SOSP’13
Proceedings of the Twenty-Fourth ACM
Symposium on Operating Systems Principles

Sponsored by:
ACM SIGOPS in cooperation with USENIX

Supported by:
Cisco, Murty Family Foundation, NSF, Apple, CRA-W, Facebook, NetApp, Microsoft Research, Akamai, Google, HP, Intel, Symantec, VMWare, Bromium, IBM, and Oracle
Forward

Dear Colleagues,

Welcome to the Proceedings of the 24th ACM Symposium on Operating Systems Principles (SOSP 2013), held at the Nemacolin Woodlands Resort, Farmington, Pennsylvania, USA. This year’s program includes 30 papers, and touches on a wide range of computer systems topics, from kernels to big data, from responsiveness to correctness, and from devices to data centers. The program committee made every effort to identify and include some of the most creative and thought-provoking ideas in computer systems today. Each accepted paper was shepherded by a program committee member to make sure the papers are as readable and complete as possible. We hope you will enjoy the program as much as we did in selecting it.

Perhaps the most exciting change this year is that for the first time in the history of SOSP, the entire conference proceedings are Open Access. SIGOPS has been at the forefront of the efforts within ACM to promote this new publication model; we hope that having Open Access proceedings will allow the symposium’s research to reach and inspire the widest possible audience. All of the SOSP 2013 authors have also opted to take advantage of a new rights management option offered by ACM—to retain copyright and ownership of their papers. We are optimistic about these new developments and look forward to seeing how they evolve in the years to come. Special thanks to Jeanna Matthews, SIGOPS Chair; without her tireless efforts, none of these changes would have happened.

We are most grateful to the authors of the 157 papers who chose to submit their work to SOSP. Their ideas and efforts are the basis of the conference’s success. Selecting the program out of so many quality submissions was a difficult task. A program committee consisting of 28 leading scholars in the broad area of computer systems conducted a three-stage reviewing process and online discussion; final selections were made during a physical meeting in Austin, TX, which was attended by a core of 12 PC members. Each submission received at least three PC reviews, with a maximum of eight. All in all, 759 reviews were produced. The PC made every effort to provide detailed and constructive feedback, which should help authors to further improve their work. Author anonymity was maintained throughout the reviewing and selection process; PC members were removed from the deliberations of any paper with which they had a conflict of interest (co-author, same institution, recent collaborator, former student/adviser).

Following the lead of SOSP 2009 and 2011, this year’s conference offers a full slate of workshops on the Sunday before the main event. These workshops cover a range of topics related to operating systems: distributed systems (LADIS), programming languages (PLOS), dependability (HotDep), Power (HotPower), and future memory systems (INFLOW). Two additional forums cover topics of broad interest to our community. The Diversity workshop supports traditionally underrepresented students in systems research, and the TRIOS Conference is an experiment focused on the timely publication of research in our fast-moving field.

We are especially thankful this year for our generous corporate and governmental donors. These donors make it possible to host SOSP in an environment that is conducive to collegial interaction and, this year, they have provided funds for full or partial travel grants to over 130 students from a wide range of countries and institutions.

SOSP is a great conference mostly because it attracts so many high-quality submissions, and we would like to again thank all the authors who submitted. We also thank the PC members for the tremendous amount of work they did: reviewing the submissions, providing feedback, and shepherding the accepted submissions. We are grateful to the external reviewers who provided an additional perspective on a few papers. SOSP has always been organized by volunteer efforts from a host of people; we would like to thank all the
following people who have dedicated so much of their time to the conference: John MacCormick (treasurer), Babu Pillai (local arrangements), Garth Gibson and Michael Swift (sponsorships), Ramakrishna Kotla (publicity), Michael Freedman (workshops), Kimberly Keeton (scholarships), Haibo Chen and Li Zhuang (registration), Allen Clement and Bryan Ford (posters), Emin Gün Sirer (WIPs), Marcos K. Aguilera (BOFs), Wolfgang Richter (conference website), Joseph Trent (submissions website), and Joan Digney (printed materials editor). We would also like to thank Eddie Kohler for continuing to maintain and improve his HotCRP conference management system, far and away the best reviewing system we’ve seen.

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.

Michael Kaminsky                      Mike Dahlin
Intel Labs                          Google and the University of Texas at Austin
General Chair                      Program Chair
Contents

Juggling Chainsaws

The Scalable Commutativity Rule: Designing Scalable Software for Multicore Processors ............ 1

Austin T. Clements, M. Frans Kaashoek, Nickolai Zeldovich, Robert Morris (MIT CSAIL), Eddie Kohler (Harvard)

Speedy Transactions in Multicore In-Memory Databases ................................................................. 18

Stephen Tu, Wenting Zheng (MIT), Eddie Kohler (Harvard), Barbara Liskov, Samuel Madden (MIT)

Everything You Always Wanted to Know about Synchronization but Were Afraid to Ask .......... 33

Tudor David, Rachid Guerraoui, Vasiletos Trigonakis (EPFL)

Time is of the Essence

Dandelion: A Compiler and Runtime for Heterogeneous Systems ...................................................... 49

Christopher J Rossbach, Yuan Yu, Jon Currey, Jean-Philippe Martin, Dennis Fetterly (Microsoft Research Silicon Valley)

Sparrow: Distributed, Low Latency Scheduling .................................................................................. 69

Kay Ousterhout, Patrick Wendell, Matei Zaharia, Ion Stoica (UC Berkeley)

Timecard: Controlling User-Perceived Delays in Server-Based Mobile Applications ................. 85

Lenin Ravindranath (MIT), Jitendra Padhye, Ratul Mahajan (Microsoft Research), Hari Balakrishnan (MIT)

Seed Corn

Fast Dynamic Binary Translation for the Kernel ..................................................................................... 101

Piyus Kedia, Sorav Bansal (IIT Delhi)

VirtuOS: An Operating System with Kernel Virtualization ............................................................... 116

Ruslan Nikolaev, Godmar Baek (Virginia Polytechnic Institute)

From L3 to seL4: What Have We Learnt in 20 Years of L4 Microkernels? ................................. 133

Kevin Elphinstone, Gernot Heiser (NICTA & UNSW)
Everything in its Place

Replication, History, and Grafting in the Ori File System.................................................. 151
Ali Mashtizadeh, Andrea Bittau, Yifeng Frank Huang, David Mazières (Stanford University)

An Analysis of Facebook Photo Caching............................................................................ 167
Qi Huang, Ken Birman, Robbert van Renesse (Cornell University), Wyatt Lloyd (Princeton University), Sanjeev Kumar, Harry C. Li (Facebook Inc.)

IOFlow: A Software-Defined Storage Architecture......................................................... 182
Eno Thereska, Hitesh Ballani, Greg O'Shea, Thomas Karagiannis, Antony Rowstron (Microsoft Research), Tom Talpey (Microsoft), Richard Black (Microsoft Research), Timothy Zhu (Carnegie Mellon University)

From ARIES to MARS: Transaction Support for Next-Generation, Solid-State Drives......... 197
Joel Coburn, Trevor Bunker, Meir Schwarz, Rajesh K. Gupta, Steven Swanson (University of California, San Diego)

Whoops

Asynchronous Intrusion Recovery for Interconnected Web Services................................. 213
Ramesh Chandra, Taesoo Kim, Nickolai Zeldovich (MIT CSAIL)

Optimistic Crash Consistency............................................................................................. 228
Vijay Chidambaram, Thanumalayan Sankaranarayana Pillai, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau (University of Wisconsin-Madison)

Do Not Blame Users for Misconfigurations....................................................................... 244
Tianyin Xu, Jiaqi Zhang, Peng Huang, Jing Zheng, Tianwei Sheng (UC San Diego), Ding Yuan (University of Toronto), Yuanyuan Zhou (UC San Diego), Shankar Pasupathy (NetApp Inc)

Towards Optimization-Safe Systems: Analyzing the Impact of Undefined Behavior...... 260
Xi Wang, Nickolai Zeldovich, M. Frans Kaashoek, Armando Solar-Lezama (MIT CSAIL)

Data, Data, Everywhere

Transaction Chains: Achieving Serializability with Low Latency in Geo-Distributed Storage Systems................................................................. 276
Yang Zhang, Russell Power, Siyuan Zhou, Yair Sovran (NYU), Marcos K. Aguilera (Microsoft Research), Jinyang Li (NYU)

SPANStore: Cost-Effective Geo-Replicated Storage Spanning Multiple Cloud Services..... 292
Zhe Wu, Michael Butkiewicz, Dorian Perkins (UC Riverside), Ethan Katz-Bassett (USC), Harsha V. Madhyastha (UC Riverside)

Consistency-Based Service Level Agreements for Cloud Storage................................... 309
Douglas B. Terry, Vijayan Prabhakaran, Ramakrishna Kotla, Mahesh Balakrishnan, Marcos K. Aguilera (Microsoft Research), Hussam Abu-Libdeh (Cornell University)
Right Makes Might

Tango: Distributed Data Structures over a Shared Log ................................................................. 325
Mahesh Balakrishnan, Dahlia Malkhi, Ted Wobber, Ming Wu, Vijayan Prabhakaran (Microsoft Research), Michael Wei (UCSD), John D. Davis (Microsoft Research), Sriram Rao (Microsoft), Tao Zou (Cornell University), Aviad Zuck (Tel-Aviv University)

Verifying Computations with State .................................................................................................. 341
Benjamin Braun (UT Austin), Ariel J. Feldman (University of Pennsylvania), Zuocheng Ren, Srinath Setty, Andrew J. Blumberg, Michael Walfish (UT Austin)

There Is More Consensus In Egalitarian Parliaments .................................................................... 358
Iulian Moraru, David G. Andersen (Carnegie Mellon University), Michael Kaminsky (Intel Labs)

N' Sync

ROOT: Replaying Multithreaded Traces with Resource-Oriented Ordering ........................................... 373
Zev Weiss, Tyler Harter, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau (University of Wisconsin-Madison)

PARROT: A Practical Runtime for Deterministic, Stable, and Reliable Threads ................................. 388
Heming Cui (Columbia University), Jiri Simsa (Carnegie Mellon University), Yi-Hong Lin, Hao Li (Columbia University), Ben Blum (Carnegie Mellon University), Xinan Xu, Junfeng Yang (Columbia University), Garth A. Gibson, Randal E. Bryant (Carnegie Mellon University)

RaceMob: Crowdsourced Data Race Detection .................................................................................. 406
Baris Kasikci, Cristian Zamfir, George Candea (EPFL)

Data into Information

Discretized Streams: Fault-Tolerant Streaming Computation at Scale ............................................. 423
Matei Zaharia, Tathagata Das, Haoyuan Li, Timothy Hunter, Scott Shenker, Ion Stoica (UC Berkeley)

Naiad: A Timely Dataflow System ....................................................................................................... 439
Derek G. Murray, Frank McSherry, Rebecca Isaacs, Michael Isard, Paul Barham, Martin Abadi (Microsoft Research)

A Lightweight Infrastructure for Graph Analytics ............................................................................... 456
Donald Nguyen, Andrew Lenharth, Keshav Pingali (University of Texas at Austin)

X-Stream: Edge-Centric Graph Processing using Streaming Partitions ............................................ 472
Amitabha Roy, Ivo Mihailovic, Willy Zwaenepoel (EPFL)
Sponsored by

in cooperation with USENIX

Supported by

**Platinum Level**
- Cisco
- Murty Family Foundation
- National Science Foundation

**Gold Level**
- Apple
- CRA-W
- Facebook
- NetApp
- Microsoft Research

**Silver Level**
- Akamai
- Google
- HP
- Intel
- Symantec
- VMWare

**Bronze Level**
- Bromium
- IBM
- Oracle
# Organizers

<table>
<thead>
<tr>
<th>Role</th>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Chair</td>
<td>Michael Kaminsky</td>
<td>Intel Labs</td>
</tr>
<tr>
<td>Program Chair</td>
<td>Mike Dahlin</td>
<td>Google and UT Austin</td>
</tr>
<tr>
<td>Treasurer</td>
<td>John MacCormick</td>
<td>Dickinson College</td>
</tr>
<tr>
<td>Local Arrangements</td>
<td>Babu Pillai</td>
<td>Intel Labs</td>
</tr>
<tr>
<td>Sponsorships Chair</td>
<td>Garth Gibson</td>
<td>Carnegie Mellon University</td>
</tr>
<tr>
<td>Sponsorships Co-Chair</td>
<td>Michael Swift</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td>Publicity Chair</td>
<td>Ramakrishna Kotla</td>
<td>MSR</td>
</tr>
<tr>
<td>Workshops Chair</td>
<td>Michael J. Freedman</td>
<td>Princeton University</td>
</tr>
<tr>
<td>Scholarship Chair</td>
<td>Kimberly Keeton</td>
<td>HP Labs</td>
</tr>
<tr>
<td>Registration Chairs</td>
<td>Haibo Chen</td>
<td>Shanghai Jiao Tong University</td>
</tr>
<tr>
<td></td>
<td>Li Zhuang</td>
<td>MSR Asia</td>
</tr>
<tr>
<td>Poster Chairs</td>
<td>Allen Clement</td>
<td>Max Planck Institute for Software Systems</td>
</tr>
<tr>
<td></td>
<td>Bryan Ford</td>
<td>Yale University</td>
</tr>
<tr>
<td>WIP Chair</td>
<td>Emin Gün Sirer</td>
<td>Cornell University</td>
</tr>
<tr>
<td>BOF Chair</td>
<td>Marcos K. Aguilera</td>
<td>MSR</td>
</tr>
<tr>
<td>Conference Webmaster</td>
<td>Wolfgang Richter</td>
<td>Carnegie Mellon University</td>
</tr>
<tr>
<td>Submissions Webmaster</td>
<td>Joseph Trent</td>
<td>UT Austin</td>
</tr>
<tr>
<td>Printed Materials Editor</td>
<td>Joan Digney</td>
<td>Carnegie Mellon University</td>
</tr>
<tr>
<td>Program Committee</td>
<td>David Andersen</td>
<td>Carnegie Mellon University</td>
</tr>
<tr>
<td></td>
<td>Herbert Bos</td>
<td>VU</td>
</tr>
<tr>
<td></td>
<td>George Candea</td>
<td>EPFL</td>
</tr>
<tr>
<td></td>
<td>Haibo Chen</td>
<td>Shanghai Jiao Tong University</td>
</tr>
<tr>
<td></td>
<td>Mike Dahlin</td>
<td>Google and UT Austin</td>
</tr>
<tr>
<td></td>
<td>Kevin Elphinstone</td>
<td>UNSW and NICTA</td>
</tr>
<tr>
<td></td>
<td>Bryan Ford</td>
<td>Yale University</td>
</tr>
<tr>
<td></td>
<td>Steven Hand</td>
<td>University of Cambridge</td>
</tr>
<tr>
<td></td>
<td>Jon Howell</td>
<td>MSR</td>
</tr>
<tr>
<td></td>
<td>Rebecca Isaacs</td>
<td>MSR</td>
</tr>
<tr>
<td></td>
<td>Anthony Joseph</td>
<td>UC Berkeley</td>
</tr>
<tr>
<td></td>
<td>Sam King</td>
<td>UIUC</td>
</tr>
<tr>
<td></td>
<td>Eddie Köhler</td>
<td>Harvard</td>
</tr>
<tr>
<td></td>
<td>Arvind Krishnamurthy</td>
<td>University of Washington</td>
</tr>
<tr>
<td></td>
<td>Philip Levis</td>
<td>Stanford University</td>
</tr>
<tr>
<td></td>
<td>Shan Lu</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td></td>
<td>James Mickens</td>
<td>MSR</td>
</tr>
<tr>
<td></td>
<td>Robert Morris</td>
<td>MIT</td>
</tr>
<tr>
<td></td>
<td>John Ousterhout</td>
<td>Stanford University</td>
</tr>
<tr>
<td></td>
<td>Robbert van Renesse</td>
<td>Cornell University</td>
</tr>
<tr>
<td></td>
<td>Timothy Roscoe</td>
<td>ETH Zurich</td>
</tr>
<tr>
<td></td>
<td>Michael Swift</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td></td>
<td>Michael Walfish</td>
<td>UT Austin</td>
</tr>
<tr>
<td></td>
<td>Helen Wang</td>
<td>MSR</td>
</tr>
<tr>
<td></td>
<td>John Wilkes</td>
<td>Google</td>
</tr>
<tr>
<td></td>
<td>Nickolai Zeldovich</td>
<td>MIT</td>
</tr>
<tr>
<td></td>
<td>Lidong Zhou</td>
<td>MSR Asia</td>
</tr>
<tr>
<td></td>
<td>Yuanyuan Zhou</td>
<td>UC San Diego</td>
</tr>
</tbody>
</table>
**Student Scholarship Committee**  
Atul Adya (Google)  
Angela Demke Brown (University of Toronto)  
Edouard Bugnion (EPFL)  
Steve Gribble (University of Washington)  
Kimberly Keeton (HP Labs)  
Derek Murray (MSR-Silicon Valley)  
George Porter (UC San Diego)

**Professional Scholarship Committee**  
Dilma Da Silva (Qualcomm)  
Cary Gray (Wheaton College)  
Geoff Kuenning (Harvey Mudd College)  
Jeanna Matthews (Clarkson University)

**Poster Program Committee**  
Sorav Bansal (Indian Institute of Technology Delhi)  
Ivan Beschastnikh (University of British Columbia)  
Allen Clement (Max Planck Institute for Software Systems)  
Dilma Da Silva (Qualcomm)  
Joseph Devietti (University of Pennsylvania)  
Bryan Ford (Yale University)  
Michio Honda (NEC Europe Ltd.)  
James Mickens (MSR)  
Don Porter (Stony Brook University)  
Mark Silberstein (Technion)  
John Wilkes (Google)  
Yu Zhang (University of Science and Technology of China)

**WIP Program Committee**  
Remzi Arpaci-Dusseau (University of Wisconsin)  
Michael J. Freedman (Princeton University)  
Emin Gün Sirer (Cornell University)  
Bernard Wong (University of Waterloo)