- 10:30 - 11:00 **Coffee Break**
- 11:00 - 11:30 **Talk IV**  
Prof. Dr. Philipp Hacker, Chair for Law and Ethics of the Digital Society, European New School of Digital Studies, European University Viadrina  
**Legal Constraints for Algorithmic Fairness**
- 11:30 - 12:00 **Talk V**  
Dr. Alexander Böhm, Chief Architect, SAP HANA/Cloud  
**Reliable, Scalable, Fast: SAP HANA's Journey to the Cloud**
- 12:00 - 12:15 **Closing Remarks**

# Hasso Plattner Institute for Digital Engineering

The Hasso Plattner Institute for Digital Engineering in Potsdam is Germany's university excellence center for IT-Systems Engineering. HPI is the only university institution in Germany to offer the Bachelor's and Master's degree in "IT-Systems Engineering" – a practical and engineering-oriented alternative to a conventional computer science study program. Current enrollment is at approximately 500 students. It carries out research noted for its high standard of excellence in its twelve topic areas. Research work is also conducted at the HPI Research School for Ph.D. candidates as well as at its branches in Cape Town, Haifa and Nanjing.

The HPI School of Design Thinking is Europe's first innovation school for university students. It is based on the Stanford d.school model and offers 240 places annually for a supplementary study program. Since 2012 the Hasso Plattner Institute has offered Massive Open Online Courses (MOOCs) at its own interactive platform, openHPI. The courses on IT topics have so far reached more than 170,000 users from over 180 countries. Free via the Internet, learners can access didactically prepared multi-media course materials and learn in close exchange with other course participants through social media.





# HPI Research Symposium

The HPI Research Symposium is the annual symposium of the HPI Research Schools. It outlines new trends in the area of Service-oriented Computing and Data Science and Engineering, featuring invited speakers and talks by selected research school members.

The HPI Research Symposium: Connecting Industry and Academia As the HPI Research Schools are an interdisciplinary effort of the HPI research groups, the HPI Research Symposium covers a wide range of topics related to service-oriented computing, data science and data engineering. Excellent speakers - both from industry and academia - leaders in their respective field of research, are invited to talk about their latest projects and resulting outcomes.



# The HPI Research Schools

## Research School

### Data Science and Engineering

The increasing abundance of data in science and in industry creates many challenges and opportunities. Data science has grown to be a foundational discipline in information technology, allowing new insights from data and creating ever more intelligent applications. Simultaneously, it is becoming increasingly difficult to collect, clean and deliver the vast amounts of data and apply and maintain complex data science processes. Targeting these challenges, the discipline of data engineering has become equally foundational.

The 2019 newly established research school "Data Science and Engineering" unites top PhD students in all areas of data-driven research and technology, including scalable storage, stream processing, data cleaning, machine learning and deep learning, text processing, data visualization and more. We apply our research to many different use cases across the participating interdisciplinary research groups, joining forces whenever possible.

## Research School

### Service-Oriented Systems Engineering

In October 2005, the HPI started its Research School on "Service-Oriented Systems Engineering", a graduate school based on the model of the DFG (German Research Foundation) "Graduiertenkolleg". The Vision of the Research School Design and implementation of service-oriented architectures impose numerous research questions from the fields of software engineering, system analysis and modeling, adaptability, and application integration. Service-Oriented Systems Engineering represents a symbiosis of best practices in object orientation, component-based development, distributed computing, and business process management. It provides integration of business and IT concerns. Service-Oriented Systems Engineering denotes a current research topic in the field of IT-Systems Engineering with high potential in academic research as well as in industrial application. Supported by an internationally renowned grant, Ph.D. students at our college participate in joint activities such as lectures, seminars, winter schools and workshops.

## The Members of the Research School

The professors of the HPI with their research groups are supporting pillars for our Ph.D. school. With its interdisciplinary structure, the research college on Service-Oriented Systems Engineering interconnects the HPI research groups and fosters close and fruitful collaborations. In context of the Research School, the different groups at HPI work on the following topics:

Enterprise Platform and Integration Concepts, Prof. Dr. h.c. Hasso Plattner  
Internet Technologies and Systems, Prof. Dr. Christoph Meinel  
Digital Health - Connected Healthcare, Prof. Dr. Bert Arnrich  
Human Computer Interaction, Prof. Dr. Patrick Baudisch  
Digital Health - Personalized Medicine, Prof. Dr. Erwin Böttinger  
Artificial Intelligence and Intelligent Systems, Prof. Dr. Gerard de Melo  
Computergraphics Systems, Prof. Dr. Jürgen Döllner  
Cybersecurity - Enterprise Security, Prof. Dr. Christian Dörr  
Algorithm Engineering, Prof. Dr. Tobias Friedrich  
System Engineering and Modeling, Prof. Dr. Holger Giese  
Software Architecture, Prof. Dr. Robert Hirschfeld  
IT Entrepreneurship, Prof. Dr. Katharina Hölzle  
Cybersecurity - Identity Management, Prof. Dr. Anja Lehmann  
Digital Health - Machine Learning, Prof. Dr. Christoph Lippert  
Information Systems, Prof. Dr. Felix Naumann  
Operating Systems and Middleware, Prof. Dr. Andreas Polze  
Data Engineering Systems, Prof. Dr. Tilmann Rabl  
Data Analytics and Computational Statistics, Prof. Dr. Bernhard Renard  
Design Thinking and Innovation Research, Prof. Dr. Falk Uebernickel  
Business Process Technolog, Prof. Dr. Mathias Weske  
School of Design Thinking, Prof. Ulrich Weinberg

On the website of the Research School, please find latest information about the Ph.D. students, their research interests, joint projects, and events:

<https://hpi.de/en/research/research-schools>

# The HPI International Research Schools

## HPI Research School at Nanjing University (NJU)

Nov 17 2011 marked the inauguration of the "HPI Research School at Nanjing University". HPI director Prof. Christoph Meinel welcomed the first five Chinese doctoral students and their supervising professors. The Ph.D. students that originally hail from the computer science department raised the total head count of the HPI Research School to 50. Together, the highly talented young scientists from all across the globe tackle various issues in the field of service-oriented IT systems. The main research areas of the Chinese Ph.D. students are "Mass Data Analytics and Knowledge Discovery". They will do their work in the former capitol of China, Nanjing, which is situated about 300 kilometers northwest of Shanghai. Since the HPI Research School in Nanjing began its work in 2011, 18 scholarship holders have successfully completed their PhD theses.

Prof. Dr. Song Fangmin is head of the research school at Nanjing University.



## HPI Research School at Technion - Israel Institute of Technology

At Israel's technical university in Haifa, 6 students are currently doing research at the "HPI Research School at Technion". Its goal is to allow a multidisciplinary group of researchers in the Computer Science and Electrical Engineering departments to rise up to the new challenges presented by the shift to Chip Multiprocessor Computing (CMP). The students are mentored jointly by professors from the Computer Science and Electrical Engineering departments. Since HPC-Technion began its work in 2010, 30 scholarship holders have successfully completed their PhD theses.

Prof. Dr. Idit Keidar is head of the research school at Technion.

### HPI Research School at University of Cape Town (UCT)

The Hasso Plattner Institute Research School at UCT, in Information and Communications Technology for Development (ICT4D), serves as a gateway for educating sub-Saharan African information technologists, and in the grand scheme, to become a leading centre of excellence in designing and understanding computational solutions (both algorithmic and user centric) and systems to address the UN Sustainable Development Goals. To this end, HPI-UCT shall promote and encourage: The education of the brightest PhD students from sub-Saharan Africa. The aim is to offer attractive funding to this pool of candidates for whom few other funding channels are available for study at South African universities. Development of strong research collaborations with others in the Hasso-Plattner Institute (HPI) as a leading world establishment in information technology. There are currently 16 postgraduates from different African countries working at UCT. Since HPI Research School at UCT began its work in 2009, 18 scholarship holders have successfully completed their PhD theses.

Prof. Dr. Melissa Densmore is head of the research school at UCT.

### HPI Research School at University of California (UCI)

The Hasso Plattner Institute (HPI) is very pleased to announce the opening of a new Research School branch for doctoral students at the University of California, Irvine (UCI), well known for its academic achievements and high quality education. The new "HPI Research Center in Machine Learning and Data Science at UC Irvine" aims to promote research and educational activities in these two fields between the two leading universities. Through this partnership, 10 UCI scholarship holders are currently jointly supervised by 8 professors from UCI's Donald Bren School of Information and Computer Sciences while closely being integrated into HPI's research activities.

Prof. Dr. Erik Sudderth is head of the research school at UCI.

# The HPI Future SoC Lab

The HPI Future SOC (Service-Oriented Computing) Lab is a cooperation of the Hasso Plattner Institute and the industrial partners EMC, Fujitsu, SAP and Hewlett Packard Enterprise. Its mission is to enable and promote exchange and interaction between the research community and the industrial partners. The Lab provides researchers with free of charge access to a complete infrastructure of state of the art hard- and software. This infrastructure includes components, which might be too expensive for an ordinary research environment. The offerings address researchers particularly from but not limited to the areas of computer science and business information systems. Main areas of research include cloud computing, parallelization, and In-Memory technologies.

Future SOC Lab Day – Spring 2021 On Tuesday, April 20, 2021, the projects of the previous Future SOC Lab period got a chance to present the results of their research activities. Additionally, selected requesters of new projects can expand their ideas. You are welcome to hand in a project proposal and to apply for the usage of the lab's resources.

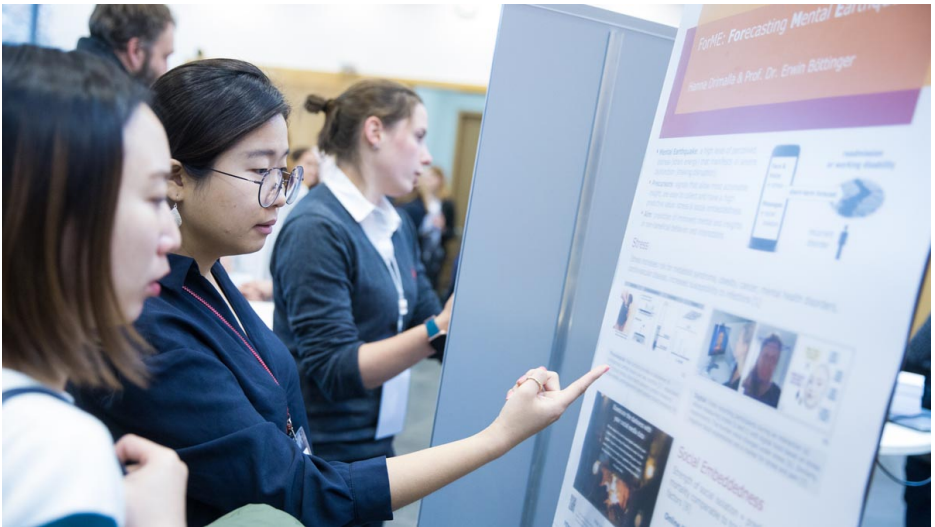
More information: <https://hpi.de/future-soc-lab>



# International Doctoral Symposium

Over the years the HPI Research School has been expanded to a state in which we are excited to – in addition to our members from Hasso Plattner Institute Potsdam, Germany – welcome colleagues from University of Cape Town (UCT), South Africa; the Technion, Israel; and Nanjing University, China. In order to foster collaboration and information exchange, we organized a Doctoral Symposium, so that Ph.D. students from HPI, UCT, Technion, and NJU will get to know each other both on the social and the academic levels. The event includes a mix of a social event, poster session, and demo presentations.

The poster session that dives into the research topics of the attending Ph.D. students, will follow a more informal and virtual event in Gather Town for exploring other people's posters and explaining yours!



# Research School Chairs



**Prof. Dr.  
Felix Naumann**

Head of Information Systems  
Group and Speaker of the HPI  
Research School

## Opening of the HPI Research Symposium

Felix Naumann studied mathematics, economy, and computer sciences at the University of Technology in Berlin. He completed his PhD thesis on “Quality-driven Query Answering” in 2000. In 2001 and 2002 he worked at the IBM Almaden Research Center on topics of data integration. From 2003 – 2006 he was assistant professor for information integration at the Humboldt-University of Berlin. Since then he holds the chair for information systems at the Hasso Plattner Institute at the University of Potsdam in Germany. He is editor-in-chief of the Information Systems journal; his research interests are in data profiling, data cleansing, and text mining.



**Prof. Dr.  
Andreas Polze**

Head of Operating Systems  
and Middleware Group and  
Speaker of the HPI Research  
School

## Welcome to the 22nd Future SOC Lab Day

Prof. Dr. Andreas Polze is the Operating Systems and Middleware Professor at the Hasso Plattner Institute at University Potsdam, Germany. He is also the speaker of the HPI Research School and member of the steering committee of HPI’s Future SOC Lab. Andreas received a doctoral degree from Freie University Berlin, Germany, in 1994 and a habilitation degree from Humboldt University Berlin in 2001, both in Computer Science. At HPI, his research focuses on architectures of operating systems, on component-based middleware, as well as on predictable distributed and cloud computing. Andreas Polze was visiting scientist with the Dynamic Systems Unit at Software Engineering Institute, at Carnegie Mellon University, Pittsburgh, USA, where he worked on real-time computing on standard middleware (CORBA) and with the Real-Time Systems Laboratory at University of Illinois, Urbana-Champaign. Current research interests include Predictable Service Computing, Adaptive System Configuration, and End-to-End Service Availability for standard middleware platforms. Together with Charité, GETEMED, and Deutsche Telekom, he has run the Fontane telemedicine project. Joint research with SAP has investigated porting HANA to new processor architectures.



# Speakers

## Roadkill: Nesting Laser-Cut Objects for Fast Assembly

Roadkill is a software tool that converts 3D models to 2D cutting plans for laser cutting—such that the resulting layouts allow for fast assembly. Roadkill achieves this by putting all relevant information into the cutting plan: (1) Thumbnails indicate which area of the model a set of parts belongs to. (2) Parts with exposed finger joints are easy to access, thereby suggesting to start assembly here. (3) Openings in the sheet act as jigs, affording assembly within the sheet. (4) Users continue assembly by inserting what has already been assembled into parts that are immediately adjacent or are pointed to by arrows. Roadkill maximizes the number of joints rendered in immediate adjacency by breaking down models into “subassemblies.” Within a subassembly, Roadkill holds the parts together using break-away tabs. (5) Users complete subassemblies according to their labels 1, 2, 3..., following 1 > 1 links to insert subassemblies into other subassemblies, until all parts come together. In our user study, Roadkill allowed participants to assemble layouts 2.4 times faster than layouts generated by a traditional pair-wise labeling of plates.



**Muhammad Abdullah**

Ph.D. Student, Hasso Plattner  
Institute, Potsdam

Muhammad Abdullah is a Ph.D. student at the Human Computer Interaction lab with Prof. Baudisch. His research focuses on enabling rapid prototyping using laser-cut parts. To facilitate this he is developing algorithms that can help speed up manual assembly of laser-cut models. He obtained his Bachelors in Electrical Engineering at the National University of Science and Technology and Masters in Computer Science and Engineering at Kyung Hee University.



## Towards Causal Bandits to Explore, Exploit, and Expire Questions for Quickly Understanding Software Faults

In this talk I will show how to prioritize questions during code inspection in order to identify and explain bugs quickly. I accomplish this via causal bandits, which combine causal inference models with Bayesian multi-armed bandits to efficiently cover the large space of bug locations. My results comprise real bugs from popular open source projects that were inspected by a crowd of 1600 programmers.

### Christian Adriano

Ph.D. Student, Hasso Plattner  
Institute, Potsdam

Christian is a Ph.D. candidate with degrees in Computer Engineering from the State University of Campinas and a Master in Software engineering from the University of California Irvine. He has a long experience in projects for the financial and energy sectors and has an active project management certification.

## Reliable, Scalable, Fast: SAP HANA's Journey to the Cloud

The database management systems market is moving to the cloud. Instead of provisioning, administering and tuning database systems on-premise, customers are looking for highly scalable, reliable, and self-tuning systems that are provided as a fully managed service. SAP HANA Cloud is SAP's cloud-based database offering.

In this talk, we highlight some of SAP HANA Cloud's key differentiators such as high-performance query processing, multi-modal processing features, multi-cloud support, as well as an integrated relational data lake. We also discuss the architectural evolution of the HANA cloud system, as well as related design aspects, and research challenges that are specific to the inherently stateful nature of database management systems.



**Dr. Alexander Böhm**

Chief Architect, SAP  
HANA/Cloud

Dr. Alexander Boehm is one of the chief architects working on SAP's HANA in-memory database management system. His focus is on performance optimization and holistic improvements of enterprise architectures, in particular application server/DBMS co-design. Prior to joining SAP in 2010, he received his PhD from the University of Mannheim, Germany, where he worked on the development of efficient and scalable applications using declarative message processing.



**Shir Cohen**

Ph.D. Student, Technion

## Not a COINcidence: Sub-Quadratic Asynchronous Byzantine Agreement WHP

King and Saia were the first to break the quadratic word complexity bound for Byzantine Agreement in synchronous systems against an adaptive adversary, and Algorand broke this bound with near-optimal resilience (first in the synchronous model and then with eventual-synchrony). Yet the question of asynchronous sub-quadratic Byzantine Agreement remained open. To the best of our knowledge, we are the first to answer this question in the affirmative. A key component of our solution is a shared coin algorithm based on a VRF. A second essential ingredient is VRF-based committee sampling, which we formalize and utilize in the asynchronous model for the first time. Our algorithms work against a delayed-adaptive adversary, which cannot perform after-the-fact removals but has full control of Byzantine processes and full information about communication in earlier rounds. Using committee sampling and our shared coin, we solve Byzantine Agreement with high probability, with a word complexity of  $O(n)$  and  $O(1)$  expected time, breaking the  $O(n^2)$  bit barrier for asynchronous Byzantine Agreement.

Shir is a PhD candidate in the Computer Science Department at the Technion, advised by Prof. Idit Keidar. Her research interests include distributed systems and algorithms. Specifically, she focuses on Byzantine behavior in distributed services and how to use cryptographic tools to make such services more resilient and efficient.

# Memory Disaggregation - Impact on Operating System and Application

Disaggregation of system resources offers various benefits, such as flexibility of provisioning, consolidation of workloads, higher limits for bursts of resource consumption. With ThymesisFlow an OpenCAPI-based memory disaggregation prototype is already available to experiment with. Several research questions arise with this new system architecture.

In this talk, we briefly tackle new challenges and the impacts on Operating systems and Applications. We will present a case for new memory placement considerations and show the first results of performance impacts on challenging workloads, such as In-Memory Databases.



**Felix Eberhardt**

Ph.D. Student, Hasso Plattner  
Institute, Potsdam

Felix Eberhardt is a researcher at the Operating Systems and Middleware Chair of Prof. Andreas Polze at the Hasso Plattner Institute in Potsdam. His research interests are performance analysis and optimization for workloads in scale-up systems. He participated in several industry projects with IBM, SAP as well as a European project (SSICLOPS).



## **Prof. Dr. Christian Dörr**

Head of Cybersecurity and Enterprise Security Group, Hasso Plattner Institute, Potsdam

# **A Quantification of Cryptojacking at Internet Scale**

As black-box machine learning models are now being deployed in many Cryptojacking - the covert, malicious mining of crypto currency on victim computers - is a recent phenomenon that has made it from the get-go into ENISA's list of top 15 Internet threats. In this talk, we look at how this new cyber criminal activity works, discuss the strategies criminals use to spread and manage their infections, and show how it fundamentally redefines the Internet threat landscape: instead of only targeting victims through their PCs and smart phones, cyber criminals have now begun to target the Internet infrastructure itself, thereby spreading out their efforts to millions of victims transparently and nearly effortless.

Christian Dörr ist Professor am Hasso Plattner Institut und der Universität Potsdam. Als Inhaber des Lehrstuhls für Cybersecurity and Enterprise Security umfasst seine Forschung sowohl technische Fragestellungen der Netzwerk- und Systemsicherheit, als auch organisatorische Aspekte, z.B. wie technische Sicherheitslösungen möglichst effektiv in Unternehmen eingebettet werden können. Das von ihm geleitete Cyber Threat Intelligence Lab entwickelt Lagebilder zu den Aktivitäten von verschiedenen Aktorengruppen sowie neuartigen Angriffsvektoren, und entwickelt Methoden um koordinierte, gezielte Angriffe in massiven Datensätzen und bei hohen Datengeschwindigkeiten aufzuspüren. Er ist Mitglied der Internet-Lagebild Expertengruppe der europäischen Sicherheitsagentur ENISA.

# Accelerating Data Analytics with the Speedata Analytics Processing Unit (APU): From Academia to a Startup



**Prof. Dr. Yoav Etsion**

Professor of Computer Science and Electrical Engineering, Speedata / Technion

As the volume of data created, collected, and consumed worldwide grows exponentially, the performance of data analytics software has become a bottleneck in many industries. State-of-the-art data analytics frameworks, including Spark, Presto, SAP HANA, Snowflake, Amazon Redshift, Google BigQuery, and Azure Synapse, struggle to process the growing data deluge.

In this talk I will present the Speedata Analytics Processing Unit (APU), the first hardware acceleration platform for data analytics workloads. Speedata's APU accelerates existing software frameworks and allows users to improve their performance/cost by orders-of-magnitude while maintaining their existing code base. In my talk I will describe the APU's highly programmable architecture and its support for high-bandwidth data fetching and processing of complex analytics queries. I will also show how the APU outperforms the open-source Apache Spark by almost 50x.

And it all started with an HPI research grant...

Yoav Etsion is a Professor of Computer Science and Electrical Engineering at Technion. He is the CTO and co-founder of Speedata, a rising startup that targets hardware acceleration of data analytics workloads. Prof. Etsion is known for his work on computer systems, non-von Neumann computer architectures, and HW/SW interoperability. Before joining Technion, Prof. Etsion was a Senior Researcher at Barcelona Supercomputing Center (BSC-CNS). Prof. Etsion received his PhD from Hebrew University in 2010.



## **Prof. Dr. Boris Glavic**

Professor of Computer  
Science, Illinois Institute of  
Technology

# **Efficient Uncertainty Tracking for Complex Queries with Attribute-level Bounds**

Incomplete and probabilistic database techniques are principled methods for coping with uncertainty in data. Unfortunately, the class of queries that can be answered efficiently over such databases is severely limited, even when advanced approximation techniques are employed. We introduce attribute-annotated uncertain databases (AU-DBs), an uncertain data model that annotates tuples and attribute values with bounds to compactly approximate an incomplete database. AU-DBs are closed under relational algebra with aggregation using an efficient evaluation semantics. Using optimizations that trade accuracy for performance, our approach scales to complex queries and large datasets, and produces accurate results.

Boris Glavic is an Associate Professor of Computer Science at the Illinois Institute of Technology where he leads the IIT DBGroup (<http://www.cs.iit.edu/dbgroup/>). Before coming to IIT, Boris spend two years as a PostDoc at the University of Toronto working with Renée J. Miller. He received his PhD from the University of Zurich in Switzerland being advised by Michael Böhlen and Gustavo Alonso. Boris is working on the systems aspects of data management and strives to build real systems that are based on solid theoretical foundations. His main research interests are data provenance and explanations, data integration and cleaning, uncertain data management, and query processing and optimization.



## Legal Constraints for Algorithmic Fairness

In recent years, algorithmic discrimination has emerged as a key problem in AI regulation. As a response, a number of criteria for implementing algorithmic fairness in machine learning have been developed in the literature. The talk will present the continuous fairness algorithm (CFA) which enables a continuous interpolation between different fairness definitions. Furthermore, legal constraints for the implementation of algorithmic fairness interventions will be discussed, touching on data protection and affirmative action law.



**Prof. Dr.  
Philipp Hacker**

Chair for Law and Ethics of  
the Digital Society, European  
New School of Digital Studies,  
European University Viadrina

Prof. Dr. Philipp Hacker, LL.M. (Yale), holds the Chair for Law and Ethics of the Digital Society at European New School of Digital Studies, located at European University Viadrina. Before joining Viadrina, he led a project on fairness in machine learning as an AXA Postdoctoral Fellow at the Faculty of Law at Humboldt University of Berlin. He was also a Max Weber Fellow at the European University Institute and an A.SK Fellow at WZB Berlin Social Science Center. His research focuses on the intersection of law and technology. In particular, he analyzes the impact of tracking technologies, Artificial Intelligence and the Internet of Things on consumer, privacy and contract law. He often cooperates with computer scientists and mathematicians, especially on questions of explainable AI and algorithmic fairness.



## **Prof. Dr. Viktor Leis**

Professor at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

# **LeanStore: A High-Performance Storage Engine for Modern Hardware**

LeanStore is a high-performance OLTP storage engine optimized for many-core CPUs and NVMe SSDs. The goal of the project is to achieve performance comparable to in-memory systems when the data set fits into RAM, while being able to fully exploit the bandwidth of fast NVMe SSDs for large data sets.

In this talk, I will present many of the performance-critical components of LeanStore, including its low-overhead buffer manager, scalable synchronization primitives, optimized B-tree indexes, and an efficient logging and recovery implementation.

Viktor Leis is a Professor at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). His research revolves around designing efficient database systems and includes core data management topics such as query processing, query optimization, index structures, and storage. Viktor received his doctoral degree in 2016 from the Technical University of Munich, where he worked on the main-memory database system HyPer. Currently, he leads a team at FAU developing LeanStore, a high-performance storage engine for fast flash storage. He is the recipient of the ACM SIGMOD Jim Gray Dissertation Award, the GI DBIS Dissertation Award, four best conference paper awards, and the IEEE TCDE Rising Star Award.

## Cost-aware Internet Security Decision Model

The current delegated or “Stupid user” implementation of Internet Security services leaves users out of security decision making and configuration processes which distorts online security’s mental models. As a result, many users do not configure optimal Internet security to meet the required level of protection. This lack of sufficient security configuration provides a fertile environment for Internet attacks. The top-down security implementation does not usually provide the user with the associated cost information, such as privacy leakage or performance degradation. Such information would enable the user to build correct mental models around security, thereby improving the Quality of Protection and Internet browsing Experience. This research investigates the use of a cost-aware security decision model and a personal Internet security configuration tool that allow users to easily configure security options mapped to underlying complex Internet security protocols. We aim to examine whether these interventions would improve users’ Quality of Experience (QoE) and security adoption. In this talk, we present the overview of the results from two studies, generative user study and internet security measurements informing the design of our model.



**Enock Mbewe**

Ph.D. Student, University of  
Cape Town

Enock sides with the unsuspecting Internet/Network users when it comes to online Security and Privacy. His motto is “Let the users take charge of their digital security!” He has 10+ years of experience providing ICT services to public and private institutions in ICT4D, cybersecurity, and general ICT training. He holds a Master of Science in Information Theory, Coding and Cryptography and a Bachelor of Science in Information and Communication Technology (ICT) obtained in Malawi. Enock is a Lecturer in ICT at Mzuzu University, Malawi’s second-largest public University. He also is a Certified Ethical Hacker and a Certified Ethical Hacking trainer. His PhD work focuses on user involvement in the Internet security decision-making and configuration processes, using a cost-aware Internet security decision model and a personal security configurator. His research interests are Internet Security, Blockchains, Machine Learning for Networking, Quality of Protection and Experience, Usable security, and Internet Security Measurements.



## Successful Systems Research

In this presentation, we give some hints on how to do successful systems research. Successful research usually starts with a clear problem or question that is relevant and measurable. Measurability is required to figure out the possibility and rate of success. Using a simple model basic assumptions can frequently be tested very quickly. The research itself needs to be properly validated. In order to make the work successful, it needs to be made available to the public in form of a publication, software artefact, or similar.

### Prof. Dr. Tilmann Rabl

Head of Data Engineering  
Systems Group, Hasso  
Plattner Institute, Potsdam

Tilmann Rabl holds the chair for Data Engineering Systems at the Hasso Plattner Institute and is Professor at the Digital Engineering Faculty of the University of Potsdam. He is also cofounder and scientific director of the startup bankmark. Tilmann Rabl received his PhD at the University of Passau in 2011. He spent 4 years at the University of Toronto as a postdoc in the Middleware Systems Research Group (MSRG). From 2015 to 2019, he was senior researcher and visiting professor at the Database Systems and Information Management (DIMA) group at Technische Universität Berlin and Vice Director of the Intelligent Analytics for Massive Data (IAM) Group at the German Research Center for Artificial Intelligence (DFKI).

## Active and Bayesian Learning of Classifier Performance

As black-box machine learning models are now being deployed in many different real-world environments there is an increasing need for accurate and robust methods to assess their performance. This is especially the case in common situations such as limited availability of data for evaluation purposes, or when a model is being used on data that is distributionally different to that on which the model was trained.

In this talk I will describe recent work in my research group on efficient methods and algorithms for assessing black-box classifier performance, leveraging ideas from both Bayesian estimation and active learning. Results across a variety of performance metrics show that these ideas are broadly useful, in particular in situations where labeled data is scarce for evaluation. The talk will conclude with a brief description of the application of these ideas to evaluation of fairness for black-box models, as well as new directions in developing frameworks that combine both algorithmic and human predictions.



**Prof. Dr.  
Padhraic Smyth**

Chancellor's Professor,  
Department of Computer  
Science, UC Irvine

Padhraic Smyth is Chancellor's Professor, Department of Computer Science, UC Irvine. He is also Co-Director of the HPI Research Center in Machine Learning and Data Science at UCI.



## Wang Yue

Ph.D. Student, Hasso Plattner  
Institute, Potsdam

## Desis: General Distributed Window Aggregation

In the last decade, the rapidly increasing number of IoT devices constitutes huge distributed sensor networks, which provide a big challenge to today's stream processing. By 2020, the installation of IoT devices is beyond 31 billion. However, the current SPEs centralize all data and regardless of the different characteristics of queries, which is a limit factor to stream processing in the sensor network and leads to unnecessary overhead, such as Apache Flink, Apache Spark Streaming, and Apache Storm. In addition, sharing computation results among multiple continuous queries can dramatically improve performance, especially for crossing intermediate results among arbitrary windows which have different types and aggregation functions. Therefore in our work, we propose a new distributed-stream oriented stream processing system, which is specific for the distributed data stream and able to process arbitrary windows as well as aggregations flexibly but does not require centralizing all queries and data in a single node. We design an optimizer that broke through the limitation of sharing intermedia results between the same window types and functions. On the contrary, it is able to cross the intermedia result among the queries which have arbitrary window types and aggregation functions.

Wang Yue is a Ph.D. student working at the Data Engineering Systems team at the HPI. Currently, he is working on stream processing systems with distributed sensor network. He is also interested in modern hardware and databases. He received his Bachelor's and Master's degrees at the Yunan University and Sichuan University in 2015 and 2018, respectively.

# An Explainable Sampling Strategy for Video Action Recognition

Frame sampling is a fundamental problem in video action recognition due to the essential redundancy in time and limited computation resources. The existing sampling strategy often employs a fixed frame selection and lacks the flexibility to deal with complex variations in videos.

In this talk, I will present an explainable, adaptive, and effective frame sampler, called Motion-guided Sampler (MGSampler). My basic motivation is that motion is an important and universal signal that can drive us to select frames from videos adaptively. Accordingly, I propose two important properties in our MGSampler design: motion sensitive and motion uniform. The MGSampler yields a new principled and holistic sample scheme, that could be incorporated into any existing video architecture. Experiments on five benchmarks demonstrate the effectiveness of the MGSampler over previously fixed sampling strategies, and also its generalization power across different backbones, video models, and datasets.



**Yuan Zhi**

Ph.D. Student, Nanjing  
University

Yuan Zhi is a Ph.D. student at Department of Computer Science and Technology, Nanjing University. She is in the second year of Doctor now. Her major study is around the video action recognition with Deep Learning in Multimedia Computing Group(MCG), NJU. She currently focuses on figuring out a dynamic sampling module to enhance the discriminative ability of trained clip-level classifiers under the guidance of her mentor Limin Wang.







## Contact

**Hasso-Plattner-Institut für Digital Engineering gGmbH**  
**Campus Griebnitzsee | University of Potsdam**

Prof.-Dr.-Helmert-Straße 2-3  
D-14482 Potsdam, Germany

Phone +49 331 5509-0  
E-Mail [hpi-info@hpi.de](mailto:hpi-info@hpi.de)  
Web <https://www.hpi.de>