Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Editorial:

Coordination of networking and computing: toward new information infrastructure and new services mode

Xiaoyun WANG[†1], Tao SUN[†‡2], Yong CUI[†3], Rajkumar BUYYA[4], Deke GUO[5], Qun HUANG[6], Hassnaa MOUSTAFA[7], Chen TIAN[8], Shangguang WANG[9]

[1]*China Mobile Communications Corporation, Beijing 100032, China*
[2]*China Mobile Research Institute, Beijing 100053, China*
[3]*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*
[4]*CLOUDS Lab, School of Computing and Information Systems, University of Melbourne, Melbourne, VIC 3010, Australia*
[5]*National Key Laboratory of Information Systems Engineering, National University of Defense Technology, Changsha 410073, China*
[6]*School of Computer Science, Peking University, Beijing 100871, China*
[7]*Intel Corporation, Santa Clara, CA 95054, USA*
[8]*State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China*
[9]*State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[†]E-mail: wangxiaoyun@chinamobile.com; suntao@chinamobile.com; cuiyong@tsinghua.edu.cn

https://doi.org/10.1631/FITEE.2430000

Computing, serving as the cornerstone of information processing, plays a pivotal role in the digital service era. The "network" and "computing," responsible for information transmission and processing respectively, traditionally belong to different stakeholders and have evolved separately. However, the recent trend toward the coordination and integration of computing and networks has garnered significant attention from both industry and academia. Concepts such as "computility network" and "computing force network" have emerged, and the International Telecommunication Union - Telecommunication Standardization (ITU-T) has initiated efforts to develop standards for the coordination of networking and computing (CNC), focusing on the architecture and framework. The coordination of computing and networks is poised to offer several immediate benefits:

1. It enables efficient resource scheduling over wide areas, leading to energy conservation. The "East Data for West Computing" national project launched by China in 2022 exemplifies this, promoting data processing and storage in the country's western regions where there is an abundance of energy, particularly renewable energy.

2. It facilitates the realization of computing power as a utility service, a vision articulated by artificial intelligence (AI) pioneer John McCarthy in the 1960s. The aim is to make computing power as readily available as household water and electricity, bridging the accessibility gap. This concept, termed "computility" by Dr. Ninghui SUN and his colleagues, is predicated on the network's ability to provide ubiquitous information processing capabilities.

3. It significantly improves user experiences. For instance, immersive communication, characterized by high bandwidth, low latency, frequent interaction, and substantial computing complexity, necessitates that both data transmission and processing should be considered in the assessment of quality of experience. The synergy between connectivity and computing allows for optimal trade-offs, such as the selection of

‡ Corresponding author
ORCID: Tao SUN, https://orcid.org/0009-0003-3491-8813
© Zhejiang University Press 2024

suitable computing nodes and network transmission paths.

This special feature, including six papers—one perspective paper, one review paper, and four research articles, revisits the longstanding question of whether it is feasible to design the system as a networked supercomputer. Managing computing resources across end devices, edge clouds, and central clouds while optimizing trade-offs for end-to-end services is essential. This approach enables the exchange of computation for transmission, where enhanced computing power can compensate for the need for ultra-low latency in network transmission. Through strategic collaboration and integration, the network and computing can synergistically amplify their capabilities. To harness the aforementioned benefits, it is crucial to address three key aspects:

1. Identification of networking and computing requirements and resources involves determining the computing power and network transmission capabilities needed by applications. It is essential to understand how to identify and model these demands, especially for heterogeneous and distributed computing resources. The challenge lies in ensuring that raw data or processing logic can be efficiently allocated to use available computing resources.

2. Joint and timely awareness of computing and networking resources necessitates the collection of load information related to network or computing capacities. Due to differing monitoring approaches, computing power metrics often derive from information technology infrastructure components such as cloud services, virtual machines, or containers. In contrast, network connectivity metrics typically depend on end-to-end data path transmission measurements.

3. Coordinated scheduling of computing and networking resources requires systematic optimization to facilitate collaboration among end devices, edge computing, and cloud services. This involves establishing a unified “brain” for the network to enable a distributed computing system where resource providers collaboratively tackle multidimensional scheduling challenges.

In addition to these core aspects, the practical implementation of such systems may involve tasks such as application task decomposition. Furthermore, it is vital to assess whether network and computing resources can collaborate in the entire information lifecycle, including generation, transmission, processing, and consumption. In the context of digital transformation, CNC heralds a promising new research domain. Numerous challenges, as outlined above, await discovery and resolution.

The architectural design of the system is pivotal in realizing these new functionalities. Xiaoyun WANG et al. introduced the concept of computing-aware network (CAN), a novel framework that incorporates an awareness plane to facilitate wide-area computing and network coordination. This framework identifies three pivotal technologies: computing-aware traffic steering (CATS), elastic broadcast, and wide-area high-throughput transmission. The Internet Engineering Task Force (IETF) has initiated discussions on CATS to explore potential scenarios and architectural designs, although the application of computing information in routing requires further investigation. Elastic broadcast is tailored for one-to-many communication in wide area networks (WANs), essential for AI model training and inference across data centers. While high-throughput transmission is not a novel concept, its application in WANs poses challenges due to long-distance delays, packet loss, and server limitations.

Energy efficiency is paramount due to the environmental and cost implications of high power consumption. Federated learning, increasingly popular in recent years, requires energy-efficient approaches, especially in edge computing contexts. Kang YAN et al. provided a comprehensive survey on energy-efficient strategies in federated learning at the edge, including system and energy consumption models. They explored three categories of strategies, including learning-based, resource allocation, and client selection, and discussed several potential research directions for energy-efficient federated learning.

To effectively provide computing services or solve complex problems, a collaborative approach to task decomposition is necessary. This includes task offloading among user devices, edge networks, and cloud data centers, fostering collaboration across cloud–edge–device infrastructures. Xiaojun BAI et al. differentiated tasks into delay-sensitive and delay-tolerant categories, employing a continuous-time Markov chain to model the system. Their research highlighted a

trade-off between average delay and blocking rate, with efforts focused on optimizing the access threshold for edge network buffers.

When dealing with multiple computing and network resource providers, selecting the appropriate provider is crucial for task completion. Yuexia FU et al. used a reputation model based on the beta distribution function to evaluate the credibility of resource providers and introduced a performance-based reputation update model. This approach was formulated as a constrained multi-objective optimization problem, with feasible solutions identified through a modified fast and elitist non-dominated sorting genetic algorithm. Their extensive simulations confirmed the validity and effectiveness of their model in enhancing user satisfaction and resource utilization.

The essence of CNC lies in the joint consideration of computing and network resources for resource allocation. Xueying HAN et al. developed an intelligent resource allocation method that integrates deep reinforcement learning with graph neural networks. This method addresses the challenge of routing in a computing force network by optimizing both network and computing resources across various network topologies, even with structural changes.

In various scenarios and fields, the synergy between computing power and networking is vital. Yizhuo CAI et al. focused on enhancing communication efficiency of federated learning in six-generation networks, examining both traditional and peer-to-peer federated learning architectures. They demonstrated that optimizing computing power scheduling based on a real-time resource status can significantly improve the performance.

The papers in this special feature cover various aspects of CNC, whether in Internet Protocol (IP) or cellular networks like the fifth-generation mobile communication technologies. This feature has garnered substantial support from both academic and industry sectors but remains in its infancy. Many questions and technical directions warrant further detailed investigation.

We extend our heartfelt gratitude to all contributors, reviewers, and the journal's editors, for their support and contributions.

Xiaoyun WANG is a chief scientist at China Mobile. She was a recipient of multiple National Science and Technology Progress Awards, the National Innovation and Excellence Award, and the Chinese Youth Science and Technology Award. Her research interests include technology strategy, system architecture, and networking technology.

Tao SUN received his BS degree in automation in 2003 and PhD degree in control science and engineering in 2008, both from Tsinghua University, China. He is a chief expert of China Mobile. He has more than 10 years of experience on mobile network architecture design and IP technology research and standardization, and was a vice chair of 3GPP SA2 (system architecture). His research interests include 6G architecture, IP network evolution, and coordination of computing and network.

Yong CUI received his BE and PhD degrees in computer science and engineering from Tsinghua University in 1999 and 2004, respectively. He is a full professor with the Department of Computer Science and Technology at Tsinghua University. He has published over 100 papers in refereed conferences and journals, earning several best paper awards. He has co-authored seven Internet standard documents (RFCs) related to IPv6 technologies. He has served on the editorial boards of *IEEE TPDS*, *IEEE TCC*, and *IEEE Int Comput*, and is currently a working group co-chair in the IETF. His main research interests include mobile cloud computing and network architecture.

Rajkumar BUYYA is a Redmond Barry Distinguished Professor and Director of the Cloud Computing and Distributed Systems (CLOUDS) Laboratory at the University of Melbourne, Australia. He is the author or co-author of over 800 publications, including seven textbooks. He is among the most cited authors in the fields of computer science and software engineering worldwide, with an *h*-index of 166, a *g*-index of 360, and more than 146 600 citations as of 2023. His research interests include cloud computing, distributed systems, service-oriented computing, and energy efficient computing.

Deke GUO received his BE degree from the Department of Industrial Engineering at Beijing University of Aeronautics and Astronautics, China and his PhD degree from the School of Information System and Management at the National University of Defense Technology (NUDT), China. He is currently a professor with the College of System Engineering, NUDT. His research interests include distributed systems, software-defined networking, data center networking, wireless and mobile systems, and interconnection networks.

Qun HUANG received his BS degree in computer science from Peking University (PKU), China in 2011 and his PhD degree from The Chinese University of Hong Kong, China in 2015. He is currently an assistant professor with the School of Computer Science at PKU. Prior to the current position, he was affiliated with the Institute of Computing Technology, Chinese Academy of Sciences, from Sept. 2017 to May 2020, and with Huawei from Sept. 2015 to Sept. 2017. His research interests include network measurement, networking systems, and distributed systems.

Hassnaa MOUSTAFA obtained her MS degree in distributed systems from University of Paris XI and PhD degree in wireless and mobile networks from Telecom Paris Tech. She is a principal engineer at Intel Corporation, focusing on edge computing and AI solutions for IoT segments and network edge infrastructure. In her prior roles at Intel, she spearheaded car-to-cloud solutions for connected and autonomous vehicles, along with connectivity technologies in various IoT domains. Before her tenure at Intel, she served as a senior R&D engineer at Orange in France, where she was instrumental in developing wireless network solutions for the EMEA region and led projects aimed at optimizing video and multimedia services over wireless networks. She has over 80 publications and over 300 filed patents, of which over 100 have been granted. Her research interests include edge computing, edge AI for converged network and IoT services, AI for the network and edge infrastructure, and media and IoT services and protocols.

Chen TIAN obtained his BS, MS, and PhD degrees from the Department of Electronics and Information Engineering at Huazhong University of Science and Technology (HUST), China, in 2000, 2003, and 2008, respectively. From 2012 to 2013, he was a postdoctoral researcher with the Department of Computer Science at Yale University. From 2013 to 2016 he was an associate professor with the School of Electronics Information and Communications at HUST. Currently, he is a professor with the State Key Laboratory for Novel Software Technology, Nanjing University, China. His research interests include data center networks, network function virtualization, distributed systems, Internet streaming, and urban computing.

Shangguang WANG received his PhD degree from Beijing University of Posts and Telecommunications (BUPT), China, in 2011. He is currently a professor with the School of Computer Science, BUPT. He is serving as chair of the IEEE Technical Community on Services Computing and a vice chair of the IEEE Technical Community on Cloud Computing. He has served or is serving as a general chair or program chair of more than 10 IEEE conferences, an advisor or associate editor of several journals including *IEEE Trans Serv Comput*, *J Cloud Comput*, *J Softw Pract Exp*, and *Int J Web Grid Serv.* He is a fellow of IET and a senior member of IEEE. His research interests include service computing, edge computing, and satellite computing.