

15th USENIX Symposium on Operating Systems Design and Implementation (OSDI '21)

July 14–16, 2021

Wednesday, July 14

Optimizations and Scheduling for Machine Learning

Pollux: Co-adaptive Cluster Scheduling for Goodput-Optimized Deep Learning1
Aurick Qiao, *Petuum, Inc. and Carnegie Mellon University*; Sang Keun Choe and Suhas Jayaram Subramanya, *Carnegie Mellon University*; Willie Neiswanger, *Petuum, Inc. and Carnegie Mellon University*; Qirong Ho, *Petuum, Inc.*; Hao Zhang, *Petuum, Inc. and UC Berkeley*; Gregory R. Ganger, *Carnegie Mellon University*; Eric P. Xing, *MBZUAI, Petuum, Inc., and Carnegie Mellon University*

Oort: Efficient Federated Learning via Guided Participant Selection19
Fan Lai, Xiangfeng Zhu, Harsha V. Madhyastha, and Mosharaf Chowdhury, *University of Michigan*

PET: Optimizing Tensor Programs with Partially Equivalent Transformations and Automated Corrections37
Haojie Wang, Jidong Zhai, Mingyu Gao, Zixuan Ma, Shizhi Tang, and Liyan Zheng, *Tsinghua University*; Yuanzhi Li, *Carnegie Mellon University*; Kaiyuan Rong and Yuanyong Chen, *Tsinghua University*; Zhihao Jia, *Carnegie Mellon University and Facebook*

Privacy Budget Scheduling55
Tao Luo, Mingen Pan, Pierre Tholoniati, Asaf Cidon, and Roxana Geambasu, *Columbia University*; Mathias Lécuyer, *Microsoft Research*

Storage

Modernizing File System through In-Storage Indexing75
Jinhyung Koo, Junsu Im, Jooyoung Song, and Juhyung Park, *DGIST*; Eunji Lee, *Soongsil University*; Bryan S. Kim, *Syracuse University*; Sungjin Lee, *DGIST*

Nap: A Black-Box Approach to NUMA-Aware Persistent Memory Indexes93
Qing Wang, Youyou Lu, Junru Li, and Jiwu Shu, *Tsinghua University*

Rearchitecting Linux Storage Stack for μ s Latency and High Throughput113
Jaehyun Hwang and Midhul Vuppapalapati, *Cornell University*; Simon Peter, *UT Austin*; Rachit Agarwal, *Cornell University*

Optimizing Storage Performance with Calibrated Interrupts129
Amy Tai, *VMware Research*; Igor Smolyar, *Technion – Israel Institute of Technology*; Michael Wei, *VMware Research*; Dan Tsafir, *Technion – Israel Institute of Technology and VMware Research*

ZNS+: Advanced Zoned Namespace Interface for Supporting In-Storage Zone Compaction147
Kyuha Han, *Sungkyunkwan University and Samsung Electronics*; Hyunho Gwak and Dongkun Shin, *Sungkyunkwan University*; Joo-Young Hwang, *Samsung Electronics*

Data Management

DMon: Efficient Detection and Correction of Data Locality Problems Using Selective Profiling163
Tanvir Ahmed Khan and Ian Neal, *University of Michigan*; Gilles Pokam, *Intel Corporation*; Barzan Mozafari and Baris Kasikci, *University of Michigan*

CLP: Efficient and Scalable Search on Compressed Text Logs183
Kirk Rodrigues, Yu Luo, and Ding Yuan, *University of Toronto and YScope Inc.*

Polyjuice: High-Performance Transactions via Learned Concurrency Control 199
Jiachen Wang, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai AI Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*; Ding Ding, *Department of Computer Science, New York University*; Huan Wang, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai AI Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*; Conrad Christensen, *Department of Computer Science, New York University*; Zhaoguo Wang and Haibo Chen, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai AI Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*; Jinyang Li, *Department of Computer Science, New York University*

Retrofitting High Availability Mechanism to Tame Hybrid Transaction/Analytical Processing219
Sijie Shen, Rong Chen, Haibo Chen, and Binyu Zang, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai Artificial Intelligence Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*

Thursday, July 15

Operating Systems and Hardware

The nanoPU: A Nanosecond Network Stack for Datacenters 239
Stephen Ibanez, Alex Mallery, Serhat Arslan, and Theo Jepsen, *Stanford University*; Muhammad Shahbaz, *Purdue University*; Changhoon Kim and Nick McKeown, *Stanford University*

Beyond malloc efficiency to fleet efficiency: a hugepage-aware memory allocator257
A.H. Hunter, *Jane Street Capital*; Chris Kennelly, Paul Turner, Darryl Gove, Tipp Moseley, and Parthasarathy Ranganathan, *Google*

Scalable Memory Protection in the PENCIL Enclave275
Erhu Feng, Xu Lu, Dong Du, Bicheng Yang, and Xueqiang Jiang, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*; Yubin Xia, Binyu Zang, and Haibo Chen, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai AI Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*

NrOS: Effective Replication and Sharing in an Operating System295
Ankit Bhardwaj and Chinmay Kulkarni, *University of Utah*; Reto Achermann, *University of British Columbia*; Irina Calciu, *VMware Research*; Sanidhya Kashyap, *EPFL*; Ryan Stutsman, *University of Utah*; Amy Tai and Gerd Zellweger, *VMware Research*

Security and Privacy

AddrA: Metadata-private voice communication over fully untrusted infrastructure313
Ishtiyaque Ahmad, Yuntian Yang, Divyakant Agrawal, Amr El Abbadi, and Trinabh Gupta, *University of California, Santa Barbara*

Bringing Decentralized Search to Decentralized Services331
Mingyu Li, Jinhao Zhu, and Tianxu Zhang, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai AI Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*; Cheng Tan, *Northeastern University*; Yubin Xia, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai AI Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*; Sebastian Angel, *University of Pennsylvania*; Haibo Chen, *Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University; Shanghai AI Laboratory; Engineering Research Center for Domain-specific Operating Systems, Ministry of Education, China*

Finding Consensus Bugs in Ethereum via Multi-transaction Differential Fuzzing 349
Youngeok Yang, *Seoul National University*; Taesoo Kim, *Georgia Institute of Technology*; Byung-Gon Chun, *Seoul National University and FriendliAI*

MAGE: Nearly Zero-Cost Virtual Memory for Secure Computation 367
Sam Kumar, David E. Culler, and Raluca Ada Popa, *University of California, Berkeley*

Zeph: Cryptographic Enforcement of End-to-End Data Privacy 387
Lukas Burkhalter, Nicolas Küchler, Alexander Viand, Hossein Shafagh, and Anwar Hithnawi, *ETH Zürich*

Friday, July 16

Correctness

DistAI: Data-Driven Automated Invariant Learning for Distributed Protocols 405
Jianan Yao, Runzhou Tao, Ronghui Gu, Jason Nieh, Suman Jana, and Gabriel Ryan, *Columbia University*

GoJournal: a verified, concurrent, crash-safe journaling system423
Tej Chajed, *MIT CSAIL*; Joseph Tassarotti, *Boston College*; Mark Theng, *MIT CSAIL*; Ralf Jung, *MPI-SWS*; M. Frans Kaashoek and Nikolai Zeldovich, *MIT CSAIL*

STORM: Refinement Types for Secure Web Applications	441
Nico Lehmann and Rose Kunkel, <i>UC San Diego</i> ; Jordan Brown, <i>Independent</i> ; Jean Yang, <i>Akita Software</i> ; Niki Vazou, <i>IMDEA Software Institute</i> ; Nadia Polikarpova, Deian Stefan, and Ranjit Jhala, <i>UC San Diego</i>	
Horcrux: Automatic JavaScript Parallelism for Resource-Efficient Web Computation	461
Shaghayegh Mardani, <i>UCLA</i> ; Ayush Goel, <i>University of Michigan</i> ; Ronny Ko, <i>Harvard University</i> ; Harsha V. Madhyastha, <i>University of Michigan</i> ; Ravi Netravali, <i>Princeton University</i>	
SANRAZOR: Reducing Redundant Sanitizer Checks in C/C++ Programs	479
Jiang Zhang, <i>University of Southern California</i> ; Shuai Wang, <i>HKUST</i> ; Manuel Rigger, Pinjia He, and Zhendong Su, <i>ETH Zurich</i>	
Graph Embeddings and Neural Networks	
Dorylus: Affordable, Scalable, and Accurate GNN Training with Distributed CPU Servers and Serverless Threads	495
John Thorpe, Yifan Qiao, Jonathan Eyolfson, and Shen Teng, <i>UCLA</i> ; Guanzhou Hu, <i>UCLA and University of Wisconsin, Madison</i> ; Zhihao Jia, <i>CMU</i> ; Jinliang Wei, <i>Google Brain</i> ; Keval Vora, <i>Simon Fraser</i> ; Ravi Netravali, <i>Princeton University</i> ; Miryung Kim and Guoqing Harry Xu, <i>UCLA</i>	
GNNAdvisor: An Adaptive and Efficient Runtime System for GNN Acceleration on GPUs	515
Yuke Wang, Boyuan Feng, Gushu Li, Shuangchen Li, Lei Deng, Yuan Xie, and Yufei Ding, <i>University of California, Santa Barbara</i>	
Marius: Learning Massive Graph Embeddings on a Single Machine	533
Jason Mohoney and Roger Waleffe, <i>University of Wisconsin-Madison</i> ; Henry Xu, <i>University of Maryland, College Park</i> ; Theodoros Rekatsinas and Shivaram Venkataraman, <i>University of Wisconsin-Madison</i>	
P³: Distributed Deep Graph Learning at Scale	551
Swapnil Gandhi and Anand Padmanabha Iyer, <i>Microsoft Research</i>	