IEEE HPSR™

IEEE International Conference on High Performance Switching and Routing

5-7 June 2023

Albuquerque, NM, USA

IEEE HPSR 2023

IEEE

IEEE ComSoc
IEEE Communications Society

IEEE HPSR'23 Committee

General Chair:

Eirini Eleni Tsiropoulou, University of New Mexico, USA

General Co-chair:

Abdelhamid Mellouk, L'Université Paris-Est Créteil, France

TPC Co-chairs:

Dimitrios Koutsonikolas, Northeastern University, USA

Carla Fabiana Chiasserini, Politecnico di Torino, Italy

Gang Li, Deakin University, Australia

Workshop Co-Chairs:

Anna Maria Vegni, Roma Tre University, Italy

Aris Leivadeas, ÉTS Montréal, Canada

Tutorial Chair:

Chiara Boldrini, IIT-CNR, Italy

Posters & Demos Chair:

Xiang Sun, University of New Mexico, USA

Panel Chair:

Symeon Papavassiliou, National Technical University of Athens, Greece

Industry Program Chair:

Xi Li, NEC Laboratories Europe, Germany

Publication Co-chairs:

Pantelis Frangoudis, TU Wien, Austria

Andrea Morichetta, TU Wien, Austria

Publicity Co-Chairs:

Suzan Bayhan, University of Twente, Netherlands

Sumitro Bhaumik, TCS Research, India

Zhu Wang, Northwestern Polytechnical University, China

Web Co-chairs:

Md Sadman Siraj, University of New Mexico, USA

Panagiotis Charatsaris, National Technical University of Athens, Greece

Financial Chair:

Bruce Worthman, IEEE Communications Society, USA

ComSoc Project Manager:

Shirley P. Cisneros, IEEE Communications Society, USA

IEEE HPSR'23 Conference Program

Day 1: Monday, June 5, 2023

Registration: 8:00 AM – 9:00 AM

Introduction, Welcome: 9:00 AM – 9:30 AM

Online Workshops: 9:30 AM – 11:30 AM

Next Generation Connectionless IP Router Architecture with Switching Delay for URLLC Services

Yoichiro Ueno and Akihiko Tsukahara (Tokyo Denki University, Japan); Noriharu Miyaho (Tokyo Denki University & Japan, Japan)

Abstract: Considering the recent demand for low-latency communication services, such as URLLC (Ultra-Reliable Low-Latency Communication), for 5G and beyond, there are still several unresolved and challenging issues. To address these challenges, we propose an innovative connectionless (CL) IP router architecture designed for use in both edge nodes and core transit nodes. To ensure a minimal IP switching delay and guarantee QoS, a high-speed switching mechanism alone is insufficient. A novel approach is required to mitigate the variable switching delays that conventionally arise due to fluctuations in outgoing traffic volume. Therefore, in order to achieve both minimal switching delay and stable QoS, a high-speed switching mechanism and an innovative solution to address conventionally occurring variable switching delays caused by fluctuations in outgoing traffic volume are both necessary. To achieve minimal switching delay and ensure stable QoS, a high-speed switching mechanism alone is insufficient. A novel idea is also necessary to prevent variable switching delay caused by fluctuations in outgoing traffic volume. Furthermore, we provide a detailed explanation of the essential mechanism required to implement the proposed CL IP switch. This innovative router architecture is designed to enable URLLC services with less than 1 millisecond end-to-end delay and less than 1 microsecond switching delay for long-distance real-time communication services. Recent advances in CMOS technology and optoelectronic device technology integration make it feasible for a wide range of URLLC network services for 5G and beyond. To confirm the functionality of the proposed router, we designed an equivalent IP switch architecture and evaluated its essential function for minimal switching size using commercially available FPGA and VHDL (VHSIC hardware description language) for critical path and circuit synthesis and simulation.

Anonymising Video Data Collection at the Edge Using DeepDish

Roman Kolcun, William Pan and Richard Mortier (University of Cambridge, United Kingdom (Great Britain))

Abstract: Increase in popularity of various deep learning methodologies has led to a huge increase in demand for training data. These training data often contain video featuring people but these videos often cannot be used due to privacy concerns as people are identifiable. We present an extension to the DeepDish object tracking system running on a Raspberry Pi. DeepDish performs object detection in the Raspberry Pi itself, avoiding the need to ship raw video data off the device. Our extension adds online face detection and blurring to preserve privacy. We implement and evaluate four different face detection algorithms, achieving 83% accuracy using YuNet, while increasing the latency by 75%. We then implement and evaluate three different object tracking algorithms to reduce the latency increase by running the simpler object tracking algorithm, reducing the number of calls to the expensive face detection algorithm. The result is a significant decrease in latency at the cost of a small decrease in accuracy. We also implemented and compared three different face obfuscation algorithms, one of which achieves differential privacy.

Intent Expression Through Natural Language Processing in an Enterprise Network

Elie El-Rif (Ecole de Technologie Superieure, Canada); Aris Leivadeas (École de Technologie Supérieure, Canada); Matthias Falkner (Cisco, Canada)

Abstract: The unprecedented expansion of network communications and the addition of new applications and services have made the network configuration a riddle for network engineers. At the same time, the daily interaction of the users with various network applications and the heterogeneous requirements they impose, necessitates a more direct communication between these two entities. Intent Based Networking (IBN) is a new technology that enables this by automating the network configuration, while allowing the users to directly interact with the network. The latter can be established through a natural language form, letting the users to freely express their high-level network requirements via a normal conversational approach. To this end, in this paper, we introduce ETS-Chatbot by extending the functionalities of a chatbot in order to be used in an IBN context. We position our work in an enterprise environment and we try to capture the high-level intent and map it into a model that can be understood by the network. The obtained results reveal that our framework can accurately classify the intent of a user with high precision, giving a strategic advantage to the enterprise.

Constraints-Aware Training (CAT) to Enable Software-Hardware Co-Design for Memristor-Based Analog Neuromorphic Chip

Zhimin Tang (Xiamen University, USA); Rujie Zhao (Southern Illinois University Carbondale, USA); Linkai Luo (Xiamen University, USA); Haibo Wang (Southern Illinois University Carbondale); Chao Lu (Southern Illinois University Carbondale, USA)

Abstract: Due to their scalability and energy efficiency, memristor-based analog neuromorphic chips (MANCs) have significant advantages in edge computing. Researchers use software to train network parameters, then in hardware deployment, weights are translated into memristor conductances and biases are provided by digital-to-analog converters (DACs). Analog operational amplifiers are utilized for neural signal summation. Nowadays, efficient design, training, and implementation of MANCs have not been thoroughly explored. Particularly, MANC co-design considering software training and hardware deployment has not been studied. To address this issue, we propose Constraints-Aware Training (CAT), which includes hardware constraints of memristor devices, operational amplifiers, and DACs in the training process, thereby enabling co-design between offline training and hardware deployment. To evaluate the proposed methodology, a 4-layer fully connected network (FC-4) and a convolutional network (LeNet-5) are trained by CAT on MNIST and CIFAR-10 datasets, and then deployed in hardware circuits. Experimental results demonstrate that incorporating hardware constraints during offline training through CAT enables convenient and successful MANC hardware solutions.

The Perfect Match: Selecting Approximate Multipliers for Energy-Efficient Neural Network Inference

Ourania Spantidi (Southern Illinois University, USA); Iraklis Anagnostopoulos (Southern Illinois University Carbondale, USA)

Abstract: Reconfigurable approximate multipliers have been proposed as a way to improve the energy efficiency of neural network inference. However, selecting the optimal combination of approximate modes is a challenging problem due to the trade-off between energy savings and accuracy loss. In this paper, we propose a methodology for selecting the best triad of approximate multipliers to form a reconfigurable approximate multiplier that can satisfy a maximum accuracy drop threshold and achieve the highest possible energy savings. We use formal methods to produce a Pareto-front of solutions that satisfy the accuracy constraint and maximize energy savings. Experimental results show that our methodology can achieve significant gains in energy with negligible drops in accuracy when compared to the baseline.

A Statistical Approach to Improve CNN Classification Accuracy

Vasileios Pentsos (Southern Illinois University Carbondale of, USA); Spyros Tragoudas (Southern Illinois University, Carbondale, USA)

Abstract: Convolutional neural networks (CNNs) have achieved state-of-the-art performance in image classification tasks. However, they may underperform for specific classes, resulting in misclassifications. To address this issue, the proposed method involves two steps that use the Mann-Whitney U test on generated image distributions. The method is evaluated on the publicly available dataset CIFAR100, utilizing ResNet-50 as the baseline network. The results show that the proposed method is effective in significantly improving the classification accuracy of low-accuracy image classes while preserving the high-accuracy classes.

Detection and Quantization of Data Drift in Image Classification Neural Networks

Danushka Senarathna (Southern Illinois University, USA); Spyros Tragoudas (Southern Illinois University, Carbondale, USA); Kiriti Nagesh Gowda and Mike Schmit (Advanced Micro Devices, USA)

Abstract: An unforeseen change in the input data is called drift and may impact the accuracy of machine-learning models. A novel scheme for diagnosing data drift in the input stream of image classification neural networks is presented. The proposed drift detection and quantization method uses a threshold dictionary for the prediction probabilities of each class in the neural network model. The method is applicable to any drift type in images such as noise, and weather effects, among others. Experimental results on various data sets, drift types, and neural network models show that the proposed method estimates the drift magnitude with high accuracy, especially when the level of drift impacts the model's performance significantly.

Tutorial 1: 9:30 AM – 12:00 PM

An Overview of a Distributed Post-5G Network Architecture within the EU SLICES-RI Research Infrastructure, Prof. Raymond Knopp, Prof. Adlen Ksentini, Dr. Damien Saucez, and Dr. Nikos Makris

Abstract

We provide an overview of a blueprint for a disaggregated real-time post-5G network architecture that can be deployed on commodity networking and computing equipment. This activity is part of the European SLICES-RI research infrastructure and constitutes a distributed experimental post-5G playground for academic research purposes to be deployed across several EU countries. The blueprint is meant to be reproducible and to evolve in a collaborative manner. It makes use of open-source solutions such as SD-Fabric, Aether, OpenAirInterface, Nephio and others. The tutorial focuses on key networking components such as software-defined edge fabric, P4-based switching implementing the 5G user-plane function (UPF), cloud-native 5G radio-access (RAN) and core network functions (5GC), O-RAN near real-time RAN intelligent controllers (nRT-RIC), O-RAN Open-fronthaul interfaces and multi-cluster orchestration solutions. The tutorial will demonstrate live deployment and operation of a radio-edge site in Sophia Antipolis, France making use of some of the above technologies.

Lunch Break: 12:00 PM – 1:00 PM

Tutorial 2: 1:00 PM – 3:30 PM

The Role Of Data Engineering In Network Automation, Dr. Engin Zeydan and Dr. Josep Mangues-Bafalluy

Abstract

To address the complex issues that larger and highly integrated networks face in the design, analysis, deployment and management phases, recent advances in data science and engineering technologies in both academia and industry have spurred the adoption of various Artificial Intelligence (AI)/Machine Learning (ML) platforms and frameworks in telecommunication network infrastructures. In this tutorial, we aim to provide a comprehensive and thorough overview of the recent advances in data engineering frameworks and link the capabilities of the data engineering ecosystem with a possible connection to future telecommunication systems in the context of network management and orchestration. Some special features of this tutorial are: a clear link between the data engineering ecosystem (including data connection, data ingestion, data processing & analysis, data storage, data monitoring & visualization and data management & orchestration frameworks) and recent developments in networking, an overview of standardization efforts in network management and orchestration and how these can be related to data engineering frameworks, the relationship to data science frameworks, ML platforms used in the industry, and related data engineering use cases for telecommunications networks will be discussed. Two examples on log management in NFV service orchestration and AI/ML-driven scaling of digital service will also be demonstrated.

Recess: 3:30 PM – 3:45 PM

Tutorial 3: 3:45 PM – 6:15 PM

Introduction to Networking Technologies for High-Performance Computing, Prof. Dhabaleswar K. (DK) Panda and Prof. Hari Subramoni (Online)

Abstract

InfiniBand (IB), High-speed Ethernet (HSE), RoCE, Omni-Path, EFA, Tofu, Slingshot, and Aquila technologies are generating a lot of excitement towards building next generation High-End Computing (HEC) systems including clusters, data- centers, file systems, storage, cloud computing and Big Data (Hadoop, Spark, HBase and Memcached) environments. This tutorial will provide an overview of these emerging technologies, their offered architectural features, their current market standing, and their suitability for designing HEC systems. It will start with a brief overview of IB, HSE, RoCE, Omni- Path, EFA, Tofu, Slingshot, and Aquila. In-depth overview of the architectural features of IB, HSE (including iWARP and RoCE), and Omni-Path, their similarities and differences, and the associated protocols will be presented. An overview of the emerging NVLink, NVLink2, NVSwitch, Slingshot, Tofu, Aquila architectures will also be given. Next, an overview of the

OpenFabrics stack which encapsulates IB, HSE, and RoCE (v1/v2) in a unified manner will be presented. An overview of libfabrics stack will also be provided. Hardware/software solutions and the market trends behind these networking technologies will be highlighted. Sample performance numbers of these technologies and protocols for different environments will be presented. Finally, hands-on exercises will be carried out for the attendees to gain first-hand experience of running experiments with high-performance networks.

Welcome Reception: 7:00 PM – 8:30 PM

Posters and Demos (Posters and Demos will remain accessible during the whole duration of the conference)

Demo papers

P5: Event-Driven Policy Framework for P4-Based Traffic Engineering
Panagiotis Famelis, Georgios P. Katsikas and Vasilios Katopodis (UBITECH, Greece); Carlos Natalino (Chalmers University of Technology, Sweden); Lluis Gifre Renom, Ricardo Martinez and Ricard Vilalta (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Spain); Dimitrios Klonidis (UBITECH, Greece); Paolo Monti (Chalmers University of Technology, Sweden); Daniel King and Adrian Farrel (Old Dog Consulting, United Kingdom (Great Britain))

Abstract: We present P5; an event-driven policy framework that allows network operators to realize end-to-end policies on top of P4-based data planes in an intuitive and effective manner. We demonstrate how P5 adheres to a service-level agreement (SLA) by applying P4-based traffic engineering with latency constraints.

Cloud Native Federated Learning for Streaming: An Experimental Demonstrator
Sergio Barrachina-Muñoz (Centre Tecnològic Telecomunicacions Catalunya, Spain); Engin Zeydan (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Spain); Luis Blanco (Centre Technologic de Telecomunicacions de Catalunya, Spain); Luca Vettori (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Spain); Farhad Rezazadeh (UPC & CTTC, Spain); Josep Mangues-Bafalluy (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Spain)

Abstract: This paper presents a demonstration of implementing Federated Learning (FL) for streaming applications using cloud-native technology. By adopting a decentralized approach, FL method can improve convergence time, reduce communication overhead, and increase network energy efficiency compared to a centralized management. The cloud-native FL architecture in the

testbed comprises three different Kubernetes (K8) sites (two for FL Analytical Engines (AEs)/clients for local training and updates and one for aggregation server for FL training) and several components (video streaming clients and servers, 5G gNodeB). After explaining the testbed infrastructure and setup, this demo showcases how real-time visualization of network parameters during FL training, and how video streaming is enhanced through proactive Central Processing Unit (CPU) scaling facilitated by the improved resource prediction features of FL sites deployed across distributed locations.

Poster papers

POSTER: Droppable Wireless Mesh Network for Intelligent Mine Rescue System
Patrick Duane, Sihua Shao, Mostafa Hassanalian, Vasileios Androulakis, Hassan Khaniani and Pedram Roghanchi (New Mexico Tech, USA)

Abstract: In the event of a mine emergency, the inherently unstable environment poses significant risks and challenges to rescue efforts. The use of intelligent systems employing robots to aid in mine rescues has emerged as a promising approach. However, the communication infrastructure within mines is often compromised or insufficient to handle the network traffic demands of these systems following an incident. Consequently, there is a need for a temporary, deployable communication network capable of supporting both time-sensitive environmental monitoring (e.g., toxic or flammable gases) and high-throughput data transmission, such as video or 3D mapping.

POSTER: Split Learning for Image Classification in Internet of Drones Networks
Jingjing Yao (Texas Tech University, USA)

Abstract: Internet of drones (IoD), where drones act as the Internet of things (IoT) devices, has attracted much attention for its application in traffic surveillance, object tracking, and disaster rescue. These applications rely heavily on machine learning (ML) techniques for image classification. The traditional method of training ML models in IoD networks involves sending all data to a centralized ground base station (BS), which can result in privacy and security concerns. Split learning, an approach that separates the training model into a client model and a server model, can mitigate these concerns by allowing the client model to reside in the drones while the server model is in the BS, preserving data privacy without sharing raw data. In this study, we explore the application of split learning in IoD networks and conduct simulations to evaluate our designed algorithm. Simulation results demonstrate that different separations do not significantly impact split learning accuracy, increasing the number of layers on clients can lead to longer training times, communication overhead is a significant bottleneck in split learning for IoD networks, client numbers do not significantly affect accuracy, and training time slightly increases with an increasing number of clients.

Day 2: Tuesday, June 6, 2023

Registration: 8:00 AM – 8:30 AM

HPSR Opening Remarks: 8:30 AM – 9:00 AM

Keynote 1: 9:00 AM – 10:00 AM

Speaker: Dr. Shivendra Panwar
Topic: Latency is the new Bandwidth

Abstract:
The data rates of both wired and wireline links have increased relentlessly over the last several decades. Wireless access rates used to trail those for wireline access rates, but of late have started catching up, so much so that they can be viewed as essentially equal. For most applications, including mobile applications, bandwidth availability is not viewed as a serious constraint anymore. 5G is delivering tens of megabits per second to users, and will soon provide more. The next driver of advances in networking is expected to be the need for reliable low latency connectivity, rather than bandwidth alone. These applications include XR (Augmented Reality, Virtual Reality and Mixed Reality), wirelessly controlled robots and haptic communications. The latency requirements for such applications vary from tens of milliseconds down to the sub-millisecond range. While the latency requirements for these applications can be met by carefully engineered wired and wireless communications, typically in controlled indoor environments, it is still a challenge to provide them over cellular networks. This talk will focus on the emerging challenge of providing reliable low latency broadband communications over cellular networks.

Biography:
SHIVENDRA S. PANWAR received the Ph.D. degree in electrical and computer engineering from the University of Massachusetts, Amherst and a B.Tech. from the Indian Institute of Technology, Kanpur. He is currently a Professor of Electrical and Computer Engineering at the NYU Tandon School of Engineering. He is also the Director of the New York State Center for Advanced Technology in Telecommunications, the Co-Founder of the New York City Media Lab, and a member of NYU Wireless. He has been a Visiting Scientist at the IBM T.J. Watson Research Center and a

consultant to Bell Laboratories. His research interests include the performance analysis and design of networks. His current research focuses on cross-layer research issues in wireless networks, and multimedia transport over networks. He has coauthored a textbook, TCP/IP Essentials: A Lab-Based Approach (Cambridge University Press). He was a winner of the IEEE Communication Society's Leonard Abraham Prize in the Field of Communication Systems for 2004, the ICC Best Paper Award in 2016, and the Sony Research Award. He was also co-awarded the 2011 Multimedia Communications Best Paper Award. He has served as the Secretary for the Technical Affairs Council of the IEEE Communications Society. He is an IEEE Fellow and a Fellow of the National Academy of Inventors (2022).

Recess: 10:00 AM – 10:30 AM

Technical Session 1: 10:30 AM – 12:00 PM

Programmable Control and Data Plane

Paper 1 - ADMBIFA: Accurate Detection and Mitigation of Blended Interest Flooding Attacks in NDN Networks

Authors: Yanan Zhang (Tianjin University of Science & Technology, China); Xin Guo (Tianjin University of Science and Technology, China); Maode Ma (Qatar University, Qatar); Yiying Zhang (Tianjin University of Science and Technology, China)

Abstract: The Named Data Network (NDN) is a future network architecture to transit conventional host-centric networks into data-centric networks. A routing attack, represented by Interest Flooding Attack (IFA), is a major security concern in NDNs. IFAs seriously harm NDNs by sending excessive fallacious Interest packets to overwhelm the routers to serve legitimate users. It is easy to launch, but difficult to defend, especially when legitimate Interest packets are blended with the IFAs as bIFAs. In this paper, we propose an accurate approach to detect and mitigate IFAs and bIFAs, named Accurate Detection and Mitigation of Blended Interest Flooding Attacks (ADMBIFA). It utilizes fuzzy logic to detect IFAs and bIFAs. And then it identifies malicious Interest prefixes with K-means. Finally, it restrains attackers to mitigate the damage to the networks. The simulation results show that the ADMBIFA is very sensitive to detect attacks with an ability to mitigate IFAs and bIFAs. And it can effectively reduce the negative impacts on NDNs.

Paper 2 - Towards Greener Data Centers via Programmable Data Plane

Authors: Garegin Grigoryan (Alfred University, USA); Minseok Kwon (Rochester Institute of Technology, USA)

Abstract: The energy demands for data centers are increasing and are expected to grow exponentially. Reducing the energy consumption of data centers decreases operational expenses, as well as their

carbon footprint. We design techniques to reduce data center power consumption by leveraging Software-Defined Networking (SDN) and programmable data plane concepts. Relying solely on in-data plane registers, our proposed system P4Green consolidates traffic in the least number of network switches and shifts workloads to the servers with the available renewable energy. Unlike existing SDN-based solutions, P4Green's operation does not depend on a centralized controller, making the system scalable and failure-resistant. Our proof-of-concept simulations show that traffic consolidation can reduce data centers' aggregation switch usage by 36% compared to standard data center load balancing techniques, while workload control can boost renewable energy consumption for 46% of the daily traffic.

Paper 3 - ReCoCo: Reinforcement Learning-Based Congestion Control for Real-Time Applications
Authors: Dena Markudova and Michela Meo (Politecnico di Torino, Italy)

Abstract: Real-time communication (RTC) platforms have seen a considerable surge in popularity in recent years, largely due to the COVID-19 pandemic which facilitated remote work. To ensure adequate Quality of Experience (QoE) for users, a good congestion control algorithm is needed. RTC applications use UDP, so congestion control is done on the application layer, leaving way for advanced algorithms. In this paper, we propose ReCoCo, a solution for congestion control in RTC applications based on Reinforcement learning (RL). ReCoCo gains information about the network conditions at the receiver-side, such as receiving rate, one-way delay and loss ratio and predicts the available bandwidth in the next time bin. We train ReCoCo on 9 bandwidth trace files that cover a vast array of network types. We try different algorithms, states and parameters, training both specific and general models. We find that ReCoCo outperforms the de-facto standard heuristic algorithm Google Congestion Control (GCC) in both specialized and general models. We also make observations on the difficulty of generalization when using RL.

Paper 4 - ADSeq-5GCN: Anomaly Detection from Network Traffic Sequences in 5G Core Network Control Plane
Authors: Zixu Tian, Rajendra Patil and Mohan Gurusamy (National University of Singapore, Singapore); Joshua McCloud (Cisco, Singapore)

Abstract: The service-based architecture (SBA) of 5G Core (5GC) introduces significant landscape changes to the modern communication and network system, and the network slicing enables different Network Functions (NFs) to meet diverse service requirements. However, with the broadening interface, some key NFs may become more vulnerable to internal hostile NFs or external malicious entities, which pose severe threats to the control-plane components in the 5GC network (5GCN). In this paper, we propose ADSeq-5GCN, a network-level anomaly detection framework based on modeling network traffic sequences. Our framework focuses on the control plane of 5GCN, where the network traffic is captured and analyzed for anomalies. We use a sequence model, Bidirectional Long Short Terms Memory (Bi-LSTM) networks, to learn normal NF-to-NF interactions and detect anomalies based on incorrect service event prediction. We evaluate our proposed framework on a

5GCN testbed with Free5GC and UERANSIM under various scenarios. Our results demonstrate the overwhelming performance of our proposed framework over the baseline model.

Paper 5 - HINT: Supporting Congestion Control Decisions with P4-Driven In-Band Network Telemetry

Authors: Alessio Sacco and Antonino Angi (Politecnico di Torino, Italy); Flavio Esposito (Saint Louis University, USA); Guido Marchetto (Politecnico di Torino, Italy)

Abstract: Years of research on congestion controls have highlighted how end-to-end and in-network protocols might perform poorly in some contexts. Recent advances in data plane network programmability could also bring advantages in transport protocols, enabling mining and processing in-network congestion signals. However, the new machine learning-based congestion control class has only partially used data from the network, favoring a more sophisticated model design but neglecting possibly precious pieces of data. In this paper, we present HINT, an in-band network telemetry architecture designed to provide insights into network congestion to the end-host TCP algorithm during the learning process. In particular, the key idea is to adapt switches' behavior via P4 and instruct them to insert simple device information, such as processing delay and queue occupancy, directly into transferred packets. Initial experimental results show that this approach comes with a little network overhead but can improve the visibility and consequently the accuracy of TCP decisions of the end-host. At the same time, the programmability of both switches and hosts also enables customization of the default behavior as the user's needs change.

Lunch Break: 12:00 PM – 1:00 PM

Technical Session 2: 1:00 PM – 2:30 PM

Switching and Routing

Paper 1 - Do Switches Still Need to Deliver Packets in Sequence?

Authors: Ufuk Usubutun (New York University, USA); Fraida Fund (NYU Tandon School of Engineering, USA); Shivendra Panwar (New York University & Tandon School of Engineering, USA)

Abstract: Internet switches become harder and costlier to build for higher line rates and switch capacities. In-sequence delivery of packets has traditionally been a constraint on switch designs because TCP loss detection was considered vulnerable to out-of-sequence arrivals. For this reason, extremely efficient and simple designs, such as the Load Balanced Birkhoff-von Neumann Switch, were considered impractical. However, we reevaluate this constraint considering modern TCP implementations with loss detection algorithms like Recent Acknowledgment (RACK) that are more resilient to out-of-order arrivals. In a set of testbed experiments representative of wide area core

networks, we evaluated the performance of TCP flows traversing a load balanced switch that reorders some packets within a flow. We show that widely deployed and standard TCP implementations of the last decade achieve similar performance when traversing a load balanced switch as they do when there is no reordering. Furthermore, we also verified that an increase in the line rate leads to favorable conditions for time based loss detection methods, such as the one used in RACK. Our results, if further validated, suggest that switch designs that were previously thought to be unsuitable can potentially be utilized, thanks to the relaxation of the in-sequence delivery constraint.

Paper 2 - Multicast Service Chaining Model Guaranteeing Reliability with Multiple Sources

Authors: Shintaro Ozaki, Takehiro Sato and Eiji Oki (Kyoto University, Japan)

Abstract: Service chaining provides network services to users by processing packets with a series of virtualized network functions (VNFs). This paper proposes a multi-source multicast service chaining model that guarantees the reliability of services with flexible routing and VNF placement. In order to obtain cost-efficient feasible solutions, we introduce an algorithm that iteratively solves an integer linear programming problem and an algorithm that conducts the VNF placement with a concept of the betweenness centrality. Numerical results show that the proposed model allocates the network and computation resources with the reduction in the total cost compared to the existing model.

Paper 3 - Instant Queue Occupancy Used for Automatic Traffic Scheduling in Data Center Networks

Authors: Muhammad Shahid Iqbal (National Yang-Ming Chiao Tung University, Taiwan); Chien Chen (National Yang Ming Chiao Tung University, Taiwan)

Abstract: Datacenter applications desire low latency for short messages to provide a better user experience. The goal of datacenter networks is to minimize flow completion time (FCT), especially for short flows. Multiple scheduling disciplines have been proposed to minimize FCT for short flows. In this paper, we develop a Dynamic Longer Stay Less Priority (D-LSLP) which looks at the current queue occupancy to adjust the demotion threshold for the strict priority queues. Initially, D-LSLP considers every flow as a short flow, and with the passage of time, the flow is demoted to the next priority queue similar to MLFQ. We leverage the programmable nature of the P4 switches and use instant queue occupancy to automatically adjust the demotion threshold for the highest priority queue. This enables D-LSLP to increase or decrease its highest priority threshold based on the instantly available queue status. It enables short flows to complete in a couple of higher-priority queues, while large flows after remaining active for a certain amount of time in higher-priority queues are demoted. D-LSLP allows multiple traffic patterns to coexist without the need to manually tune the thresholds for different traffic patterns. Furthermore, it reduces the tail drop in the highest priority queue when a large number of short flows overwhelm the queue/link. The performance evaluation shows that it works well with different traffic patterns without operator intervention.

Paper 4 - Optimized SRv6 Multicasting for Network Assisted Publish-Subscribe Systems

Authors: Hyunseok Chang (Nokia Bell Labs, USA); Fang Hao (Bell Labs, Nokia, USA); Murali Kodialam (Nokia Bell Labs, USA); T. V Lakshman (Bell Labs, Nokia, USA); Sarit Mukherjee (Nokia Bell Labs, USA); Matteo Varvello (Nokia, Bell Labs, USA)

Abstract: In the new industrial Internet, a wide variety of industrial applications are expected to rely on highperformance data communication between a multitude of sensors and actuators that are deployed on a large scale. Publish-subscribe-based communication model is well-suited to handle such large-scale data gathering and dissemination among data sources and sinks. To support publish-subscribe-based data delivery, the newly standardized Segmented Routing over IPv6 (SRv6) can provide nondisruptive network programming primitives for building and maintaining network-efficient, shareable data distribution trees within the network. We study optimal algorithms for setting up different types of multicasting in the SRv6-capable network. In particular, we show, both theoretically and experimentally, that splitting multicast streams into multiple sub-streams, as well as using endto-end application-layer coding without any network participation can provide significant benefits in terms of multicast throughput compared to traditional single stream multicasting.

Paper 5 - Toward QoE-Based Routing Path Selection

Authors: Umakant Kulkarni and Yufeng Chen (Purdue University, USA); Patrick Melampy (Juniper Networks, USA); Sonia Fahmy (Purdue University, USA)

Abstract: The increasing popularity of video streaming and conferencing services have altered the nature of Internet traffic. In this paper, we take a first step toward quantifying the impact of this changing nature of traffic on the Quality of Experience (QoE) of popular video streaming and conferencing applications. We first analyze the traffic characteristics of these applications and of backbone links, and show how simple multipath routing may adversely impact application QoE. To mitigate this problem, we propose a new routing path selection approach, inspired by the TCP timeout computation algorithm, that uses both the average and variation of path load. Preliminary results show that this approach improves application QoE by on average 14% and packet latency by 11% for video streaming and conferencing applications, respectively.

Recess: 2:30 PM – 2:45 PM

Technical Session 3: 2:45 PM – 4:15 PM

Wireless Networks

Paper 1 - A Multi-Table Programmable Parser for Satellite Networks

Authors: Jin Zhang, Daoye Wang, Kai Liu and Jianhua Lu (Tsinghua University, China)

Abstract: A programmable packet parser with online configuration, which can identify protocols and extract key fields on demand at runtime, is essential to realize protocol upgrading for satellite networks. Furthermore, low complexity is urgently needed for the packet parser due to limited resources. However, traditional programmable packet parsers, with large storage redundancy, can only be configured offline. In this paper, we propose a Multi-Table Programmable Parser (MTPP), consisting of five tables and corresponding protocol independent fixed logic to represent a parser graph. Based on MTPP, online configuration is achieved by changing table entries at runtime. The numerical analysis under typical parse graphs demonstrates that MTPP uses less storage resource than that in PISA-based parser. The prototype on Xilinx FPGA shows that the response time of MTPP and the networks is 1.635μs and 3.537ms and the size of configuration information is 88.7% less than that of PISA-based parser.

Paper 2 - CTGAN-Assisted CNN for High-Resolution Wireless Channel Delay Estimation

Authors: Liyan Xu, Lei Feng and Li Wenjing (Beijing University of Posts and Telecommunications, China)

Abstract: The estimation accuracy of first-arrival-path (FAP) delay plays a vital role in positioning performance. We investigate the limitations of traditional cross-correlation (CC) algorithms in delay estimation. Our work proposes a FAP delay estimation mechanism using conditional tabular generative adversarial network (CTGAN) assisted convolutional neural network (CNN). The mechanism uses the CC algorithm to extract the delay feature in the wireless signal as input and finally outputs the FAP delay. For communication scenarios where it is difficult to obtain a large amount of training data, we use CTGAN to assist CNN training to improve the accuracy of FAP delay estimation. A series of simulation experiments were presented to evaluate the performance of CTGAN-assisted CNN and compare it with traditional high-resolution delay estimation algorithms. The results show that CNN performs well in weak LOS signals and dense multipath situations. It can still maintain high precision in the case of insufficient data.

Paper 3 - Experiment-Driven Platform for Link Quality Estimation in IEEE 802.11 WLANs

Authors: Thierry Arrabal (INSA Lyon, France); Marija Stojanova (ENS Paris, France); Isabelle Guérin Lassous (Université Claude Bernard Lyon 1 - LIP, France); Joris Picot (École Normale Supérieure de Lyon LIP Université de Lyon France, France)

Abstract: Experimental link quality assessment of recent Wi-Fi networks remains a challenge due to the rapid development of the Wi-Fi technology, the lack of availability of public datasets, and the difficulty to build such datasets. This paper addresses all three issues by first providing a publicly-available dataset using a custom-made Wi-Fi~5 experimental testbed. We then present an open-source framework for estimating the Frame Delivery Ratio (FDR) of a Wi-Fi link using the experimental data. The proposed solution relies on a small number of input features to build an estimation model of high accuracy, with an R2 coefficient of 0.89 and a mean absolute error of 0.06.

Paper 4 - Modelling the Performance of High Capacity Access Networks for the Benefit of End-Users and Public Policies

Authors: Antonio Capone and Maurizio Dècina (Politecnico di Milano, Italy); Aldo Milan and Marco Petracca (AGCOM, Italy)

Abstract: This paper addresses the challenge of modeling the performance of broadband access networks while maintaining technology-neutrality and accuracy in measurable quality. We discuss how public policies have been so far limited by models mainly based on the maximum nominal speed of the access networks, while the widespread use of measurement tools like "speed test" have influenced the perceived user quality. We present a performance modelling approach based on the extension of well-known traffic models that accurately characterizes the performance of broadband access networks. The model was validated with data from two network operators and used to guide choices and performance targets in Italian public policies for the development of broadband access networks.

Paper 5 - TSOA: Two-State Offloading Algorithm from Users to Co-Located Vehicular Microclouds

Authors: Bo-Jun Qiu and Jyh-Cheng Chen (National Yang Ming Chiao Tung University, Taiwan); Falko Dressler (TU Berlin, Germany)

Abstract: Offloading in edge computing scenarios is considered a prime solution to reduce computational time and also energy resources of the user equipment. This paper focuses on computation offloading from the user equipment to co-located vehicular microclouds. We derive the closed-form system metrics and cross-validate them with simulation. Through a comprehensive observation of vehicular microcloud behavior, the proposed two- state offloading algorithm (TSOA) provides the optimal offloading configuration in both planning and operating states. Finally, our evaluation demonstrates that the proposed TSOA performs optimally among the three offloading schemes.

Recess: 4:15 PM – 4:30 PM

Tutorial 4: 4:30 PM – 6:00 PM

Network Softwarization at the Edge With SD-WAN, Prof. Sebastian Troia and Prof. Guido Maier

Abstract

This tutorial addresses the Software-Defined Wide Area Network (SD-WAN) technology, which has recently conquered the enterprise-networking market all over the world. SD-WAN is regarded as very promising for the next-generation WANs, especially by the Communication Service Providers (CSPs)

as a new highly effective solution they can offer to their customers. SD-WAN brings the advantages of SDN to the WAN, applying the concept of separation among data and control plane. The main goal is to provide dynamic, fast and reliable interconnections between the sites of an organization, such as headquarters, data-centers, branch offices, that are geographically distributed over a wide area. A communication infrastructure with a national or international or even global extension can thus be provided to the tenants as an overlay network over heterogeneous public WANs. SD-WAN reduces the costs, but has to preserve the same quality of service of alternative, more expensive technologies, such as MPLS. We will present a detailed overview of SD-WAN by addressing the most important use cases, such as enterprise branch-to-headquarter and headquarter-to-data-center switching interconnection. In particular, we will focus on the network architecture requirements in order to obtain an agile and efficient control plane. Afterwards, we will describe the decision techniques that can be implemented inside the SD-WAN controller, making a comparison between traditional and Machine-Learning solutions. During the tutorial we will display some testbeds and present a live demo.

Dinner: 6:30 PM – 8:30 PM - - Announcement of Best Paper Award

Day 3: Wednesday, June 7, 2023

Registration: 8:00 AM – 9:00 AM

Keynote 2: 9:00 AM – 10:00 AM

Speaker: Ian F. Akyildiz
Topic: A New CubeSat Design with Reconfigurable Multi-band Radios for Dynamic Spectrum Access in Internet of Space Things

Abstract:
Small satellites, or CubeSats, are envisioned as a promising solution for future satellite communication networks because of their low costs and short deployment cycle. Currently, CubeSats communicate at conventionally allocated satellite communication frequencies. However, with the increase in the number of CubeSats, CubeSat-enabled communication systems, and many new use

cases, new spectrum bands and a more efficient spectrum usage are needed. In this talk, a novel CubeSat design with reconfigurable multi-band radios for communication in dynamic frequencies is proposed. The multi-band radio design is realized by two complementary approaches, namely, an electronics-based and a photonics-based approach. The multi-band communication covers a wide range from radio frequencies (2–30 GHz), millimeter wave (30–300 GHz), Terahertz band (up to 10 THz), and optical frequencies (with typical bands of 850 nm/350 THz, 1300 nm/230 THz, and 1550 nm/193 THz). A thorough link budget analysis is conducted to demonstrate the potential of the proposed multi-band architecture for space information networks. Key parameters in the satellite constellation design are investigated to explore the feasibility of deployment at different altitudes in the exosphere orbit (500 km and above). Furthermore, software-defined networking (SDN), and network function virtualization (NFV) have been incorporated to effectively separate the abstraction of functionalities from the hardware by decoupling the data forwarding plane from the control plane, such separation is of prime importance given the limited onboard processing on CubeSats. Additionally, key parameters in the constellation design including the coverage footprint and number of CubeSats as well as orbital planes, etc. are investigated for feasibility and deployment studies at different altitudes in the exosphere orbit.

Biography:

Ian F. Akyildiz received his BS, MS, and PhD degrees in Electrical and Computer Engineering from the University of Erlangen-Nürnberg, Germany, in 1978, 1981 and 1984, respectively. Currently, he is the President and CTO of the Truva Inc. since March 1989. He serves on the Advisory Board of the Technology Innovation Institute (TII) in Abu Dhabi, United Arab Emirates since June 1, 2020. He is the Ken Byers Chair Professor Emeritus in Telecommunications, Past Chair of the Telecom group at the ECE and the Director of the Broadband Wireless Networking Laboratory between (1985-2020) at the at the Georgia Institute of Technology. Dr. Akyildiz had many international affiliations during his career. He established many research centers in Spain, South Africa, Finland, Saudi Arabia, Germany, Russia, India, Cyprus, etc. He is the Founder and Editor in Chief of the newly established of the ITU (International Telecommunication Union) Journal on Future and Evolving Technologies (ITU-J FET) since August 2020, and is the Editor-in-Chief Emeritus of Computer Networks Journal (Elsevier) (1999-2019), the founding Editor-in-Chief Emeritus of the Ad Hoc Networks Journal (Elsevier) (2003-2019), the founding Editor-in-Chief Emeritus of the Physical Communication (PHYCOM) Journal (Elsevier) (2008-2017), and the founding Editor-in-Chief Emeritus of the Nano Communication Networks (NANOCOMNET) Journal (Elsevier) (2010-2017). He is an IEEE Fellow (1996) and ACM Fellow (1997) and received numerous awards from IEEE and ACM and other professional organizations, including Humboldt Award from Germany. His current research interests are in Metaverse, Extended Reality, 6G/7G Wireless Systems, TeraHertz Communication, Reconfigurable Intelligent Surfaces, and Underwater Communication. According to Google Scholar as of November, his h-index is 134 and the total number of citations to his papers is 138+K.

Recess: 10:00 AM – 10:30 AM

Technical Session 4: 10:30 AM – 12:00 PM

Machine Learning & Networks

Paper 1 - Network Traffic Prediction with Attention-Based Spatial-Temporal Graph Network
Authors: Yufei Peng, Yingya Guo, Hao Run and Junda Lin (Fuzhou University, China)

Abstract: Network traffic prediction plays a significant role in network management. Previous network traffic prediction methods mainly focus on the temporal relationship between network traffic, and used time series models to predict network traffic, ignoring the spatial information contained in traffic data. Therefore, the prediction accuracy is limited, especially in long-term prediction. To improve the prediction accuracy of the dynamic network traffic in the long term, we propose an Attention-based Spatial-Temporal Graph Network (ASTGN) model for network traffic prediction to better capture both the temporal and spatial relations between the network traffic. Specifically, in ASTGN, we exploit an encoder-decoder architecture, where the encoder encodes the input network traffic and the decoder outputs the predicted network traffic sequences, integrating the temporal and spatial information of the network traffic data through the Spatio-Temporal Embedding module. The experimental results demonstrate the superiority of our proposed method ASTGN in long-term prediction.

Paper 2 - LOBIN: In-Network Machine Learning for Limit Order Books
Authors: Xinpeng Hong, Changgang Zheng, Stefan Zohren and Noa Zilberman (University of Oxford, United Kingdom (Great Britain))

Abstract: Machine learning is driving the evolution of algorithmic trading, but the demands for fast execution speed remain. Although both aim for driving higher profitability, embedding more powerful machine learning approaches and lowering trading's latency are hard to achieve simultaneously. Offloading machine learning inference to programmable network devices, also referred to as in-network machine learning, provides a delicate balance between the two ends of this trade-off. In this paper, we present LOBIN, providing machine learning based market prediction by building limit order books within programmable switches, using high-frequency market data feeds. Compared with server-based solutions, LOBIN predicts future stock price movements with lower latency, higher throughput, and a minor impact on machine-learning performance.

Paper 3 - D3T: Double Deep Q-Network Decision Transformer for Service Function Chain Placement

Authors: Binghui Wu and Dongbo Chen (National University of Singapore, Singapore); Venkata Abhishek Nalam (Singapore Institute of Technology, Singapore); Mohan Gurusamy (National University of Singapore, Singapore)

Abstract: In this paper, we propose an effective algorithm D3T: Double Deep Q-Network Decision Transformer) to optimize the SFC placement. The algorithm is designed using a Decision Transformer (DT) that is assisted by a Double Deep Q-Network (DDQN). We employ a DDQN model as the baseline algorithm to generate offline training data. The trajectory data in the Experience Reply Memory of DDQN will be processed into sequences and modeled by a transformer. The algorithm's objective function considers end-to-end delay and rejection ratio as the objectives. Specifically, D3T combines the transformer with DRL to give an optimal solution. The results presented demonstrate the effectiveness of the proposed solution.

Paper 4 - FTG-Net: Hierarchical Flow-To-Traffic Graph Neural Network for DDoS Attack Detection (Online)

Authors: Luca Barsellotti (University of Modena and Reggio Emilia (UNIMORE), Italy); Lorenzo De Marinis (Scuola Superiore Sant'Anna, Italy); Filippo Cugini and Francesco Paolucci (CNIT, Italy)

Abstract: Distributed Denial of Service (DDoS) is one of the most common cyber-attacks and caused several damages in recent years. Such attacks can be executed either through the orches- tration of multiple devices that synchronously send requests or through specific patterns followed by a single device to force the victim to keep resources overrun. It becomes crucial to develop robust techniques to promptly detect those two kinds of DDoS attacks and mitigate their consequences. Most of the existing Machine Learning (ML) methods are based on flow and traffic information aggregations expressed in the form of independent vectors of statistical data, ignoring topological connections. Few recent solutions try to exploit the structural information of the network to improve the classification results. In particular, Graph Neural Network (GNN) based models can process traffic-level or flow-level relationships, represented as graphs, to detect malicious patterns. The objective of this paper is to combine the relationships at both the traffic-level and the flow-level by developing a two-level hierarchical graph representation and a GNN model able to process it, maximizing the information brought by the traffic structure and removing the necessity of stateful features. Experiments on the CIC-IDS2017 dataset show that the per- formances are comparable to the state-of-the-art solutions even using only the traffic structure. The code can be accessed at https://github.com/lucabarsellotti/FTG-Net.

Paper 5 - A Two-Stage Cooperative Reinforcement Learning Scheme for Energy-Aware Computational Offloading

Authors: Marios Avgeris (Carleton University, Canada); Meriem Mechennef and Aris Leivadeas (École de Technologie Supérieure, Canada); Ioannis Lambadaris (Carleton University, Canada)

Abstract: In the 5G/6G era of networking, computational offloading, i.e., the act of transferring resource-intensive computational tasks to separate external devices in the network proximity, constitutes a paradigm shift for mobile task execution on Edge Computing infrastructures. However, in order to provide firm Quality of Service (QoS) assurances for all the involved users, meticulous planning of the offloading decisions should be made, which potentially involves inter-site task transferring. In this paper, we consider a multi-user, multi-site Multi-Access Edge Computing (MEC) infrastructure, where mobile devices (MDs) can offload their tasks to the available edge sites (ESs). Our goal is to minimize end-to-end delay and energy consumption, which constitute the sum cost of the considered system, and comply with the MDs' application requirements. To this end, we introduce a two-stage Reinforcement Learning (RL)-based mechanism, where the MDs-to-ES task offloading and the ES-to-ES task transferring decisions are iteratively optimized. The proper operation, effectiveness and efficiency of our proposed offloading mechanism is assessed under various evaluation scenarios.

Lunch Break: 12:00 PM – 1:00 PM

Technical Session 5: 1:00 PM – 2:30 PM

Terrestrial and Aerial Mobile Networks

Paper 1 - Energy-Efficient Federated Learning in Internet of Drones Networks

Authors: Jingjing Yao (Texas Tech University, USA); Xiang Sun (University of New Mexico, USA)

Abstract: Internet of drones (IoD), where drones act as the Internet of things (IoT) devices, makes IoT networks much more flexible and responsive because of the high mobility of drones. Machine learning (ML) techniques can be applied in IoD to facilitate multiple applications such as object tracking and traffic surveillance, where ML data samples are collected and analyzed in the edge servers at the ground base station (BS). However, aggregating all data samples incurs huge wireless network traffic and potential data privacy leakage. Federated learning (FL) is then proposed to address these challenges by performing local training in drones and aggregating model parameters at the BS without sharing raw data samples. The FL performance in IoD networks is greatly affected by limited drone batteries which power FL local training, wireless data transmission, and drones' movements. This paper hence investigates the energy-efficient FL in IoD networks to optimize CPU frequencies of drones' on-board computing units such that the total energy consumption of all the drones in the FL process can be minimized, while satisfying the FL training time requirement. We formulate the problem as a non-linear programming problem and then design an algorithm with polynomial time complexity to derive the optimum solution. Extensive simulations are conducted to demonstrate the performance of our proposed algorithm.

Paper 2 - Self-Sovereign Identity Management for Hierarchical Federated Learning in Vehicular Networks

Authors: Engin Zeydan and Josep Mangues-Bafalluy (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Spain); Suayb S. Arslan (Massachusetts Institute of Technology & TUBITAK, USA); Yekta Turk (Aselsan, Turkey)

Abstract: There has been a rapid increase in the number of connected vehicles with a huge amount of data exchange between these vehicles that needs to be communicated, processed and analyzed reliably and efficiently. For secure and decentralized authentication, self-sovereign identity (SSI) management in vehicular networks have attracted attention in recent years. Hierarchical deployment frameworks, on the other hand, can provide secure and efficient knowledge sharing for vehicular networks with heterogeneous and geographically distributed vehicles and infrastructure in 6G networks. In this paper, we explore the joint use of hierarchical federated learning, as a collaborative machine learning framework, and hierarchical SSI management in vehicular networks, highlighting its advantages, limitations. At the end of the paper, we also provide two illustrative use cases.

Paper 3 - DOSM: Demand-Prediction Based Online Service Management for Vehicular Edge Computing Networks

Authors: Anum Talpur and Mohan Gurusamy (National University of Singapore, Singapore)

Abstract: In this work, we investigate an online service management problem in vehicular edge computing networks. To satisfy the varying service demands of mobile vehicles, a service management framework is required to make decisions on the service lifecycle to maintain good network performance. We describe the service lifecycle consists of creating an instance of a given service (scale-out), moving an instance to a different edge node (migration), and/or termination of an underutilized instance (scale-in). In this paper, we propose an efficient online algorithm to perform service management in each time slot, where performance quality in the current time slot, the service demand in future time slots, and the minimal observed delay by vehicles and the minimal migration delay are considered while making the decisions on service lifecycle. Here, the future service demand is computed from a gated recurrent unit (GRU)-based prediction model, and the network performance quality is estimated using a deep reinforcement learning (DRL) model which has the ability to interact with the vehicular environment in real-time. The choice of optimal edge location to deploy a service instance at different times is based on our proposed optimization formulations. Simulation experiments using real-world vehicle trajectories are carried out to evaluate the performance of our proposed demand-prediction based online service management (DOSM) framework against different state-of-the-art solutions using several performance metrics.

Paper 4 - Urban Air Mobility: Vision, Challenges and Opportunities

Authors: Debjyoti Sengupta and Sajal K. Das (Missouri University of Science and Technology, USA)

Abstract: Urban Air Mobility (UAM) involving a large number of aerial vehicles, piloted or autonomous, is envisioned as an emerging disruptive technology for next generation smart transportation, particularly addressing the mobility challenges in congested cities. Broadly, UAM can include all types of aircrafts ranging from small scale unmanned aerial vehicles (UAVs) or drones to aircrafts with passenger carrying capacity, such as personal air vehicles (PAVs). This paper highlights the UAM vision along with the underlying research challenges and opportunities in computing, networking, and services for sustainable design and implementation of such an innovative infrastructure. Important research questions include, but are not limited to, real-time autonomous scheduling, efficient route planning, aerial-to-ground and inter-vehicle communications, airspace traffic management, on-demand air mobility and resource management, quality of service and quality of experience, sensing analytics and machine learning for trustworthy decisions, optimization of operational services, and socio-economic impacts of UAM infrastructure on sustainability. This paper attempts to provide a vision with stimulating introduction and insights of this promising technology, and spark imagination to explore fundamental computing and networking challenges related to UAM networks and services.

Paper 5 - Autonomous Electric Vehicles as Mobile Green Energy Sources

Authors: Xin Liu (University at Buffalo, USA); Yangming Zhao (University of Science and Technology of China, China); Adel Sadek (State University of New York at Buffalo, USA); Chunming Qiao (University at Buffalo, USA)

Abstract: We envision a wide deployment of battery-operated edge devices such as data kiosks, to provide pervasive data collection and dissemination services. Although these data kiosks can be powered by green energy sources, their batteries may deplete overtime, and have to recharged frequently to prevent power outages and potential loss of critical data. In this paper, we propose an edge device recharging system using autonomous electric vehicles as Mobile Chargers (MCs). First, we determine the numbers and locations of Dispatching Centers (DCs) for these MCs, each of which will be responsible for recharging some edge devices in surrounding area. Then, we plan optimal routes of the MCs in order to minimize the number of MCs needed to recharge all edge devices before a deadline. To reduce the time complexity involved in optimizing the routes, we cluster the edge devices and propose efficient algorithms to plan the route of MCs for each cluster based on the relaxation and rounding of a Mixed Integer Linear Programming (MILP) model. Extensive simulations show that our approach can reduce the number of MCs required to recharge all edge devices before a deadline by up to 71.93% compared with greedy-based heuristic algorithms.

Recess: 2:30 PM – 2:45 PM

Tutorial 5: 2:45 PM – 5:15 PM

6G Network Slicing Technics: Recent Advances, Standards, and Challenges, Prof. Jiadai Wang and Prof. Jiajia Liu (Online)

Abstract

6G is designed to achieve breakthroughs in mobile networks, and is expected to have features such as seamless global coverage, full virtualization, and ubiquitous intelligence. More application scenarios with specific capabilities will emerge, which urgently require customized end-to-end service provisioning. Network slicing is widely recognized as a key enabling technology for 6G, which can divide the physical network into multiple logical networks as needed, changing the network form from “one-size-fits-all” to “one-size-per-service” to meet differentiated performance metrics and service requirements. Although the network slicing concept has been explored to some extent in 5G, it will be fully extended and perfected in 6G with the combination of ubiquitous intelligence and various innovative application scenarios. This tutorial comprehensively reviews the research work on 6G network slicing technics from three aspects: recent advances, standards, and challenges. Slicing technics are the basis of end-to-end logical isolation and service provisioning, which involves many technical domains such as access network, transport network, core network, and slicing management system. We first introduce key recent advances in slicing technics according to these different domains. In particular, we emphasize the great potential of artificial intelligence techniques for network slicing. In addition, cross-domain and cross-vendor standards can provide guidance for slicing technics, which is necessary for the wide application and commercialization of 6G network slicing. Finally, we highlight the challenges faced by network slicing and envision its future development.

Closing Remarks: 5:15 PM – 5:30 PM