Guest Editorial Special Issue on Advances in Artificial Intelligence and Machine Learning for Networking

Prosper Chemouil, *Fellow, IEEE*, Pan Hui, *Fellow, IEEE*, Wolfgang Kellerer, *Senior Member, IEEE*, Noura Limam, *Member, IEEE*, Rolf Stadler, *Senior Member, IEEE*, and Yonggang Wen, *Fellow, IEEE*

I. Introduction

ARTIFICIAL Intelligence (AI) and Machine Learning (ML) approaches have emerged in the networking domain with great expectation. They can be broadly divided into AI/ML techniques for network engineering and management, network designs for AI/ML applications, and system concepts. AI/ML techniques for networking and management improve the way we address networking. They support efficient, rapid, and trustworthy engineering, operations, and management. As such, they meet the current interest in softwarization and network programmability that fuels the need for improved network automation in agile infrastructures, including edge and fog environments. Network design and optimization for AI/ML applications addresses the complementary topic of supporting AI/ML-based systems through novel networking techniques, including new architectures and algorithms. The third topic area is system implementation and open-source software development.

This evolution draws particular attention to interdisciplinary approaches. Researchers in communication networks apply ML and AI concepts to optimize and automate network architecture, control, and management. Similarly, AI experts collaborate with networking researchers to optimize network support for architecture and design of data communication and processing for AI purposes.

This special issue is a follow-up to the JSAC's Special Issue on Artificial Intelligence and Machine Learning for Networking and Communications published in June 2019 [1]. It has been organized by the same core team of researchers. In addition to the strong response from the community to the earlier call (we received over 150 submissions), we observed a further increase in interest around AI and ML concepts for networking and related areas. New workshops and venues have been organized at premier conferences, including SIGOMM, Infocom, ICNP and ICCN, and several journals in the field issued calls. This development motivated us to offer an outlet for the best-quality research.

For this Special Issue, we made some adjustments to the scope of [1]. Primarily, we wanted to give this issue a clearer **networking focus**, since other JSAC calls have covered wireless communication, mostly targeted toward the physical layer and the MAC layer. While the interest in AI concepts for networking is growing rapidly, we found that the topic area is still less visible than wireless research in quality venues. Specifically, we believe that the development around network softwarization and automation in synergy with AI/ML is promising. In addition, network slicing—a major component of 5G technology, particularly in radio access networks—can be addressed in a novel way using AI/ML techniques.

Prosper Chemouil is with the Center for Studies and Research in Computer Science and Communication (CEDRIC), 75003 Paris, France (e-mail: prosper.chemouil@ieee.org).

Pan Hui is with the Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland, and also with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: panhui@cse.ust.hk).

Wolfgang Kellerer is with the Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany (e-mail: kellerer@tum.de).

Noura Limam is with the Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: noura.limam@uwaterloo.ca).

Rolf Stadler is with the Department of Computer Science, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (e-mail: stadler@kth.se).

Yonggang Wen is with School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ygwen@ntu.edu.sg).

Digital Object Identifier 10.1109/JSAC.2020.3003065

II. Paper Statistics

In response to the call of this issue, 72 papers were submitted by 340 authors from 28 countries and six continents. Counting the first authors, 23 papers were submitted from China including Hong Kong, 22 papers from Europe (EU, UK, Switzerland, and Norway), and 11 from the United States.

Eight papers were judged to be out of scope and thus were rejected early. The remaining submissions received at least three external reviews each and four reviews on average. After the first review round, the authors of 16 papers were given the opportunity to revise their papers and address the reviewers' concerns. All revised manuscripts were then shepherded by one or two guest editors. Finally, 15 papers were accepted for inclusion in this Special Issue, which corresponds to an acceptance ratio of 21%. From the accepted papers, five originate from China including Hong Kong, five from Europe, four from the United States, and one from Russia. The time between the initial submission and the publication of the papers in this Special Issue has been 11 months.

III. The Selected Papers

We introduce the papers grouped according to the network-related topics that they address. A key area of concern in networking is **traffic engineering and routing**. Two selected papers make specific contributions to this area. In “Significant Sampling for Shortest Path Routing: A Deep Reinforcement Learning Solution,” Shao *et al.* address the problem of finding the optimal sampling frequency for a centralized monitoring and control system in a dynamic network environment [2]. Using a deep reinforcement learning approach, the authors present a framework that is model-free and thus more flexible than analytically tractable solutions. The second paper is titled “CFR-RL: Traffic Engineering with Reinforcement Learning in SDN” and is authored by Zhang *et al.* [3]. The focus of this work is on identifying and minimizing the number of flows to be rerouted so as to optimize link utilization. Using the SDN architectural framework, they develop a reinforcement learning solution that reroutes the most critical flows. The evaluation shows that this approach favorably compares to known heuristics for many network configurations.

Intimately related to traffic engineering is **traffic estimation and scheduling**, with four accepted papers that address issues in this domain. In “RouteNet: Leveraging Graph Neural Networks for Network Modeling and Optimization in SDN,” Rusek *et al.* propose RouteNet, a novel network model based on Graph Neural Networks (GNNs) [4]. This model reflects the complex relationship between network topology, routing, and input traffic to produce estimates of packet delay and loss distributions per source–destination pair. RouteNet leverages the ability of GNNs to learn and model graph-structured information, and, as a result, it can generalize to arbitrary topologies, routing schemes, and traffic intensities.

Three papers deal specifically with traffic estimation and scheduling for wireless network environments. In “A Meta-Learning Scheme for Adaptive Short-Term Network Traffic Prediction,” by He *et al.*, the authors focus on a meta-learning approach for predicting traffic intensity to improve cellular network scheduling [5]. The proposed approach consists of a master policy and several subpolicies. These are called predictors and are trained on specific types of traffic. The authors employ a Long Short-Term Memory (LSTM) neural network and compare it with an Auto-Regressive Integrated Moving Average (ARIMA) model on data sets of video and nonvideo traffic. The paper “ASR—Adaptive Similarity-Based Regressor for Uplink Data Rate Estimation in Mobile Networks,” by Nikolov *et al.*, presents a passive data-rate estimation method for wireless modems [6]. It makes use of an online Support-Vector Regression algorithm that learns from a fixed-size cache of samples. Using measurements from a private testbed and traces from several commercial networks, the authors demonstrate that the cache size can be small while still allowing for accurate estimation. Finally, in “Peekaboo: Learning-based Multipath Scheduling for Dynamic Heterogeneous Environments,“ Wu *et al.* present an adaptive multipath scheduling algorithm named Peekaboo [7]. This algorithm learns and adopts scheduling decisions based on the changing characteristics of the heterogeneous paths and is effective specifically in wireless networks.

Two papers focus on improving and maintaining **Quality of Experience (QoE) for video streaming**, one proposing a collaborative, distributed framework, the second presenting a novel adaptive bitrate algorithm. In the paper “Personalized QoE Improvement for Networking Video Service” by Gao *et al.*, the authors study a personalized QoE improvement scheme for a video service from a multidimensional perspective, including user-awareness, device-awareness, and context-awareness [8]. They propose an efficient deep learning model for personalized characteristics extraction and employ a federated learning architecture with privacy protection to realize QoE improvement from sparse data. The paper titled “Quality-aware Neural Adaptive Video Streaming with Lifelong Imitation Learning,“ by Huang *et al.*, presents a novel method for adaptive video streaming to improve QoS [9]. The proposed solution uses a neural network-based action selection approach and a QoE metric. The authors apply imitation learning to learn more efficiently with fewer data samples.

Two papers contain contributions to the area of **Mobile Edge Networks**. In “Reinforcement Learning Based Optimal Computing and Caching in Mobile Edge Network,” Qian *et al.* present a hierarchical reinforcement learning (RL) algorithm for jointly pushing and caching content in mobile edge computing (MEC) networks [10]. An infinite-horizon average-cost Markov decision process (MDP) problem is formulated to maximize bandwidth utilization and decrease the amount of data transmitted. By predicting future requests, an optimal pushing and caching policy is designed to fully utilize the network bandwidth. The second paper investigates task offloading among mobile nodes, which is titled “Design of *a 5G Network Slice Extension with MEC UAVs Managed with Reinforcement Learning*” and is authored by Faraci *et al.* [11]. The authors propose to extend 5G network slices with Unmanned Aerial Vehicle (UAV)-enabled Multiaccess Edge Computing (MEC) in support of delay-constrained applications. The framework leverages reinforcement learning to maximize the performance of the system and extend the battery life of UAVs by turning on onboard MEC computing elements and offloading tasks when needed.

Cache admission and eviction policies for networked environments are studied in three papers. One of them performs the investigation in the context of the mobile edge and is introduced above [10]. The second paper focuses on content delivery networks and is titled “*Learning-Based Cache Admission for Content Delivery*” [12]. In this contribution, Kirilin *et al.* combine Monte Carlo sampling and direct policy search to design RL-Cache, a cache-admission algorithm, as a simple front end for a CDN server. RL-Cache considers a broad set of features including request recency, frequency, and size to maximize the cache hit rate. The authors evaluate their implementation on image, video and web traces from Akamai’s production CDN. By learning the different caching strategies needed for various traffic classes and cache sizes, RL-Cache outperforms or matches the hit rate achieved by state-of-the-art algorithms. The paper also evaluates feature importance, hyper-parameter sensitivity, and the ability of RL-Cache to be trained and executed on different traces at different locations. The third paper, which is introduced earlier, deals with a sample cache whose policies are designed to maximize the accuracy of traffic estimation [6].

One of the accepted papers addresses the problem of **failure recovery**. In particular, in "DeepPR: Progressive Recovery for Interdependent VNFs with Deep Reinforcement Learning," Ishigaki *et al.* focus on the progressive recovery from failures in networks with Virtual Network Functions (VNFs) [13], and propose a deep reinforcement learning technique that achieves near-optimal solutions for certain networks and that is robust to adversarial failures.

Finally, three papers investigate aspects of **distributed learning** in networked systems. The paper titled "Robust Coreset Construction for Distributed Machine Learning," by Lu *et al.*, introduces two coreset construction algorithms, which produce a small subset that approximates an original dataset for efficient learning [14]. Through comparing the robustness of known coreset construction algorithms, the authors identify a new class of clustering algorithms that yield superior robustness. They establish theoretical conditions under which the generated coreset provides a guaranteed approximation of the original dataset, and they develop a centralized and a distributed algorithm based on these conditions. The algorithms are evaluated using diverse datasets. The following two papers study **privacy** topics in distributed networked systems. In "FORESEEN: Towards Differentially Private Deep Inference for Intelligent Internet of Things," by Lyu *et al.*, the authors propose a data-privacy-preserving framework for deep learning suited for large-scale IoT applications [15]. To ensure privacy, network nodes and the cloud collaboratively perform noisy training using perturbation techniques. To meet the constraints of accuracy, memory, and energy in IoT end devices, models with mixed-precision are constructed. The paper titled "Analyzing User-Level Privacy Attack Against Federated Learning," by Song *et al.*, focuses on attacks by a malicious server [16]. The authors introduce an attack framework called mGAN-AI that is based on Generative Adversarial Networks (GANs). The framework allows the malicious server to recover private user data. In addition, the authors extend the framework to counter anonymization strategies against mGAN-AI.

Several works included in this Special Issue were conducted within the context of **softwarization**, i.e., software-defined networking (SDN), Virtual Network Functions (VNFs), and so on [3], [4], [13]. The core contribution of some papers, such as [2], lies in theoretical analysis and simulation, while others put strong emphasis on experimentation and evaluation involving many data sets [6].

Regarding specific **machine-learning methods** that are introduced to address networking problems, many papers propose reinforcement learning techniques [2], [3], [7], [10]–[13], all of which, except [11], use model-free reinforcement learning. More specifically, [10], [12], and [13] apply a value-based approach, often referred to as Q-learning, while [2] and [3] follow a policy-based Approach; [7] applies a multiarmed bandit model as the basis for reinforcement learning, while all other above-referenced papers rely on Markov Decision Processes (MDPs).

Many works apply deep learning methods, which are based on neural networks [2]–[5], [8]–[10], [13], [15], [16]. Some papers use methods based on Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) [5], [8]. One work uses a Graph Neural Network (GNN) to reflect relationships in networked systems [4], and one uses Generative Adversarial Networks (GANs) to address privacy issues [16].

Two papers rely on a clustering technique as the key aspect of the proposed solution [8], [14]. The contribution in [6] is based on an online Support Vector Regression (SVR) scheme. The authors of [9] use the concept of life-long learning as part of their approach. The paper [4] proposes a meta-learning scheme that allows to switch among sub-policies. Finally, [16] applies federated learning to build a solution for user privacy in case of malicious server.

IV. Conclusion

While a clear evolution is observed in terms of depth and maturity of the works selected for this issue, compared to the papers published in [1], we believe that much work remains to obtain a fundamental understanding of the best use of AI/ML methods for engineering and operating networks and networked systems. Many of the contributions in this issue, though important, remain "point solutions," and it is not clear how they can generalize in a broader context. Furthermore, it seems to us that most of the research challenges we presented in [1] remain to be fully investigated: What is an effective representation of network data that enables efficient learning? What is the best architecture for learning from network data with respect to various target technologies like softwarized core networks, edge networks, cloud RANs, etc.? Which are the test cases and data sets researchers can and should use to benchmark proposed AI/ML solutions? How can we best integrate various AI-driven processes, including monitoring, prediction, decision making, and control in a networked system, both from an architectural and an algorithmic point of view? And finally: How can AI/ML concepts be applied to significantly further automation in network environments that are increasingly powerful but also increasingly complex.

Acknowledgment

The Guest Editors would like to thank all authors who submitted papers to this Special Issue. We are especially grateful to the many reviewers who provided timely and thorough reviews that helped improve the selected papers in the revision round. We thank Raouf Boutaba, JSAC EiC, as well as Patrick Thiran, JSAC Senior Editor and mentor for this issue, for their support and advice. Finally, we acknowledge the help of Janine Bruttin, JSAC Executive Editor, and Lauren Briede, Editorial Support and Production Assistant, who helped with logistical and production issues.

References

[1] P. Chemouil *et al.*, "Guest editorial: Special issue on artificial intelligence and machine learning for networking and communications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1185–1191, Jun. 2020.

[2] Y. Shao, A. Rezaee, S. C. Liew, and V. W. S. Chan, "Significant sampling for shortest path routing: A deep reinforcement learning solution," *IEEE J. Sel. Areas Commun.*, early access, Jun. 8, 2020, doi: 10.1109/JSAC.2020.3000364.

[3] J. Zhang, M. Ye, Z. Guo, C. Y. Yen, and H. J. Chao, "CFR-RL: Traffic engineering with reinforcement learning in SDN," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000371.

[4] K. Rusek, J. Suárez-Varela, P. Almasan, P. Barlet-Ros, and A. Cabellos-Aparicio, "RouteNet: Leveraging graph neural networks for network modeling and optimization in SDN," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000405.

[5] Q. He, A. Moayyedi, G. Dán, G. P. Koudouridis, and P. Tengkvist, "A meta-learning scheme for adaptive short-term network traffic prediction," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000408.

[6] G. Nikolov, M. Kuhn, A. McGibney, and B.-L. Wenning, "ASR-adaptive similarity-based regressor for uplink data rate estimation in mobile networks," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000414.

[7] H. Wu, O. Alay, A. Brunström, S. Ferlin, and G. Caso, "Peekaboo: Learning-based multipath scheduling for dynamic heterogeneous environments," *IEEE J. Sel. Areas Commun.*, early access, Jun. 8, 2020, doi: 10.1109/JSAC.2020.3000365.

[8] Y. Gao, X. Wei and L. Zhou, "Personalized QoE improvement for networking video service," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000395.

[9] T. Huang, C. Zhou, X. Yao, R.-X. Zhang, C. Wu, and L. Sun, "Quality-aware neural adaptive video streaming with lifelong imitation learning," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000363.

[10] Y. Qian, R. Wang, J. Wu, B. Tan, and H. Ren, "Reinforcement learning based optimal computing and caching in mobile edge network," *IEEE J. Sel. Areas Commun.*, early access, Jun. 8, 2020, doi: 10.1109/JSAC.2020.3000396.

[11] G. Faraci, C. Grasso, and G. Schembra, "Design of a 5G network slice extension with MEC UAVs managed with reinforcement learning," *IEEE J. Sel. Areas Commun.*, to be published.

[12] V. Kirilin, A. Sundarrajan, S. Gorinsky, and R. K. Sitaraman, "RL-Cache: Learning-based cache admission for content delivery," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000415.

[13] G. Ishigaki, S. Devic, R. Gour, and J. P. Jue, "DeepPR: Progressive recovery for interdependent VNFs with deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000402.

[14] H. Lu, M. J. Li, T. He, S. Wang, V. Narayanan, and K. S. Chan, "Robust coreset construction for distributed machine learning," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000373.

[15] L. Lyu, J. C. Bezdek, J. Jin, and Y. Yang, "FORESEEN: Towards differentially private deep inference for intelligent Internet of Things," *IEEE J. Sel. Areas Commun.*, early access, Jun. 8, 2020, doi: 10.1109/JSAC.2020.3000374.

[16] M. Song *et al.*, "Analyzing user-level privacy attack against federated learning," *IEEE J. Sel. Areas Commun.*, early access, Jun. 5, 2020, doi: 10.1109/JSAC.2020.3000372.

Prosper Chemouil (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in control theory from the École Centrale de Nantes in 1976 and 1978, respectively. He is a Research Adjunct at the Center for Studies and Research in Computer Science and Communication (CEDRIC), Cnam. He is retired as the Research Director on Future Networks at Orange Laboratories. His research interests include the design and management of new networks and technologies and their impact on network architecture, traffic engineering and control, and performance, with significant involvement in standardization at ITU-T for 25 years. In this context, he then investigated the potential of various artificial intelligence techniques for network management in the early 1990s. For several years, he has been then focusing on cognitive management and network softwarization. In 2016, he became the Co-Chair of the IEEE SDN Initiative and a member of the IEEE ComSoc Industry Communities Board, representing the SDN/NFV/Cloud area. Within the IEEE SDN Initiative, he launched NetSoft, the IEEE International Conference on Network Softwarization, in 2015. He has been involved as the General or the TPC Co-Chair in many events dealing with network performance and management. He has also served as a Board Member, an Associate Editor, and the Guest Editor of various journals, including the *IEEE Communications Magazine*, the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, the IEEE NETWORKS, and *Annals of Telecommunications*. He was a recipient of several awards, such as the Blondel Medal in 1996, the Ampère Medal in 2003, the Salah Aidarous Memorial Award in 2014, and Arne Jensen Lifetime Achievement Award in 2015. He was nominated as French Senior Engineer of the Year in 1995 and became a fellow of France Telecom Research and Development in 1998. He is a fellow of SEE, the French Society of Electrical and Electronical Engineers.

Pan Hui (Fellow, IEEE) received the bachelor's and M.Phil. degrees from the University of Hong Kong and the Ph.D. degree from the Computer Laboratory, University of Cambridge. He was an Adjunct Professor of social computing and networking with Aalto University from 2012 to 2016. He was a Senior Research Scientist and then a Distinguished Scientist for Telekom Innovation Laboratories (T-labs), Germany, from 2008 to 2015. His industrial profile also includes his research at Intel Research Cambridge and Thomson Research Paris from 2004 to 2006. He is currently the Nokia Chair Professor in data science and a Professor of computer science with the University of Helsinki, and the Director of the HKUST-DT System and Media Laboratory, Hong Kong University of Science and Technology. He has published more than 300 research articles and with more than 18 000 citations. He has 30 granted and filed European and U.S. patents in the areas of augmented reality and mobile computing. He has founded and chaired several IEEE/ACM conferences/workshops, and he has been serving on the Organizing and Technical Program Committee of numerous top international conferences, including ACM SIGCOMM, MobiSys, MobiCom, CoNext, IEEE Infocom, ICNP, ICDCS, IJCAI, AAAI, ECAI, ICWSM, and WWW. He was an Associate Editor of the IEEE TRANSACTIONS ON CLOUD COMPUTING, and a Guest Editor of the *IEEE Communication Magazine* and the *ACM Transactions on Multimedia Computing*, *Communications, and Applications*. He has been an Associate Editor of leading journals such as the IEEE TRANSACTIONS ON MOBILE COMPUTING since 2014. He is an ACM Distinguished Scientist and a member of Academia Europaea.

Wolfgang Kellerer (Senior Member, IEEE) received the Dipl.-Ing. (master's) and Dr.-Ing. (Ph.D.) degrees from the Technical University of Munich, Germany, in 1995 and 2002, respectively. He is currently a Full Professor and heading the Chair of communication networks with the Department of Electrical and Computer Engineering, Technical University of Munich. He was with NTT DOCOMO's European Research Laboratories for ten years in leading positions, contributing to research and standardization of LTE-A and 5G technologies. In 2001, he was a Visiting Researcher with the Information Systems Laboratory, Stanford University, Stanford, CA, USA. His research has resulted in more than 200 publications and 35 granted patents. He was awarded with an ERC Consolidator Grant from the European Commission for his research project FlexNets "*Quantifying Flexibility in Communication Networks*" in 2015. He also serves as an Associate Editor for the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT. He is on the Editorial Board of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS and the IEEE NETWORKING LETTERS. He is a member of ACM and VDE ITG.

Noura Limam (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University Pierre and Marie Curie, Paris VI, in 2002 and 2007, respectively. She is currently a Research Assistant Professor of computer science with the University of Waterloo, Waterloo, ON, Canada. She has been serving on the Technical Program Committees and Organization Committees of several IEEE conferences. Her contributions are in the area of network and service management. Her current research interests are in network softwarization and cognitive network management. She is a recipient of the 2008 IEEE Communications Society Fred W. Ellersick Prize.

Rolf Stadler (Senior Member, IEEE) received the M.Sc. degree in mathematics and the Ph.D. degree in computer science from the University of Zurich. Before joining KTH in 2001, he held positions at the IBM Zurich Research Laboratory, Columbia University, and ETH Zürich. His group has made contributions to real-time monitoring, resource management, and self-management for large-scale networked systems. He is currently a Professor with the KTH Royal Institute of Technology, Stockholm, Sweden, and the Head of the Department of Network and Systems Engineering. His current interests include advanced monitoring techniques and data-driven methods for network engineering and management. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT (TNSM) from 2014 to 2017.

Yonggang Wen (Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer science (minor in *Western Literature*) from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2008. He has worked extensively in learning-based system prototyping and performance optimization for large-scale networked computer systems. He is currently a Professor with the School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU), Singapore, where he also serves as the Associate Dean (Research) with the College of Engineering. His work in *Multi-Screen Cloud Social TV* has been featured by global media (more than 1600 news articles from more than 29 countries) and received 2013 ASEAN ICT awards (*Gold Medal*). His research interests include cloud computing, green data center, distributed machine learning, blockchain, big data analytics, multimedia networks, and mobile computing. His recent work on *Cloud3DView*, as the only academia entry, has won 2016 ASEAN ICT awards (*Gold Medal*) and 2015 Datacentre Dynamics awards 2015—APAC (*Oscar* Award of data center industry). He is the Sole Winner of the 2016 Nanyang awards in Innovation and Entrepreneurship at NTU. He was a co-recipient of multiple best papers awards, including the 2019 IEEE TCSVT Best Paper Award, the 2015 IEEE Multimedia Best Paper Award, the 2016 IEEE Globecom, the 2016 IEEE Infocom MuSIC Workshop, the 2015 EAI/ICST Chinacom, the 2014 IEEE WCSP, the 2013 IEEE Globecom, and the 2012 IEEE EUC. He received the 2016 IEEE ComSoc MMTC Distinguished Leadership Award. He serves and had served on the Editorial Board for multiple transactions and journals, including the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *IEEE Wireless Communication Magazine*, the IEEE COMMUNICATIONS SURVEY AND TUTORIALS, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, IEEE ACCESS JOURNAL, and *Ad Hoc Networks* (Elsevier), and was elected as the Chair for the IEEE ComSoc Multimedia Communication Technical Committee from 2014 to 2016.