In-Network Computing: Emerging Trends for the Edge-Cloud Continuum

Deze Zeng Nirwan Ansari Marie-José Montpetit Eve M. Schooler Daniele Tarchi

Recent advances in virtualization technologies (e.g., unikernels, containers, and virtual machines), the move to data driven approaches and softwarized networking technologies (e.g., SDN, NFV, and data plane programming) have invigorated a new focus on combining computation and communication in distributed systems. At the same time, the proliferation of edge computing, a complementary and at times alternative solution to centralized cloud computing, has resulted in an edge-cloud continuum of applications and services. Edge computing offers more proximate resources closer to where the service is needed and supports emerging applications (e.g., self-driving vehicles, autonomous systems, and VR/AR) with stringent requirements of fast response, low delay, high bandwidth, trust-sensitivity, and/or continued operation, despite intermittent or incomplete connectivity. These trends have converged into the concept of Computing in the Network (COIN), the integration of in-network computation and network processing in a common framework. COIN naturally fits within the edge-cloud continuum, where expanded resource distribution and tightly integrated computing-networking capabilities are provisioned from the edge to the cloud infrastructure including at points in between.

Adding computation in the network nodes, beyond packet forwarding, enables a more collaborative processing, better capability discovery and improved integration of communication and computation resource management and allocation to respond to diverse application needs in an increasingly data driven environment. The network provides more than a simple connection and can be considered as an essential constituent of a distributed application. COIN melts the boundary between the networking domain and the computing domain. The network behavior can be dynamically adapted to the required task computation requirement and scheduling, following specific requirements such as application level features, data discovery, operator policies, user preferences, traffic engineering and congestion avoidance, data availability, and geo-location.

COIN is still an emerging trend and many open issues are yet to be addressed in order to understand how the edge-cloud continuum should evolve to accommodate the challenges of an increasingly computerized and connected world. The joint optimization of communication and computation requires further investigation beyond the focus on high-level resource abstraction and header-level processing. Current enabling technologies such as data plane programming when combined with unikernel, containers, NFV, and SDN, warrant further investigations into novel services including the use of artificial intelligence in networking. An edge-cloud application may consist of a large number of container-based microservices with short millisecond lifetime for load balancing but requiring a set of common shared virtualized network functions (VNFs) with comparatively long lifetime for management; this implies the need for both edge and cloud services to collaborate. Up to now, the networking domain and computing domain have been developed independently, thus making this collaboration difficult. Fine-grain optimization needs to address the mismatch between end-to-end semantics and computing-aware hop-by-hop decision making based on local execution of service-specific functions or telemetry results. It is also required to abstract both the transport and the remote function invocation, and to provide uniform interfaces for transparent orchestration of services and applications.

This special issue mainly targets emerging directions in COIN, especially edge-cloud challenges to its deployment. We received 37 high quality submissions, of which 11 articles were eventually accepted after a rigorous review process. These articles cover several different aspects of COIN, including architecture design, performance analysis, system telemetry, resource allocation and task scheduling, and application implementation. A brief description of each of the articles follows.

In the first article "Revisiting Network Telemetry in COIN: A Case for Runtime Programmability", the authors realized that COIN based applications require flexible network telemetry data to drive efficient resource allocation decisions. They exploited programmable switches to process queries using in-network resources. They proposed a reactive runtime query scheduling framework to address the core challenges such as dynamic traffic loads, diverse and dynamic queries, and network-wide execution. They implemented their solution and reported the evaluation results, demonstrating that dynamically scheduling operations in programmable switches indeed improve the scalability of network-wide telemetry with respect to dynamic traffic and query loads.

In the article "Computing Meets Network: COIN-Aware Offloading for Data-Intensive Blind Source Separation", the authors applied COIN to an acoustic source separation application. In order to exploit spare computing power on distributed network nodes, they introduced a Network Joint Independent Component Analysis (NJICA) framework. NJICA first decomposes monolithic applications into a set of distributed components that can be offloaded to distributed network nodes. In addition, they also proposed an ahead-of-time data extraction strategy and a message-based transport layer for message aggregation and differentiation. They also implemented NJICA and evaluated its performance efficiency via real workloads.

In "Edge-Based Virtual Reality over 6G Terahertz Channels", the authors applied stochastic network calculus to analyze the end-to-end delay for edge computing based VR applications over 6G Terahertz channels. The computing node is modeled as a tandem system consisting of a processing and a transmission subsystems, based on which the VR service's reliability and delay are analyzed and evaluated. This article provides theoretical performance analysis framework for a critical COIN application.

"Empowering Edge Intelligence by Air-Ground Integrated Federated Learning" also uses 6G networks but focuses on edge intelligence. In order to implement ubiquitous intelligence locally in 6G networks, the authors proposed a framework called Air-Ground Integrated Federated Learning, which leverages unmanned aerial vehicles (UAVs), balloons, and airships with flying base stations to cooperatively train a common model using local data via federated learning.

UAVs are also central to the next article, "Intelligent Task Offloading and Energy Allocation in the UAV-Aided Mobile Edge-Cloud Continuum". Here, the authors considered an edge-cloud continuum consisting of UAVs, which serve as mobile edge nodes to execute tasks offloaded from the ground IoT devices. UAVs are powered via laser-based wireless power transfer. A joint task offloading and energy allocation problem with consideration of task diversity and environment dynamics using an algorithm based on federated deep reinforcement learning is investigated.

In the article "Edge In-Network Computing Meets Blockchain: A Multi-Domain Heterogeneous Resource Trust Management Architecture", the authors addressed open challenges in resource sharing, scheduling and service credibility in edge computing empowered by in-network computing. In this architecture, the edge-side heterogeneous resource trust management in multi-domain scenarios is studied. They proposed a blockchain-based multi-domain heterogeneous resource trust management architecture, which combines a smart contract-based incentive mechanism and a differential evolution based microservice orchestration strategy for edge-side resource sharing. The architecture has been applied to a smart grid use case to demonstrate its efficiency on service delay reduction.

Another blockchain related contribution, "A Blockchain Enabled SLA Compliance for Crowdsourced Edge based Network Function Virtualization", uses crowdsourced edge devices to provision edge resources and at the same time addresses issues with respect to service availability, service quality and data privacy. The authors explored how blockchain can be used to design a framework for monitoring service-level agreement compliance of edge devices in offering VNF services.

The softwarization of networking is investigated in "A Software-Defined Networking Enabled Approach for Edge-Cloud Computing in Internet of Things". In this article, the authors proposed to use edge-cloud orchestrated computing and SDN to guarantee secure and intelligent services for Internet-of-Things (IoT) applications. The control plane and data plane are decoupled using SDN mechanisms to enable the management of the IoTs through the centralized controller. The SDN-enabled architecture leverages blockchain as an immutable metadata store to ensure network ownership and authenticity. Reinforcement learning is applied to schedule network resources and improve resource utilization in IoT.

Artificial Intelligence and its impact on edge computing and networking are delineated in "Intelligence-Empowered Mobile Edge Computing: Framework, Issues, Implementation, and Outlook". The authors aimed to provide a comprehensive survey on edge intelligence (EI). Instead of solely relying on the cloud, EI can sink some of the cloud's processing capabilities into the edge side in order to provide real-time response while enabling more intelligent services with high performance. In addition, the data leakage and loss of users' private information can also explicitly be alleviated by design. The article notably presents an overview of EI based on existing literature. It finally discusses a typical application scenario as well as its specific embodiment in EI to shed light on some potential challenges to the transition of EI from theory to practice.

EI and AI oriented research is also the focus of "An Adaptive Neural Architecture Search Design for Collaborative Edge-Cloud Computing". The widespread of EI asks for the adoption of deep neural network at the edge. A new network architecture paradigm called neural architecture search (NAS) surveys and assesses different network architectures to determine which high-quality architecture meets a design goal. The authors proposed an adaptive NAS in the collaborative edge-cloud scenario. There, the central cloud is responsible for searching architectures with the proxy dataset, and the edge is used for training the resulting architecture with a target dataset.

Finally in the article "A Cross-Domain Augmentation-Based AI Learning Framework for In-Network Gesture Recognition", the authors presented an in-network gesture recognition framework called CAL, which trains an AI-model to enable RFID-based gesture recognition. To achieve time-efficient recognition, knowledge distillation is applied to obtain a light and accurate model, which is deployed at the edge side to shorten the recognition latency.

All these articles provide an overview of the COIN landscape and elicit the promise of the approach in the future of networking. We hope that this special issue will foster further research in the COIN field to improve system performance and availability and to expedite the development of highly efficient data-driven systems in networking, IoT and end-user applications.

Biographies

Deze Zeng [M'14] is a full professor in the School of Computer Science, China University of Geosciences, Wuhan, China. His current research interests mainly focus on edge computing, cloud computing, and future networking technologies. He has authored three books and over 100 papers in refereed journals and conferences in these areas. He serves on the editorial boards of *IEEE Transactions on Sustainable Computing*, *Journal of Network and Computer Applications*, *Frontiers of Computer Sciences*, and *IEEE Open Journal of the Computer Society*.

Nirwan Ansari [S'78, M'83, SM'94, F'09] is a distinguished professor of electrical and computer engineering at the New Jersey Institute of Technology (NJIT) and is also a Fellow of the National Academy of Inventors. His current research focuses on green communications and networking, cloud computing, drone-assisted networking, and various aspects of broadband networks. Some of his recognitions include several excellence in teaching awards, a few best paper awards, NCE Excellence in Research Award, several ComSoc TC technical recognition awards, NJ Inventors Hall of Fame Inventor of the Year Award, Thomas Alva Edison Patent Award, Purdue University Outstanding Electrical and Computer Engineering Award, NCE 100 Medal, and NJIT Excellence in Research Prize and Medal. He has also been granted more than 40 U.S. patents.

Marie-José Montpetit [SM'98] is an affiliate professor at Concordia University in Montreal, Canada and the co-chair of the Computing in the Network Research Group (COINRG) in the Internet Research Task Force (IRTF). She also leads the digital technology group in a Montreal startup in distributed data driven systems. She was previously a researcher in MIT dealing with video and network performance. Her research interests are mainly in IoT, edge and fog networking, and distributed systems resource allocation techniques.

Eve M. Schooler [M'83, SM'20, F'21] is a principal engineer and Director of Emerging IoT Networks at Intel, where her current work focuses on developing architectural building blocks to evolve the Internet toward a sustainable edge-cloud infrastructure. She serves in leadership roles in the IETF standards organization, to enable deterministic networks, the Internet of Things, and COIN. She has authored over 60 publications and is an inventor on 30+ granted or pending patents. She is an IEEE Fellow and the co-recipient of the 2020 IEEE Internet Award for her work on control protocols for Internet telephony and multimedia teleconferencing.

Daniele Tarchi [S'99, M'05, SM'12] is an associate professor at the University of Bologna, Italy. He is the author of approximately 120 papers, among which almost 50 are in international journals. His research interests are mainly focused on wireless communications and networking, resource allocation techniques, Fog and Edge Computing solutions. He has been an IEEE Senior Member since 2012. He is an Editorial Board member for *IEEE Wireless Communications Letters*, *IEEE Transactions on Vehicular Technology* and *IET Communications*. He has been a symposium co-chair for IEEE WCNC 2011, IEEE Globecom 2014, IEEE Globecom 2018 and IEEE ICC 2020, and a workshop co-chair at IEEE ICC 2015.