

USENIX ATC '14:
2014 USENIX Annual Technical Conference
June 19–20, 2014
Philadelphia, PA

Message from the Program Co-Chairs. vii

Thursday, June 19

Big Data

ShuffleWatcher: Shuffle-aware Scheduling in Multi-tenant MapReduce Clusters1
Faraz Ahmad, *Teradata Aster and Purdue University*; Srimat T. Chakradhar, *NEC Laboratories America*;
Anand Raghunathan and T. N. Vijaykumar, *Purdue University*

Violet: A Storage Stack for IOPS/Capacity Bifurcated Storage Environments13
Douglas Santry and Kaladhar Voruganti, *NetApp, Inc.*

ELF: Efficient Lightweight Fast Stream Processing at Scale25
Liting Hu, Karsten Schwan, Hrishikesh Amur, and Xin Chen, *Georgia Institute of Technology*

Exploiting Bounded Staleness to Speed Up Big Data Analytics37
Henggang Cui, James Cipar, Qirong Ho, Jin Kyu Kim, Seunghak Lee, Abhimanu Kumar, Jinliang Wei,
Wei Dai, and Gregory R. Ganger, *Carnegie Mellon University*; Phillip B. Gibbons, *Intel Labs*; Garth A. Gibson
and Eric P. Xing, *Carnegie Mellon University*

Making State Explicit for Imperative Big Data Processing49
Raul Castro Fernandez, *Imperial College London*; Matteo Migliavacca, *University of Kent*; Evangelia Kalyvianaki,
City University London; Peter Pietzuch, *Imperial College London*

Virtualization

OSv—Optimizing the Operating System for Virtual Machines61
Avi Kivity, Dor Laor, Glauber Costa, Pekka Enberg, Nadav Har’El, Don Marti, and Vlad Zolotarov,
Clouddius Systems

Gleaner: Mitigating the Blocked-Waiter Wakeup Problem for Virtualized Multicore Applications73
Xiaoning Ding, *New Jersey Institute of Technology*; Phillip B. Gibbons and Michael A. Kozuch, *Intel Labs*
Pittsburgh; Jianchen Shan, *New Jersey Institute of Technology*

HYPERHELL: A Practical Hypervisor Layer Guest OS Shell for Automated In-VM Management85
Yangchun Fu, Junyuan Zeng, and Zhiqiang Lin, *The University of Texas at Dallas*

XvMotion: Unified Virtual Machine Migration over Long Distance97
Ali José Mashtizadeh, *Stanford University*; Min Cai, Gabriel Tarasuk-Levin, and Ricardo Koller, *VMware, Inc.*;
Tal Garfinkel; Sreekanth Setty, *VMware, Inc.*

GPUvm: Why Not Virtualizing GPUs at the Hypervisor?109
Yusuke Suzuki, *Keio University*; Shinpei Kato, *Nagoya University*; Hiroshi Yamada, *Tokyo University of*
Agriculture and Technology; Kenji Kono, *Keio University*

A Full GPU Virtualization Solution with Mediated Pass-Through121
Kun Tian, Yaozu Dong, and David Cowperthwaite, *Intel Corporation*

(Thursday, June 19, continues on p. iv)

Storage

- vCacheShare: Automated Server Flash Cache Space Management in a Virtualization Environment133**
Fei Meng, *North Carolina State University*; Li Zhou, *Facebook*; Xiaosong Ma, *North Carolina State University and Qatar Computing Research Institute*; Sandeep Uttamchandani, *VMware Inc.*; Deng Liu, *Twitter*
- Missive: Fast Application Launch From an Untrusted Buffer Cache145**
Jon Howell, Jeremy Elson, Bryan Parno, and John R. Douceur, *Microsoft Research*
- A Modular and Efficient Past State System for Berkeley DB.157**
Ross Shaull, *NuoDB*; Liuba Shriru, *Brandeis University*; Barbara Liskov, *MIT/CSAIL*
- SCFS: A Shared Cloud-backed File System.169**
Alysson Bessani, Ricardo Mendes, Tiago Oliveira, and Nuno Neves, *Faculdade de Ciências and LaSIGE*; Miguel Correia, *INESC-ID and Instituto Superior Técnico, University of Lisbon*; Marcelo Pasin, *Université de Neuchâtel*; Paulo Verissimo, *Faculdade de Ciências and LaSIGE*
- Accelerating Restore and Garbage Collection in Deduplication-based Backup Systems via Exploiting Historical Information.181**
Min Fu, Dan Feng, and Yu Hua, *Huazhong University of Science and Technology*; Xubin He, *Virginia Commonwealth University*; Zuoning Chen, *National Engineering Research Center for Parallel Computer*; Wen Xia, Fangting Huang, and Qing Liu, *Huazhong University of Science and Technology*
- ## Hardware and Low-level Techniques
- The TURBO Diaries: Application-controlled Frequency Scaling Explained.193**
Jons-Tobias Wamhoff, Stephan Diestelhorst, and Christof Fetzer, *Technische Universität Dresden*; Patrick Marlier and Pascal Felber, *Université de Neuchâtel*; Dave Dice, *Oracle Labs*
- Implementing a Leading Loads Performance Predictor on Commodity Processors205**
Bo Su, *National University of Defense Technology*; Joseph L. Greathouse, Junli Gu, and Michael Boyer, *AMD Research*; Li Shen and Zhiying Wang, *National University of Defense Technology*
- HaPPy: Hyperthread-aware Power Profiling Dynamically211**
Yan Zhai, *University of Wisconsin*; Xiao Zhang and Stephane Eranian, *Google Inc.*; Lingjia Tang and Jason Mars, *University of Michigan*
- Scalable Read-mostly Synchronization Using Passive Reader-Writer Locks219**
Ran Liu, *Fudan University and Shanghai Jiao Tong University*; Heng Zhang and Haibo Chen, *Shanghai Jiao Tong University*
- Large Pages May Be Harmful on NUMA Systems231**
Fabien Gaud, *Simon Fraser University*; Baptiste Lepers, *CNRS*; Jeremie Decouchant, *Grenoble University*; Justin Funston and Alexandra Fedorova, *Simon Fraser University*; Vivien Quéma, *Grenoble INP*
- Efficient Tracing of Cold Code via Bias-Free Sampling.243**
Baris Kasikci, *École Polytechnique Fédérale de Lausanne (EPFL)*; Thomas Ball, *Microsoft*; George Candea, *École Polytechnique Fédérale de Lausanne (EPFL)*; John Erickson and Madanlal Musuvathi, *Microsoft*

Friday, June 20, 2014

Distributed Systems

- Gestalt: Fast, Unified Fault Localization for Networked Systems**255
Radhika Niranjana Mysore, *Google*; Ratul Mahajan, *Microsoft Research*; Amin Vahdat, *Google*;
George Varghese, *Microsoft Research*
- Insight: In-situ Online Service Failure Path Inference in Production Computing Infrastructures**269
Hiep Nguyen, Daniel J. Dean, Kamal Kc, and Xiaohui Gu, *North Carolina State University*
- Automating the Choice of Consistency Levels in Replicated Systems**281
Cheng Li, *Max Planck Institute for Software Systems (MPI-SWS)*; Joao Leitão, *NOVA University of Lisbon/ CITI/NOVA-LINCS*; Allen Clement, *Max Planck Institute for Software Systems (MPI-SWS)*; Nuno Preguiça and Rodrigo Rodrigues, *NOVA University of Lisbon/CITI/NOVA-LINCS*; Viktor Vafeiadis, *Max Planck Institute for Software Systems (MPI-SWS)*
- Sirius: Distributing and Coordinating Application Reference Data**293
Michael Bevilacqua-Linn, Maulan Byron, Peter Cline, Jon Moore, and Steve Muir, *Comcast Cable*
- In Search of an Understandable Consensus Algorithm**305
Diego Ongaro and John Ousterhout, *Stanford University*

Networking

- GASPP: A GPU-Accelerated Stateful Packet Processing Framework**321
Giorgos Vasiliadis and Lazaros Koromilas, *FORTH-ICS*; Michalis Polychronakis, *Columbia University*;
Sotiris Ioannidis, *FORTH-ICS*
- Panopticon: Reaping the Benefits of Incremental SDN Deployment in Enterprise Networks**333
Dan Levin, *Technische Universität Berlin*; Marco Canini, *Université catholique de Louvain*; Stefan Schmid, *Technische Universität Berlin and Telekom Innovation Labs*; Fabian Schaffert and Anja Feldmann, *Technische Universität Berlin*
- Programmatic Orchestration of WiFi Networks**347
Julius Schulz-Zander, Lalith Suresh, Nadi Sarrar, and Anja Feldmann, *Technische Universität Berlin*;
Thomas Hühn, *DAI-Labor and Technische Universität Berlin*; Ruben Merz, *Swisscom*
- HACK: Hierarchical ACKs for Efficient Wireless Medium Utilization**359
Lynne Salameh, Astrit Zhushi, Mark Handley, Kyle Jamieson, and Brad Karp, *University College London*
- Pythia: Diagnosing Performance Problems in Wide Area Providers**371
Partha Kanuparth, *Yahoo Labs*; Constantine Dovrolis, *Georgia Institute of Technology*
- BISmark: A Testbed for Deploying Measurements and Applications in Broadband Access Networks**383
Srikanth Sundaresan, Sam Burnett, and Nick Feamster, *Georgia Institute of Technology*; Walter de Donato, *University of Naples Federico II*

Security and Correctness

- Application-Defined Decentralized Access Control**395
Yuanzhong Xu and Alan M. Dunn, *The University of Texas at Austin*; Owen S. Hofmann, *Google, Inc.*; Michael Z. Lee, Syed Akbar Mehdi, and Emmett Witchel, *The University of Texas at Austin*
- MiniBox: A Two-Way Sandbox for x86 Native Code**409
Yanlin Li, *CyLab/Carnegie Mellon University*; Jonathan McCune and James Newsome, *CyLab/Carnegie Mellon University and Google, Inc.*; Adrian Perrig, *CyLab/Carnegie Mellon University*; Brandon Baker and Will Drewry, *Google, Inc.*

(Friday, June 20, continues on p. vi)

Static Analysis of Variability in System Software: The 90,000 #ifdefs Issue	421
Reinhard Tartler, Christian Dietrich, Julio Sincero, Wolfgang Schröder-Preikschat, and Daniel Lohmann, <i>Friedrich-Alexander-Universität Erlangen-Nürnberg</i>	
Yat: A Validation Framework for Persistent Memory Software	433
Philip Lantz, Subramanya Dullloor, Sanjay Kumar, Rajesh Sankaran, and Jeff Jackson, <i>Intel Labs</i>	
Medusa: Managing Concurrency and Communication in Embedded Systems	439
Thomas W. Barr and Scott Rixner, <i>Rice University</i>	
Flash	
Reliable Writeback for Client-side Flash Caches	451
Dai Qin, Angela Demke Brown, and Ashvin Goel, <i>University of Toronto</i>	
Flash on Rails: Consistent Flash Performance through Redundancy	463
Dimitris Skourtis, Dimitris Achlioptas, Noah Watkins, Carlos Maltzahn, and Scott Brandt, <i>University of California, Santa Cruz</i>	
I/O Speculation for the Microsecond Era	475
Michael Wei, <i>University of California, San Diego</i> ; Matias Bjørling and Philippe Bonnet, <i>IT University of Copenhagen</i> ; Steven Swanson, <i>University of California, San Diego</i>	
OS I/O Path Optimizations for Flash Solid-state Drives	483
Woong Shin, Qichen Chen, Myoungwon Oh, Hyeonsang Eom, and Heon Y. Yeom, <i>Seoul National University</i>	
FlexECC: Partially Relaxing ECC of MLC SSD for Better Cache Performance	489
Ping Huang, <i>Virginia Commonwealth University and Huazhong University of Science and Technology</i> ; Pradeep Subedi, <i>Virginia Commonwealth University</i> ; Xubin He, <i>Virginia Commonwealth University</i> ; Shuang He and Ke Zhou, <i>Huazhong University of Science and Technology</i>	
Nitro: A Capacity-Optimized SSD Cache for Primary Storage	501
Cheng Li, <i>Rutgers University</i> ; Philip Shilane, Fred Douglass, Hyong Shim, Stephen Smaldone, and Grant Wallace, <i>EMC Corporation</i>	