23rd USENIX Symposium on Networked Systems Design and Implementation (NSDI'26)

Renton, Washington, USA
4-6 May 2026

Volume 1 of 4

ISBN: 979-8-3313-3898-5

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2026) by Usenix Association
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact Usenix Association
at the address below.

Usenix Association
2560 Ninth Street, Suite 215
Berkeley, California, 94710

https://www.usenix.org/

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

23rd USENIX Symposium on Networked Systems Design and Implementation (NSDI '26)

May 4–6, 2026
Renton, WA, USA

Monday, May 4

All Your Networks Are Belong to ML

Unum: A New Framework for Network Control **1**
Jiayi Chen, *UT Austin;* Nihal Sharma, *Capital One;* Debajit Chakraborty, Saurabh Agarwal, Jeffrey Zhou, Aditya Akella, and Sanjay Shakkottai, *UT Austin*

BURST: Seeking High-performance, Interoperability and Scalability in Soft-RDMA **19**
Huijun Shen, *Hunan University;* Zelong Yue, Jian Yang, Zhuo Jiang, Lang An, Yulin Chen, Yong Zhang, Luochangqi Ding, Xiaolong Zhong, Zhihong Wang, Jie Ding, Hongyu Wu, and Jianxi Ye, *ByteDance Inc.;* Xijin Yin, Xingyu Zhang, Xingyu Guo, and Guo Chen, *Hunan University*

PolicyCache: Intra-flow Learning in Congestion Control. **35**
Han Tian, Han Wang, and Wenbo Li, *University of Science and Technology of China;* Xudong Liao, Decang Sun, and Wenxue Li, *Hong Kong University of Science and Technology;* Donghui Chen, Bin Huang, and Senbo Fu, *Huawei Technologies Co., Ltd.;* Junxue Zhang, *University of Science and Technology of China;* Dian Shen, *Southeast University;* Kai Chen, *Hong Kong University of Science and Technology*

FastServe: Iteration-Level Preemptive Scheduling for Large Language Model Inference **57**
Bingyang Wu, Yinmin Zhong, Zili Zhang, Shengyu Liu, Fangyue Liu, Yuanhang Sun, Gang Huang, Xuanzhe Liu, and Xin Jin, *School of Computer Science, Peking University*

Symi: Efficient Mixture-of-Experts Training via Model and Optimizer State Decoupling **75**
Athinagoras Skiadopoulos, *Stanford University;* Mark Zhao, *University of Colorado Boulder;* Swapnil Gandhi, *Stanford University and NVIDIA;* Thomas Norrie, *OpenAI;* Shrijeet Mukherjee, *NVIDIA;* Christos Kozyrakis, *Stanford University and NVIDIA*

The Care and Feeding of Networked Systems

Towards Performance Robustness for Microservices **93**
Divyanshu Saxena and Gaurav Vipat, *UT Austin;* Jiaxin Lin, *Cornell University;* Jingbo Wang, *Purdue University;* Isil Dillig, Sanjay Shakkottai, and Aditya Akella, *UT Austin*

Pilot Execution: Simulating Failure Recovery *In Situ* for Production Distributed Systems **113**
Zhenyu Li, *University of Virginia;* Angting Cai, *University of California San Diego;* Chang Lou, *University of Virginia*

Blade: Adaptive Wi-Fi Contention Control for Next-Generation Real-Time Communication **131**
Fengqian Guo, *Tencent;* Yuhan Zhou, *Peking University and Tencent;* Longwei Jiang and Congcong Miao, *Tencent;* Yuxin Liu, *University at Buffalo, SUNY;* Chenren Xu, *Peking University;* Hancheng Lu, *Institute of Artificial Intelligence, China;* Chang Wen Chen, *The Hong Kong Polytechnic University;* Yaxiong Xie, *University at Buffalo, SUNY;* Honghao Liu, *Tencent*

Observability Is Eating Your Cores: Fine-Grained Analysis of Microservice Metrics with IPU-Hosted Sketches . . **153**
Alessandro Cornacchia, *King Abdullah University of Science and Technology;* Theophilus A. Benson, *Carnegie Mellon University;* Muhammad Bilal and Marco Canini, *King Abdullah University of Science and Technology*

***MirrorNet*: High-fidelity and Scalable Network Emulation for Software-defined WAN** **175**
Congcong Miao, *Tencent;* Yuejie Wang, *Peking University;* Jianming Wang, Xuefeng Ji, Guozhi Shan, Sirui Li, Pan Fang, and Yanke Zhang, *Tencent;* Jialin Li, *National University of Singapore;* Xianneng Zou, *Tencent;* Guyue Liu, *Peking University*

Mobile and Embedded Systems

From Bits to Tokens: Knowledge-Driven Generative Communication of Multimodal Data 191
Xingyu Chen, *University of California San Diego;* Zihao Feng, *University of Southern California;* Wuqiong Zhao, *University of California San Diego;* Jianrong Ding, *Chinese University of Hong Kong;* Ke Sun, *University of Michigan Ann Arbor;* Xinyu Zhang, *University of California San Diego*

KeepON: Supporting Deterministic Traffic on Standard NICs 209
Chuanyu Xue, *University of Connecticut;* Tianyu Zhang, *University of Iowa;* Andrew Loveless, *NASA Johnson Space Center;* Song Han, *University of Connecticut*

RASC: Enhancing Observability & Programmability in Smart Spaces 237
Anna Karanika, Kai-Siang Wang, Han-Ting Liang, Shalni Sundram, and Indranil Gupta, *University of Illinois Urbana-Champaign*

Decoding RSSI Compression in RFID: Dynamic RCS Modeling and Tag-Intrinsic Power Metrics for Reliable Backscatter Networks 253
Jia Liu, Yifei Ma, Xingyu Chen, and Haipeng Dai, *Nanjing University;* He Huang, *Soochow University;* Zihao Lin and Wei Zheng, *Nanjing University;* Junzhao Du, *Xidian University;* Guihai Chen, *Nanjing University*

BBC: Enabling BLE to Support Bluetooth Classic 267
Hsun-Wei Cho and Kang G. Shin, *University of Michigan*

Overfitting the Internet

SpliDT: Partitioned Decision Trees for Scalable Stateful Inference at Line Rate 283
Murayyiam Parvez, *Purdue University;* Annus Zulfiqar, *University of Michigan;* Roman Beltiukov, *UCSB;* Shir Landau Feibish, *University of Haifa;* Walter Willinger, *Northwestern University;* Arpit Gupta, *UCSB;* Muhammad Shahbaz, *University of Michigan*

Morphe: High-Fidelity Generative Video Streaming with Vision Foundation Model 301
Tianyi Gong, Zijian Cao, and Zixing Zhang, *The Chinese University of Hong Kong, Shenzhen, and Shenzhen Future Network of Intelligence Institute;* Jiangkai Wu and Xinggong Zhang, *Peking University;* Shuguang Cui and Fangxin Wang, *The Chinese University of Hong Kong, Shenzhen, and Shenzhen Future Network of Intelligence Institute*

DroidSpeak: KV Cache Sharing Across Fine-tuned Model Variants 319
Yuhan Liu, Yuyang Huang, Jiayi Yao, Shaoting Feng, Zhuohan Gu, Kuntai Du, Hanchen Li, Yihua Cheng, and Junchen Jiang, *University of Chicago;* Shan Lu, Madan Musuvathi, and Esha Choukse, *Microsoft*

Checkmate: Zero Performance Overhead Model Checkpointing via Network Gradient Replication 339
Ankit Bhardwaj, *Tufts University;* Weiyang Wang, Jeremy Carin, and Adam Belay, and Manya Ghobadi, *Massachusetts Institute of Technology*

Cloudy with a Chance of Tenants

HybridMesh: A Hardware-software Hybrid Approach for Accelerating Service Mesh Ingress 359
Myoungsung You, *University of Seoul;* Jaehyun Nam, *Dankook University;* Minjae Seo, *ETRI;* Taejune Park, *Chonnam National University;* Seungwon Shin, *KAIST*

Di-PS: System-Algorithm Co-Design for Asynchronous and Heterogeneous Cross-cluster LLM Training at Scale . . .377
Shengwei Li, *National Key Laboratory of Parallel and Distributed Computing;* Qiaoling Chen, *Shanghai Artificial Intelligence Laboratory and Nanyang Technology University;* Zhiquan Lai, *National Key Laboratory of Parallel and Distributed Computing;* Penglong Jiao, Wenwen Qu, Kun Cai, Jiaxing Li, Peng Sun, and Xingcheng Zhang, *Shanghai Artificial Intelligence Laboratory;* Xiaoge Deng, Dongsheng Li, and Kai Lu, *National Key Laboratory of Parallel and Distributed Computing;* Tianwei Zhang, *Nanyang Technological University*

Stimpack: An Adaptive Rendering Optimization System for Scalable Cloud Gaming 397
Jin Heo, *Dolby Laboratories;* Vic Wang, Ketan Bhardwaj, and Ada Gavrilovska, *Georgia Institute of Technology*

HydraServe: Minimizing Cold Start Latency for Serverless LLM Serving in Public Clouds 415
Chiheng Lou, Sheng Qi, and Chao Jin, *School of Computer Science, Peking University;* Dapeng Nie, Haoran Yang, and Yu Ding, *Alibaba Group;* Xuanzhe Liu and Xin Jin, *School of Computer Science, Peking University*

Storage Systems and Architecture

FalconFS: Distributed File System for Large-Scale Deep Learning Pipeline 431
Jingwei Xu, *Shanghai Jiao Tong University and Huawei Technologies;* Junbin Kang, *Huawei Technologies;* Mingkai Dong, *Shanghai Jiao Tong University;* Mingyu Liu, Lu Zhang, Shaohong Guo, and Ziyan Qiu, *Huawei Technologies;* Mingzhen You and Ziyi Tian, *Shanghai Jiao Tong University;* Anqi Yu, Tianhong Ding, and Xinwei Hu, *Huawei Technologies;* Haibo Chen, *Shanghai Jiao Tong University and Huawei Technologies*

DistVS: Large-scale Vector Search with Compute-Memory Disaggregation 449
Peiqi Yin, *The Chinese University of Hong Kong;* Xiao Yan, *Wuhan University;* Shiyuan Deng, *Huawei Cloud;* Hui Li, Yifan Zhu, and Xiangyu Zhi, *The Chinese University of Hong Kong;* Jingqi Mao, Ran Xu, and Wenliang Zhang, *Huawei Cloud;* James Cheng, *The Chinese University of Hong Kong*

Lemonshark: Asynchronous DAG-BFT With Early Finality 469
Michael Yiqing Hu, Alvin Hong Yao Yan, Yihan Yang, Xiang Liu, and Jialin Li, *National University of Singapore*

CacheCatalyst: Enhancing Web Caching for the Latency-Constrained Internet. 493
Mohammad Hosseini, *Shahid Beheshti University;* Sina Darabi, *Università della Svizzera italiana;* Hannaneh B. Pasandi, *University of California, Berkeley;* Patrick Eugster, *Università della Svizzera italiana;* Mahmood Choopani, *Shahid Beheshti University*

Clouds with Benefits

ZooRoute: Enhancing Cloud-Scale Network Reliability via Candidate Path Provisioning and Overlay Proactive Rerouting 509
Xiaoqing Sun, *Alibaba Cloud;* Xing Li, *Zhejiang University and Alibaba Cloud;* Xionglie Wei, Tian Pan, Ju Zhang, Bowen Yang, Yi Wang, Ye Yang, Yu Qi, Le Yu, Chenhao Jia, Zhanlong Zhang, Xinyu Chen, Xiaobo Xue, Jianyuan Lu, Shize Zhang, Enge Song, and Yang Song, *Alibaba Cloud;* Rong Wen, *Fudan University and Alibaba Cloud;* Biao Lyu, *Alibaba Cloud and Hangzhou Alibaba Cloud Feitian Information Technology and Hangzhou Alibaba Feitian Information Technology;* Yang Xu, *Fudan University;* Shunmin Zhu, *Alibaba Cloud and Hangzhou Feitian Cloud*

Remote TCP Connection Offload and Applications 527
Shuo Li, Steven Chien, Tianyi Gao, and Michio Honda, *University of Edinburgh*

Oscar: *O*(1)-Step Convergence And Readily-deployable Congestion Control 543
Zhaochen Zhang, Feiyang Xue, and Rui Ning, *Nanjing University;* Keqiang He, *Shanghai Jiao Tong University;* Gianni Antichi, *Politecnico di Milano & Queen Mary University of London;* Jiaqi Gao, *unaffiliated;* Zhimeng Yin, *City University of Hong Kong;* Kexin Liu, Rui Li, Zhengqi Cui, Zhehao Lin, Peirui Cao, Guihai Chen, and Chen Tian, *Nanjing University*

Bifrost: Alibaba's Next-Generation VPC Network with High-Performance Multipath Reliable Transport. 571
Zihao Fan, *Shanghai Jiao Tong University and Alibaba Cloud;* Xing Li, *Zhejiang University and Alibaba Cloud;* Ye Yang, *Alibaba Cloud;* Bo Jiang, *Shanghai Jiao Tong University;* Bowen Yang, Yilong Lv, Yuke Hong, Yinian Zhou, Junnan Cai, Jiayue Xu, Yunrui Hu, Zhao Gao, Ke Sun, Yimin Liu, Xiangdong Zhang, Enge Song, Jianyuan Lu, Xiaoqing Sun, Shize Zhang, Haonan Li, Mingxin Li, Changgang Zheng, Yang Song, Jun Liang, and Biao Lyu, *Alibaba Cloud;* Rong Wen, *Fudan University and Alibaba Cloud;* Zhigang Zong and Shunmin Zhu, *Alibaba Cloud*

Attack of the Bubbles: Straggler-Resilient Pipeline Parallelism for Large Model Training 591
Tianyuan Wu, Lunxi Cao, Hanfeng Lu, and Xiaoxiao Jiang, *Hong Kong University of Science and Technology;* Yinghao Yu, Siran Yang, Guodong Yang, Jiamang Wang, Lin Qu, and Liping Zhang, *Alibaba Group;* Wei Wang, *Hong Kong University of Science and Technology*

Managing Complexity at Scale

R-TCP: A Framework to Optimize TCP Performance Over Rate-Limiting Networks. 609
Shengtong Zhu, *The Chinese University of Hong Kong;* Yan Liu and Lingfeng Guo, *Independent Researcher;* Jack Yiu Bun Lee, *The Chinese University of Hong Kong*

Slowpoke: End-to-end Throughput Optimization Modeling for Microservice Applications 631
Yizheng Xie, Di Jin, and Oğuzhan Çölkesen, *Brown University;* Vasiliki Kalavri and John Liagouris, *Boston University;* Nikos Vasilakis, *Brown University*

CrossCheck: Input Validation for WAN Control Systems 647
Alexander Krentsel, *UC Berkeley and Google;* Rishabh Iyer, *UC Berkeley;* Isaac Keslassy, *Technion and UC Berkeley;* Bharath Modhipalli, *Google;* Sylvia Ratnasamy, *UC Berkeley and Google;* Anees Shaikh and Rob Shakir, *Google*

Detecting and Diagnosing Errors in Serving Archived Web Pages 669
Jingyuan Zhu, *University of Michigan;* Huanchen Sun and Harsha V. Madhyastha, *University of Southern California*

Skyline: A Cloud Centric Internet Monitoring Engine 685
Shixian Guo, *ByteDance;* Ziqian Liu, *The University of Hong Kong;* Yangyang Bai, Yuan Chen, Kefei Liu, Qi Zhang, Songlin Liu, Yang Lv, Jianwei Hu, Gen Li, Zhenyang Zhong, Sisi Wen, Yongbin Dong, Feng Luo, Anjian Chen, Rui Han, Jiale Feng, Lingpei Meng, Siwan Chen, Hang Li, Shuai Xu, Juntao Zhong, and Chaoran Hu, *ByteDance;* Yibo Huang, *University of Michigan;* Yiming Qiu, *The University of Hong Kong*

Trust but Verify

Feedback-guided Adaptive Testing of Distributed Systems Designs 701
Ao Li, *Carnegie Mellon University;* Ankush Desai, *Amazon Web Services;* Rohan Padhye, *Carnegie Mellon University*

From Intention to Practice: Towards Systematic Validation of NIDS Rule Enforcement 721
Huan Liu, *Huazhong University of Science and Technology;* Haoyu Chen, *Zhejiang Lab;* Biang Xu, *Huazhong University of Science and Technology and Jinyinhu Laboratory;* Jingyao Zhou, *Huazhong University of Science and Technology;* Bin Yuan, *Huazhong University of Science and Technology and Songshan Laboratory;* Qiankun Zhang, *Huazhong University of Science and Technology;* Deqing Zou, *Huazhong University of Science and Technology and Jinyinhu Laboratory;* Hai Jin, *Huazhong University of Science and Technology*

Count-Based Abstractions for Performance Verification of Contention Points 743
Amir Seyhani, *University of Waterloo;* Aarti Gupta and David Walker, *Princeton University;* Mina Tahmasbi Arashloo, *University of Waterloo*

Iceberg: Automated Verification of DNS Authoritative Engines via Just-in-Time Summarization 761
Yuxing Xiang, Rilin Huang, Naiqian Zheng, and Xin Jin, *Peking University*

EYWA: Automating Model-Based Testing using LLMs 777
Rajdeep Mondal, Rathin Singha, Todd Millstein, and George Varghese, *UCLA;* Ryan Beckett and Siva Kesava Reddy Kakarla, *Microsoft Research*

Tuesday, May 5

Learning at Line Rate

Making Logic a First-Class Citizen in Generative ML for Networking 801
Hongyu Hè, Minhao Jin, and Maria Apostolaki, *Princeton University*

JITServe: SLO-aware LLM Serving with Imprecise Request Information 825
Wei Zhang, Zhiyu Wu, and Yi Mu, *University of Illinois, Urbana-Champaign;* Rui Ning, *unaffiliated;* Banruo Liu, *University of Illinois Urbana-Champaign;* Nikhil Sarda, *Google;* Myungjin Lee, *Cisco Research;* Fan Lai, *University of Illinois Urbana-Champaign*

RollPacker: Taming Long-Tail Rollouts for RL Post-Training with Tail Batching 849
Wei Gao, Yuheng Zhao, Dakai An, Tianyuan Wu, and Lunxi Cao, *Hong Kong University of Science and Technology;* Shaopan Xiong, Ju Huang, Weixun Wang, Siran Yang, Wenbo Su, Jiamang Wang, Lin Qu, and Bo Zheng, *Alibaba Group;* Wei Wang, *Hong Kong University of Science and Technology*

FENIX: Enabling In-Network DNN Inference with FPGA-Enhanced Programmable Switches 867
Xiangyu Gao, *Tsinghua University;* Tong Li, *Renmin University of China;* Yinchao Zhang, *Tsinghua University;* Ziqiang Wang, *Southeast University and Tsinghua University;* Xiangsheng Zeng, *Huazhong University of Science and Technology;* Su Yao, *Tsinghua University and BNRist;* Ke Xu, *Tsinghua University and Zhongguancun Laboratory*

Measuring at Scale

DDoS Detection at the Scale of One Hundred Tbps 883
Yunming Xiao, *The Chinese University of Hong Kong, Shenzhen, and Tencent;* Xijun Luo, Youliang Jiang, Aike Wang, Hu Chen, and Zhibin Zhou, *Tencent;* Heng Yu and Jiahao Cao, *Tsinghua University;* Yong Jiang, *Tsinghua Shenzhen International Graduate School;* Jilong Wang and Mingwei Xu, *Tsinghua University;* Yan Chen, *Northwestern University;* Congcong Miao, *Tencent*

SketchPipe: Toward Accurate Sketch-based Network Measurement on Multi-Pipeline Switches with Splitless Sketch Placement 899
Xiang Chen, Longlong Zhu, and Linying Zheng, *Zhejiang University;* Hongyang Du, *The University of Hong Kong;* Dong Zhang, *Fuzhou University;* Jianshan Zhang, *Minjiang University;* Xuan Liu, *Yangzhou University;* Qun Huang, *Peking University;* Dusit Niyato, *Nanyang Technological University;* Haifeng Zhou and Chunming Wu, *Zhejiang University;* Hongyan Liu, *Fuzhou University;* Kui Ren, *Zhejiang University*

MORP4: A Dynamic Network Telescope 915
Iliana Xygkou, Jithin K. Sojan, Dhruv Rauthan, and Feng Zhu, *Georgia Institute of Technology;* Thomas Holterbach, *Georgia Institute of Technology and University of Strasbourg;* Shane Alcock, *Alcock Network Intelligence Ltd;* Brian Flanagan, Ahmed Saeed, and Alberto Dainotti, *Georgia Institute of Technology*

BayWatch: Practical Internet-Scale Topology Monitoring with Dynamic Bayesian Estimation 933
Zhongxu Guan, *Tsinghua University and Tsinghua Shenzhen International Graduate School;* Shuai Wang, Li Chen, and Zhaoteng Yan, *Zhongguancun Laboratory;* Jiaye Lin and Dan Li, *Tsinghua University;* Yong Jiang, *Tsinghua Shenzhen International Graduate School;* Yingxin Wang and Ziqian Liu, *China Telecom Cybersecurity Technology Co.,Ltd*.

From Silicon to Networked Systems

SyncWise: Error-Aware Time Synchronization for Reconfigurable Data Center Networks 951
Yiming Lei, *Max Planck Institute for Informatics;* Jialong Li, *Shenzhen University of Advanced Technology;* Zhengqing Liu, *Imperial College London;* Raj Joshi, *Harvard University;* Yiting Xia, *Max Planck Institute for Informatics*

Building A CSFQ-Inspired Transport for Switched CXL Memory Pooling 969
Zerui Guo, *University of Wisconsin-Madison;* Emily Shriver, *Intel;* Ming Liu, *University of Wisconsin-Madison*

OpenOptics: Enabling Open Research and Implementation of Optical Data Center Networks 987
Yiming Lei and Federico De Marchi, *Max Planck Institute for Informatics;* Jialong Li, *Shenzhen University of Advanced Technology;* Raj Joshi, *Harvard University;* Shu-Ting Wang, *UC San Diego;* Xiaoqi Chen, *Purdue University;* Balakrishnan Chandrasekaran, *Vrije Universiteit Amsterdam;* Yiting Xia, *Max Planck Institute for Informatics*

PlanB: Efficient Software IPv6 Lookup with Linearized B+-Tree 1005
Zhihao Zhang, *Alibaba Cloud, NICE Lab, XMU, and Tsinghua University;* Lanzheng Liu, Chen Chen, and Huiba Li, *Alibaba Cloud;* Jiwu Shu, *Tsinghua University;* Windsor Hsu, *Alibaba Cloud;* Yiming Zhang, *NICE Lab, SJTU and NICE Lab, XMU*

Smart Pipes, Smarter Learning

FLARE: Anomaly Diagnostics for Divergent LLM Training in GPU Clusters of Thousand-Plus Scale 1021
Weihao Cui, *Shanghai Jiao Tong University and National University of Singapore;* Ji Zhang, *Independent Researcher;* Han Zhao, *Shanghai Jiao Tong University;* Chao Liu, *Independent Researcher;* Jian Sha, *Tsinghua University;* Bo Sang, *Ant Group;* Bingsheng He, *National University of Singapore;* Minyi Guo and Quan Chen, *Shanghai Jiao Tong University*

TurboTest: Learning When Less is Enough through Early Termination of Internet Speed Tests 1037
Haarika Manda, *UC Santa Barbara;* Manshi Sagar, Yogesh, and Kartikay Singh, *IIT Delhi;* Cindy Zhao, *UC Santa Barbara;* Tarun Mangla, *IIT Delhi;* Phillipa Gill, *Google;* Elizabeth Belding and Arpit Gupta, *UC Santa Barbara*

Defeating Slow-and-Low Threats via Diffusion Model-based Generative Inference 1053
Seyed Mohammad Mehdi Mirnajafizadeh and Prashant Khanduri, *Wayne State University;* DaeHun Nyang, *Ewha Womans University;* Rhongho Jang, *Wayne State University*

SwiftEP: Accelerating MoE Inference with Buffer Fusion and TMA Offloading 1073
Xingyi Li, *unaffiliated;* Yadong Liu and Xiaojie Huang, *Tencent;* Yiran Zhang, Shuai Wang, and Shangguang Wang, *unaffiliated;* Zhehao Lin and Yinben Xia, *Tencent;* Chang Yu, *Nanjing University;* Qihang Liu, Xuan Zhang, Hao Lu, Xiang Li, Zekun He, Yachen Wang, and Xianneng Zou, *Tencent*

Geminet: Learning the Duality-based Topology-Agnostic Update Operator for Lightweight Traffic Engineering in Changing Topologies 1091
Ximeng Liu, *Shanghai Jiao Tong University and Zhongguancun Academy;* Zhuoran Liu, *Shanghai Jiao Tong University;* Yingming Mao, *Xi'an Jiaotong University and Shanghai Innovation Institute;* Yatao Li, *Zhongguancun Academy and Zhongguancun Institute of Artificial Intelligence;* Shizhen Zhao and Xinbing Wang, *Shanghai Jiao Tong University*

Debugging Distributed Systems

Eroica: Online Performance Troubleshooting for Large-scale Model Training 1113
Yu Guan, Zhiyu Yin, Haoyu Chen, Sheng Cheng, Chaojie Yang, and Kun Qian, *Alibaba Group;* Tianyin Xu, *University of Illinois Urbana-Champaign;* Pengcheng Zhang, Yang Zhang, Hanyu Zhao, Yong Li, Dennis Cai, and Ennan Zhai, *Alibaba Group*

Supercharging Packet-level Network Simulation of Large Model Training via Memoization and Fast-Forwarding 1131
Fei Long, *Tsinghua University;* Kaihui Gao and Li Chen, *Zhongguancun Laboratory;* Dan Li and Yiwei Zhang, *Tsinghua University;* Fei Gui, *Zhongguancun Laboratory;* Yitao Xing, Wenjia Wei, and Bingyang Liu, *Huawei*

Iris: Expressive Traffic Analysis for the Modern Internet 1153
Thea Rossman, Diana Qing, Gerry Wan, and Zakir Durumeric, *Stanford University*

Who Watches the Watchers? On the Reliability of Softwarizing Cloud Application Management 1171
Jiawei Tyler Gu, Zhen Tang, Yiming Su, Bogdan Alexandru Stoica, Xudong Sun, and William X. Zheng, *University of Illinois Urbana-Champaign;* Yue Zhang and Akond Rahman, *Auburn University;* Chen Wang, *IBM Research;* Tianyin Xu, *University of Illinois Urbana-Champaign*

From Source to Solution: Tackling Packet Losses in Large-scale Cloud Gaming Systematically and Precisely . . . 1191
Jing Wang, Xiao Kong, Yunzhe Ni, Nian Wen, Jiaxing Zhang, Congcong Miao, and Honghao Liu, *Tencent Inc.*

Scalable Storage Systems

PD3: Prefetching Data with DPUs for Disaggregated Memory 1207
Sidharth Sankhe, Felix Zhang, and Umayrah Chonee, *University of Toronto;* Sherman Lim, *National University of Singapore;* Jiasheng Hu, *University of Toronto;* Jialin Li, *National University of Singapore;* Qizhen Zhang, *University of Toronto*

UpFuzz: Detecting Data Format Incompatibility Bugs during Distributed Storage System Upgrade 1225
Ke Han and Sruthi P C, *Purdue University;* Yayu Wang, *The University of British Columbia;* Yaoxu Song and Bishal Basak Papan, *Purdue University;* Junwen Yang, *Meta;* Pedro Fonseca and Yongle Zhang, *Purdue University*

Libra: Flexible Request Partitioning and Scheduling for Serving Unbalanced and Dynamic LLM Workloads . . . 1243
Chaoyi Ruan, *National University of Singapore;* Yinhe Chen, Dongqi Tian, and Yandong Shi, *University of Science and Technology of China;* Yongji Wu, *UC Berkeley;* Jialin Li, *National University of Singapore;* Cheng Li, *University of Science and Technology of China and Institute of Artificial Intelligence, Hefei Comprehensive National Science Center*

Come Hell or Still Water: Alleviating Tail Latency in Cloud Block Store 1259
Chaolei Hu, *Tsinghua University and Alibaba Cloud;* Kun Qian, Erci Xu, Yifan Shen, Haoran Zhang, Xue Li, Yuesheng Gu, and Lingjun Zhu, *Alibaba Cloud;* Fengyuan Ren, *Tsinghua University;* Ennan Zhai, *Alibaba Cloud*

Wallet: Confidential Serverless Computing 1275
Patrick Sabanic, Masanori Misono, Teofil Bodea, Julian Pritzi, Michael Hackl, Dimitrios Stavrakakis, and Pramod Bhatotia, *Technical University of Munich*

Living Large in Shared Clouds

Octopus: Enhancing CXL Memory Pods via Sparse Topology 1303
Yuhong Zhong, *Columbia University;* Fiodar Kazhamiaka, Pantea Zardoshti, Shuwei Teng, and Rodrigo Fonseca, *Microsoft Azure;* Mark D. Hill, *University of Wisconsin–Madison;* Daniel S. Berger, *Microsoft Azure and University of Washington*

RLBoost: Harvesting Preemptible Cloud Resources for Cost-Efficient Reinforcement Learning on LLMs 1323
Yongji Wu, *UC Berkeley;* Xueshen Liu, *University of Michigan;* Haizhong Zheng, *Carnegie Mellon University;* Juncheng Gu, *Google;* Beidi Chen, *Carnegie Mellon University;* Z. Morley Mao, *University of Michigan;* Arvind Krishnamurthy, *Google and University of Washington;* Ion Stoica, *UC Berkeley*

KUBEDIRECT: Unleashing the Full Power of the Cluster Manager for Serverless Computing 1341
Sheng Qi, Zhiquan Zhang, Xuanzhe Liu, and Xin Jin, *Peking University*

FlexLLM: Token-Level Co-Serving of LLM Inference and Finetuning with SLO Guarantees 1359
Gabriele Oliaro, *Carnegie Mellon University;* Xupeng Miao, *Purdue University;* Xinhao Cheng, *Carnegie Mellon University;* Vineeth Kada, *Anthropic PBC;* Mengdi Wu, Ruohan Gao, and Yingyi Huang, *Carnegie Mellon University;* Remi Delacourt, *Mistral AI;* April Yang, *Carnegie Mellon University;* Yingcheng Wang, *Purdue University;* Colin Unger, *Stanford University;* Zhihao Jia, *Carnegie Mellon University and Amazon Web Services*

Reliability Through Verification

Express Lane to Efficiency and Reliability: Multi-Dimensional Control in Meta's Express Backbone Network . . 1381
Faisal Iqbal, Vitaly Neganov, Brian Chang, Rong Rong, Yuanjun Yao, Marek Denis, Qin Zhang, Alexandru Manea, Anton Marchenko, and Ulas Kozat, *Meta;* Aditya Akella, *The University of Texas at Austin;* Ying Zhang, *Meta*

Sparse Checkpointing for Fast and Reliable MoE Training 1395
Swapnil Gandhi, *Stanford University;* Christos Kozyrakis, *Stanford University and NVIDIA*

MAE: More Adaptive Video Encoder for Consistent Low Latency in High-Quality Real-Time Communication. . .1417
Hua Meng, Yufan Zhuang, Yasna Noushirvani, Xiangjie Huang, and Zili Meng, *Hong Kong University of Science and Technology*

In Link We Trust: BFT at the Speed of CFT using Switches. 1431
Lior Zeno, Naama Ben-David, and Mark Silberstein, *Technion – Israel Institute of Technology*

Virtualization Meets the Network

Controlling Arbitrary Internet Queues with Titrate 1453
Anchengcheng Zhou and Joshua Lau, *Princeton University;* P. Brighten Godfrey, *University of Illinois Urbana–Champaign and Broadcom;* Maria Apostolaki, *Princeton University*

Managing Congestion Control Heterogeneity on the Internet with *Approximate Performance Isolation*. 1469
Ayush Mishra, *ETH Zurich;* Archit Bhatnagar, *University of Michigan;* Yixuan Zhang, *Tsinghua University;* Ben Leong, *National University of Singapore;* Ya Gao, *Wuxi Institute of Technology;* Raj Joshi, *Harvard University*

The GOODPUT System: A Machine Learning-Driven Optimization Framework for Dynamic Spectrum Control in Heterogeneous WLANs 1485
Yu Wang and Robert Dick, *University of Michigan*

CStar Gateway: Augmenting Public Cloud Infrastructure for Heterogeneous Network Function Virtualization . . . 1501
Haonan Li, Tian Pan, Jin Ke, Baohai Hu, Changgang Zheng, Enge Song, Zhi Xu, Ye Yang, Bowen Yang, Donglin Lai, Yisong Qiao, Bengbeng Xue, Jianyuan Lu, Xiaoqing Sun, Shize Zhang, Zihao Fan, Mingxin Li, Yang Song, Jun Liang, Xionglie Wei, and Biao Lyu, *Alibaba Cloud;* Rong Wen, *Fudan University and Alibaba Cloud;* Zhigang Zong, *Alibaba Cloud;* Jiao Zhang and Tao Huang, *Purple Mountain Laboratories;* Shunmin Zhu, *Alibaba Cloud*

Clouds, Carefully Engineered

DistRS: Disaggregated Reward Service for RLVR with Batch-Level Constraint 1517
Ruidong Zhu, *School of Computer Science, Peking University;* Mingcong Han, *ByteDance Seed;* Yinmin Zhong, *School of Computer Science, Peking University;* Wencong Xiao, *ByteDance Seed;* Xuanzhe Liu and Xin Jin, *School of Computer Science, Peking University*

MuxTune: Efficient Multi-Task LLM Fine-Tuning in Multi-Tenant Datacenters via Spatial-Temporal Backbone Multiplexing. 1533
Chunyu Xue and Yi Pan, *Shanghai Jiao Tong University;* Weihao Cui, *Shanghai Jiao Tong University and National University of Singapore;* Quan Chen and Shulai Zhang, *Shanghai Jiao Tong University;* Bingsheng He, *National University of Singapore;* Minyi Guo, *Shanghai Jiao Tong University*

CCC: Re-architecting Delay-based Congestion Control in Datacenter Networks 1553
Wanchun Jiang, Haoyang Li, Kai Wang, Yujie Hu, and Xiao Han, *Central South University;* Danfeng Shan, *Xi'an Jiaotong University;* Fengyuan Ren, *Tsinghua University;* Jiawei Huang and Jianxin Wang, *Central South University*

Starfish: A Topology-Routing Co-Design for Small-Scale Data Centers 1571
Anchengcheng Zhou, *Princeton University;* Vipul Harsh, *Conviva;* Sangeetha Abdu Jyothi, *UC Irvine;* Maria Apostolaki, *Princeton University;* Brighten Godfrey, *UIUC*

HEDGE: Traffic Engineering with Probabilistic Link Capacities 1591
Arjun Devraj, *Cornell University;* Bill Owens, *NYSERNet;* Umesh Krishnaswamy, *Microsoft;* Ying Zhang, *Meta;* Rachee Singh, *Cornell University*

When Systems Run Themselves

CascadeNet: Generating Network Traffic with High-Fidelity Temporal Patterns 1613
Runwei Lu, Yanran Deng, Ruixuan Li, and Jinting Liu, *New York University Shanghai;* Yuejie Wang, *Peking University;* Xinyu Li, *Carnegie Mellon University;* Deming Xu, *New York University Shanghai;* Han Tian, *University of Science and Technology of China;* Kai Chen, *Hong Kong University of Science and Technology;* Guyue Liu, *Peking University*

Diagnosing and Repairing Distributed Routing Configurations Using Selective Symbolic Simulation 1635
Rulan Yang, Gao Han, Hanyang Shao, Xiaoqiang Zheng, Xing Fang, Ziyi Wang, and Lizhao You, *Xiamen University;* Ruiting Zhou, *Southeast University;* Linghe Kong, *Shanghai Jiao Tong University;* Ennan Zhai, *Alibaba Cloud;* Qiao Xiang and Jiwu Shu, *Xiamen University*

A Composable Emulation Framework for Whitebox Switches 1653
Congcong Miao, *Tencent;* Xianneng Zou, *Tencent and Tsinghua Shenzhen International Graduate School;* Chuwen Zhang, *Tsinghua University;* Shiping Yang, Qihang Liu, Zhijie Yan, and Yanke Zhang, *Tencent;* Yong Jiang, *Tsinghua Shenzhen International Graduate School;* Qiao Xiang, *Xiamen University;* Xin Jin, *Peking University;* Zili Meng, *Hong Kong University of Science and Technology;* Ang Chen, *University of Michigan*

Harp: Improving VPC Network Availability via Efficient Failure Detection and Rerouting in Tencent Cloud 1669
Jiayu Hu, Feng Jin, and Xianping Zhou, *Tencent;* Kai Zhang, *Fudan University;* Zhen Shen and Yongkang Luo, *Tencent*

Themis: Detecting Distributed Concurrency Bugs through RPC-Driven Race-Directed Test Generation and Fuzzing 1685
Hongchen Cao and Jingzhu He, *ShanghaiTech University;* Ting Dai, *InsightFinder AI;* Guoliang Jin, *North Carolina State University*

From Fiber to Radio

Bridging Storage and Execution: A Semantic Virtual Bus for On-Demand Application Streaming. 1707
Jun Lu, *Central South University;* Jialin Li, *National University of Singapore;* Yaoxue Zhang and Ju Ren, *Tsinghua University*

QCON: Seamless QoE-Aware 5G Streaming via Multi-Connectivity 1721
Goodsol Lee, *Seoul National University;* Junhong Min, *University of Colorado Boulder;* Seyeon Kim, *Korea University;* Juheon Yi, *Seoul National University;* Kwang Taik Kim and Mung Chiang, *Purdue University;* Sangtae Ha, *University of Colorado Boulder;* Kyunghan Lee and Saewoong Bahk, *Seoul National University*

Law: Towards Consistent Low Latency in 802.11 Home Networks. 1741
Yibin Shen and Zili Meng, *Hong Kong University of Science and Technology*

Enabling SLO-Aware 5G Multi-Access Edge Computing with SMEC 1759
Xiao Zhang and Daehyeok Kim, *University of Texas at Austin*

Learning to Tune Optical WANs: A Field Deployment of Noise Models in Optical Networks. 1777
Bhaskar Kataria and Howard Hua, *Cornell University;* Andrea D Amico, *NEC Labs;* Bill Owens, *NYSERNet;* Rachee Singh, *Cornell University*

Wednesday, May 6

When ML Meets Its Network

Enabling AI Network Cross-Layer Design and Operations with Arcadia: A Simulation Platform at Scale 1791
Zhaodong Wang, Satyajeet Singh Ahuja, Xu Zhang, Yuhui Zhang, Max Noormohammadpour, Gregory R. Steinbrecher, Thomas Fuller, Xin Liu, Kevin Quirk, Mikel Jimenez Fernandez, Abhinav Triguna, Yan Cai, and Steve Politis, *Meta Platforms;* Petr Lapukhov and Naader Hasani, *Nvidia;* Ying Zhang, *Meta Platforms*

Phantora: Maximizing Code Reuse in Simulation-based Machine Learning System Performance Estimation . . . 1809
Jianxing Qin, *Duke University;* Jingrong Chen, *Uber;* Xinhao Kong, *NVIDIA;* Yongji Wu, *University of California, Berkeley;* Tianjun Yuan, *Duke University;* Liang Luo, Zhaodong Wang, and Ying Zhang, *Meta;* Tingjun Chen, Alvin R. Lebeck, and Danyang Zhuo, *Duke University*

PrvTel: Lightweight Models for Private and Accurate Telemetry Data Retention. 1827
Yajie Zhou, *University of Maryland;* Fuheng Zhao, *University of Utah;* Eric Wang, *University of Maryland;* Ayse K. Coskun, *Boston University;* Divyakant Agrawal and Amr El Abbadi, *University of California, Santa Barbara;* Zaoxing Liu, *University of Maryland*

ServeGen: Workload Characterization and Generation of Large Language Model Serving in Production 1845
Yuxing Xiang, *Peking University and Alibaba Group;* Xue Li and Kun Qian, *Alibaba Group;* Yan Zhang, *Peking University;* Wenyuan Yu and Ennan Zhai, *Alibaba Group;* Xin Jin, *Peking University;* Jingren Zhou, *Alibaba Group*

Networked Resources, Virtualized

REAL: Emulating Control Plane at Simulator's Cost 1861
Ze Xia, Hao Li, Jinyu Fu, Xin Wan, Yihan Dang, and Danfeng Shan, *Xi'an Jiaotong University;* Li Chen, *harnets.ai;* Peng Zhang, *Xi'an Jiaotong University*

ZOC: Elastic and Cost-Efficient Virtual SmartNIC Architecture for Cloud Physical Machines 1879
Naixuan Guan, Xiaokang Hu, Yisheng Xie, Xishi Qiu, Chaojie Liu, Yuchao Cao, Banghao Ying, Dianchen Tian, Yu Zhou, Yangzeyu Zhang, Hujun Ge, Yibin Shen, and Jiesheng Wu, *Alibaba Cloud Computing*

SLATE: Service Layer Traffic Engineering 1895
Gangmuk Lim, *University of Illinois Urbana–Champaign;* Aditya Prerepa, *University of Illinois Urbana–Champaign and xAI;* Brighten Godfrey and Radhika Mittal, *University of Illinois Urbana–Champaign*

A Fast Solver-Free Algorithm for Traffic Engineering in Large-Scale Data Center Network 1919
Yingming Mao, *Xi'an Jiaotong University and Shanghai Innovation Institute;* Qiaozhu Zhai, *Xi'an Jiaotong University;* Ximeng Liu, *Shanghai Jiao Tong University;* Zhen Yao and Xia Zhu, *Huawei;* Yuzhou Zhou, *Xi'an Jiaotong University*

Hardware, Mobility, and Systems

Ava: Towards Agentic Video Analytics with Vision Language Models. 1939
Yuxuan Yan, *Zhejiang University;* Shiqi Jiang, *Microsoft Research;* Ting Cao, *Tsinghua University;* Yifan Yang, *Microsoft Research;* Qianqian Yang and Yuanchao Shu, *Zhejiang University;* Yuqing Yang and Lili Qiu, *Microsoft Research*

Remembrall: Leaning into Memory for Accurate Video Analytics on System-on-Chip GPUs 1959
Murali Ramanujam, Yinwei Dai, Kyle Jamieson, and Ravi Netravali, *Princeton University*

Themis: Scheduling-Aware Buffer Management for HBM-Based Hybrid Buffers 1975
Zhiyu Zhang, Minkun Xue, Kan Yu, Ruyi Yao, Shili Chen, and Hao Mei, *Fudan University;* Zixuan Chen, *China Telecom;* Weiyi Chen, Yibo Fan, and Yang Xu, *Fudan University*

SmartNIC-enabled Live Migration for Storage-Optimized VMs with Pyrocumulus 1995
Jiechen Zhao, *University of Toronto and Microsoft Research Asia;* Ran Shu, Lei Qu, Ziyue Yang, and Rui Ma, *Microsoft Research Asia;* Derek Chiou, *Microsoft and UT Austin;* Natalie Enright Jerger, *University of Toronto;* Peng Cheng and Yongqiang Xiong, *Microsoft Research Asia*

Neurons Meet Routers

Uber's Failover Architecture: Reconciling Reliability and Efficiency in Hyperscale Microservice Infrastructure . . . 2011
Mayank Bansal, Milind Chabbi, Kenneth Bøgh, Srikanth Prodduturi, Kevin Xu, Amit Kumar, David Bell, Ranjib Dey, Yufei Ren, Sachin Sharma, Juan Marcano, Shriniket Kale, and Subhav Pradhan, *Uber Technologies;* Ivan Beschastnikh, *University of British Columbia;* Miguel Covarrubias, Chien-Chih Liao, Sandeep Koushik Sheshadri, Wen Luo, Kai Song, Ashish Samant, Sahil Rihan, Nimish Sheth, Albert Greenberg, and Uday Kiran Medisetty, *Uber Technologies*

SYMPHONY: Enabling Compute-Memory Disaggregation in LLM Serving Systems 2027
Saurabh Agarwal and Bodun Hu, *UT-Austin;* Anyong Mao, *UW-Madison;* Aditya Akella, *UT-Austin;* Shivaram Venkataraman, *UW-Madison*

Defending against Traffic Analysis Attacks with Flexible In-Network Obfuscation. 2043
Guorui Xie and Qing Li, *Pengcheng Laboratory;* Zhenning Shi, *Tsinghua Shenzhen International Graduate School;* Gianni Antichi, *Politecnico di Milano;* Yijia Zhu, *Xidian University;* Kejun Li and Changxing Weng, *Pengcheng Laboratory;* Sebastiano Miano, *Politecnico di Milano;* Yong Jiang, *Tsinghua Shenzhen International Graduate School and Pengcheng Laboratory;* Mingwei Xu, *Tsinghua University*

ForestColl: Throughput-Optimal Collective Communications on Heterogeneous Network Fabrics. 2065
Liangyu Zhao, *University of Washington;* Saeed Maleki, *Independent Researcher;* Yuanhong Wang, *Tsinghua University;* Zezhou Wang, *University of Washington;* Ziyue Yang, *Microsoft Research;* Hossein Pourreza, *Microsoft;* Arvind Krishnamurthy, *University of Washington*

Matryoshka: Realizing Hyperscale Data Center Network Design for the AI Era 2095
Yan Cai, *Meta;* Jialong Li, *Max Planck Institute for Informatics;* Kutalmis Akpinar, Tianxiang Li, Hany Morsy, Jason Wilson, and Sunil Khaunte, *Meta;* Yiting Xia, *Max Planck Institute for Informatics;* Ying Zhang, *Meta*

Self-Managing Networks

PlanetServe: A Decentralized, Scalable, and Privacy-Preserving Overlay for Democratizing Large Language Model Serving 2111
Fei Fang, Yifan Hua, Shengze Wang, Ruilin Zhou, Yi Liu, and Chen Qian, *University of California, Santa Cruz;* Xiaoxue Zhang, *University of Nevada, Reno*

MoCE: A Mixture-of-Context Aware Experts Framework for Troubleshooting Internet-scale Services 2131
Vipul Harsh, *Conviva and Carnegie Mellon University;* Sayan Sinha, *Conviva and Georgia Tech;* Henry Milner, *Conviva;* B. Aditya Prakash, *Conviva and Georgia Tech;* Vyas Sekar and Hui Zhang, *Conviva and Carnegie Mellon University*

Heuristic Analysis from Source Code via Symbolic-Guided Optimization 2151
Pantea Karimi, *MIT;* Siva Kesava Reddy Kakarla and Ryan Beckett, *Microsoft Research;* Santiago Segarra, *Rice University;* Pooria Namyar, *Microsoft Research;* Mohammad Alizadeh, *MIT;* Behnaz Arzani, *Microsoft Research*

CCEval: Accurately and Confidently Evaluating Performance Metrics of Congestion Control Algorithms for Datacenter Networks 2175
Tianfeng Liu, Kaihui Gao, and Li Chen, *Zhongguancun Laboratory;* Dan Li, *Tsinghua University;* Jin Guang and Xinyun Chen, *The Chinese University of Hong Kong, Shenzhen;* Vincent Liu, *University of Pennsylvania;* Zhiyong Chen and Yiwei Zhang, *Tsinghua University;* Ni Jin, *Zhongguancun Laboratory and Beijing University of Posts and Telecommunications;* Ran Zhang, *Zhongguancun Laboratory*

Predict, Prune, Play: Efficient Video Playback Optimization Under Device Diversity and Drift 2191
Harsha Sharma, *Massachusetts Institute of Technology and Amazon;* Pouya Hamadanian and Arash Nasr-Esfahany, *Massachusetts Institute of Technology;* Zahaib Akhtar, *Amazon and North Carolina State University;* Mohammad Alizadeh, *Massachusetts Institute of Technology*

Security in Networked Systems

Over-Threshold Multiparty Private Set Intersection for Collaborative Network Intrusion Detection 2211
Onur Eren Arpaci, Raouf Boutaba, and Florian Kerschbaum, *University of Waterloo*

Secure Vickrey Auctions for Online Advertising. 2227
Archit Bhatnagar, *University of Michigan;* Yunming Xiao, *The Chinese University of Hong Kong, Shenzhen;* Ang Chen and Amrita Roy Chowdhury, *University of Michigan*

EZ-SAVE: Evaluation of Easy-to-Deploy Source Address Validation Policies. 2247
Nicholas Scaglione and Justin Furuness, *University of Connecticut;* Yossi Gilad, *Hebrew University of Jerusalem;* Hemi Leibowitz, *The College of Management Academic Studies;* Cameron Morris and Bing Wang, *University of Connecticut;* Kotikalapudi Sriram, *National Institute of Standards and Technology* (*NIST*)*;* Amir Herzberg, *University of Connecticut*

cc-pipe: Breaking Systemic Bottlenecks in RPKI Data Supply Chain with Concurrent and Conflict-Free Pipelines 2267
Chenhui Yu, *Computer Network Information Center, Chinese Academy of Sciences; School of Computer Science and Technology, University of Chinese Academy of Sciences;* Yanbiao Li, *Computer Network Information Center, Chinese Academy of Sciences; School of Computer Science and Technology, University of Chinese Academy of Sciences; Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences;* Hui Zou, Yuxuan Chen, Shiyi Liu, and Gaogang Xie, *Computer Network Information Center, Chinese Academy of Sciences; School of Computer Science and Technology, University of Chinese Academy of Sciences*

A Systematic Threat Analysis and Practical Attacks on Automated Frequency Coordination Systems. 2285
Yilu Dong and Tianchang Yang, *The Pennsylvania State University;* Arupjyoti Bhuyan, *Idaho National Laboratory;* Syed Rafiul Hussain, *The Pennsylvania State University*

Designing Robust Networked Systems

Mortise: Auto-tuning Congestion Control to Optimize QoE via Network-Aware Parameter Optimization 2303
Yixin Shen, *Tsinghua University, Bytedance Inc., and Zhongguancun Laboratory;* Ruihua Chen, *Tsinghua University;* Bo Wang, *Tsinghua University and Zhongguancun Laboratory;* Jing Chen, Haochen Zhang, and Minhu Wang, *Tsinghua University;* Yan Liu, *Bytedance Inc.;* Mingwei Xu, *Tsinghua University and Zhongguancun Laboratory;* Zili Meng, *Hong Kong University of Science and Technology*

Harvesting Spare CPU Resources in Container Systems 2323
Adam Hall and Anirudh Sarma, *Georgia Institute of Technology;* Esha Choukse, *Microsoft Azure Research;* Umakishore Ramachandran, *Georgia Institute of Technology;* Sameh Elnikety, *Microsoft Research*

Fractal: Fault-Tolerant Shell-Script Distribution 2339
Zhicheng Huang, Ramiz Dundar, and Yizheng Xie, *Brown University;* Konstantinos Kallas, *University of California, Los Angeles;* Nikos Vasilakis, *Brown University*

eXpressSFU: Toward Super-Scalable Video Conferencing with SmartNICs. 2355
Tuan Tran and S. M. H. Hosseini, *University of Colorado Boulder;* Seyeon Kim, *Korea University;* Kyunghan Lee, *Seoul National University;* Nam Bui, *University of Colorado Denver;* Dirk Grunwald and Sangtae Ha, *University of Colorado Boulder*

Hot Data, Cold Data

ZipLLM: Efficient LLM Storage via Model-Aware Synergistic Data Deduplication and Compression. 2371
Zirui Wang, Tingfeng Lan, and Zhaoyuan Su, *University of Virginia;* Juncheng Yang, *Harvard University;* Yue Cheng, *University of Virginia*

Latency-Aware Caching with Delayed Hits: From Bursty Traffic to Pipeline Architectures 2389
Nadav Keren, Gil Einziger, and Gabriel Scalosub, *Ben Gurion University of The Negev*

Cortex: Achieving Low-Latency, Cost-Efficient Remote Data Access For LLM via Semantic-Aware Knowledge Caching 2407
Chaoyi Ruan, *National University of Singapore;* Chao Bi, *University of Science and Technology of China;* Kaiwen Zheng, *University of Toronto;* Ziji Shi, *National University of Singapore;* Xinyi Wan, *Sea AI Lab and National University of Singapore;* Jialin Li, *National University of Singapore*

Unleashing The Potential of Datacenter SSDs by Taming Performance Variability. 2423
Gohar Irfan Chaudhry, *MIT CSAIL;* Ankit Bhardwaj, *Tufts University;* Zhenyuan (Zain) Ruan and Adam Belay, *MIT CSAIL*

Virtualization Revisited

Agentix: An Efficient Serving Engine for LLM Agents as General Programs 2443
Michael Luo, *University of California, Berkeley, and Google DeepMind;* Xiaoxiang Shi, *Shanghai Jiao Tong University;* Colin Cai, Tianjun Zhang, Justin Wong, and Yichuan Wang, *University of California, Berkeley;* Chi Wang, Yanping Huang, and Zhifeng Chen, *Google DeepMind;* Joseph E. Gonzalez and Ion Stoica, *University of California, Berkeley*

QUEUE-MEM: Energy-Efficient Hardware Storage for Advanced Network Function Acceleration 2461
Mariano Scazzariello, *RISE Research Institutes of Sweden;* Tommaso Caiazzi, *Roma Tre University;* Hamid Ghasemirahni, Dejan Kostić, and Marco Chiesa, *KTH Royal Institute of Technology*

Hierarchical Integration of WebAssembly in Serverless for Efficiency and Interoperability 2481
Mohammadamin Baqershahi, Changyuan Lin, and Visal Saosuo, *University of British Columbia;* Paul Chen, *Huawei Technologies Canada;* Mohammad Shahrad, *University of British Columbia*

Mitigating CPU Frontend for Complex Data Plane Applications 2499
Yihan Dang, Hao Li, Ze Xia, Jiajun Luan, and Peng Zhang, *Xi'an Jiaotong University*

Cloud Systems, Shared and Scalable

FAST: An Efficient Scheduler for All-to-All GPU Communication 2515
Yiran Lei, *Carnegie Mellon University and MangoBoost;* Dongjoo Lee, *MangoBoost;* Liangyu Zhao, *University of Washington;* Daniar Kurniawan, Chanmyeong Kim, Heetaek Jeong, Changsu Kim, and Hyeonseong Choi, *MangoBoost;* Liangcheng Yu, *University of Pennsylvania;* Arvind Krishnamurthy, *University of Washington;* Justine Sherry, *Carnegie Mellon University;* Eriko Nurvitadhi, *MangoBoost*

HeteCCL: Synthesizing Near-Optimal Collective Communication Schedules for Heterogeneous GPU Clusters . . 2533
Chenyang Hei, Fuliang Li, and Jiayi Li, *Northeastern University;* Jiamin Cao, *Alibaba Cloud;* Chengxi Gao, *Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences;* Xiuzhu Sha, Tongrui Liu, and Dengke Zhang, *Northeastern University;* Ennan Zhai, *Alibaba Cloud;* Xingwei Wang, *Northeastern University*

Medley: Optimizing Midgress Bandwidth for Commercial Live Streaming CDNs 2553
Haiping Wang, *Fudan University and ByteDance;* Wanxin Shi, *Fudan University;* Sandesh Dhawaskar Sathyanarayana, Shu Shi, and Feng Qian, *ByteDance;* Yinghao Yang and La Zuo, *Fudan University and ByteDance;* Hebin Yu and Ruixiao Zhang, *ByteDance;* Ruoshi Sun, *Fudan University;* Yajie Peng, Xiaofei Pang, Ruili Fang, Zhenpeng Zhu, and Weiqi Chen, *ByteDance;* Yang Xu, *Fudan University*

Offloading Cloud Network Services at Production Scale with SONiC DASH SmartSwitch 2569
Shaofeng Wu, *The Chinese University of Hong Kong and Microsoft Research Asia;* Zhixiong Niu, *Microsoft Research Asia;* Riff Jiang, Lawrence Lee, Junhua Zhai, Ze Gan, Vasundhara Volam, Prabhat Aravind, Prince Sunny, Prince George, Qi Luo, Evan Langlais, Soumya Tiwari, Venkat Satish Katta, Weixi Chen, Rishiraj Hazarika, Sachin Jain, Deven Jagasia, Michal Zygmunt, Avijit Gupta, Neeraj Motwani, and Pranjal Shrivastava, *Microsoft;* Qiang Su, *The Chinese University of Hong Kong;* Anil Reddy Pannala, Kristina Moore, James Grantham, Anupam Pandey, Xin Liu, Guohan Lu, Gerald De Grace, Rishabh Tewari, Lihua Yuan, Erica Lan, Deepak Bansal, and Dave Maltz, *Microsoft;* Yongqiang Xiong, *Microsoft Research Asia;* Hong Xu, *The Chinese University of Hong Kong*

AnyPro: Preference-Preserving Anycast Optimization based on Strategic AS-Path Prepending 2587
Minyuan Zhou, *Nanjing University and Alibaba Cloud;* Yuning Chen, *University of California, Merced, and Alibaba Cloud;* Jiaqi Zheng, *Nanjing University;* Yifei Xu, *University of California, Los Angeles, and Alibaba Cloud;* Pan Hu, Yongping Tang, Wendong Yin, Jie Lin, Qingyan Yu, and Yuanchao Su, *Alibaba Group;* Guihai Chen and Wanchun Dou, *Nanjing University;* Songwu Lu, *University of California, Los Angeles;* Wan Du, *University of California, Merced*

Distributed Data Systems

HCDN: Coordinated Stream Scheduling for Cost-Effective Live Video Delivery 2603
Liying Wang, *Peking University;* Jing Liu, *ByteDance;* Yuhan Zhou and Chengke Wang, *Peking University;* Mingming Lu, Qingyue Li, Song Geng, Linsen Wang, Kaida Hu, Haoyuan Huang, Shimao Tian, Ri Lu, and Mingfei Hao, *ByteDance;* Chenren Xu, *Peking University; and Key Laboratory of High Confidence Software Technologies, Ministry of Education (PKU);* Shu Shi, *ByteDance*

XLL: Cross-Layer Logging for Data Deduplication in Consensus-Based Storage 2619
John Shawger, Arnav Jhingran, Andrea Arpaci-Dusseau, and Remzi Arpaci-Dusseau, *University of Wisconsin – Madison*

HyperEdge: An Edge CDN Infrastructure for Cost Efficient Video Streaming 2635
Dehui Wei, *National University of Singapore;* Jiao Zhang, *Beijing University of Posts and Telecommunications, and Purple Mountain Laboratories;* Haozhe Li, Rui Han, Zhichen Xue, Yajie Peng, Xiaofei Pang, and Yan Ma, *ByteDance;* Jialin Li, *National University of Singapore*

Co-Designing Traffic Control with NVMe-oF for Disaggregated Storage: A Comparative Study of Switched and Switchless SAN Architectures 2651
Chendong Wang, Joontaek Oh, and Ming Liu, *University of Wisconsin–Madison*

Cost-effective and Reliable Global Internet Peering with Programmable Switches 2671
Congcong Miao, *Tencent and National University of Singapore;* Zhiyi Yao, Jianchao Lv, and Jinglin Wang, *Tencent;* Shihan Lin, *University of Michigan;* Xinyi Zhang, *CNIC CAS, China;* Yunming Xiao, *The Chinese University of Hong Kong, Shenzhen;* Wei Guo, Jiwu Bu, Yachen Wang, and Xianneng Zou, *Tencent;* Yong Jiang, *Tsinghua Shenzhen International Graduate School;* Marco Canini, *KAUST;* Gaogang Xie, *CNIC CAS, China*

Sandboxes, Containers, and Beyond

KrakenGuard: Towards Fine-Grained eBPF Isolation 2685
Jainil Patel, *IIT Roorkee;* Lucas Graeff Buhl-Nielsen, *Quantco;* Adrien Ghosn, *Microsoft;* Marios Kogias, *Imperial College London*

OneSidedMW: Managing Disaggregated Memory Efficiently, Flexibly, and Securely with RNIC Offloading 2705
Zixuan Wang, Jinyu Gu, Xingda Wei, and Yubin Xia, *Shanghai Jiao Tong University*

Syntra: Synthesizing Cross-Layer Controllers for Low-Latency Video Streaming 2721
Jia Pan, *The University of Texas at Austin;* Anup Agarwal, *Carnegie Mellon University;* Işıl Dillig and Venkat Arun, *The University of Texas at Austin*

Net-P4ct: Enhanced WAN Bandwidth Fair Sharing Using P4 Programmable Switches 2739
Haoran Chen and Mingwei Cui, *Bytedance;* Yihan Zou, Yihang Miao, Suhan Jiang, Damu Ding, Lirong Lai, Ming Gao, Rui Jiang, Shengyuan He, Anjian Chen, Jiaming Shi, Junjie Wan, Yandong Duan, Ruomin Fang, Hongyu Wu, and Yongping Tang, *ByteDance;* Qiao Kang, *unaffiliated;* Guangrui Wu and Xiyun Xu, *ByteDance*

FRCC: Towards Provably Fair and Robust Congestion Control 2755
Anup Agarwal, *Carnegie Mellon University;* Venkat Arun, *University of Texas at Austin;* Srinivasan Seshan, *Carnegie Mellon University*