

USENIX ATC '19: 2019 USENIX Annual Technical Conference

July 10–12, 2019
Renton, WA, USA

Refreshing ATC – USENIX ATC '19 Program Co-Chairs Message xi

Real-World, Deployed Systems

The Design and Operation of CloudLab 1

Dmitry Duplyakin, Robert Ricci, Aleksander Maricq, Gary Wong, Jonathon Duerig, Eric Eide, Leigh Stoller, Mike Hibler, David Johnson, and Kirk Webb, *University of Utah*; Aditya Akella, *University of Wisconsin–Madison*; Kuangching Wang, *Clemson University*; Glenn Ricart, *US Ignite*; Larry Landweber, *University of Wisconsin–Madison*; Chip Elliott, *Raytheon*; Michael Zink and Emmanuel Cecchet, *University of Massachusetts Amherst*; Snigdhaswin Kar and Prabodh Mishra, *Clemson University*

Everyone Loves File: File Storage Service (FSS) in Oracle Cloud Infrastructure 15

Bradley C. Kuszmaul, Matteo Frigo, Justin Mazzola Paluska, and Alexander (Sasha) Sandler, *Oracle Corporation*

Zanzibar: Google’s Consistent, Global Authorization System 33

Ruoming Pang, Ramon Caceres, Mike Burrows, Zhifeng Chen, Pratik Dave, Nathan Germer, Alexander Golynski, Kevin Graney, and Nina Kang, *Google*; Lea Kissner, *Humu, Inc.*; Jeffrey L. Korn, *Google*; Abhishek Parmar, *Carbon, Inc.*; Christina D. Richards and Mengzhi Wang, *Google*

IASO: A Fail-Slow Detection and Mitigation Framework for Distributed Storage Services 47

Biswaranjan Panda and Deepthi Srinivasan, *Nutanix Inc.*; Huan Ke, *University of Chicago*; Karan Gupta and Vinayak Khot, *Nutanix Inc.*; Haryadi S. Gunawi, *University of Chicago*

Runtimes

PARTISAN: Scaling the Distributed Actor Runtime 63

Christopher S. Meiklejohn and Heather Miller, *Carnegie Mellon University*; Peter Alvaro, *UC Santa Cruz*

Unleashing the Power of Learning: An Enhanced Learning-Based Approach for Dynamic Binary Translation 77

Changheng Song, *Fudan University*; Wenwen Wang, Pen-Chung Yew, and Antonia Zhai, *University of Minnesota*; Weihua Zhang, *Fudan University*

Transactuans: Where Transactions Meet the Physical World 91

Aritra Sengupta, Tanakorn Leesatapornwongsa, and Masoud Saecida Ardekani, *Samsung Research*; Cesar A. Stuardo, *University of Chicago*

Not So Fast: Analyzing the Performance of WebAssembly vs. Native Code 107

Abhinav Jangda, Bobby Powers, Emery D. Berger, and Arjun Guha, *University of Massachusetts Amherst*

Filesystems

Extension Framework for File Systems in User space 121

Ashish Bijlani and Umakishore Ramachandran, *Georgia Institute of Technology*

FlexGroup Volumes: A Distributed WAFL File System 135

Ram Kesavan, *Google*; Jason Hennessey, Richard Jernigan, Peter Macko, Keith A. Smith, Daniel Tennant, and Bharadwaj V. R., *NetApp*

EROFS: A Compression-friendly Readonly File System for Resource-scarce Devices 149

Xiang Gao, *Huawei Technologies Co., Ltd.*; Mingkai Dong, *Shanghai Jiao Tong University*; Xie Miao, Wei Du, and Chao Yu, *Huawei Technologies Co., Ltd.*; Haibo Chen, *Shanghai Jiao Tong University / Huawei Technologies Co., Ltd.*

QZFS: QAT Accelerated Compression in File System for Application Agnostic and Cost Efficient Data Storage . . . 163

Xiaokang Hu and Fuzong Wang, *Shanghai Jiao Tong University, Intel Asia-Pacific R&D Ltd.*; Weigang Li, *Intel Asia-Pacific R&D Ltd.*; Jian Li and Haibing Guan, *Shanghai Jiao Tong University*

Big-Data Programming Models & Frameworks

Apache Nemo: A Framework for Building Distributed Dataflow Optimization Policies 177
Youngseok Yang and Jeongyoon Eo, *Seoul National University*; Geon-Woo Kim, *Viva Republica*; Joo Yeon Kim, *Samsung Electronics*; Sanha Lee, *Naver Corp.*; Jangho Seo, Won Wook Song, and Byung-Gon Chun, *Seoul National University*

Tangram: Bridging Immutable and Mutable Abstractions for Distributed Data Analytics 191
Yuzhen Huang, Xiao Yan, Guanxian Jiang, Tatiana Jin, James Cheng, An Xu, Zhanhao Liu, and Shuo Tu, *The Chinese University of Hong Kong*

STRADS-AP: Simplifying Distributed Machine Learning Programming without Introducing a New Programming Model 207
Jin Kyu Kim and Abutalib Aghayev, *Carnegie Mellon University*; Garth A. Gibson, *Carnegie Mellon University, Vector Institute, University of Toronto*; Eric P. Xing, *Petuum Inc, Carnegie Mellon University*

SOPHIA: Online Reconfiguration of Clustered NoSQL Databases for Time-Varying Workloads 223
Ashraf Mahgoub, *Purdue University*; Paul Wood, *Johns Hopkins University*; Alexander Medoff, *Purdue University*; Subrata Mitra, *Adobe Research*; Folker Meyer, *Argonne National Lab*; Somali Chaterji and Saurabh Bagchi, *Purdue University*

Security #1: Kernel

libmpk: Software Abstraction for Intel Memory Protection Keys (Intel MPK) 241
Soyeon Park, *Georgia Institute of Technology*; Sangho Lee, *Microsoft Research*; Wen Xu, *Georgia Institute of Technology*; Hyungon Moon, *Ulsan National Institute of Science and Technology*; Taesoo Kim, *Georgia Institute of Technology*

Effective Static Analysis of Concurrency Use-After-Free Bugs in Linux Device Drivers 255
Jia-Ju Bai, *Tsinghua University*; Julia Lawall, *Sorbonne Université/Inria/LIP6*; Qiu-Liang Chen and Shi-Min Hu, *Tsinghua University*

LXDs: Towards Isolation of Kernel Subsystems 269
Vikram Narayanan, *University of California, Irvine*; Abhiram Balasubramanian, Charlie Jacobsen, Sarah Spall, Scott Bauer, and Michael Quigley, *University of Utah*; Aftab Hussain, Abdullah Younis, Junjie Shen, Moinak Bhattacharyya, and Anton Burtsev, *University of California, Irvine*

JumpSwitches: Restoring the Performance of Indirect Branches In the Era of Spectre 285
Nadav Amit, *VMware Research*; Fred Jacobs, *VMware*; Michael Wei, *VMware Research*

Parallelism & Synchronization

Multi-Queue Fair Queuing 301
Mohammad Hedayati, *University of Rochester*; Kai Shen, *Google*; Michael L. Scott, *University of Rochester*; Mike Marty, *Google*

BRAVO—Biased Locking for Reader-Writer Locks 315
Dave Dice and Alex Kogan, *Oracle Labs*

Mitigating Asymmetric Read and Write Costs in Cuckoo Hashing for Storage Systems 329
Yuanyuan Sun, Yu Hua, Zhangyu Chen, and Yuncheng Guo, *Huazhong University of Science and Technology*

Programmable I/O Devices

NICA: An Infrastructure for Inline Acceleration of Network Applications 345
Haggai Eran, *Technion—Israel Institute of Technology & Mellanox Technologies*; Lior Zeno, Maroun Tork, Gabi Malka, and Mark Silberstein, *Technion—Israel Institute of Technology*

E3: Energy-Efficient Microservices on SmartNIC-Accelerated Servers 363
Ming Liu, *University of Washington*; Simon Peter, *The University of Texas at Austin*; Arvind Krishnamurthy, *University of Washington*; Phitchaya Mangpo Phothilimthana, *University of California, Berkeley*

INSIDER: Designing In-Storage Computing System for Emerging High-Performance Drive 379
Zhenyuan Ruan, Tong He, and Jason Cong, *UCLA*

Cognitive SSD: A Deep Learning Engine for In-Storage Data Retrieval	395
<i>Shengwen Liang and Ying Wang, State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing; University of Chinese Academy of Sciences; Youyou Lu and Zhe Yang, Tsinghua University; Huawei Li and Xiaowei Li, State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing; University of Chinese Academy of Sciences</i>	
Graph Processing Frameworks	
SIMD-X: Programming and Processing of Graph Algorithms on GPUs	411
<i>Hang Liu, University of Massachusetts Lowell; H. Howie Huang, George Washington University</i>	
LUMOS: Dependency-Driven Disk-based Graph Processing	429
<i>Keval Vora, Simon Fraser University</i>	
NeuGraph: Parallel Deep Neural Network Computation on Large Graphs	443
<i>Lingxiao Ma and Zhi Yang, Peking University; Youshan Miao, Jilong Xue, Ming Wu, and Lidong Zhou, Microsoft Research; Yafei Dai, Peking University</i>	
Pre-Select Static Caching and Neighborhood Ordering for BFS-like Algorithms on Disk-based Graph Engines	459
<i>Eunjae Lee, UNIST; Junghyun Kim, TmaxOS; Keunhak Lim, Nexon; Sam H. Noh, UNIST; Jiwon Seo, Hanyang University</i>	
Virtualization Flavors	
From Laptop to Lambda: Outsourcing Everyday Jobs to Thousands of Transient Functional Containers	475
<i>Sadjad Fouladi, Francisco Romero, Dan Iter, and Qian Li, Stanford University; Shuvo Chatterjee, unaffiliated; Christos Kozyrakis, Matei Zaharia, and Keith Winstein, Stanford University</i>	
Hodor: Intra-Process Isolation for High-Throughput Data Plane Libraries	489
<i>Mohammad Hedayati, Spyridoula Gravani, Ethan Johnson, John Criswell, and Michael L. Scott, University of Rochester; Kai Shen and Mike Marty, Google</i>	
A Retargetable System-Level DBT Hypervisor	505
<i>Tom Spink, Harry Wagstaff, and Björn Franke, University of Edinburgh</i>	
MTS: Bringing Multi-Tenancy to Virtual Networking	521
<i>Kashyap Thimmaraju and Saad Hermak, Technische Universität Berlin; Gabor Retvari, BME HSNLab; Stefan Schmid, Faculty of Computer Science, University of Vienna</i>	
Security #2: Isolation	
StreamBox-TZ: Secure Stream Analytics at the Edge with TrustZone	537
<i>Heejin Park and Shuang Zhai, Purdue ECE; Long Lu, Northeastern University; Felix Xiaozhu Lin, Purdue ECE</i>	
CoSMIX: A Compiler-based System for Secure Memory Instrumentation and Execution in Enclaves	555
<i>Meni Orenbach, Technion; Yan Michalevsky, Anjuna Security; Christof Fetzer, TU Dresden; Mark Silberstein, Technion</i>	
Secured Routines: Language-based Construction of Trusted Execution Environments	571
<i>Adrien Ghosn, James R. Larus, and Edouard Bugnion, EPFL</i>	
Supporting Security Sensitive Tenants in a Bare-Metal Cloud	587
<i>Amin Mosayyebzadeh, Boston University; Apoorve Mohan, Northeastern University; Sahil Tikale, Boston University; Mania Abdi, Northeastern University; Nabil Schear, MIT Lincoln Laboratory; Trammell Hudson, Two Sigma; Charles Munson, MIT Lincoln Laboratory; Larry Rudolph, Two Sigma; Gene Cooperman and Peter Desnoyers, Northeastern University; Orran Krieger, Boston University</i>	
Exotic Kernel Features	
Asynchronous I/O Stack: A Low-latency Kernel I/O Stack for Ultra-Low Latency SSDs	603
<i>Gyusun Lee, Seokha Shin, and Wonsuk Song, Sungkyunkwan University; Tae Jun Ham and Jae W. Lee, Seoul National University; Jinkyu Jeong, Sungkyunkwan University</i>	
M³x: Autonomous Accelerators via Context-Enabled Fast-Path Communication	617
<i>Nils Asmussen, Michael Roitzsch, and Hermann Härtig, Technische Universität Dresden, Germany; Barkhausen Institut, Dresden, Germany</i>	

(continued on next page)

Deduplication

SmartDedup: Optimizing Deduplication for Resource-constrained Devices 633
Qirui Yang, Runyu Jin, and Ming Zhao, *Arizona State University*

Data Domain Cloud Tier: Backup here, backup there, deduplicated everywhere! 647
Abhinav Duggal, Fani Jenkins, Philip Shilane, Ramprasad Chinthekindi, Ritesh Shah, and Mahesh Kamat, *Dell EMC*

Exotic Kernel Features #2

GAIA: An OS Page Cache for Heterogeneous Systems 661
Tanya Brokhman, Pavel Lifshits, and Mark Silberstein, *Technion—Israel Institute of Technology*

Transkernel: Bridging Monolithic Kernels to Peripheral Cores 675
Liwei Guo, Shuang Zhai, Yi Qiao, and Felix Xiaozhu Lin, *Purdue ECE*

Detecting Asymmetric Application-layer Denial-of-Service Attacks In-Flight with FINELAME 693
Henri Maxime Demoulin, Isaac Pedisich, Nikos Vasilakis, Vincent Liu, Boon Thau Loo, and Linh Thi Xuan Phan, *University of Pennsylvania*

SemperOS: A Distributed Capability System 709
Matthias Hille, *Technische Universität Dresden*; Nils Asmussen, *Technische Universität Dresden; Barkhausen Institut*;
Pramod Bhatotia, *University of Edinburgh*; Hermann Härtig, *Technische Universität Dresden; Barkhausen Institut*

Key-Value Stores

Pragh: Locality-preserving Graph Traversal with Split Live Migration 723
Xiating Xie, Xingda Wei, Rong Chen, and Haibo Chen, *Shanghai Jiao Tong University*

ElasticBF: Elastic Bloom Filter with Hotness Awareness for Boosting Read Performance in Large Key-Value Stores 739
Yongkun Li, Chengjin Tian, Fan Guo, Cheng Li, and Yinlong Xu, *University of Science and Technology of China*

SILK: Preventing Latency Spikes in Log-Structured Merge Key-Value Stores 753
Oana Balmau, Florin Dinu, and Willy Zwaenepoel, *University of Sydney*; Karan Gupta and Ravishankar Chandhiramoorthi, *Nutanix Inc.*; Diego Didona, *IBM Research—Zurich*

Unification of Temporary Storage in the NodeKernel Architecture 767
Patrick Stuedi, *IBM Research*; Animesh Trivedi, *Vrije Universiteit*; Jonas Pfefferle, *IBM Research*; Ana Klimovic, *Stanford University*; Adrian Schuepbach and Bernard Metzler, *IBM Research*

Solid-State & Hard Disk Drives

Evaluating File System Reliability on Solid State Drives 783
Shehbaz Jaffer, Stathis Maneas, Andy Hwang, and Bianca Schroeder, *University of Toronto*

Alleviating Garbage Collection Interference Through Spatial Separation in All Flash Arrays 799
Jaeho Kim, *Virginia Tech*; Kwanghyun Lim, *Cornell University*; Youngdon Jung and Sungjin Lee, *DGIST*; Changwoo Min, *Virginia Tech*; Sam H. Noh, *UNIST*

Practical Erase Suspension for Modern Low-latency SSDs 813
Shine Kim, *Seoul National University and Samsung Electronics*; Jonghyun Bae, *Seoul National University*;
Hakbeom Jang, *Sungkyunkwan University*; Wenjing Jin and Jeonghun Gong, *Seoul National University*; Seungyeon Lee, *Samsung Electronics*; Tae Jun Ham and Jae W. Lee, *Seoul National University*

Track-based Translation Layers for Interlaced Magnetic Recording 821
Mohammad Hossein Hajkazemi, *Northeastern University*; Ajay Narayan Kulkarni, *Seagate Technology*; Peter Desnoyers, *Northeastern University*; Timothy R Feldman, *Seagate Technology*

Networking

Your Coflow has Many Flows: Sampling them for Fun and Speed833
Akshay Jajoo, Y. Charlie Hu, and Xiaojun Lin, *Purdue University*

PostMan: Rapidly Mitigating Bursty Traffic by Offloading Packet Processing849
Panpan Jin, Jian Guo, and Yikai Xiao, *National Engineering Research Center for Big Data Technology and System, Key Laboratory of Services Computing Technology and System, Ministry of Education, School of Computer Science and Technology, Huazhong University of Science and Technology, China*; Rong Shi, *The Ohio State University, USA*; Yipei Niu and Fangming Liu, *National Engineering Research Center for Big Data Technology and System, Key Laboratory of Services Computing Technology and System, Ministry of Education, School of Computer Science and Technology, Huazhong University of Science and Technology, China*; Chen Qian, *University of California Santa Cruz, USA*; Yang Wang, *The Ohio State University, USA*

R2P2: Making RPCs first-class datacenter citizens863
Marios Kogias, George Prekas, Adrien Ghosn, Jonas Fietz, and Edouard Bugnion, *EPFL*

Lancet: A self-correcting Latency Measuring Tool881
Marios Kogias, *EPFL*; Stephen Mallon, *University of Sydney*; Edouard Bugnion, *EPFL*

Non-Volatile Memory

Pangolin: A Fault-Tolerant Persistent Memory Programming Library897
Lu Zhang and Steven Swanson, *UC San Diego*

Pisces: A Scalable and Efficient Persistent Transactional Memory913
Jinyu Gu, Qianqian Yu, Xiayang Wang, Zhaoguo Wang, Binyu Zang, Haibing Guan, and Haibo Chen, *Shanghai Jiao Tong University*

Scheduling Things

EDGEWISE: A Better Stream Processing Engine for the Edge929
Xinwei Fu, Talha Ghaffar, James C. Davis, and Dongyoon Lee, *Virginia Tech*

Analysis of Large-Scale Multi-Tenant GPU Clusters for DNN Training Workloads947
Myeongjae Jeon, *UNIST and Microsoft Research*; Shivaram Venkataraman, *University of Wisconsin and Microsoft Research*; Amar Phanishayee and Junjie Qian, *Microsoft Research*; Wencong Xiao, *Beihang University and Microsoft Research*; Fan Yang, *Microsoft Research*

Storage Failure & Recovery

Lessons and Actions: What We Learned from 10K SSD-Related Storage System Failures961
Erci Xu, *Ohio State University*; Mai Zheng, *Iowa State University*; Feng Qin, *Ohio State University*; Yikang Xu and Jiesheng Wu, *Alibaba Group*

Who's Afraid of Uncorrectable Bit Errors? Online Recovery of Flash Errors with Distributed Redundancy977
Amy Tai, *Princeton University and VMware Research*; Andrew Kryczka and Shobhit O. Kanaujia, *Facebook*; Kyle Jamieson and Michael J. Freedman, *Princeton University*; Asaf Cidon, *Columbia University*

Dayu: Fast and Low-interference Data Recovery in Very-large Storage Systems993
Zhufan Wang and Guangyan Zhang, *Tsinghua University*; Yang Wang, *The Ohio State University*; Qinglin Yang, *Tsinghua University*; Jiayi Zhu, *Alibaba Cloud*

OPTR: Order-Preserving Translation and Recovery Design for SSDs with a Standard Block Device Interface1009
Yun-Sheng Chang and Ren-Shuo Liu, *National Tsing Hua University*

Machine Learning Applications & System Aspects

Optimizing CNN Model Inference on CPUs1025
Yizhi Liu, Yao Wang, Ruofei Yu, Mu Li, Vin Sharma, and Yida Wang, *Amazon*

Accelerating Rule-matching Systems with Learned Rankers1041
Zhao Lucis Li, *University of Science and Technology China*; Chieh-Jan Mike Liang and Wei Bai, *Microsoft Research*; Qiming Zheng, *Shanghai Jiao Tong University*; Yongqiang Xiong, *Microsoft Research*; Guangzhong Sun, *University of Science and Technology China*

(continued on next page)

MArk: Exploiting Cloud Services for Cost-Effective, SLO-Aware Machine Learning Inference Serving1049
Chengliang Zhang, Minchen Yu, and Wei Wang, *Hong Kong University of Science and Technology*; Feng Yan, *University of Nevada, Reno*

Cross-dataset Time Series Anomaly Detection for Cloud Systems1063
Xu Zhang, *Microsoft Research, Nanjing University*; Qingwei Lin, Yong Xu, and Si Qin, *Microsoft Research*; Hongyu Zhang, *The University of Newcastle*; Bo Qiao, *Microsoft Research*; Yingnong Dang, Xinsheng Yang, Qian Cheng, Murali Chintalapati, Youjiang Wu, and Ken Hsieh, *Microsoft*; Kaixin Sui, Xin Meng, Yaohai Xu, and Wenchi Zhang, *Microsoft Research*; Furao Shen, *Nanjing University*; Dongmei Zhang, *Microsoft Research*