Guest Editorial: Special Section on Data Analytics and Machine Learning for Network and Service Management—Part I

Nur Zincir-Heywood, *Member, IEEE*, Giuliano Casale, *Member, IEEE*, David Carrera, *Member, IEEE*, Lydia Y. Chen, *Senior Member, IEEE*, Amogh Dhamdhere, *Member, IEEE*, Takeru Inoue, *Member, IEEE*, Hanan Lutfiyya, *Senior Member, IEEE*, and Taghrid Samak, *Member, IEEE*

I. Introduction

NETWORK and Service analytics can harness the immense stream of operational data from clouds, to services, to social and communication networks. In the era of big data and connected devices of all varieties, analytics and machine learning have found ways to improve reliability, configuration, performance, fault and security management. In particular, we see a growing trend towards using machine learning, artificial intelligence and data analytics to improve operations and management of information technology services, systems and networks.

Research is therefore needed to understand and improve the potential and suitability of data analytics and machine learning in the context of services, systems and network management. This will provide deeper understanding and better decision making based on largely collected and available operational and service data. It will also present opportunities for improving machine learning and data analytics algorithms and methods on aspects such as reliability, dependability and scalability, as well as demonstrate the benefits of these methods in management and control systems. Moreover, there is an opportunity to define novel platforms that can harness the vast operational data and advanced data analysis algorithms to drive management decisions in networks, data centers, and clouds.

Manuscript received November 14, 2020; accepted November 14, 2020. Date of current version December 9, 2020. The associate editor coordinating the review of this article and approving it for publication was F. De Turck.
(Corresponding author: Nur Zincir-Heywood.)

Nur Zincir-Heywood is with the Department of Computer Science, Dalhousie University, Halifax, NS B3H 1W5, Canada (e-mail: zincir@cs.dal.ca).

Giuliano Casale is with the Department of Computing, Imperial College London, London SW7 2AZ, U.K.

David Carrera is with the Department of Computer Science, Barcelona Supercomputing Center, 08034 Barcelona, Spain.

Lydia Y. Chen is with the Department of Computer Science, Delft University of Technology, 2628 Delft, The Netherlands.

Amogh Dhamdhere is with Amazon Web Services, Amazon.com Inc., Seattle, WA 98108 USA.

Takeru Inoue is with the NTT Network Innovation Laboratories, NTT Corporation, Yokosuka 239-0847, Japan.

Hanan Lutfiyya is with the Department of Computer Science, Western University, London, ON N6A 5B7, Canada.

Taghrid Samak is with the Google Global Networking, Google Inc., Mountain View, CA 94043 USA.

Digital Object Identifier 10.1109/TNSM.2020.3038736

This special section of IEEE Transactions on Network and Service Management presents novel research tackling the above challenges. It is the fourth special section in this area to appear in this series, after issues published in [1], [2], [3]. The collection of works we present illustrates recent trends, novel solutions and approaches to leverage Data analytics and Machine Learning in Network and Service management, as well as to extract insights from data that can guide system operators and network managers in their daily activities.

The special sections consists of two parts. In Part I, presented here, we have accepted 15 papers out of 83 papers submitted to the open call for novel contributions addressing the underlying challenges of *Data Analytics and Machine Learning for Network and Service management*. Part II will be published in the next issue (March 2021).

II. Special Section Overview

The special section papers span four central areas of *Data Analytics and Machine Learning for Management:* (i) Data Analytics and Machine Learning for Network Management in general, (ii) Data Analytics and Machine Learning for Service Management, (iii) Data Analytics and Machine Learning for Mobile Networks, and (iv) Data Analytics and Machine Learning for Social Network Platforms.

A. Data Analytics and Machine Learning for Network Management in General

Four papers in this special section focus on data analytics and machine learning for management of networks.

In "Deep Reinforcement Adversarial Learning Against Botnet Evasion Attacks," Apruzzese *et al.* [item 1) in the Appendix] investigate network traffic and characteristics to interpret botnet evasion attacks. They propose a framework that can protect botnet detectors from adversarial attacks through Deep Reinforcement Learning mechanisms.

In "Exploring Network-Wide Flow Data With Flowyager," Saidi *et al.* [item 2) in the Appendix] extend the issue of improving the response time for a priori unknown network-wide network flow queries. They propose Flowyager, a system that is built on top of existing traffic capture utilities, enables querying the resulting datasets from hundreds of routers across sites and over time.

In "ViCrypt to the Rescue: Real-Time, Machine-Learning-Driven Video-QoE Monitoring for Encrypted Streaming Traffic," Wassermann *et al.* [item 3) in the Appendix] investigate real-time, passive Quality of Experience monitoring of HTTP Adaptive Video Streaming from the Internet Service Provider perspective. They present a monitoring solution based on a machine learning model that is able to infer Key Video Quality of Experience Indicators such as stalling, initial delay, video resolution, and average video bitrate.

In "Collaborative Flow Control in the DARPA Spectrum Collaboration Challenge," Mennes *et al.* [item 4) in the Appendix] consider an intelligent, dynamic spectrum allocation mechanism, where different network technologies collaboratively optimize the spectrum usage. The authors present a flow control mechanism based on policies to implement collaboration in a Quality of Service driven way.

B. Data Analytics and Machine Learning for Service Management

Three papers in this special section focus on data analytics and machine learning for management of services.

In "You Only Run Once: Spark Auto-Tuning From a Single Run," Buchaca *et al.* [item 5) in the Appendix] seek to examine tuning configurations for Apache Spark data processing framework that uses information from one run of a Spark workload. The authors validate their method against state-of-the-art features and techniques showing its scalability, prediction accuracy and increased speedup.

In "Diminishing Returns and Deep Learning for Adaptive CPU Resource Allocation of Containers," Abdullah *et al.* [item 6) in the Appendix] introduce a machine learning approach for automatically allocating optimal CPU resources to the containers. The authors evaluate the performance of their approach using real workloads on a Docker-based containerized infrastructure demonstrating the effectiveness in reducing the completion time of the jobs and compare it to commonly used static CPU allocation methods.

In "HitAnomaly: Hierarchical Transformers for Anomaly Detection in System Log," Huang *et al.* [item 7) in the Appendix] explore log-based anomaly detection methods for service and system maintenance. The authors focus in particular on a log-based anomaly detection model that utilizes a hierarchical transformer structure to capture the semantic information in both log template sequences and parameter values.

C. Data Analytics and Machine Learning for Mobile Networks

Five papers in this special section focus on data analytics and machine learning for management of mobile and radio networks and services.

In "Few-Shot Learning and Self Training for eNodeB Log Analysis for Service Level Assurance in LTE Networks," Aoki *et al.* [item 8) in the Appendix] illustrate the use of machine learning approaches to characterize different states of base stations. The authors investigate the impact of a low number of labeled data based on key performance indicators that base stations generate, demonstrating their benefits using a live Long-Term Evolution network.

In "Estimating Pole Capacity From Radio Network Performance Statistics by Supervised Learning," Gijo Ìn *et al.* [item 9) in the Appendix] present a comprehensive analysis of different supervised learning algorithms for estimating cell and user throughput from the collected radio network mesasurements. The analysis is performed on two radio access technologies, namely High Speed DownLink Packet Access and Long Term Evolution, demonstrating the most appropriate options for network dimnesioning. The performance is demonstrated on datasets collected from live cellular networks.

In "Mobility Management With Transferable Reinforcement Learning Trajectory Prediction," Zhao *et al.* [item 10) in the Appendix] define a proactive mobility management approach based on group user trajectory prediction. The authors combine the Long-Short Term Memory networks with Reinforcement Learning to automate the model training procedure. They develop a group user trajectory predictor to reduce prediction calculation overheads of users with similar movement patterns. They show the impact of their approach on a virtual reality service migration scheme built on the top of the handover mechanism using trajectory predictions.

In "Machine Learning-Based Radio Coverage Prediction in Urban Environments," Mohammadjafari *et al.* [item 11) in the Appendix] explore the optimal location of radio transmitters in order to maximize the radio coverage in a geographic area. They benchmark various machine learning models to estimate the relationship between transmitter location and the resulting power coverage. They also investigate the feature construction methods and the effect of training data size.

In "Machine Learning Based Recommender Systems to Achieve Self-Coordination Between SON Functions," Bag *et al.* [item 12) in the Appendix] propose an online Recommender Systems to model the dynamics between Self-Organizing Network functions of cellular networks. The authors jointly implement two intertwined Self-Organizing Network functions, namely Inter Cell Interference Coordination and Coverage and Capacity Optimization, using a machine learning based approach. This approach implicitly handles the conflicts and achieves the desired trade-off between coverage and capacity.

D. Data Analytics and Machine Learning for Social Network Platforms

Three papers in this special section focus on Data Analytics and Machine Learning for social networks and platforms.

In "Critical Impact of Social Networks Infodemic on Defeating Coronavirus COVID-19 Pandemic: Twitter-Based Study and Research Directions," Mourad *et al.* [item 13) in the Appendix] present a large-scale study based on approximately one million COVID-19 related tweets collected over a period of two months. The paper adopts an approach based on data analytics for analyzing social media posts on the COVID-19 pandemic, with a focus on false and misleading information and redirecting users to irrelevant topics. Coverage of existing research is discussed and used to provide guidelines and research directions for potential solutions and social networks management strategies during crisis periods.

In "Spotting Political Social Bots in Twitter: A Use Case of the 2019 Spanish General Election," Pastor-Galindo *et al.* [item 14) in the Appendix] propose a machine learning based approach to analyze the presence and behavior of social bots on Twitter. The authors classify the involved users as social bots or humans, and analyze their interactions from a quantitative and qualitative perspectives.

In "Data Fusion Oriented Graph Convolution Network Model for Rumor Detection," Yu *et al.* [item 15) in the Appendix] present a rumor detection model based on Graph Convolution Networks. The proposed model leverages both static and dynamic features to detect rumor propogation on real life social media, Sina Weibo dataset.

ACKNOWLEDGMENT

We sincerely thank the authors for contributing their papers and the reviewers for their thorough assessment and their work to improve the quality and presentation of each paper. We are very grateful to the Editor-in-Chief Filip De Turck for his continuous support throughout the process and to Janine Bruttin and Catherine Van Sciver for their help with the administrative tasks associated to this special section.

RELATED WORKS
APPENDIX

1) G. Apruzzese, M. Andreolini, M. Marchetti, A. Venturi, and M. Colajanni, "Deep reinforcement adversarial learning against botnet evasion attacks," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 1975–1987, Dec. 2020, doi: 10.1109/TNSM.2020.3031843.
2) S. J. Saidi, A. Maghsoudlou, D. Foucard, G. Smaragdakis, I. Poese, and A. Feldmann, "Exploring network-wide flow data with flowyager," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 1988–2006, Dec. 2020, doi: 10.1109/TNSM.2020.3034278.
3) S. Wassermann, M. Seufert, P. Casas, L. Gang, and K. Li, "ViCrypt to the rescue: Real-time, machine-learning-driven video-QOE monitoring for encrypted streaming traffic," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2007–2023, Dec. 2020, doi: 10.1109/TNSM.2020.3036497.
4) R. Mennes *et al.*, "Collaborative flow control in the DARPA spectrum collaboration challenge," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2024–2038, Dec. 2020, doi: 10.1109/TNSM.2020.3031078.
5) D. Buchaca, F. Portella, C. Costa, and J. L. Berral, "You only run once: Spark auto-tuning from a single run," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2039–2051, Dec. 2020, doi: 10.1109/TNSM.2020.3034824.
6) M. Abdullah, W. Iqbal, F. Bukhari, and A. Erradi, "Diminishing returns and deep learning for adaptive CPU resource allocation of containers," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2052–2063, Dec. 2020, doi: 10.1109/TNSM.2020.3033025.
7) S. Huang *et al.*, "HitAnomaly: Hierarchical transformers for anomaly detection in system log," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2064–2076, Dec. 2020, doi: 10.1109/TNSM.2020.3034647.
8) S. Aoki, K. Shiomoto, and C. Lam Eng, "Few-shot learning and self training for eNodeB log analysis for service level assurance in LTE networks," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2077–2089, Dec. 2020,
9) C. Gijón, M. Toril, S. Luna-Ramírez, J. L. Bejarano-Luque, and M-L-Marí-Altozano, "Estimating pole capacity from radio network performance statistics by supervised learning," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2090–2101, Dec. 2020, doi: 10.1109/TNSM.2020.3031333.
10) Z. Zhao *et al.*, "Mobility management with transferable reinforcement learning trajectory prediction," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2102–2116, Dec. 2020, doi: 10.1109/TNSM.2020.3034482.
11) S. Mohammadjafari, S. Roginsky, E. Kavurmacioglu, M. Cevik, J. Ethier, and A. Basar, "Machine learning-based radio coverage prediction in urban environments," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2117–2130, Dec. 2020, doi: 10.1109/TNSM.2020.3035442.
12) T. Bag, S. Garg, D. F. P. Rojas, and A. Mitschele-Thiel, "Machine Learning based recommender systems to achieve self-coordination between SON functions," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2131–2144, Dec. 2020, doi: 10.1109/TNSM.2020.3024895.
13) A. Mourad, A. Srour, H. Harmanani, C. Jenainati, and M. Arafeh, "Critical impact of social networks infodemic on defeating coronavirus COVID-19 pandemic: Twitter-based study and research directions," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2145–2155, Dec. 2020, doi: 10.1109/TNSM.2020.3031034.
14) J. Pastor-Galindo *et al.*, "Spotting political social bots in Twitter: A use case of the 2019 Spanish general election," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2156–2170, Dec. 2020, doi: 10.1109/TNSM.2020.3031573.
15) K. Yu, H. Jiang, T. Li, S. Han, and X. Wu, "Data fusion oriented graph convolution networkmodel for rumor detection," *IEEE Trans. Netw. Service Manag.*, vol. 17, no. 4, pp. 2171–2181, Dec. 2020, doi: 10.1109/TNSM.2020.3033996.

REFERENCES

[1] G. Casale, Y. Diao, H. Lutfiyya, P. Owezarski, and D. Raz, "Guest editors' introduction: Special issue on big data analytics for management," *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 3, pp. 578–580, Sep. 2016.
[2] G. Casale, Y. Diao, M. Mellia, R. Ranjan, and N. Zincir-Heywood, "Guest editorial: Special section on advances in big data analytics for management," *IEEE Trans. Netw. Service Manag.*, vol. 15, no. 1, pp. 10–12, Mar. 2018.
[3] D. Carrera, G. Casale, T. Inoue, H. Lutfiyya, J. Wang, and N. Zincir-Heywood, "Guest editorial: Special issue on novel techniques in big data analytics for management," *IEEE Trans. Netw. Service Manag.*, vol. 16, no. 3, pp. 797–799, Sep. 2019.

Nur Zincir-Heywood (Member, IEEE) received the Ph.D. degree in computer science and engineering from Ege University, Turkey, in 1998. She is a Full Professor of Computer Science with Dalhousie University, Canada. She is a co-editor of the book *Recent Advances in Computational Intelligence in Defense and Security* (Springer). Her research interests include machine learning and data mining techniques for network management and cybersecurity, topics on which she has published over 200 fully reviewed papers. She is a recipient of several best paper awards as well as the supervisor for the recipient of the IFIP/IEEE IM 2013 Best Ph.D. Dissertation Award in Network Management. She received the 2017 Women Leaders in the Digital Economy Award. She is an Associate Editor of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and is the General Co-Chair of the 16th International Conference on Network and Service Management 2020. She has been a Co-Organizer for the IEEE/IFIP International Workshop on Analytics for Network and Service Management since 2016.

Giuliano Casale (Member, IEEE) joined the Department of Computing, Imperial College London in 2010, where he is currently a Senior Lecturer of Modeling and Simulation. Previously, he worked as a Scientist with SAP Research U.K. and as a consultant in the capacity planning industry. He teaches and does research in performance engineering and cloud computing, topics on which he has published more than 130 refereed papers. He is a recipient of multiple awards, recently the Best Paper Award at ACM SIGMETRICS 2017. He has served as the Program Co-Chair for several conferences in the area of performance engineering, such as ACM SIGMETRICS/Performance and IEEE MASCOTS. He serves on the editorial boards of IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and ACM TOMPECS and as the Current Chair of ACM SIGMETRICS.

David Carrera (Member, IEEE) received the M.S. and Ph.D. degrees from the Technical University of Catalonia in 2002 and 2008, respectively, where he is an Associate Professor with the Computer Architecture Department. He is also the Head of the "DataCentric Computing" Research Group, Barcelona Supercomputing Center (BSC). His research interests are focused on the performance management of data center workloads. In 2015, he was awarded an ERC Starting Grant for the project HiEST and ICREA Academia Award, and an ERC Proof of Concept Grant ("Hi-OMICS") in 2017 to explore the commercialization of an SDI orchestrator for genomics workloads. He has participated in several EU-funded projects and has led the team at BSC that has developed the Aloja Project (aloja.bsc.es) and the servIoTicy platform (servioticy.com). He is the PI for several industrial projects and collaborations with IBM, Microsoft, and Cisco, among others. He was a summer intern with IBM Watson, Hawthorne, NY, USA, in 2006, and a Visiting Research Scholar with IBM Watson, Yorktown, NY, USA, in 2012. He received the IBM Faculty Award in 2010. He is an ACM Member.

Lydia Y. Chen (Senior Member, IEEE) received the B.A. degree from National Taiwan University in 2002, and the Ph.D. degree from Pennsylvania State University in 2006. She is an Associate Professor with the Department of Computer Science, Technology University Delft (TU Delft). Prior to joining TU Delft, she was a Research Staff Member with IBM Zurich Research Lab from 2007 to 2018. Her research interests center around dependability management, resource allocation, and privacy enhancement for large-scale data processing systems and services. More specifically, her work focuses on developing stochastic and machine learning models and applying these techniques to application domains, such as datacenters and AI systems. She has published more than 80 papers in journals, e.g., IEEE TRANSACTIONS ON DISTRIBUTED SYSTEMS, IEEE TRANSACTIONS ON SERVICE COMPUTING, and conference proceedings, e.g., INFOCOM, Sigmetrics, DSN, and Eurosys. She was a co-recipient of the Best Paper Awards at CCgrid'15 and eEnergy'15. She received TU Delft Professor Fellowship in 2018. She was the Program Co-Chair for Middleware Industry Track 2017 and IEEE ICAC 2019 and the Track Vice-Chair for ICDCS 2018. She has served on the editorial boards of IEEE TRANSACTIONS ON SERVICE COMPUTING and IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT.

Amogh Dhamdhere (Member, IEEE) received the B.E. degree in computer science from Mumbai University, Mumbai, India, in 2002, and the Ph.D. degree in computer science from the College of Computing, Georgia Institute of Technology, Atlanta, in 2009. He is currently a Principal Research Scientist with Amazon Web Services, USA. Before joining to Amazon Web Services, he was a Researcher with the Cooperative Association for Internet Data Analysis, USA. His current research focuses on the structure and dynamics of the Internet topology, interdomain traffic characteristics, Internet economics, and the management/troubleshooting of IP networks.

Takeru Inoue (Member, IEEE) received the B.E. and M.E. degrees in engineering science and the Ph.D. degree in information science from Kyoto University, Japan, in 1998, 2000, and 2006, respectively. He joined Nippon Telegraph and Telephone Corporation Laboratories in 2000 and is currently a Senior Researcher. He was an ERATO Researcher with the Japan Science and Technology Agency from 2011 to 2013. His research interests widely cover algorithmic approaches in computer networks. He is a member of the Institute of Electronics, Information, and Communication Engineers.

Hanan Lutfiyya (Senior Member, IEEE) is a Professor and the Chair of the Department of Computer Science, Western University, Canada. Her research interests include Internet of Things, software engineering, self-adaptive and self-managing systems, autonomic computing, monitoring and diagnostics, mobile systems, policies, and clouds. She was a recipient of the UWO Faculty Scholar Award in 2006. She is currently an Associate Editor of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, and has recently served as the Program Co-Chair for the IEEE/IFIP Network Operations and Management Symposium and the IEEE International Conference on Network and Service Management. She is currently on the steering committee for the Ontario Celebration of Women in Computing Conference. She is a past member of the Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grant Committee, and a past member and the Chair of an NSERC Strategic Grants Committee. She was a member of the Computer Science Accreditation Council.

Taghrid Samak (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from Alexandria University, Egypt, and the Doctoral degree in computer science from DePaul University, Chicago. She is currently pursuing the Juris Doctorate degree with the University of San Francisco School of Law. She worked as a Teaching Assistant, a Research Assistant, and then a Lecturer with DePaul University. In her free time, she volunteers as a mentor for various women in computing organizations. She is a Senior Data Analyst with Google. She applies statistical modeling for a diversity of network applications from capacity planning to wireless networks. Prior to Google, she worked as a Research Scientist with Lawrence Berkeley National Laboratory where her research focused on applying data analysis and machine learning to enable cross-discipline scientific discovery, from modeling application behavior in large-scale systems, to enabling statistical analysis for genomics. She is the Co-Founder and a Steering Committee Member of the Arab Women in Computing Organization.