OSDI '18:
13th USENIX Symposium on
Operating Systems Design and Implementation

October 8–10, 2018
Carlsbad, CA, USA

Understanding Failures

Capturing and Enhancing *In Situ* System Observability for Failure Detection 1
Peng Huang, *Johns Hopkins University;* Chuanxiong Guo, *ByteDance Inc.;* Jacob R. Lorch and Lidong Zhou, *Microsoft Research;* Yingnong Dang, *Microsoft*

REPT: Reverse Debugging of Failures in Deployed Software 17
Weidong Cui and Xinyang Ge, *Microsoft Research Redmond;* Baris Kasikci, *University of Michigan;* Ben Niu, *Microsoft Research Redmond;* Upamanyu Sharma, *University of Michigan;* Ruoyu Wang, *Arizona State University;* Insu Yun, *Georgia Institute of Technology*

Finding Crash-Consistency Bugs with Bounded Black-Box Crash Testing 33
Jayashree Mohan, Ashlie Martinez, Soujanya Ponnapalli, and Pandian Raju, *University of Texas at Austin;* Vijay Chidambaram, *University of Texas at Austin and VMware Research*

An Analysis of Network-Partitioning Failures in Cloud Systems 51
Ahmed Alquraan, Hatem Takruri, Mohammed Alfatafta, and Samer Al-Kiswany, *University of Waterloo*

Operating Systems

LegoOS: A Disseminated, Distributed OS for Hardware Resource Disaggregation 69
Yizhou Shan, Yutong Huang, Yilun Chen, and Yiying Zhang, *Purdue University*

The benefits and costs of writing a POSIX kernel in a high-level language 89
Cody Cutler, M. Frans Kaashoek, and Robert T. Morris, *MIT CSAIL*

Sharing, Protection, and Compatibility for Reconfigurable Fabric with AmorphOS 107
Ahmed Khawaja, Joshua Landgraf, and Rohith Prakash, *UT Austin;* Michael Wei and Eric Schkufza, *VMware Research Group;* Christopher J. Rossbach, *UT Austin and VMware Research Group*

Adaptive Dynamic Checkpointing for Safe Efficient Intermittent Computing 129
Kiwan Maeng and Brandon Lucia, *Carnegie Mellon University*

Scheduling

Arachne: Core-Aware Thread Management 145
Henry Qin, Qian Li, Jacqueline Speiser, Peter Kraft, and John Ousterhout, *Stanford University*

Principled Schedulability Analysis for Distributed Storage Systems using Thread Architecture Models . . .161
Suli Yang, *Ant Financial Services Group;* Jing Liu, Andrea C. Arpaci-Dusseau, and Remzi H. Arpaci-Dusseau, *UW-Madison*

***μ*Tune: Auto-Tuned Threading for OLDI Microservices 177**
Akshitha Sriraman and Thomas F. Wenisch, *University of Michigan*

RobinHood: Tail Latency Aware Caching – Dynamic Reallocation from Cache-Rich to Cache-Poor 195
Daniel S. Berger and Benjamin Berg, *Carnegie Mellon University;* Timothy Zhu, *Pennsylvania State University;* Siddhartha Sen, *Microsoft Research;* Mor Harchol-Balter, *Carnegie Mellon University*

(continued on next page)

Data

Noria: dynamic, partially-stateful data-flow for high-performance web applications 213
Jon Gjengset, Malte Schwarzkopf, Jonathan Behrens, and Lara Timbó Araújo, *MIT CSAIL;* Martin Ek, *Norwegian University of Science and Technology;* Eddie Kohler, *Harvard University;* M. Frans Kaashoek and Robert Morris, *MIT CSAIL*

Deconstructing RDMA-enabled Distributed Transactions: Hybrid is Better! 233
Xingda Wei, Zhiyuan Dong, Rong Chen, and Haibo Chen, *Shanghai Jiao Tong University*

Dynamic Query Re-Planning using QOOP 253
Kshiteej Mahajan, *UW-Madison;* Mosharaf Chowdhury, *U. Michigan;* Aditya Akella and Shuchi Chawla, *UW-Madison*

Focus: Querying Large Video Datasets with Low Latency and Low Cost 269
Kevin Hsieh, *Carnegie Mellon University;* Ganesh Ananthanarayanan and Peter Bodik, *Microsoft;* Shivaram Venkataraman, *Microsoft / UW-Madison;* Paramvir Bahl and Matthai Philipose, *Microsoft;* Phillip B. Gibbons, *Carnegie Mellon University;* Onur Mutlu, *ETH Zurich*

Verification

Nickel: A Framework for Design and Verification of Information Flow Control Systems 287
Helgi Sigurbjarnarson, Luke Nelson, Bruno Castro-Karney, James Bornholt, Emina Torlak, and Xi Wang, *University of Washington*

Verifying concurrent software using movers in CSPEC 307
Tej Chajed and Frans Kaashoek, *MIT CSAIL;* Butler Lampson, *Microsoft;* Nickolai Zeldovich, *MIT CSAIL*

Proving confidentiality in a file system using DISKSEC 323
Atalay Ileri, Tej Chajed, Adam Chlipala, Frans Kaashoek, and Nickolai Zeldovich, *MIT CSAIL*

Proving the correct execution of concurrent services in zero-knowledge 339
Srinath Setty, *Microsoft Research;* Sebastian Angel, *University of Pennsylvania;* Trinabh Gupta, *Microsoft Research and UCSB;* Jonathan Lee, *Microsoft Research*

Reliability

The FuzzyLog: A Partially Ordered Shared Log 357
Joshua Lockerman, *Yale University;* Jose M. Faleiro, *UC Berkeley;* Juno Kim, *UC San Diego;* Soham Sankaran, *Cornell University;* Daniel J Abadi, *University of Maryland, College Park;* James Aspnes, *Yale University;* Siddhartha Sen, *Microsoft Research;* Mahesh Balakrishnan, *Yale University / Facebook*

Maelstrom: Mitigating Datacenter-level Disasters by Draining Interdependent Traffic Safely and Efficiently 373
Kaushik Veeraraghavan, Justin Meza, Scott Michelson, Sankaralingam Panneerselvam, Alex Gyori, David Chou, Sonia Margulis, Daniel Obenshain, Shruti Padmanabha, Ashish Shah, and Yee Jiun Song, *Facebook;* Tianyin Xu, *Facebook and University of Illinois at Urbana-Champaign*

Fault-Tolerance, Fast and Slow: Exploiting Failure Asynchrony in Distributed Systems 391
Ramnatthan Alagappan, Aishwarya Ganesan, Jing Liu, Andrea C. Arpaci-Dusseau, and Remzi H. Arpaci-Dusseau, *University of Wisconsin - Madison*

Taming Performance Variability 409
Aleksander Maricq and Dmitry Duplyakin, *University of Utah;* Ivo Jimenez and Carlos Maltzahn, *University of California Santa Cruz;* Ryan Stutsman and Robert Ricci, *University of Utah*

(continued on next page)

File Systems

Pocket: Elastic Ephemeral Storage for Serverless Analytics 427
Ana Klimovic and Yawen Wang, *Stanford University;* Patrick Stuedi, Animesh Trivedi, and Jonas Pfefferle, *IBM Research;* Christos Kozyrakis, *Stanford University*

Sharding the Shards: Managing Datastore Locality at Scale with Akkio 445
Muthukaruppan Annamalai, Kaushik Ravichandran, Harish Srinivas, Igor Zinkovsky, Luning Pan, Tony Savor, and David Nagle, *Facebook;* Michael Stumm, *University of Toronto*

Write-Optimized and High-Performance Hashing Index Scheme for Persistent Memory. 461
Pengfei Zuo, Yu Hua, and Jie Wu, *Huazhong University of Science and Technology*

FlashShare: Punching Through Server Storage Stack from Kernel to Firmware for Ultra-Low Latency SSDs. 477
Jie Zhang, Miryeong Kwon, Donghyun Gouk, Sungjoon Koh, and Changlim Lee, *Yonsei University;* Mohammad Alian, *UIUC;* Myoungjun Chun, *Seoul National University;* Mahmut Taylan Kandemir, *Penn State University;* Nam Sung Kim, *UIUC;* Jihong Kim, *Seoul National University;* Myoungsoo Jung, *Yonsei University*

Debugging

Orca: Differential Bug Localization in Large-Scale Services 493
Ranjita Bhagwan, Rahul Kumar, Chandra Sekhar Maddila, and Adithya Abraham Philip, *Microsoft Research India*

Differential Energy Profiling: Energy Optimization via Diffing Similar Apps. 511
Abhilash Jindal and Y. Charlie Hu, *Purdue University and Mobile Enerlytics, LLC*

wPerf: Generic Off-CPU Analysis to Identify Bottleneck Waiting Events 527
Fang Zhou, Yifan Gan, Sixiang Ma, and Yang Wang, *The Ohio State University*

Sledgehammer: Cluster-Fueled Debugging 545
Andi Quinn, Jason Flinn, and Michael Cafarella, *University of Michigan*

Machine Learning

Ray: A Distributed Framework for Emerging AI Applications. 561
Philipp Moritz, Robert Nishihara, Stephanie Wang, Alexey Tumanov, Richard Liaw, Eric Liang, Melih Elibol, Zongheng Yang, William Paul, Michael I. Jordan, and Ion Stoica, *UC Berkeley*

TVM: An Automated End-to-End Optimizing Compiler for Deep Learning 579
Tianqi Chen and Thierry Moreau, *University of Washington;* Ziheng Jiang, *University of Washington, AWS;* Lianmin Zheng, *Shanghai Jiao Tong University;* Eddie Yan, Haichen Shen, and Meghan Cowan, *University of Washington;* Leyuan Wang, *UC Davis, AWS;* Yuwei Hu, *Cornell;* Luis Ceze, Carlos Guestrin, and Arvind Krishnamurthy, *University of Washington*

Gandiva: Introspective Cluster Scheduling for Deep Learning. 595
Wencong Xiao, *Beihang University & Microsoft Research;* Romil Bhardwaj, Ramachandran Ramjee, Muthian Sivathanu, and Nipun Kwatra, *Microsoft Research;* Zhenhua Han, *The University of Hong Kong & Microsoft Research;* Pratyush Patel, *Microsoft Research;* Xuan Peng, *Huazhong University of Science and Technology & Microsoft Research;* Hanyu Zhao, *Peking University & Microsoft Research;* Quanlu Zhang, Fan Yang, and Lidong Zhou, *Microsoft Research*

Pretzel: Opening the Black Box of Machine Learning Prediction Serving Systems 611
Yunseong Lee, *Seoul National University;* Alberto Scolari, *Politecnico di Milano;* Byung-Gon Chun, *Seoul National University;* Marco Domenico Santambrogio, *Politecnico di Milano;* Markus Weimer and Matteo Interlandi, *Microsoft*

(continued on next page)

Networking

Splinter: Bare-Metal Extensions for Multi-Tenant Low-Latency Storage 627
Chinmay Kulkarni, Sara Moore, Mazhar Naqvi, Tian Zhang, Robert Ricci, and Ryan Stutsman, *University of Utah*

Neural Adaptive Content-aware Internet Video Delivery. 645
Hyunho Yeo, Youngmok Jung, Jaehong Kim, Jinwoo Shin, and Dongsu Han, *KAIST*

Floem: A Programming System for NIC-Accelerated Network Applications 663
Phitchaya Mangpo Phothilimthana, *University of California, Berkeley;* Ming Liu and Antoine Kaufmann, *University of Washington;* Simon Peter, *The University of Texas at Austin;* Rastislav Bodik and Thomas Anderson, *University of Washington*

Security

Graviton: Trusted Execution Environments on GPUs 681
Stavros Volos and Kapil Vaswani, *Microsoft Research;* Rodrigo Bruno, *INESC-ID / IST, University of Lisbon*

ZebRAM: Comprehensive and Compatible Software Protection Against Rowhammer Attacks 697
Radhesh Krishnan Konoth, *Vrije Universiteit Amsterdam;* Marco Oliverio, *University of Calabria/Vrije Universiteit Amsterdam;* Andrei Tatar, Dennis Andriesse, Herbert Bos, Cristiano Giuffrida, and Kaveh Razavi, *Vrije Universiteit Amsterdam*

Karaoke: Distributed Private Messaging Immune to Passive Traffic Analysis. 711
David Lazar, Yossi Gilad, and Nickolai Zeldovich, *MIT CSAIL*

Obladi: Oblivious Serializable Transactions in the Cloud 727
Natacha Crooks, *The University of Texas at Austin;* Matthew Burke, Ethan Cecchetti, Sitar Harel, Rachit Agarwal, and Lorenzo Alvisi, *Cornell University*

Graphs and Data

ASAP: Fast, Approximate Graph Pattern Mining at Scale 745
Anand Padmanabha Iyer, *UC Berkeley;* Zaoxing Liu and Xin Jin, *Johns Hopkins University;* Shivaram Venkataraman, *Microsoft Research / University of Wisconsin;* Vladimir Braverman, *Johns Hopkins University;* Ion Stoica, *UC Berkeley*

RStream: Marrying Relational Algebra with Streaming for Efficient Graph Mining on A Single Machine 763
Kai Wang, *UCLA;* Zhiqiang Zuo, *Nanjing University;* John Thorpe, *UCLA;* Tien Quang Nguyen, *Facebook;* Guoqing Harry Xu, *UCLA*

Three steps is all you need: fast, accurate, automatic scaling decisions for distributed streaming dataflows 783
Vasiliki Kalavri, John Liagouris, Moritz Hoffmann, and Desislava Dimitrova, *ETH Zurich;* Matthew Forshaw, *Newcastle University;* Timothy Roscoe, *ETH Zurich*

Flare: Optimizing Apache Spark with Native Compilation for Scale-Up Architectures and Medium-Size Data 799
Gregory Essertel, Ruby Tahboub, and James Decker, *Purdue University;* Kevin Brown and Kunle Olukotun, *Stanford University;* Tiark Rompf, *Purdue University*