2026 IEEE/ACM International Symposium on Quality of Service (IWQoS 2026)

Istanbul, Turkiye
29 June - 2 July 2026

Pages 1–611

IEEE Catalog Number: CFP26IWQ-POD
ISBN: 979-8-3195-1704-3

Copyright © 2026, IEEE

All Rights Reserved

Copyright and Reprint Permissions:

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

*** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.

IEEE Catalog Number:	CFP26IWQ-POD
ISBN (Print-On-Demand):	979-8-3195-1704-3
ISBN (Online):	979-8-3195-1703-6
ISSN:	1548-615X

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
E-mail:	curran@proceedings.com
Web:	www.proceedings.com

CURRAN ASSOCIATES INC. proceedings .com

TABLE OF CONTENTS

Toward Communication-Sensing Integrated Cloud-Network Monitoring: Real-Time Vibration Event Recognition on Underground Optical Cables 1
Wanchang Jiang, Yuan Huang

CEINR: Critical Element-aware In-Network Recovery for Distributed Machine Learning 7
Jixing Yang, Shuran Zhang, Zehua Guo, Songshi Dou, Jianye Wang

Dynamic Malicious Encrypted Traffic Detection via Temporal Knowledge Graph in Cloud-Edge Networks 13
Qilong Huang, Zihao Wang, Yuxiang Hu, Le Tian, Yufei Zhu, Menglong Li

MiniQ: Minimizing Tail Latency for Cloud SFCs with Queuing-aware Execution Planning 19
Weijia Yu, Wei Zhang, Guoyuan Li, Wendi Feng, Jianjun Chang

FedRIDA: Coordinated Participation and Update Control for Time-Efficient Federated Learning 25
Gang Yan, Yujie Jia

RAPIDS: Reliable Adaptive Priority-based Intelligent Delivery Streaming for Enhanced Error Resilience in Real-time Video 35
Dayou Zhang, Xiaolin Wang, Wenlong Chen, Zijian Cao, Hao Zhu, Fangxin Wang

PRO: Deterministic Load Balancing for AI Datacenter Network 45
Yiran Zhang, Qingkai Meng, Yadong Liu, Xiangyu Han, Xiao Ma, Shangguang Wang

Speak Your Network: Automated Network Emulation Construction with Cost-Efficient Multi-Agent Orchestration 55
Daolin Zou, Jie Xu, Long Luo, Hongfang Yu

RABR-Co: Retention-Guided SVC ABR with Joint Buffer Tuning for Non-linear Video Streaming 61
Lihuan Hui, Wang Yang, Yanbo Wang

Cloud-Compatible and Scalable Private Messaging via Authentication Piggybacking 71
Jingwen Lu, Meng Hao, Yuan Zhang, Yaqing Song, Tianhao Yang

ReasonCache: Accelerating Large Reasoning Model Serving through KV Cache Sharing 81
Kaiwen Chen, Xin Tan, Minchen Yu, Jingzong Li, Hong Xu

Ensuring QoS Stability in Dynamic WANs through Selectively Optimized Range Routing 87
Chong Wu, Zehua Guo, Songshi Dou, Minghao Ye

StarRound: An Efficient Multiple Geographic Region-Avoidance Mechanism for High Dynamic LEO Satellite Networks 93
Yuxuan Weng, Qian Wu, Zeqi Lai, Chenwei Gu, Hewu Li, Jihao Li, Qi Zhang, Weisen Liu, Jun Liu Yuanjie Li

DistSRE: Synthesizing Runtime Executions with Automated Co-Optimization for Distributed LLM Training 103
Xiuzhu Sha, Chenyang Hei, Fuliang Li, Chengxi Gao, Rongfei Zeng, Xingwei Wang

Bias in the Shadows: Explore Shortcuts in Encrypted Network Traffic Classification 113
Chuyi Wang, Xiaohui Xie, Tongze Wang, Yong Cui

IntentVQA: A QoE-Aware Cloud-End Collaborative System for Video QA 119
Yishuo Zhang, Lanshan Zhang, Yaya Wei, Ziyan Zhong, Xiaohui Xie, Ziyi Wang, Wendong Wang

PSM: Timely and Resource-Efficient Sketch Migration in Network Measurement 129
Xun Zhou, Hongyan Liu, Xiang Chen, Zhengyan Zhou, Zihao Cheng, Di Wang, Chunming Wu

MACH: A Matrix-Accelerated Classifier for High-throughput Packet Processing on GPUs 135
Zhengyu Liao, Shiyou Qian, Jian Cao, Guangtao Xue, Zhonglong Zheng, Minglu Li

Minimizing the Delay Disparity for Cross-Region Virtual Reality Gaming in Satellite Edge Computing 141
Tong Sheng, Zichuan Xu, Haocheng Zhou, Qiufen Xia

The Imitation Game: Using Large Language Models to Disrupt Chinese Chat-Based Cybercrime 151
Yifan Yao, Baojuan Wang, Jinhao Duan, Kaidi Xu, Chuankai Guo, Zhibo Sun, Yue Zhang

Switch-Transparent Load Balancing for RDMA Data Centers: A Host-Only Approach 157
Tao Zhang, Ao Zhang, Yuchen Wang, Haozhi Tang, Linfei Dong, Ziyi Liu, Hui Yu, Siyuan Fan, Hui Yin Fangmin Li

Chariot: Accelerating Distributed Protocols with Data-Path Accelerator in DPUs 167
Yukai Huang, Jingqi Feng, Chunpu Huang, Sicheng Liang, Ming Yan, Jie Wu

Chronus: Accurate and Memory-Efficient Sketching for High-Cumulative-RTT Flow Detection 177
Hanwen Zhang, He Huang, Yu-E Sun, Chuanwei Li

LADS: Towards Low-Latency Atomicity Interval for Cross-shard Transactions via Dynamic Queue Scheduling 187
Xiaoke Tang, Huawei Huang, Jian Zheng

Pluto: Fast and Accurate Persistent Flow Detection in High-Speed Networks via Adaptive Protection 197
Weihe Li, Jiawei Huang, Zhaoyi Li, Tianyue Chu, Qichen Su, Jianxin Wang

Joint Optimization of MEC Offloading and Traffic Signal Control via Diffusion-based Hierarchical DRL 207
Rui Wang, Tong Li, Zhu Xiao, Shiceng Zhang, Zhuo Tang, Kenli Li

RCAgentBench: An Agent-Oriented Benchmark for Multimodal Root Cause Analysis in Microservices 217
Hengyue Jiang, Zexin Wang, Xiaohui Nie, Di Gao, Jingjing Li, Changhua Pei

ChatGosen: A Network Configuration Synthesis Approach with Semantic-Computation Decoupling 227
Chunyuan Liu, Zhaokun Tan, Fuliang Li, Bocheng Liang, Jiaxing Shen, Lei Yang, Xingwei Wang

SE-STGAE: A Shapelet-Enhanced Spatiotemporal Graph Autoencoder for Early Anomaly Detection 237
En Wang, Fujia Zhou, Hao Du, Wenbin Liu, Jie Wu

ISP or Customer? Inferring the Ownership of Public IPs of Non-Cooperative Satellite Internet via Internet Measurements 247
Hong He, Enhuan Dong, Jiahai Yang, Wenjian Zhang, Guanglei Song, Kexin Qiang, Jiaxing Li, Hui Zhang Xiaowen Quan

H2I: A Handover-State Encoding-Based Data Inheritance Method for Mobile Crowdsensing 257
Haifeng Jiang, Siyuan Yin, Chaogang Tang, Shuo Xiao, Huaming Wu, Ruidong Li

ReMu: Bridging Fidelity and Flexibility in High-Mobility Network Emulation at Microsecond Scale 263
Mingtai Lv, Xuyan Jiang, Chengxin Liu, Huan Zhou, Gaofeng Lv, Jinshu Su, Xiangrui Yang

OpScope: Exploiting Operation-Driven Visual Scope for QoS-Stable Cloud Gaming 269
Taolei Wang, Chao Li, Jing Wang, Xiaofeng Hou, Minyi Guo

Odysseus: A Context-Level Pre-training Framework for Out-of-Distribution Encrypted Traffic Classification 275
Wenqi Dong, Xinjie Lin, Longtao He, Gaopeng Gou, Zhen Li, Junzheng Shi, Jianshuo Liu, Gang Xiong

BGP-Shadow: Accelerating BGP Verification via Programmable Data Planes 285
Chuan Lin, Hongyan Liu, Dong Zhang

StreamDuet: Bandwidth Efficient Multi-Drone Video Analytics with Iterative Streaming 295
Haihan Zhang, Weijun Wang, Haipeng Dai, Ruiben Zhou, Liang Mi, Mi Zhou, Guihai Chen

P4XC: A Unified Compiler Framework for Network Dataplane with Heterogeneous Processors 305
Zhuang Ling, TianYing Tang, Haoyu Song, Zhikang Chen, Shuxin Liu, Bin Liu

GCA-BULF: A Bottom-Up Framework for Short-Term Load Forecasting Using Grouped Critical Appliances 315
Yunhao Yao, Jinwei Fang, Puhan Luo, Zhiqiang Wang, Jiahui Hou, Xiang-Yang Li

Zero-Cost Merging Transitioning for Large-Scale Erasure-Coded Storage Systems 325
Canghai Yang, Kan Zhong, Zhiwang Yu, Wanyi Guo, Chaoxia Qin, Duo Liu

SwitchNN: In-Network CNN Inference for Edge-Assisted Smart Roadside Networks 336
Jianqiang Zhong, Jingpu Duan, Xiaoxi Zhang, Shuo Liu, Haoran Xu, Wenfei Wu, Yu Wu, Deke Guo Bingyang Liu, Xu Chen

ClusterFi: Enabling Concurrent WiFi Backscatter Communication via Collided Signal Clustering 346
Weiqi Wu, Wei Gong, Jingwei Sun, Guangzhong Sun

Efficient and High-Precision Online Backdoor Detection for Unknown Attack Domains 356
Tie Xiao, Jiahui Hou, Yunyi Huang, Jie Zhang

Breaking the Accuracy-Latency Trade-Off in Sketch Compression for Network Measurement 366
Hui Yu, Tao Zhang, Siyuan Fan, Ziyi Liu, Zhihong Wang, Yunsheng Liu, Linfei Dong, Haozhi Tang Ao Zhang, Hui Yin, Fangmin Li

Asymmetric-Aware Hybrid Granularity Load Balancing in RDMA-enabled Data Center Networks 376
Jiacheng Qu, Shaojun Zou, Yi Jiang, Zirong Liu, Rui Chen

HyperWeave: QoS-aware GPU Overcommitment for Deep Learning Training Jobs 386
Long Qin, Hailiang Chen, Yong Peng, Lujia Yin, Miao Zhang, Yiming Zhang

Scylla: Scheduling Multiple Latency-Sensitive Applications in the Edge-Cloud Continuum 396
Yinan Cao, Etienne Rivière, Ramin Sadre

Velo-NC: Verified Worst-Case End-to-End Queueing-Delay Bounds Under Network Dynamics 406
Shangsen Li, Changhao Qiu, Lailong Luo, Bangbang Ren, Rui Chen, Deke Guo

Few-Packet Traffic Classification under Dynamic Network Conditions 416
Xujing Li, Zhuochen Fan, Yuchao Zhang, Yi Wang

SpecOffload: Unlocking Latent GPU Capacity for LLM Inference on Resource-Constrained Devices 426
Xiangwen Zhuge, Shen Xu, Fan Dang, Xuan Ding, Danyang Li, Tianxiang Hao, Qiang Ma, Xin Miao Yahui Han, Zheng Yang

MPFV: Lightweight Multipath Forwarding Validation for SRv6 Networks 436
Xiaoliang Wang, Mengyao Liu, Longyu Duan, Ziwei Li, Gang Zhao, Wenlong Chen

Chain-Based Service Function Chain Placement: Optimizing Latency in Multi-Topology Networks 446
Shuo Quan, Zixuan Chen, Jie Wu

HeraClass: Towards Open-World Network Flow Classification via Traffic-Language Mapping 456
Ni Jin, Libin Liu, Yukai Miao, Li Chen, Dan Li, Xizheng Wang, Xiuting Xu, Baojiang Cui

Semantic Raft: A Fault-Tolerant Multi-Agent Inference Framework for Reliable LLM Services 466
Yidong Su, Yangyang Wang, Yucong Dong, Zichen Xu

DNSMedic: Automated Precision Diagnosis and Repair Guidance for DNS Misconfigurations 476
Ziyi Wang, Kaiqiang Hu, Lei Tian, Mochun Long, Haizhou Du, Qiao Xiang, Mengrui Zhang, Letian Zhu

CANListener: A Lightweight and Adaptive Intrusion Detection System for In-vehicle CAN Bus 486
Jiacheng Jin, Yihe Zhang, Chong Zhang, Feng Lin, Xiaohang Wang, Li Lu

Learning from Experience: Real-time and Efficient Network Measurement on Programmable Switches 496
Pulun Gao, Guoju Gao, Yu-E Sun, He Huang, Yue Kan, Yang Du

CoInfer: Breaking the Resource Wall for Collaborative Edge Inference 502
Yongmin Zhang, Mingyi Dong, Jinrui Zhang, Pengyu Huang, Shuai Zhang

FlyPS: A Flexible Multi-job Placement Scheme with Communication Scheduling in GPU Clusters 512
Jianfeng Bao, Gongming Zhao, Hongli Xu, Lixin Deng, Junhong Lu, Hao Zheng, Wenpeng Zhu

GaLB: Gap-Aware Load Balancing for High-Performance AI Training in RDMA Networks 518
Jinbin Hu, Zijing Zhong, Rui Zhi, Jin Wang

Decentralized But Not Immune: Empirical and Formal Security Analysis of Messaging Networks 524
Hao Li, Rui Wang, Hongfeng Sun, Zhenxiang Chen, Shanqing Guo

ConvexLens: Overcoming Frequency Stagnation in Millisecond Workloads via Temporal Co-location 530
Zhenglian Li, Weihong Wu, Jingzhao Xie, Hetian Li, Jiayi He, Gang Sun, Hongfang Yu

WiFi-based Human Pose Estimation via Intermediate Pose Synthesis and Part Grouping 540
Yi Fang, Shigeng Zhang, Xuan Liu, Zhiwei Zheng, Song Guo

The Enduring Potential of Locality: Efficient Routing for Fine-Grained Cluster Deduplication 550
Cai Deng, Xiangyu Zou, Philip Shilane, Haoliang Tan, Shiyi Li, Wen Xia

FedPDR: Server-Side Post-Plateau Refinement for Federated Learning with Latent Diffusion 560
Xiaowei Liu, Hui Wang, Huan Zhou, Zhenyu Zhang, Peng Sun, Qi Gao, Tingting Wu, Zhiwen Yu

Adaptive Multi-Path Mamba Knowledge Distillation Framework for Industrial Defect Detection 566
Shengjie Li, Jiancheng Chi, Pengfei Wang, Lei Wang, Xiaobo Zhou, Ning Chen, Tie Qiu

RAN Parameters Under the Microscope: A Deep Dive into 5G Handover Performance in Dense Urban Network 576
Jiahui Xu, Chengxiang Lin, Yang Wang, Hongtao Guan

TAILOR: Token-Aware Partitioning and Routing for Edge–Cloud Transformer Inference 586
Xiaoyao Huang, Remington R. Liu, Jie Wu

DP-MPC: Offloading Secure Multi-Party Computation to Programmable Data Planes 592
Xun Zhou, Yu Zhou, Ming Chen, Jianhao Li, Di Wang, Xiang Chen, Chunming Wu, Hongyan Liu

Scalable Simulation-based Configuration Verification of DCNs via Destination-Independent Compression 602
Mengrui Zhang, Xiaoqiang Zheng, Letian Zhu, Xing Fang, Lizhao You, Zhang Liu, Ziyang Yao, Yang Wang Rui Wen, Zhi Zhang, Ronghua Sun, Yuanhui Zhong, Haihua Li, Fei Yuan, Qiao Xiang

Entangled Photon Source Pooling and Entanglement Distribution for Quantum Networks 612
Junyuan Shi, Yangming Zhao, Dong Zhang, Bingheng Yan, Chen Tian, Chunming Qiao

DOSA: Optimizing DVFS Policy for Stochastic Workloads with Analytical Solutions 622
Xuanhao Feng, Qiang Wang

Achieving Fast and High Throughput Data Exchange for Serverless Computing Systems via Switch-Native Serialization/Deserialization 632
Gonglong Chen, Baiyan Ke, Yuxin Xu, Shenghong Xiong, Haolin Pan, Jiamei Lv, Wenxing Ge, Yuxi Wang Chengzhong Xu, Kejiang Ye

BEP: A Bayesian-Entropy Pruning Framework for Efficient IPv6 Active Network Discovery 642
Yuxuan Liao, Tao Yang, Zhiping Cai

TD-RCA: Topology-Aware and Dual-Perspective Decoupling for Root Cause Analysis in Microservices 652
Shiming He, Bin Tang, Mengyao Wei, Lingyun Xiang, Kun Xie

Desi: Revisiting Signature Verification in Blockchain-based Storage System 662
Ziwei Li, Songsong Xu, Tong Li, Xiaoliang Wang, Bo Wu, Yangfei Guo, Shenglin Jiang, Ke Xu

Beagle: Auto-tuning Performance Diagnosis for Live Video Streaming 668
Chandan Bothra, Leonardo Teixeira, Zahaib Akhtar, Sanjay G Rao, Bruno Ribeiro

Congestion-Resilient Auditable V2X Alert Scheduling with Publicly Checkable Evidence 678
Yongming Zhang, Lei Liu, He Xu

Square: Towards Sharing Internet-Scale GPU Clouds Fairly and Efficiently 687
Ningxin Su, Freeman Cheng, Baochun Li

ALPS: ACK-induced Latency-based Packet Spraying for Multipath Transmission 693
Jinyu Xiao, Haoyu Song, Zhikang Chen, Ying Wan

Renewables Power the Orbit? Achieving Sustainable Space Edge Computing via QoS-Aware Offloading 703
Xiaoyi Fan, Yi Ching Chou, Hao Fang, Long Chen, Haoyuan Zhao, Ershun Du, Chongqing Kang, Zhe Chen Jiangchuan Liu

REACT: Toward Real-Time, End-to-End, Adaptive Cross-Layer Restoration for IP-Over-Optical Networks 713
Siyong Huang, Lei Tian, Mochun Long, Qingyu Song, Lizhao You, Qiang Su, Lu Tang, Wanjian Feng Fei Yuan, Qiao Xiang, Jiwu Shu

DelAct: A Replayable Boundary Runtime for Auditable and Governed LLM Agent Workflows 723
Yuanbo Zhang, Hanlong Liao, Deke Guo, Guoming Tang

NeuroForensics: Unmasking Covert Backdoor Attack via Endogenous Defense 729
Feirun Huang, Hui Cai, Biyun Sheng, Lu Zhao, Yanmeng Wang, Jian Zhou

It Takes Two: Embracing Sparsity and Speculative Decoding for Efficient LLM Inference 735
Haolin Chu, Changyu Chen, Wei Liu, Jian Luan, Jiabin Deng, Liang Liu, Huadong Ma, Xiaolong Zheng

UPServe: Backend Agnostic Proxy for Black-box Heterogeneous LLM Scheduling 745
Haorui Wan, Chenyang Hei, Fuliang Li, Chengxi Gao, Yuhan Jia, Tongrui Liu, Xingwei Wang

HawkEye: Hierarchical Intent Inference for Critical Targets on Edge-Cloud Co-Assisted UAVs 755
Ningzhe Liu, Junqi LYu, Kaixin Chang, Yantao Zhang, Zhijun Ding

Biphasic Sketch: Multi-Attribute Stream Summarization for Arbitrary Attribute Combinations 761
Niuniu Zhang, Lailong Luo, Zelin Wei, Qianzhen Zhang, Yan Li, Deke Guo

StarkServe: A Framework for Elastic Serverless LLM Inference at the Extreme Edge 771
Mingxuan Zhang, Xiaofeng Hou, Jiacheng Liu, Chang Liu, Xinkai Wang, Chao Li, Minyi Guo, Han Yang

NimbleChain: Automatic Timeout Tuning for PBFT-based Blockchain Systems 777
Huahui Xia, Kailang Zhu, Tong Li, Jinchuan, Keman Huang, Wuqiong Pan, Xiaoyong Du

SketchPlan: Full-Visibility Sketch-Based Telemetry with Limited Programmable Switch Coverage 783
Jinbo Sun, Haifeng Sun, Jintao He, Qun Huang, Sa Wang, Yungang Bao

Joint Optimization of Video Recommendation and Cooperative Edge Caching for Maximizing Profit 793
Haoqiu Luo, Youling Zeng, Yufan Shen, Yue Zeng, Liying Li, Qianpiao Ma, Peijin Cong, Junlong Zhou

Duet: Towards Efficient and Accurate Sketch-Based Measurement on DPUs 803
Yang Wu, Songlin Chen, Fuliang Li, Man Hou, Jiaxing Shen, Xingwei Wang

Ygg: Tree-Based Collaborative Speculative Decoding with Token-Only Transmission 813
Sijia Li, Yumeng Liang, Ruiqi Li, Jianjiang Li, Jie Wu

Tremerity-Fi: Non-Contact Daily-Life Tremor Severity Assessment by Commercial mmWave Radar 823
Xiao Li, Jingyang Hu, Shang Gao, Yichao Gao, Yiyu Xin, Hongbo Jiang, Xiang-Yang Li

Temporal Predictive Joint Hybrid Beamforming Framework for Mitigating Channel Aging in High-Mobility MIMO-OFDM Systems 833
Peilun Jia, Hanwei Zhang, Mingbo Zhang, Xiaoqiang Zhu, Yuan Liu, Chenyang Wang

Later is More: Trading Tolerable Latency to Meet Stringent Jitter Requirement in Time-Sensitive Networking 843
Shaodong Huang, Jiawei Huang, Yifeng Liu, Yiting Wang, Qichen Su, Yilin Zhao, Junjie Tang, Yufan Hu Jiaxin Jiang, Yan Liu, Xiaojuan Lu

Multi-View Flow Representation and Relation Graph Modeling for Network Traffic Classification 853
Huanhuan Gu, Qingjun Xiao, Mingyi Wang, Jiaming Zhang, Qianmu Li

LocMore: Locating More Bursty Latency-critical Jobs on Resource-constrained Nodes 863
Chang Liu, Xiaofeng Hou, Cheng Xu, Xinkai Wang, Jiacheng Liu, Chao Li, Minyi Guo

GBNN: In-Network Gradient Boosting Neural Network 873
Hongyi Liu, Shaowei Xu, Shengrui Lin, Hongyan Liu, Feng Liu, Libin Xu, Jing Chen, Dong Zhang Chunming Wu

LVProv: A Lightweight Blockchain-Anchored Verifiable Provenance Scheme for High-Dimensional Data 879
Xixun Yu, Yuhang Xiao, Huazhong Liu, Zheng Yan, Laurence T. Yang

SBA-Route: Domain-Based Routing for Scalable and Low-Latency 5G Distributed Control Planes 889
Zhuoran Ma, Yanbiao Li, Xin Wang, Xinyi Zhang, Xian Yu, Shiyi Liu, Kun Xie, Gaogang Xie

LAMP5G: A Link-Aware Multipath QUIC Scheduler for Heterogeneous and Disruption-Prone 5G Cellular Networks 899
Hongmiao Yu, Yuanhao Chang, Jorge Perez Ortiz, Quentin De Coninck, K. K. Ramakrishnan

O3-Sketch: Memory-Efficient Online Chaotic Flow Detection in High-Speed Networks 905
Hanwen Zhang, He Huang, Yu-E Sun, Chuanwei Li

Structure-Aware GPU Scheduling with Cycle Fusion and Feedback Adaptation for Multi-Task Training 915
Luwen Guo, Guangping Xu, Tingting Zhang, Mingyuan Ding, Lei Yang, Yunjie Li

LARSS: A Hardware-Software Co-designed Framework for Load-Aware Receive Side Scaling 925
Peng Zhao, Guang Cheng, Hua Wu, Deyu Zhao, Yuyu Zhao, Xiaoyan Hu

NAST: In-Network Aggregation with Worker Selection for Accelerating Distributed Training 935
Jianfeng Bao, Peng Yang, Gongming Zhao, Huihui Tang, Hongli Xu, Qianpiao Ma

CodedFRER: Enhancing Reliability for FRER Protocol via Erasure Codes in Error-Prone Networks 945
Shi Dun, Haolan Jia, Chuang Gan, Yuchong Hu, Tianyun Long, Dan Feng

RiCE: Precise Remote In-Network Congestion Elimination in Inter-Datacenter RDMA Networks 955
Hongyi Li, Chengyuan Huang, Zhiqiang Li, Guangyu Zhao, Lu Lu, Zirui Wan, Jiaqing Dong, Zhuo Tang Guihai Chen, Chen Tian

QoS-driven Network Soft Slicing through Token-based Hierarchical Scheduling 965
Xiaoyang Fu, Zehua Guo, Songshi Dou

FedCod: An Efficient Coded Communication Protocol for Cross-Silo Federated Learning 975
Peishen Yan, Jun Li, Hao Wang, Yang Hua, Tao Song, Lu Peng, Haibing Guan

Heterogeneous Tasks Offloading in Vehicular Edge Computing: A Federated Meta Deep Reinforcement Learning Approach 985
Yaorong Huang, Jingtao Luo, Xuechao Wang

CF-BAD: Coarse-to-Fine Granularity BGP Anomaly Detection for Prefix Hijacking 991
Zheng Wu, Shiyu Liu, Jiong Jin, Fu Xiao, Gaogang Xie

Taming GPU Inference Variability with Risk-Aware Scheduling and Dynamic Kernel Segments 997
Mingyuan Ding, Guangping Xu, Kang Liu, Junhui Li, Lei Yang, Yunjie Li

DepartKVS: KV Separation for Redundant Replicas in Strongly Consistent Distributed Stores 1007
Qiren Zhao, Guangping Xu, Kai Shi, Yunjie Li, Lei Yang

ALCHEMY: Reusing Congestion Control Wisdom for Adaptive QUIC Transport 1017
Yanbing Li, Jashanjot Singh Sidhu, Abdelhak Bentaleb

RGB: Resource Graph Base for Integrating Tri-Elements in Computing Power Networks 1023
Wenjie Wang, Hongyu Tao, Yu Zhao, Hao Tang, Cheng Wang

CT-Sketch: Persistent Item Lookup Based on Collision Statistics and Thresholds 1033
Chen Yang, Lailong Luo, Yuliang Lu, Qianzhen Zhang, Guozheng Yang, Yi Xie

Dynamic Hot Expert Replication with Load and Topology-Aware Joint Gating for Distributed MoE Inference 1043
Jing Yang, Huaqing Tu, Gongming Zhao, Hongli Xu, Baoqing Wang, Long Li

Mitigating Cold Starts in Container-Ephemeral Large Language Model Serving 1049
Wei Jin, Bo Peng, Jianguo Yao

Purification and Routing for Makespan Minimization in Quantum Networks 1059
Yangyu Wang, Yangming Zhao, Hongli Xu

AdaMoA: Enhancing Mixture-of-Agents via Task-Adaptive Architecture Optimization 1069
Wenhui Lv, Hongli Xu, Yang Xu, Jianchun Liu, Zihuai Xu

CoPT: Collaborative Machine-Learning-Based Mechanism for Adaptive Congestion Control Parameter Tuning in Data Center Networks 1079
Ziwen Yang, Ruya Gu, Duokun Xu, Shuyong Zhu, He Liu, Yanmin Jia, Xiaohe Hu, Yujun Zhang

AuEx: Automatic Exploration for Fresh Creative Ads via Large-Scale Deep Reinforcement Learning 1089
Yi Zhao, Qi Tan, Min Zhu, Liehuang Zhu

CBDDoSCLIP: A Lightweight Multimodal Framework for Carpet-Bombing DDoS Attack Detection 1099
Haotian Meng, Tao Jia, Yongming Xu, Zongyuan Zhang, Tianyang Duan, Xinggong Zhang

Dual-Sketch: A Route-Oblivious Framework for Distributed Sparse Superspreader Detection 1105
Zeyu Yang, Shu Li, Meijie Du, Yong Wang

DMKG: DDoS Defense Enhanced by Multimodal Knowledge Graph Data 1115
Jialin Niu, Mingzhe Xing, Lei Zhang, Da An, Yong Cui

FedQCab: A Common Plug-and-Play Layer-wise Quantization Tool for Communication-Efficient Federated Learning 1121
Haizhou Du, Tonghao Chen

TSN-FLARE: A Flexible and Resilient Telemetry Framework for Time-Sensitive Networking 1127
Wenliang Ma, Qian Yang, Gang Sun, Wei Quan

Beyond Up and Down: Analyzing Temporal Phenomena in IPv6 Probing Responses 1133
Yudong Lian, Jinfeng Peng, Pengfei Xue, Chengxi Xu, Fan Shi, Min Zhang, Jiatang Zhao, Quming Peng

CDWF: Few-Shot Learning for Cross-Domain Multi-Tab Website Fingerprinting 1139
Xinlei Ju, Zhen Li, Yang Li, Lihua Yin, Zhong Guan, Gaopeng Gou, Mi Zhou, Junzheng Shi, Gang Xiong

Distributed or Centralized: Edge-Cloud Collaborative Active Learning 1149
Chenxin Mao, Shangyu Liu, Chen Gong, Ting Zhou, Zhenzhe Zheng, Huiyang Xu, Fan Wu, Guihai Chen

Fair Scheduling for Lossless RDMA Networks 1155
Yazhu Zhao, Zehua Guo, Xiaoyang Fu, Chong Wu, Yang Su

ISP-Path Alignment Can Outweigh Protocol: An HTTP/3 Case Study in Suzhou, China 1157
Yuxuan Li, Jiaxian Tu, Lifan Mei

MAS-CP: Multi-Agent Communication Prediction for Distributed Deep Neural Network Training 1159
Zijie Huang, Zixuan Chen, Kai Shi

Inferring Starlink Latency Structure from Public RIPE Atlas Measurements 1161
Jiaxian Tu, Yuxuan Li, Shenghe Xu, Lifan Mei

An Adaptive Multi-Metric Forwarding Strategy for Vehicular Named Data Networking with Hybrid Contention and Opportunistic Delivery 1163
Zhuoxin Yuan, Rui Hou, Wei Li, Yuanai Xie, Wanneng Shu

Prediction-Based Adaptive Edge Caching Update Strategy for the Internet of Vehicles 1165
Sirui Ruan, Rui Hou, Wei Li, Yuanai Xie, Wanneng Shu

Janus: Edge Content Negotiation for Agent-Oriented Web QoS 1167
Zixuan Chen, Zijie Huang

Exploring Energy Saving Opportunities for Data Center Cooling Systems 1169
Jianuo Li, Zehua Guo, Xiaoyang Fu

Zephyr: Switch-Aided Weighted Congestion Control for Differentiated Inference Workloads 1171
Zijia Xu, Tian Pan, Junkai Huang, Wenxuan Zhao, Chenlin Ge, Hao Ouyang, Guohao Ruan, Tao Huang

SOPSmith: Forging Executable SOPs for LLM-Driven GPU Cluster Network Diagnosis 1177
Guoyao Yu, Xiaoqing Sun, Ziqi Huang, Yangyang Shi, Xiong Xiao, Yang Song, Xing Li, Biao Lyu Zhenguang Liu, Qinming He

Beyond Random Probing: Intelligent IPv6 Discovery via Ensemble Learning and Contextual Expansion 1183
Haoxiang Yang, Gang Ren, Wenying Jiang, Wei Zhang

SpecNet-Agent: Network-Aware Speculation Control for QoS in Agentic Generative AI Services 1189
Le Gai, Lin He, Chentian Wei, Zedong Jia, Daguo Cheng, Ying Liu

Privacy-aware Hand Gesture Recognition using Textile-based Sensing Glove 1194
Omur Fatmanur Erzurumluoglu, Mehmet Ali Sarikaya, Asli Tuncay Atalay, Ozgur Atalay, Gökhan Ince

QoS, Privacy, and Security Challenges in Deploying Metaverse-Based Virtual Reality Learning for Children with Dyslexia: A Case Study 1202
Reem Salem, Omar Alfandi, Mousa Al-Kfairy

Location-Aware SerDes Configuration for Robust Scale-Up AI Rack Systems 1210
Nan Wang, Rudong Shi, Yang Wu

A Flexible-Topology Rack-Scale Interconnect Architecture for Heterogeneous LLM Workloads 1215
Shiyu Hua, Nan Wang, Bo Chen, Albert Hu, Shangming Liu, Rudong Shi

Noah: Tile-Level Interval Analysis for NPU Performance Modeling 1220
Mengqi Fu, Rulan Yang, Mengrui Zhang, Ziyi Wang, Qingyu Song, Yuanxun Kang, Longhui Zhang Qiao Xiang