

# USENIX ATC '17: 2017 USENIX Annual Technical Conference

## Contents

### Kernel

- Lock-in-Pop: Securing Privileged Operating System Kernels by Keeping on the Beaten Path** ..... 1  
Yiwen Li, Brendan Dolan-Gavitt, Sam Weber, and Justin Cappos, *New York University*
- Fast and Precise Retrieval of Forward and Back Porting Information for Linux Device Drivers** ..... 15  
Julia Lawall, Derek Palinski, Lukas Gnirke, and Gilles Muller, *Sorbonne Universités/UPMC/Inria/LIP6*
- Optimizing the TLB Shutdown Algorithm with Page Access Tracking** ..... 27  
Nadav Amit, *VMware Research*
- Falcon: Scaling IO Performance in Multi-SSD Volumes** ..... 41  
Pradeep Kumar and H. Howie Huang, *The George Washington University*

### Datacenters

- deTector: a Topology-aware Monitoring System for Data Center Networks** ..... 55  
Yanghua Peng, *The University of Hong Kong*; Ji Yang, *Xi'an Jiaotong University*; Chuan Wu, *The University of Hong Kong*; Chuanxiong Guo, *Microsoft Research*; Chengchen Hu, *Xi'an Jiaotong University*; Zongpeng Li, *University of Calgary*
- Pricing Intra-Datacenter Networks with Over-Committed Bandwidth Guarantee** ..... 69  
Jian Guo, Fangming Liu, and Tao Wang, *Key Laboratory of Services Computing Technology and System, Ministry of Education, School of Computer Science and Technology, Huazhong University of Science and Technology*; John C.S. Lui, *The Chinese University of Hong Kong*
- Unobtrusive Deferred Update Stabilization for Efficient Geo-Replication** ..... 83  
Chathuri Gunawardhana, Manuel Bravo, and Luis Rodrigues, *University of Lisbon*
- Don't cry over spilled records: Memory elasticity of data-parallel applications and its application to cluster scheduling** ..... 97  
Călin Iorgulescu and Florin Dinu, *EPFL*; Aunn Raza, *NUST Pakistan*; Wajih Ul Hassan, *UIUC*; Willy Zwaenepoel, *EPFL*

### Pursuing Efficiency

- Popularity Prediction of Facebook Videos for Higher Quality Streaming** ..... 111  
Linpeng Tang, *Princeton University*; Qi Huang and Amit Puntambekar, *Facebook*; Ymir Vigfusson, *Emory University & Reykjavik University*; Wyatt Lloyd, *University of Southern California & Facebook*; Kai Li, *Princeton University*
- Squeezing out All the Value of Loaded Data: An Out-of-core Graph Processing System with Reduced Disk I/O** ..... 125  
Zhiyuan Ai, Mingxing Zhang, and Yongwei Wu, *Department of Computer Science and Technology, Tsinghua National Laboratory for Information Science and Technology (TNLIST), Tsinghua University and Research Institute of Tsinghua*; Xuehai Qian, *University of Southern California*; Kang Chen and Weimin Zheng, *Department of Computer Science and Technology, Tsinghua National Laboratory for Information Science and Technology (TNLIST), Tsinghua University, and Research Institute of Tsinghua*
- Ending the Anomaly: Achieving Low Latency and Airtime Fairness in WiFi** ..... 139  
Toke Høiland-Jørgensen, *Karlstad University*; Michał Kazior, *Tieto Poland*; Dave Täht, *TekLibre*; Per Hurtig and Anna Brunstrom, *Karlstad University*

<b>Persona: A High-Performance Bioinformatics Framework</b> . . . . .	<b>153</b>
Stuart Byma and Sam Whitlock, <i>EPFL</i> ; Laura Flueratoru, <i>University Politehnica of Bucharest</i> ; Ethan Tseng, <i>CMU</i> ; Christos Kozyrakis, <i>Stanford University</i> ; Edouard Bugnion and James Larus, <i>EPFL</i>	
<b>Let's Talk about GPUs</b>	
<b>SPIN: Seamless Operating System Integration of Peer-to-Peer DMA Between SSDs and GPUs</b> . . . . .	<b>167</b>
Shai Bergman and Tanya Brokhman, <i>Technion</i> ; Tzachi Cohen, <i>unaffiliated</i> ; Mark Silberstein, <i>Technion</i>	
<b>Poseidon: An Efficient Communication Architecture for Distributed Deep Learning on GPU Clusters</b> . . . . .	<b>181</b>
Hao Zhang, <i>Carnegie Mellon University</i> ; Zeyu Zheng, <i>Petuum Inc.</i> ; Shizhen Xu and Wei Dai, <i>Carnegie Mellon University</i> ; Qirong Ho, <i>Petuum Inc.</i> ; Xiaodan Liang, Zhiting Hu, Jinliang Wei, and Pengtao Xie, <i>Carnegie Mellon University</i> ; Eric P.Xing, <i>Petuum Inc.</i>	
<b>Garaph: Efficient GPU-accelerated Graph Processing on a Single Machine with Balanced Replication</b> . . . . .	<b>195</b>
Lingxiao Ma, Zhi Yang, and Han Chen, <i>Computer Science Department, Peking University, Beijing, China</i> ; Jilong Xue, <i>Microsoft Research, Beijing, China</i> ; Yafei Dai, <i>Institute of Big Data Technologies Shenzhen Key Lab for Cloud Computing Technology &amp; Applications, School of Electronics and Computer Engineering (SECE), Peking University, Shenzhen, China</i>	
<b>GPU Taint Tracking</b> . . . . .	<b>209</b>
Ari B. Hayes, <i>Rutgers University</i> ; Lingda Li, <i>Brookhaven National Laboratory</i> ; Mohammad Hedayati, <i>University of Rochester</i> ; Jiahuan He and Eddy Z. Zhang, <i>Rutgers University</i> ; Kai Shen, <i>Google</i>	
<b>Virtualization</b>	
<b>Optimizing the Design and Implementation of the Linux ARM Hypervisor</b> . . . . .	<b>221</b>
Christoffer Dall, Shih-Wei Li, and Jason Nieh, <i>Columbia University</i>	
<b>Multi-Hypervisor Virtual Machines: Enabling an Ecosystem of Hypervisor-level Services</b> . . . . .	<b>235</b>
Kartik Gopalan, Rohit Kugve, Hardik Bagdi, and Yaohui Hu, <i>Binghamton University</i> ; Daniel Williams and Nilton Bila, <i>IBM T.J. Watson Research Center</i>	
<b>Preemptive, Low Latency Datacenter Scheduling via Lightweight Virtualization</b> . . . . .	<b>251</b>
Wei Chen, <i>University of Colorado, Colorado Springs</i> ; Jia Rao, <i>University of Texas at Arlington</i> ; Xiaobo Zhou, <i>University of Colorado, Colorado Springs</i>	
<b>The RCU-Reader Preemption Problem in VMs</b> . . . . .	<b>265</b>
Aravinda Prasad and K Gopinath, <i>Indian Institute of Science, Bangalore</i> ; Paul E. McKenney, <i>IBM Linux Technology Center, Beaverton</i>	
<b>Security and Privacy I</b>	
<b>Bunshin: Compositing Security Mechanisms through Diversification</b> . . . . .	<b>271</b>
Meng Xu, Kangjie Lu, Taesoo Kim, and Wenke Lee, <i>Georgia Institute of Technology</i>	
<b>Glamdring: Automatic Application Partitioning for Intel SGX</b> . . . . .	<b>285</b>
Joshua Lind, Christian Priebe, Divya Muthukumaran, Dan O'Keeffe, Pierre-Louis Aublin, and Florian Kelbert, <i>Imperial College London</i> ; Tobias Reiher, <i>TU Dresden</i> ; David Goltzsche, <i>TU Braunschweig</i> ; David Eysers, <i>University of Otago</i> ; Rudiger Kapitza, <i>TU Braunschweig</i> ; Christof Fetzer, <i>TU Dresden</i> ; Peter Pietzuch, <i>Imperial College London</i>	
<b>High-Resolution Side Channels for Untrusted Operating Systems</b> . . . . .	<b>299</b>
Marcus Hähnel, <i>TU Dresden, Operating Systems Group</i> ; Weidong Cui and Marcus Peinado, <i>Microsoft Research</i>	
<b>Understanding Security Implications of Using Containers in the Cloud</b> . . . . .	<b>313</b>
Byungchul Tak, <i>Kyungpook National University</i> ; Canturk Isci, Sastry Duri, Nilton Bila, Shripad Nadgowda, and James Doran, <i>IBM TJ Watson Research Center</i>	

(continued on next page)

## Key-Value Stores and Databases

- Memshare: a Dynamic Multi-tenant Key-value Cache** . . . . . **321**  
Asaf Cidon, *Stanford University*; Daniel Rushton, *University of Utah*; Stephen M. Rumble, *Google Inc.*;  
Ryan Stutsman, *University of Utah*
- Replication-driven Live Reconfiguration for Fast Distributed Transaction Processing** . . . . . **335**  
Xingda Wei, Sijie Shen, Rong Chen, and Haibo Chen, *Shanghai Jiao Tong University*
- HiKV: A Hybrid Index Key-Value Store for DRAM-NVM Memory Systems** . . . . . **349**  
Fei Xia, *Institute of Computing Technology, Chinese Academy of Sciences*; *University of Chinese Academy of Sciences*; Dejun Jiang, Jin Xiong, and Ninghui Sun, *Institute of Computing Technology, Chinese Academy of Sciences*
- TRIAD: Creating Synergies Between Memory, Disk and Log in Log Structured Key-Value Stores** . . . . . **363**  
Oana Balmau, Diego Didona, Rachid Guerraoui, and Willy Zwaenepoel, *EPFL*; Huapeng Yuan, Aashray Arora, Karan Gupta, and Pavan Konka, *Nutanix*

## Help Me Debug

- Engineering Record And Replay For Deployability** . . . . . **377**  
Robert O'Callahan and Chris Jones, *unaffiliated*; Nathan Froyd, *Mozilla Corporation*; Kyle Huey, *unaffiliated*;  
Albert Noll, *Swisscom AG*; Nimrod Partush, *Technion*
- Proactive error prediction to improve storage system reliability** . . . . . **391**  
Farzaneh Mahdisoltani, *University of Toronto*; Ioan Stefanovici, *Microsoft Research*; Bianca Schroeder, *University of Toronto*
- Towards Production-Run Heisenbugs Reproduction on Commercial Hardware** . . . . . **403**  
Shiyu Huang, Bowen Cai, and Jeff Huang, *Texas A&M University*
- A DSL Approach to Reconcile Equivalent Divergent Program Executions** . . . . . **417**  
Luís Pina, Daniel Grumberg, Anastasios Andronidis, and Cristian Cadar, *Imperial College London*

## Networking

- Titan: Fair Packet Scheduling for Commodity Multiqueue NICs** . . . . . **431**  
Brent Stephens, Arjun Singhvi, Aditya Akella, and Michael Swift, *UW-Madison*
- MopEye: Opportunistic Monitoring of Per-app Mobile Network Performance** . . . . . **445**  
Daoyuan Wu, *Singapore Management University*; Rocky K. C. Chang, Weichao Li, and Eric K. T. Cheng, *The Hong Kong Polytechnic University*; Debin Gao, *Singapore Management University*
- Emu: Rapid Prototyping of Networking Services** . . . . . **459**  
Nik Sultana, Salvator Galea, David Greaves, Marcin Wojcik, and Jonny Shipton, *University of Cambridge*;  
Richard Clegg, *Queen Mary University of London*; Luo Mai, *Imperial College London*; Pietro Bressana and Robert Soule, *Università della Svizzera italiana*; Richard Mortier, *University of Cambridge*; Paolo Costa, *Microsoft Research*; Peter Pietzuch, *Imperial College London*; Jon Crowcroft, Andrew W Moore, and Noa Zilberman, *University of Cambridge*
- Protego: Cloud-Scale Multitenant IPsec Gateway** . . . . . **473**  
Jeongseok Son, *KAIST, Microsoft Research*; Yongqiang Xiong, *Microsoft Research*; Kun Tan, *Huawei*;  
Paul Wang and Ze Gan, *Microsoft Research*; Sue Moon, *KAIST*

## Caching along the Way

- Cache Modeling and Optimization using Miniature Simulations** . . . . . **487**  
Carl Waldspurger, Trausti Saemundson, and Irfan Ahmad, *CachePhysics, Inc.*; Nohhyun Park, *Datos IO, Inc.*
- Hyperbolic Caching: Flexible Caching for Web Applications** . . . . . **499**  
Aaron Blankstein, *Princeton University*; Siddhartha Sen, *Microsoft Research*; Michael J. Freedman, *Princeton University*

**Execution Templates: Caching Control Plane Decisions for Strong Scaling of Data Analytics** . . . . . **513**  
Omid Mashayekhi, Hang Qu, Chinmayee Shah, and Philip Levis, *Stanford University*

**cHash: Detection of Redundant Compilations via AST Hashing.** . . . . . **527**  
Christian Dietrich and Valentin Rothberg, *Leibniz Universität Hannover*; Ludwig Füracker and Andreas Ziegler,  
*Friedrich-Alexander Universität Erlangen-Nürnberg*; Daniel Lohmann, *Leibniz Universität Hannover*

## **Storage**

**Giza: Erasure Coding Objects across Global Data Centers.** . . . . . **539**  
Yu Lin Chen, *NYU & Microsoft Corporation*; Shuai Mu and Jinyang Li, *NYU*; Cheng Huang, Jin Li,  
Aaron Ogus, and Douglas Phillips, *Microsoft Corporation*

**SmartCuckoo: A Fast and Cost-Efficient Hashing Index Scheme for Cloud Storage Systems.** . . . . . **553**  
Yuanyuan Sun and Yu Hua, *Huazhong University of Science and Technology*; Song Jiang, *University of Texas,*  
*Arlington*; Qiuyu Li, Shunde Cao, and Pengfei Zuo, *Huazhong University of Science and Technology*

**Repair Pipelining for Erasure-Coded Storage** . . . . . **567**  
Runhui Li, Xiaolu Li, Patrick P. C. Lee, and Qun Huang, *The Chinese University of Hong Kong*

**PARIX: Speculative Partial Writes in Erasure-Coded Systems** . . . . . **581**  
Huiba Li, *mos.meituan.com*; Yiming Zhang, *NUDT*; Zhiming Zhang, *mos.meituan.com*; Shengyun Liu,  
Dongsheng Li, Xiaohui Liu, and Yuxing Peng, *NUDT*

## **Multicore**

**E-Team: Practical Energy Accounting for Multi-Core Systems.** . . . . . **589**  
Till Smejkal and Marcus Hähnel, *TU Dresden*; Thomas Ilsche, *Center for Information Services and High*  
*Performance Computing (ZIH) Technische Universität Dresden*; Michael Roitzsch, *TU Dresden*; Wolfgang  
E. Nagel, *Center for Information Services and High Performance Computing (ZIH) Technische Universität*  
*Dresden*; Hermann Härtig, *TU Dresden*

**Scalable NUMA-aware Blocking Synchronization Primitives** . . . . . **603**  
Sanidhya Kashyap, Changwoo Min, and Taesoo Kim, *Georgia Institute of Technology*

**StreamBox: Modern Stream Processing on a Multicore Machine.** . . . . . **617**  
Hongyu Miao and Heejin Park, *Purdue ECE*; Myeongjae Jeon and Gennady Pekhimenko, *Microsoft Research*;  
Kathryn S. McKinley, *Google*; Felix Xiaozhu Lin, *Purdue ECE*

**Everything you always wanted to know about multicore graph processing but were afraid to ask** . . . . . **631**  
Jasmina Malicevic, Baptiste Lepers, and Willy Zwaenepoel, *EPFL*

## **Security and Privacy II**

**Graphene-SGX: A Practical Library OS for Unmodified Applications on SGX** . . . . . **645**  
Chia-Che Tsai, *Stony Brook University*; Donald E. Porter, *University of North Carolina at Chapel Hill and*  
*Fortanix*; Mona Vij, *Intel Corporation*

**PrivApprox: Privacy-Preserving Stream Analytics** . . . . . **659**  
Do Le Quoc and Martin Beck, *TU Dresden*; Pramod Bhatotia, *University of Edinburgh*; Ruichuan Chen,  
*Nokia Bell Labs*; Christof Fetzer and Thorsten Strufe, *TU Dresden*

**Mercury: Bandwidth-Effective Prevention of Rollback Attacks Against Community**  
**Repositories** . . . . . **673**  
Trishank Karthik Kuppusamy, Vladimir Diaz, and Justin Cappos, *New York University*

(continued on next page)

**CAB-Fuzz: Practical Concolic Testing Techniques for COTS Operating Systems** . . . . . 689  
Su Yong Kim, *The Affiliated Institute of ETRI*; Sangho Lee, Insu Yun, and Wen Xu, *Georgia Tech*;  
Byoungyoung Lee, *Purdue University*; Youngtae Yun, *The Affiliated Institute of ETRI*; Taesoo Kim,  
*Georgia Tech*

## **Don't Forget the Memory**

**Log-Structured Non-Volatile Main Memory** . . . . . 703  
Qingda Hu, *Tsinghua University*; Jinglei Ren and Anirudh Badam, *Microsoft Research*; Jiwu Shu,  
*Tsinghua University*; Thomas Moscibroda, *Microsoft Research*

**Soft Updates Made Simple and Fast on Non-volatile Memory** . . . . . 719  
Mingkai Dong and Haibo Chen, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University*

**SmartMD: A High Performance Deduplication Engine with Mixed Pages** . . . . . 733  
Fan Guo, *University of Science and Technology of China*; Yongkun Li, *University of Science and Technology of China*; *Collaborative Innovation Center of High Performance Computing, NUDT*; Yinlong Xu, *University of Science and Technology of China*; *Anhui Province Key Laboratory of High Performance Computing, USTC*; Song Jiang, *University of Texas, Arlington*; John C. S. Lui, *The Chinese University of Hong Kong*

**Elastic Memory Management for Cloud Data Analytics** . . . . . 745  
Jingjing Wang and Magdalena Balazinska, *University of Washington*

## **File Systems**

**Improving File System Performance of Mobile Storage Systems Using a Decoupled Defragmenter** . . . . . 759  
Sangwook Shane Hahn, *Seoul National University*; Sungjin Lee, *Daegu Gyeongbuk Institute of Science and Technology*; Cheng Ji, *City University of Hong Kong*; Li-Pin Chang, *National Chiao-Tung University*; Inhyuk Yee, *Seoul National University*; Liang Shi, *Chongqing University*; Chun Jason Xue, *City University of Hong Kong*; Jihong Kim, *Seoul National University*

**Octopus: an RDMA-enabled Distributed Persistent Memory File System** . . . . . 773  
Youyou Lu, Jiwu Shu, and Youmin Chen, *Tsinghua University*; Tao Li, *University of Florida*

**iJournaling: Fine-Grained Journaling for Improving the Latency of Fsync System Call** . . . . . 787  
Daejun Park and Dongkun Shin, *Sungkyunkwan University, Korea*

**Scaling Distributed File Systems in Resource-Harvesting Datacenters** . . . . . 799  
Pulkit A. Misra, *Duke University*; Íñigo Goiri, Jason Kace, and Ricardo Bianchini, *Microsoft Research*