Guest Editorial
Multi-Tier Computing for Next Generation Wireless Networks—Part II

Kunlun Wang, *Member, IEEE*, Yang Yang, *Fellow, IEEE*, Jiong Jin, *Member, IEEE*, Tao Zhang, *Fellow, IEEE*, Arumugam Nallanathan, *Fellow, IEEE*, Chintha Tellambura, *Fellow, IEEE*, and Bijan Jabbari, *Life Fellow, IEEE*

I. Introduction

MULTI-TIER computing effectively enables flexible computation and communication resource sharing by offloading computation-intensive tasks to nearby servers along the cloud-to-thing continuum. In essence, multi-tier computing networks can distribute computing, storage, and communication functions anywhere between the cloud and the endpoint to take full advantage of the resources available along this continuum, thus extending the traditional cloud computing architecture to the edge of the network. With multi-tier computing, some application component processing, such as delay-sensitive components, can take place at the edge of the network, while other components, such as time-tolerant and computation-intensive components, can be performed in the cloud. To best meet user requirements, centralized cloud computing with extensive resources, secure environments, and powerful algorithms is still needed, but also must be complemented by distributed fog and edge computing with shared resources, accessible environments, and simple algorithms for real-time decision-making. Given heterogeneous computing resources and collaborative service architectures, future multi-tier computing networks will be capable of supporting a full range of computing and networking services for different environments and applications. Multi-tier computing enables low-latency processing by allowing data to be processed at the network edge close to end devices. It also facilitates the distribution of fog/edge nodes to collect data from end devices. Therefore, multi-tier computing effectively complements the cloud computing architecture.

Kunlun Wang is with the Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China, and also with the School of Communication and Electronic Engineering, East China Normal University, Shanghai 200241, China (e-mail: klwang@cee.ecnu.edu.cn).

Yang Yang is with Terminus Group, Beijing 100027, China, also with the Peng Cheng Laboratory, Shenzhen 518055, China, and also with Shenzhen Smart City Technology Development Group Company Ltd., Shenzhen 518046, China (e-mail: dr.yangyang@terminusgroup.com).

Jiong Jin is with the School of Science, Computing and Engineering Technologies, Swinburne University of Technology, Melbourne, VIC 3122, Australia (e-mail: jiongjin@swin.edu.au).

Tao Zhang is with the U.S. National Institute of Standards and Technology (NIST), Gaithersburg, MD 20878 USA (e-mail: taozhang1@yahoo.com).

Arumugam Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K. (e-mail: a.nallanathan@qmul.ac.uk).

Chintha Tellambura is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2W3, Canada (e-mail: ct4@ualberta.ca).

Bijan Jabbari is with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030 USA (e-mail: bjabbari@gmu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2022.3228094.

Digital Object Identifier 10.1109/JSAC.2022.3228094

This Special Issue led to a strong response and has attracted over 100 high-quality submissions from researchers around the world. This is a testament to the widespread interest in multi-tier computing research. Based on a rigorous review process, a total of 34 submissions were selected for publication in double issues. Every submission received at least three reviews, and each accepted paper went through at least one revision round.

In this guest editorial for the second part of double issues, we briefly review the research featured in the issue and highlight emerging research themes for multi-tier computing. The contributions of the papers included are categorized as follows.

II. Multi-Tier Computing for Next-Generation Wireless Networks

A. Massive Multiple Access Schemes for Multi-Tier Computing

In [A1], Wang et al. propose a non-orthogonal multiple access (NOMA)-aided joint communication, sensing, and multi-tier computing (JCSMC) framework. In this framework, a multi-functional base station (BS) carries out target sensing, while providing multi-tier computing services for nearby users.

In [A2], Shao et al. propose a Bayesian approach by letting multiple edge devices transmit two pieces of statistical information to the center device such that Bayesian estimators can be devised to tackle the misalignments of over-the-air computation, which enables efficient function computation in multi-tier computing systems.

In [A3], Zhao et al. propose a multi-tier-enabled computing framework and the associated algorithms for URLLC to support multiple end users with different QoS requirements. This technique concentrates on K-repetition Grant-Free (GF) access in light of its simplicity and well-balanced performance for practical systems.

B. Multi-Tier Computing Meets Semantic Communication

In [A4], Qin et al. derive the closed-form expressions of timeliness of information for multi-tier computing at both edge tier and fog tier, where two-stage tandem queues are exploited to abstract the task transmission and computation process. The proposed technique exploits the statistical structure of the Gauss-Markov process and derives the closed-form expression for task process-related timeliness of information

In [A5], Xiao et al. propose a novel reasoning-based implicit semantic-aware communication network, which allows multi-tier of cloud data center (CDC) and edge servers to cooperate. This technique supports efficient semantic encoding, decoding, and interpretation for end-users.

C. Multi-Tier Computing With Processing Uncertainty

In [A6], the authors formulate a Chance-Constrained Program (CCP) that guarantees the task processing delay requirement under the condition that the number of required processing cycles of a task is unknown beforehand and only known until its completion. This technique can minimize task computation energy consumption for the users while meeting the probabilistic task deadlines.

In [A7], Kim et al. propose a multi-tier computing architecture that provides a rigorous theoretical framework and studies the challenges of long-term user mobility to minimize both the service provider cost and user latency. As devices move, the proposed approach can offload the tasks to different edge servers and migrate data or state information from one edge server to another.

D. Caching for Multi-Tier Computing

In [A8], the authors investigate the issue of recommendation-enabled edge caching in multi-tier (edge-cloud) computing networks. The proposed technique integrates recommender systems and data caching to support both direct hits and soft hits. The proposed algorithm can improve the resource utilization of edge servers and alleviate the effects of soft hits.

In [A9], Tang and Lei study a multi-tier cache-aided relaying network, where the destination is randomly located in the network and it requests files from the source through the help of multi-tier cache-aided BS and relays. The proposed algorithm optimizes the caching policies among the relays and BS, to improve the network throughput.

E. Machine Learning-Aided Multi-Tier Computing

In [A10], Zeng et al. study the cost optimization for distributed Graph Neural Networks (GNNs) processing over a multi-tier heterogeneous edge network. The proposed technique theoretically reveals the structural property of quadratic submodularity implicated in GNN's unique computing pattern and serves as a foundation for future analysis and optimization on specific GNN-driven applications.

In [A11], Liu et al. propose an edge-assisted crowdsourced live video transcoding approach, where the transcoding capabilities of the edge transcoders are unknown and dynamic. Based on the derived bounds and by leveraging the contextual information of devices, two risk-aware contextual learning schemes are developed to efficiently estimate the transcoding capabilities of the edge devices.

In [A12], Deng et al. develop a joint communication and computation efficient federated learning (FL) framework for resource-limited Industrial Internet of Things (IIoT) systems. This framework integrates the deep neural networks (DNNs) partition technique into the standard FL technique, wherein IIoT devices perform local model training at the bottom layers of the objective DNN, and offload the top layers to the edge gateway side.

In [A13], the authors propose a novel technique to control the distribution of communications latency, which can support efficient multi-tier cooperative deep learning architecture in the harsh industrial environment. The proposed method is essentially an inverse process of the Gaussian Mixture Model (GMM) that controls latency distribution.

In [A14], the authors introduce the uplink and downlink fully-decoupled radio access network (FD-RAN) architecture to improve the wireless link rate via multiple BSs collaboration and power allocation solution. The proposed technique could accelerate FL-enabled multi-tier computing over wireless networks.

In [A15], the authors propose a framework to analyze the convergence behavior of multi-tier hybrid learning that leverages the server dataset and its computation capacity for collaborative model training. Different from standard FL, where stochastic gradient descent (SGD) is always computed in a parallel fashion across all clients, this proposed architecture enjoys both parallel SGD at clients and sequential SGD at the server by using the aggregated model from clients.

F. Service-Oriented Multi-Tier Computing

In [A16], the authors propose a multi-tier edge-cloud computing framework, EdgeMatrix, to maximize the throughput of the system while guaranteeing different service-level agreement (SLA) priorities. EdgeMatrix introduces the Networked Multi-agent Actor-Critic (NMAC) algorithm to redefine physical layer resources with the same quality of service. This technique can logically isolate resource units and combinations, i.e., cells and channels.

In [A7], the authors investigate the problem of service routing to achieve efficient microservice-based service provision in multi-tier computing systems. The objective is to optimize the request routing with low service delay and resource cost.

III. Conclusion

The Guest Editors hope that this Special Issue will provide valuable insights into current and future research areas on multi-tier computing technologies. They deeply appreciate the great mentoring provided by Prof. Moshe Zukerman, whose invaluable guidance was crucial to the success of the special issue. Meanwhile, the Guest Editors would like to thank both Prof. Petar Popovski and Janine Bruttin for their timely assistance in preparing the special issue. Last but not the least, they would also like to take this opportunity to thank all the authors and reviewers for their efforts in ensuring that this is a quality and relevant Special Issue.

Appendix: Related Articles

[A1] Z. Wang, X. Mu, Y. Liu, X. Xu, and P. Zhang, "NOMA-aided joint communication, sensing, and multi-tier computing systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 574–588, Mar. 2023, doi: 10.1109/JSAC.2022.3229447.

[A2] Y. Shao, D. Gunduz, and S. C. Liew, "Bayesian over-the-air computation," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 589–606, Mar. 2023, doi: 10.1109/JSAC.2022.3229428.

[A3] Z. Zhao, Q. Du, and G. K. Karagiannidis, "Improved grant-free access for URLLC via multi-tier-driven computing: Network-load learning, prediction, and resource allocation," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 607–622, Mar. 2023, doi: 10.1109/JSAC.2023.3234696.

[A4] X. Qin, Y. Li, X. Song, N. Ma, C. Huang, and P. Zhang, "Timeliness of information for computation-intensive status updates in task-oriented communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 623–638, Mar. 2023, doi: 10.1109/JSAC.2022.3229431.

[A5] Y. Xiao, Z. Sun, G. Shi, and D. Niyato, "Imitation learning-based implicit semantic-aware communication networks: Multi-layer representation and collaborative reasoning," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 639–658, Mar. 2023, doi: 10.1109/JSAC.2022.3229419.

[A6] S. Li, C. Li, Y. Huang, B. A. Jalaian, Y. T. Hou, and W. Lou, "Enhancing resilience in mobile edge computing under processing uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 659–674, Mar. 2023, doi: 10.1109/JSAC.2023.3234706.

[A7] T. Kim et al., "MoDEMS: Optimizing edge computing migrations for user mobility," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 675–689, Mar. 2023, doi: 10.1109/JSAC.2022.3229425.

[A8] C. Sun, X. Li, J. Wen, X. Wang, Z. Han, and V. C. M. Leung, "Federated deep reinforcement learning for recommendation-enabled edge caching in mobile edge-cloud computing networks," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 690–705, Mar. 2023, doi: 10.1109/JSAC.2023.3235443.

[A9] S. Tang and X. Lei, "Collaborative cache-aided relaying networks: Performance evaluation and system optimization," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 706–719, Mar. 2023, doi: 10.1109/JSAC.2023.3234693.

[A10] L. Zeng, C. Yang, P. Huang, Z. Zhou, S. Yu, and X. Chen, "GNN at the Edge: Cost-efficient graph neural network processing over distributed edge servers," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 720–739, Mar. 2023, doi: 10.1109/JSAC.2022.3229422.

[A11] X. Liu, M. Derakhshani, L. Mihaylova, and S. Lambotharan, "Risk-aware contextual learning for edge-assisted crowdsourced live streaming," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 740–754, Mar. 2023, doi: 10.1109/JSAC.2022.3229423.

[A12] X. Deng, J. Li, C. Ma, K. Wei, L. Shi, M. Ding, and W. Chen, "Low-latency federated learning with DNN partition in distributed industrial IoT networks," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 755–775, Mar. 2023, doi: 10.1109/JSAC.2022.3229436.

[A13] F. Qin, Y. Xiao, X. Sun, X. Dai, W. Zhang, and F. Shen, "Inverse-GMM: A latency distribution shaping method for industrial cooperative deep learning systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 776–788, Mar. 2023, doi: 10.1109/JSAC.2022.3229448.

[A14] Y. Xu, B. Qian, K. Yu, T. Ma, L. Zhao, and H. Zhou, "Federated learning over fully-decoupled RAN architecture for two-tier computing acceleration," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 789–801, Mar. 2023, doi: 10.1109/JSAC.2023.3236003.

[A15] K. Yang, S. Chen, and C. Shen, "On the convergence of hybrid server-clients collaborative training," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 802–819, Mar. 2023, doi: 10.1109/JSAC.2022.3229443.

[A16] S. Shen, Y. Ren, Y. Ju, X. Wang, W. Wang, and V. C. M. Leung, "*EdgeMatrix*: A resource-redefined scheduling framework for SLA-guaranteed multi-tier edge-cloud computing systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 820–834, Mar. 2023, doi: 10.1109/JSAC.2022.3229444.

[A17] S. Wang, Y. Guo, X. Liu, and A. Zhou, "Service routing in multi-tier edge computing: A matching game approach," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 835–844, Mar. 2023, doi: 10.1109/JSAC.2022.3229378.

Kunlun Wang (Member, IEEE) received the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2016.

From 2016 to 2017, he was with Huawei Technologies Company Ltd., where he was involved in energy efficiency algorithm design. From 2017 to 2019, he was with the Key Laboratory of Wireless Sensor Network and Communication, SIMIT, Chinese Academy of Sciences, Shanghai. From 2019 to 2020, he was with the School of Information Science and Technology, ShanghaiTech University. Since 2021, he has been a Professor with the School of Communication and Electronic Engineering, East China Normal University. His current research interests include B5G wireless communications, fog/edge computing networks, resource allocation, and optimization algorithm. He is the Lead Guest Editor of the IEEE Journal on Selected Areas in Communications on Multi-Tier Computing for Next Generation Wireless Networks and the Review Editor of *Signal Processing for Communications*.

Yang Yang (Fellow, IEEE) received the B.S. and M.S. degrees in radio engineering from Southeast University, Nanjing, China, in 1996 and 1999, respectively, and the Ph.D. degree in information engineering from The Chinese University of Hong Kong in 2002.

He is currently the Chief Scientist of the Internet of Things (IoT) at Terminus Group, China. He is also an Adjunct Professor with the Department of Broadband Communication, Peng Cheng Laboratory, and a Senior Consultant with the Shenzhen Smart City Technology Development Group, China. Before joining Terminus, he has held faculty positions at The Chinese University of Hong Kong, Brunel University, U.K., University College London (UCL), U.K., CAS-SIMIT, and ShanghaiTech University, China. His research interests include multi-tier computing networks, 5G/6G systems, AIoT technologies, intelligent services and applications, and advanced wireless testbeds. He has published more than 300 papers and filed more than 120 technical patents in these research areas. He has been the Chair of the Steering Committee of Asia–Pacific Conference on Communications (APCC) from 2019 to 2021. He is currently serving as the Chair for the IEEE Communications Society at 5G Industry Community and the Asia Region at Fog/Edge Industry Community.

Jiong Jin (Member, IEEE) received the B.E. degree (Hons.) in computer engineering from Nanyang Technological University, Singapore, in 2006, and the Ph.D. degree in electrical and electronic engineering from The University of Melbourne, Australia, in 2011. From 2011 to 2013, he was a Research Fellow with the Department of Electrical and Electronic Engineering, University of Melbourne. He is currently an Associate Professor with the School of Science, Computing and Engineering Technologies, Swinburne University of Technology, Melbourne, Australia. His research interests include network design and optimization, edge computing and networking, robotics and automation, cyber-physical systems, the Internet of Things, and their applications in smart manufacturing, smart transportation, and smart cities.

Tao Zhang (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from Northern Jiaotong University, Beijing, China, and the Ph.D. degree in electrical and computer engineering from the University of Massachusetts, Amherst, MA, USA.

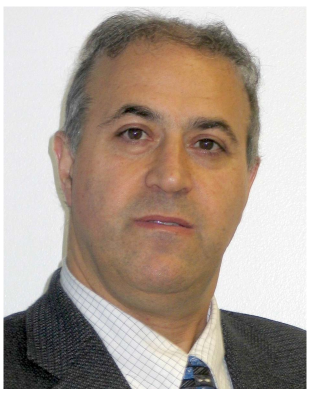
He is currently managing the Transformational Networks and Services Research Group with the U.S. National Institute of Standards and Technology (NIST). He was the CTO for the Smart Connected Vehicles Business at Cisco Systems, and the Chief Scientist and the Director of multiple research and development groups focused on wireless and vehicular networking at Telcordia Technologies (formerly Bell Communications Research). He co-founded the OpenFog Consortium and the Connected Vehicle Trade Association (CVTA) and served as a founding Board Director for them. He has been leading research, product development, and corporate strategies, which led to new technologies, products, and standards. He holds more than 100 patents awarded globally and coauthored two books *Vehicle Safety Communications: Protocols, Security, and Privacy* and *IP-Based Next Generation Wireless Networks* and more than 100 peer-reviewed articles. He is a fellow of the Society for Information Reuse and Integration and the Asia–Pacific Artificial Intelligence Academy. He has served as the CIO and a Board Governor of the IEEE Communications Society and a Distinguished Lecturer of the IEEE Vehicular Technology Society.

Arumugam Nallanathan (Fellow, IEEE) was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, from August 2000 to December 2007. He was with the Department of Informatics, Kings College London, from December 2007 to August 2017, where he was a Professor at wireless communications from April 2013 to August 2017 and a Visiting Professor from September 2017. He has been a Professor of wireless communications and the Head of the Communication Systems Research (CSR) Group with the School of Electronic Engineering and Computer Science, Queen Mary University of London, since September 2017. He has published more than 500 technical papers in scientific journals and international conferences. His research interests include artificial intelligence for wireless systems, beyond 5G wireless networks, the Internet of Things (IoT), and molecular communications. He received the IEEE Communications Society SPCE Outstanding Service Award in 2012 and the IEEE Communications Society RCC Outstanding Service Award in 2014. He was a co-recipient of the Best Paper Awards presented at the IEEE International Conference on Communications 2016 (ICC'2016), the IEEE Global Communications Conference 2017 (GLOBECOM'2017), and the IEEE Vehicular Technology Conference 2018 (VTC'2018). He was also a co-recipient of the IEEE Communications Society Leonard G. Abraham Prize in 2022. He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science Highly Cited Researcher in 2016 and 2022 and AI 2000 Internet of Things Most Influential Scholar in 2020. He has served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society and the Technical Program Chair and a member of Technical Program Committees in numerous IEEE conferences. He is an Editor-at-Large of the IEEE TRANSACTIONS ON COMMUNICATIONS and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2006 to 2011, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2006 to 2017, and the IEEE SIGNAL PROCESSING LETTERS.

Chintha Tellambura (Fellow, IEEE) received the B.Sc. degree (Hons.) from the University of Moratuwa, Sri Lanka, the M.Sc. degree in electronics from the Kings College, University of London, U.K., and the Ph.D. degree in electrical engineering from the University of Victoria, Canada.

He was with Monash University, Australia, from 1997 to 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta. He has authored or coauthored over 600 journal and conference papers with an H-index of 80 (Google Scholar). His current research interests include the design, modeling, and analysis of current and future wireless networks.

Prof. Tellambura was elected as a fellow of the Canadian Academy of Engineering in 2017. He has received best paper awards in the Communication Theory Symposium in 2012 IEEE International Conference on Communications (ICC) in Canada and 2017 ICC, France. He is the winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. He has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS from 1999 to 2011 and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2007. He was an Area Editor of *Wireless Communications Systems and Theory* from 2007 to 2012.



Bijan Jabbari (Life Fellow, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA. He is currently a Professor with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, USA. He is also an affiliated Faculty Member with Telecom Paris-Tech (ENST-Paris), France. His research interests include wireless communication and networks with particular emphasis on multi-user access, resource allocation, mobility and performance optimization, and cognitive radio. He was a recipient of the IET Fellow Grade, the IEEE Millennium Medal, and the Washington Metropolitan Area Engineer of the Year Award. He received the Outstanding Faculty Research Award from the School of Engineering, George Mason University. He was the Chairman of the IEEE Communications Society Technical Committee on Communications Switching and Routing. He was the General Chair of the IEEE GLOBECOM held in Washington, DC, USA, in 2016. He has served in various editorial capacities for a number of IEEE JOURNALS and TRANSACTIONS, including the IEEE TRANSACTIONS ON COMMUNICATIONS, a Division Editor for Wireless Communications of the *Journal of Communications and Networks*, and the Editorial Board of the PROCEEDINGS OF THE IEEE, and in different editorship capacities for several other journals.