# Guest Editorial: Special Issue on Novel Techniques in Big Data Analytics for Management

David Carrera, *Member, IEEE*, Giuliano Casale, *Member, IEEE*, Takeru Inoue, *Member, IEEE*
Hanan Lutfiyya, *Senior Member, IEEE*, Jia Wang, *Fellow, IEEE*, Nur Zincir-Heywood, *Member, IEEE*

## I. Introduction

Cloud and network analytics can harness the immense stream of operational data from clouds and networks, and can perform analytics processing to improve reliability, configuration, performance, fault and security management. In particular, we see a growing trend towards using statistical analysis, Artificial Intelligence (AI) and machine learning to improve operations and management of IT systems and networks.

Research is therefore needed to understand and improve the potential and suitability of Big Data analytics and AI in the context of systems and network management. This will not only provide deeper understanding and better decision making based on largely collected and available operational data, but present opportunities for improving data analysis algorithms and methods on aspects such as accuracy and scalability, as well as demonstrate the benefits of machine intelligence methods in system and network management and control. 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.

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

In particular, in this special issue we have accepted 12 papers out of 37 papers submitted to the open call for novel contributions addressing the underlying challenges of *Big Data Analytics for Management*.

## II. Special Issue Overview

The special issue papers span three central areas of *Big Data Analytics for Management*: (i) analysis of network topology and traffic data, (ii) machine learning for network operations, and (iii) analytics in systems and data centers.

D. Carrera is with UPC and Barcelona Supercomputing Center, Spain; G. Casale is with Imperial College London, UK; T. Inoue is with NTT Network Innovation Laboratories, Japan; H. Lutfiyya is with The University of Western Ontario, Canada; J. Wang is with AT&T Labs – Research, US; N. Zincir-Heywood, Dalhousie University, Canada.

### A. Analytics for Network Traffic Monitoring and Analysis

In "A Survey on Big Data for Network Traffic Monitoring and Analysis", D'Alconzo *et al.* [item 1) in the Appendix] present a survey to review the broader challenges and existing body of work at the interface between Big Data and network traffic monitoring and analysis. Coverage of existing research is discussed and used to provide guidelines and research directions for future work.

In "LENTA: Longitudinal Exploration for Network Traffic Analysis from Passive Data", Morichetta *et al.* [item 2) in the Appendix] examine the issue of interpreting large volumes of network data and identifying changes or anomalies in network traffic. They propose a self-learning method based on clustering to help network operators understand traffic characteristics and unexpected network activities.

In "Multivariate Multi-order Markov Prediction with Its Application in Network Traffic Management", Liu *et al.* [item 3) in the Appendix] extend the range of modeling methods for network traffic data by defining a general multivariate multiorder Markov model and a new state transition approach that improve the accuracy of multi-modal predictions. They validate the approach against real traffic data, and in comparison with eigen decomposition techniques.

In "Transparent and Service-Agnostic Monitoring of Encrypted Web Traffic", Brissaud *et al.* [item 4) in the Appendix] consider the use of analytics to address monitoring and detection of user actions based on network traffic. The authors develop a classification technique based on random forest for monitoring encrypted HTTP traffic in a non-invasive fashion, preserving user privacy.

In "Inferring functional topology from time-series of network events", Messager *et al.* [item 5) in the Appendix] seek to examine the functional dependencies between elements of a network and define a data-driven approach based on a time series of emitted events. The authors validate their method against state-of-the-art techniques showing its scalability and increased accuracy.

### B. Machine Learning for Network Operations

Four papers in this special issue focus on machine learning for management of networks.

In "Machine-Learning-Based Routing and Wavelength Assignment in Software-Defined Optical Networks", Martìn *et al.* [item 6) in the Appendix] introduce a machine learning approach for routing and wavelength assignments in optical WDM networks. The authors illustrate in a testbed the concrete

employment of their machine learning classifier, illustrating its benefit and reduced computational requirements.

In "Optimizing and updating LoRa communication parameters: a Machine Learning approach", Sandoval *et al.* [item 7) in the Appendix] explore the use of machine learning techniques for optimal network configuration in LoRAWAN networks. The authors focus in particular on global network configuration to maximize throughput and intelligent configuration dissemination to nodes using reinforcement learning.

In "Greener RAN operation through machine learning", Vallero *et al.* [item 8) in the Appendix] illustrate the use of machine learning approaches to characterize network traffic in radio access networks. The authors investigate the impact of different prediction models on energy savings and quality-of-service, illustrating their benefits for concrete implementation of energy saving strategies.

In "Resource Sharing Efficiency in Network Slicing", Marquez *et al.* [item 9) in the Appendix] explore the problem of moving away from one-size fits-all service offerings to network slices. The study uses a data-driven approach rooted in real-world measurements to understand the achievable efficiency of network slicing architectures.

### C. Analytics in Systems and Data Centers

Three papers in this special issue focus on management, security and anomaly detection in systems and data centers.

In "A Hybrid Deep Learning based Model for Anomaly Detection in Cloud Datacentre Networks", Garg *et al.* [item 10) in the Appendix] define a novel data processing model for Big data-based real-time anomaly detection. The authors combine optimization and convolutional neural networks to obtain a new anomaly detection model for cloud systems.

In "DDoS Detection System: Using a set of Classification Algorithms controlled by Fuzzy Logic in Apache Spark", Alsirhani *et al.* [item 11) in the Appendix] propose a dynamic attack detection system for distributed denial of service. The authors combine fuzzy logic with multiple classification algorithms to detect different attack patterns, studying the accuracy and classification delay trade-off.

In "ISMAEL: Using Machine Learning To Predict Acceptance of Virtual Clusters in Data Centers", Zerwas *et al.* [item 12) in the Appendix] propose a machine learning framework for admission control of virtual clusters in data centers. The paper adopts an approach based on convolutional and deep neural networks to leverage past information from past acceptance decisions.

## 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 for her help with the administrative tasks associated to this special issue.

## APPENDIX

1) A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia, P. Casas, "A Survey on Big Data for Network Traffic Monitoring and Analysis", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
2) A. Morichetta, M. Mellia, "LENTA: Longitudinal Exploration for Network Traffic Analysis from Passive Data", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
3) H. Liu, L. Yang, J. Chen, M. Ye, J. Ding, L. Kuang, "Multivariate Multi-order Markov Prediction with Its Application in Network Traffic Management", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
4) P.-O. Brissaud, J. François, I. Chrisment, T. Cholez, O. Bettan, "Transparent and Service-Agnostic Monitoring of Encrypted Web Traffic", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
5) A. Messager, G. Parisis, I. Kiss, R. Harper, P. Tee, L. Berthouze, "Inferring functional topology from time-series of network events", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
6) I. Martìn, S. Troia, J. A. Hernández, F. Musumeci, G. Maier, R. Alvizu, O. Gonzalez de Dios, "Machine-Learning-Based Routing and Wavelength Assignment in Software-Defined Optical Networks", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
7) R. M. Sandoval, A. J. Garcia-Sanchez, J. Garcia-Haro, "Optimizing and updating LoRa communication parameters: a Machine Learning approach", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
8) G. Vallero, D. Renga, M. Meo, M. Ajmone Marsan, "Greener RAN operation through machine learning", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
9) C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, X. Costa-Pérez, "Resource Sharing Efficiency in Network Slicing", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
10) S. Garg, K. Kaur, N. Kumar, G. Kaddoum, A. Zomaya, R. Ranjan, "A Hybrid Deep Learning based Model for Anomaly Detection in Cloud Datacentre Networks", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
11) A. Alsirhani, S. Sampalli, P. Bodorik, "DDoS Detection System: Using a set of Classification Algorithms controlled by Fuzzy Logic in Apache Spark", IEEE Trans. Netw. Service Manag., vol. 16, no. 3, Sep. 2019.
12) J. Zerwas, P. Kalmbach, S. Schmid, A. Blenk, "ISMAEL: Using Machine Learning To Predict Acceptance of Virtual Clusters in Data Centers", IEEE Trans. Netw. Service Manag., vol. 16, no. 3,, Sep. 2019.

## 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, 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, 2018.

## BIOGRAPHIES



**David Carrera** received the MS degree at the Technical University of Catalonia (UPC) in 2002 and his PhD from the same university in 2008. He is an associate professor at the Computer Architecture Department of the UPC. He is also the Head of the "DataCentric Computing" research group at the 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 (1.5M, 2015-2020), and ICREA Academia award (2015-2020) 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 at IBM Watson (Hawthorne, NY) in 2006, and a Visiting Research Scholar at IBM Watson (Yorktown, NY) in 2012. He received an IBM Faculty Award in 2010. He is an IEEE and ACM member.



**Giuliano Casale** joined the Department of Computing at Imperial College London in 2010, where he is currently a Senior Lecturer in modeling and simulation. Previously, he worked as a scientist at SAP Research UK 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 has served as program co-chair for several conferences in the area of performance engineering, such as ACM SIGMETRICS/Performance and IEEE MASCOTS. His research is recipient of multiple awards, recently the best paper award at ACM SIGMETRICS 2017. He serves on the editorial boards of IEEE TNSM and ACM TOMPECS and as current chair of ACM SIGMETRICS.



**Takeru Inoue** (M'04) 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 (NTT) Laboratories in 2000 and is now 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 (IEICE).



**Hanan Lutfiyya** is a Professor and the Chair of the Department of Computer Science, University of Western Ontario. Her research interests include Internet of Things, software engineering, selfadaptive 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 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. 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.



**Jia Wang** is currently a Lead Inventive Scientist at ATT Labs Research. Her research interests lie in computer networking. She works in the areas of network measurement and service performance analysis, network design and management, software define network, big data for networking. Her research projects expands from traditional IP network, routing, network security, packet classification and firewall optimization, Internet video and social networks, and more recently mobile networks and software defined networks. She received her MS and PhD degrees in Computer Science from Cornell University in May 1999 and January 2001, respectively. She is a Fellow of IEEE and a member of ACM.



**Nur Zincir-Heywood** 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. 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 a co-editor of the book Recent Advances in Computational Intelligence in Defense and Security by Springer. She is an Associate Editor of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and is the Technical Program Co-Chair of the 15th International Conference on Network and Service Management 2019. She has been a Co-Organizer for the IEEE/IFIP International Workshop on Analytics for Network and Service Management since 2016.