Foreword

Dear colleagues,

Welcome to the Proceedings of the 26th ACM Symposium on Operating Systems Principles (SOSP 2017), held at the Pudong Shangri-La Hotel in Shanghai, China. This is the first SOSP to take place outside of North America and Europe, reflecting the global community of researchers in software systems.

This year marks the 50th anniversary of SOSP, and we are pleased to continue the half-century tradition of publishing innovative and relevant research in the field of software systems. This year's program includes 39 wonderful papers (the most in SOSP history) on a wide range of relevant topics, including kernels and storage systems, privacy and security, understanding and fixing bugs, in-network computing, scalability and resource management, and data analytics. We hope you enjoy reading these papers as much as we enjoyed selecting them.

We received 232 submissions (a record), and an expert group of 55 program committee members (23 "heavy" members and 32 "light" members) reviewed these carefully. Each paper received at least 3 reviews in the first round, and about 50% of the papers received 3-4 additional reviews in later rounds. In all, the program committee produced 1131 reviews, and we solicited 12 external reviews from outside experts. Authors were given a chance to read and respond to the reviews for their papers, and these responses were then considered by the reviewers.

At the end of the review process, the program committee had a vigorous online discussion about each of the papers that made it past the first round of reviewing and chose 76 papers to discuss in person. The members of the heavy program committee met for 2 days in Ann Arbor, Michigan to discuss these papers and chose to accept just over half of them for the conference. Each accepted paper was then shepherded by a member of the program committee. Throughout the entire process, the program committee preserved author anonymity and avoided conflicts of interests. Frans Kaashoek served as the administrator of papers for which both program chairs had conflicts of interest.

Breaking with recent tradition, this year’s SOSP had two program committee chairs, something that had not happened since SOSP 1985. We hope the deliberate process that this collaboration encouraged was reflected in a better experience for authors, reviewers, and shepherds---as for us, we thoroughly enjoyed working together!

One of program committee chairs of SOSP 1985 was Andrew Birrell, who passed away this year. We remember him not only as a personal friend and mentor to many in our community, but as a friend and contributor in many ways to the entire SOSP community, including leading its transition to digitally published proceedings. Through his work at Xerox, DEC, and MSR—twice recognized in the SIGOPS Hall of Fame—Andrew will continue to be an inspiration to generations of systems researchers.

A defining characteristic of SOSP is the unspoken pact that binds authors and reviewers. The former tend to submit to SOSP only their best work, often practicing self-censorship prior to the deadline; the latter, in exchange, invest what their colleagues in other areas of CS see as a completely insane
amount of time and effort to produce extensive, high-quality feedback. That virtuous partnership was in full display this year. Our thanks go to all the authors, who have continued to send in so many high quality submissions. And our deep gratitude goes to PC members, and to the external experts often called upon for help under a tight deadline, for their care and thoroughness: with an average review of about 6.5 KBytes, they collectively produced text comparable, in quantity at least, to Anna Karenina, Brothers Karamazov, and Harry Potter and The Prisoner of Azkaban–combined.

Organizing a successful SOSP goes beyond choosing an outstanding program. We are indebted to the organizing committee members, who voluntarily took on key responsibilities with passion and accountability to ensure the success of this SOSP. It has been a great honor and pleasure to work together with such a wonderful team: Yubin Xia (Local arrangements); Gernot Heiser and Cheng Li (Posters); Jeanna Matthews and Ming Wu (Publication); Yungang Bao and Simon Peter (Publicity); Yang Chen and Patrick Lee (Registrations); Thomas Bressoud (Scholarships); Wenguang Chen and Shan Lu (Sponsorships); Srinath Setty (Treasurer); Olivier Marin and Catello De Martino (Tutorials); Hakim Weatherspoon (Workshops); and the SIGOPS Board, chaired by Robbert van Renesse.

We are grateful to the sponsors for their strong support. We are happy to report that a record number of organizations sponsored this SOSP: Platinum sponsor Huawei; Gold sponsors Alibaba, Apple, ArXanFintech, Baidu, BOPU Asset, Facebook, Intel, Microsoft, and Tencent; Silver sponsors CreditEase, DiDi, Google, SmartX, and VMware; Bronze sponsors Akamai, Amazon, FMA.AI, IBM Research, Oracle, and Uber; and HP Enterprise. The generosity of these sponsors allowed us to provide an unprecedented level of support for student travel grants and the diversity workshop. We are particularly delighted to have numerous first-time sponsors, many of them from the Asia-Pacific region. Clearly there is much interest and activity in software systems research in the Asia-Pacific region. We hope that you will find the program interesting and inspiring, and trust that the symposium will provide you with a valuable opportunity to network and share ideas with researchers and practitioners from institutions around the world.

Haibo Chen and Lidong Zhou
General Co-Chairs

Lorenzo Alvisi and Peter Chen
Program Co-Chairs
Bug hunting

DeepXplore: Automated Whitebox Testing of Deep Learning Systems

Kexin Pei (Columbia University); Yinzhi Cao (Lehigh University); Junfeng Yang, Suman Jana (Columbia University)

Pensieve: Non-Intrusive Failure Reproduction for Distributed Systems using the Event Chaining Approach

Yongle Zhang, Serguei Makarov, Xiang Ren, David Lion, Ding Yuan (University of Toronto)

Canopy: An End-to-End Performance Tracing And Analysis System

Jonathan Kaldor (Facebook); Jonathan Mace (Brown University); Michał Bejda, Edison Gao, Wiktor Kuropatwa, Joe O’Neill, Kian Win Ong, Bill Schaller, Pingjia Shan, Brendan Viscomi, Vinod Venkataraman, Kaushik Veeraraghavan, Yee Jiun Song (Facebook)

Scalability

Algorand: Scaling Byzantine Agreements for Cryptocurrencies

Yossi Gilad (MIT CSAIL / Boston University); Rotem Hemo, Silvio Micali, Georgios Vlachos, Nickolai Zeldovich (MIT CSAIL)

Scaling a file system to many cores using an operation log

Srivatsa S. Bhat, Rasha Eqbal, Austin T. Clements, M. Frans Kaashoek, Nickolai Zeldovich (MIT CSAIL)

SVE: Distributed Video Processing at Facebook Scale

Qi Huang, Petchean Ang, Peter Knowles, Tomasz Nykiel, Iaroslav Tverdokhlib, Amit Yajurvedi, Paul Dapolito IV, Xifan Yan, Maxim Bykov, Chuen Liang, Mohit Talwar, Abhishek Mathur, Sachin Kulkarni (Facebook); Matthew Burke (University of Southern California; Facebook; Cornell); Wyatt Lloyd (University of Southern California; Facebook; Princeton)

In-network computing

Eris: Coordination-Free Consistent Transactions Using In-Network Concurrency Control

Jialin Li, Ellis Michael, Dan R. K. Ports (University of Washington)

NetCache: Balancing Key-Value Stores with Fast In-Network Caching

Xin Jin (Johns Hopkins University); Xiaozhou Li (Barefoot Networks); Haoyu Zhang (Princeton University); Robert Soulé (Barefoot Networks and Università della Svizzera italiana); Jeongkeun Lee (Barefoot Networks); Nate Foster (Barefoot Networks and Cornell University); Changhoon Kim (Barefoot Networks); Ion Stoica (UC Berkeley)
KV-Direct: High-Performance In-Memory Key-Value Store with Programmable NIC

Bojie Li (USTC and Microsoft Research); Zhenyuan Ruan (UCLA and Microsoft Research); Wencong Xiao (Beihang University and Microsoft Research); Yuanwei Lu (USTC and Microsoft Research); Yongqiang Xiong (Microsoft Research); Andrew Putnam (Microsoft); Enhong Chen (USTC); Lintao Zhang (Microsoft Research)

Resource management

Resource Central: Understanding and Predicting Workloads for Improved Resource Management in Large Cloud Platforms

Eli Cortez (Microsoft); Anand Bonde (Microsoft Research); Alexandre Muzio (ITA, Brazil); Mark Russinovich, Marcus Fontoura (Microsoft); Ricardo Bianchini (Microsoft Research)

MittOS: Supporting Millisecond Tail Tolerance with Fast Rejecting SLO-Aware OS Interface

Mingzhe Hao, Huaiycng Li, Michael Hao Tong, Chrisma Pakha, Riza O. Sumindo, Cesar A. Stuardo, Andrew A. Chien, Haryadi S. Gunawi (University of Chicago)

Monotasks: Architecting for Performance Clarity in Data Analytics Frameworks

Kay Ousterhout (UC Berkeley); Christopher Canel (Carnegie Mellon University); Sylvia Ratnasamy (UC Berkeley); Scott Shenker (UC Berkeley, ICSI)

Kernels

NEVE: Nested Virtualization Extensions for ARM

Jin Tack Lim, Christoffer Dall, Shih-Wei Li, Jason Nieh (Columbia University); Marc Zyngier (ARM Ltd)

My VM is Lighter (and Safer) than your Container

Filipe Manco (NEC Europe Ltd.); Costin Lupu (University Politehnica of Bucharest); Florian Schmidt, Jose Mendes, Simon Kuenzer, Sumit Sati, Kenichi Yasukata (NEC Europe Ltd.); Costin Raiciu (University Politehnica of Bucharest); Felipe Huici (NEC Europe Ltd.)

Multiprogramming a 64 kB Computer Safely and Efficiently

Amit Levy (Stanford University); Bradford Campbell (University of Virginia); Branden Ghena (University of California, Berkeley); Daniel B Giffin (Stanford University); Pat Pannuto, Prabal Dutta (University of California, Berkeley); Philip Levis (Stanford University)

Verification

Hyperkernel: Push-Button Verification of an OS Kernel


Verifying a high-performance crash-safe file system using a tree specification

Haogang Chen, Tej Chajed, Stephanie Wang, Alex Konradi, Atalay Ileri, Adam Chlipala, M. Frans Kaashoek, Nickolai Zeldovich (MIT CSAIL)

Komodo: Using verification to disentangle secure-enclave hardware from software

Andrew Ferraiuolo (Cornell University); Andrew Baumann, Chris Hawblitzel (Microsoft Research); Bryan Parno (Carnegie Mellon University)
Potpourri

LITE Kernel RDMA Support for Datacenter Applications .................. 306
Shin-Yeh Tsai, Yiying Zhang (Purdue University)

ZygOS: Achieving Low Tail Latency for Microsecond-scale Networked Tasks ............... 325
George Prekas, Marios Kogias, Edouard Bugnion (EPFL)

ffwd: delegation is (much) faster than you think ......................... 342
Sepideh Roghanchi, Jakob Eriksson, Nilanjana Basu (University of Illinois at Chicago)

Adaptation and repair

Automatically Repairing Network Control Planes Using an Abstract Representation ........ 359
Aaron Gember-Jacobson (Colgate University); Aditya Akella (University of Wisconsin-Madison);
Ratul Mahajan (Intentionet); Hongqiang Harry Liu (Microsoft Research)

Drizzle: Fast and Adaptable Stream Processing at Scale .................................. 374
Shivaram Venkataraman, Aurojit Panda, Kay Ousterhout (UC Berkeley); Ali Ghodsi (UC Berkeley / Databricks); Michael Armbrust (Databricks); Benjamin Recht (UC Berkeley); Michael J. Franklin (University of Chicago); Ion Stoica (UC Berkeley)

Rocksteady: Fast Migration for Low-latency In-memory Storage .......................... 390
Chinmay Kulkarni, Aniraj Kesavan, Tian Zhang, Robert Ricci, Ryan Stutsman (University of Utah)

Privacy

Atom: Horizontally Scaling Strong Anonymity .................................. 406
Albert Kwon (MIT); Henry Corrigan-Gibbs (Stanford); Srinivas Devadas (MIT); Bryan Ford (EPFL)

Stadium: A Distributed Metadata-Private Messaging System .................. 423
Nirvan Tyagi (Cornell University); Yossi Gilad (MIT CSAIL/Boston University); Derek Leung (MIT CSAIL);
Matei Zaharia (Stanford); Nickolai Zeldovich (MIT CSAIL)

Prochlo: Strong Privacy for Analytics in the Crowd ............................ 441
Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan (Google Brain);
David Lie (Google Brain & University of Toronto); Mitch Rudominer, Ushasree Kode, Julien Tinnes,
Bernhard Seefeld (Google)

Storage systems

Strata: A Cross Media File System ........................................ 460
Youngjin Kwon, Henrique Fingler, Tyler Hunt, Simon Peter, Emmett Witchel (UT Austin);
Thomas Anderson (University of Washington)

NOVA-Fortis: A Fault-Tolerant Non-Volatile Main Memory File System ............. 478
Jian Xu, Lu Zhang, Amirsaman Memaripour, Akshatha Gangadharaiah, Amit Borase,
Tamires Brito Da Silva (UC San Diego); Andy Rudoff (Intel); Steven Swanson (UC San Diego)

PebblesDB: Building Key-Value Stores using Fragmented Log-Structured Merge Trees ...... 497
Pandian Raju, Rohan Kadekodi (University of Texas at Austin); Vijay Chidambaram (University of Texas at Austin and VMware Research); Ittai Abraham (VMware Research)
Security

WatchIT: Who Watches Your IT Guy? ........................................ 515
Noam Shalev, Idit Keidar (Technion); Yaron Weinsberg (Microsoft); Yosef Moatti, Elad Ben-Yehuda (IBM Research)

Secure Page Fusion with VUSion ........................................ 531
Marco Oliverio (Vrije Universiteit Amsterdam and University of Calabria); Kaveh Razavi, Herbert Bos, Cristiano Giuffrida (Vrije Universiteit Amsterdam)

The Efficient Server Audit Problem, Deduplicated Re-execution, and the Web .......................... 546
Cheng Tan, Lingfan Yu (NYU); Joshua B. Leners (Two Sigma); Michael Walfish (NYU)

Understanding failures

Log20: Fully Automated Optimal Placement of Log Printing Statements under Specified Overhead Threshold .................................................. 565
Xu Zhao, Kirk Rodrigues, Yu Luo, Michael Stumm, Ding Yuan (University of Toronto); Yuanyuan Zhou (University of California, San Diego)

Lazy Diagnosis of In-Production Concurrency Bugs ........................................ 582
Baris Kasikci (University of Michigan and Microsoft Research); Weidong Cui, Xinyang Ge, Ben Niu (Microsoft Research)

CrystalNet: Faithfully Emulating Large Production Networks ........................................ 599
Hongqiang Harry Liu, Yibo Zhu (Microsoft Research); Jitu Padhye, Jiaxin Cao, Sri Tallapragada (Microsoft); Nuno P. Lopes, Andrey Rybalchenko (Microsoft Research); Guohan Lu, Lihua Yuan (Microsoft)

Data analytics

Sub-millisecond Stateful Stream Querying over Fast-evolving Linked Data .................. 614
Yunhao Zhang, Rong Chen, Haibo Chen (Shanghai Jiao Tong University)

Optimizing Big-Data Queries Using Program Synthesis ........................................ 631
Matthias Schlaipfer (TU Wien); Kaushik Rajan, Akash Lal (Microsoft Research); Malavika Samak (MIT CSAIL)

Low-Latency Analytics on Colossal Data Streams with SummaryStore .................. 647
Nitin Agrawal, Ashish Vulimiri (Samsung Research)

Author index .......................................................... 665
Sponsors

acm sigops

in cooperation with USENIX

Supporters

HUAWEI

Alibaba Group 阿里巴巴集团
apple 阿普
Arxan Fintech
Baidu 百度
facebook
intel
Microsoft
tencent 腾讯
DiDi
Google
smartx make it simple
vmware

Akami
amazon

F=ma
IBM Research

ORACLE

UBER

Hewlett Packard Enterprise
Organizers

General Chairs
Haibo Chen, Shanghai Jiao Tong University
Lidong Zhou, Microsoft

Program Committee Chairs
Lorenzo Alvisi, Cornell University
Peter Chen, University of Michigan

Local Arrangements
Yubin Xia, Shanghai Jiao Tong University

Poster Chairs
Gernot Heiser, UNSW Australia and Data61
Cheng Li, USTC

Publication Chairs
Jeanna Matthews, Clarkson University
Ming Wu, Microsoft

Publicity Chairs
Yungang Bao, ICT, CAS
Simon Peter, University of Texas at Austin

Registration Chairs
Yang Chen, Fudan University
Patrick Lee, CUHK

Sponsorship Chairs
Wenguang Chen, Tsinghua University
Shan Lu, University of Chicago

Scholarship Chair
Thomas Bressoud, Denison University

Treasurer
Srinath Setty, Microsoft

Tutorial Chairs
Olivier Marin, NYU Shanghai
Catello Di Martino, Bell Labs

Workshop Chair
Hakim Weatherspoon, Cornell University

Program Committee
Marcos K. Aguilera, VMware Research Group
Aditya Akella, University of Wisconsin–Madison
Lorenzo Alvisi, Cornell University
Remzi Arpaci-Dusseau, University of Wisconsin–Madison
Emery Berger, UMass Amherst
George Candea, EPFL
Miguel Castro, Microsoft
Haibo Chen, Shanghai Jiao Tong University
Wenguang Chen, Tsinghua University
Peter Chen, University of Michigan
Byung-Gon Chun, Seoul National University
Landon Cox, Duke University
Mike Dahlin, Google
Peter Druschel, MPI-SWS
Jason Flinn, University of Michigan
Roxana Geambasu, Columbia University
Johannes Gehrke, Microsoft
Cristiano Giuffrida, Vrije Universiteit Amsterdam
Ashvin Goel, University of Toronto
Andreas Haeberlen, University of Pennsylvania
Tim Harris, Oracle
Gernot Heiser, UNSW Australia and Data61
Henry Hoffmann, University of Chicago
Jon Howell, Google
Rebecca Isaacs, Google
Frans Kaashoek, MIT
Manos Kapritsos, University of Michigan
Brad Karp, University College London
Eddie Kohler, Harvard University
Ramakrishna Kotla, Amazon
Arvind Krishnamurthy, University of Washington
Jay Lorch, Microsoft
Shan Lu, University of Chicago
James Mickens, Harvard University
Thomas Moscibroda, Microsoft
Madanlal Musuvathi, Microsoft
John Ousterhout, Stanford University
Simon Peter, University of Texas at Austin
Don Porter, UNC-Chapel Hill
Dan Ports, University of Washington
Michael Reiter, UNC-Chapel Hill
Rodrigo Rodrigues, IST (U. Lisboa) and INESC-ID
Chris Rossbach, The University of Texas at Austin
Marco Serafini, Qatar Computing Research Institute
Mark Silberstein, Technion-Israel Institute of Technology
Ion Stoica, UC Berkeley
Michael Swift, University of Wisconsin–Madison
Doug Terry, Amazon Web Services
Robbert van Renesse, Cornell University
Geoffrey M. Voelker, UC San Diego
Michael Walfish, New York University
Junfeng Yang, Columbia University
Ding Yuan, University of Toronto
Nickolai Zeldovich, MIT
Yuanyuan Zhou, UC San Diego

External Reviewers

Mahesh Balakrishnan, Yale University
Sanjoy Baruah, UNC-Chapel Hill
Iddo Bentov, Cornell University
Björn Brandenburg, MPI-SWS
Aaron Carroll, Apple
Alex Dimakis, The University of Texas at Austin
Ian Goldberg, University of Waterloo
Qirong Ho, Institute for Infocomm Research
Kim Keeton, Hewlett Packard Labs
Donald Kossmann, Microsoft
Ted Wobber, Google
Eric Xing, Carnegie Mellon University