POLITECNICO DI TORINO
Repository ISTITUZIONALE

Guest Editorial: Special Section on Advances in Big Data Analytics for Management

Terms of use:

(Article begins on next page)

02 May 2024

Guest Editorial: Special Issue on Advances in Big Data Analytics for Management

Giuliano Casale, *Member, IEEE*, Yixin Diao, *Fellow, IEEE*, Marco Mellia, *Senior Member, IEEE*,
Rajiv Ranjan, *Senior Member, 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, automated configuration, performance, and optimized network management in general. In this area, we have witnessed a growing trend towards using statistical analysis and machine learning techniques 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 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 analytics algorithms and methods on aspects such as accuracy and scalability. 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 deals with *Advances in Big Data Analytics for Management*. It is a second special issue, after the success of last year edition [1]. It presents recent, emerging approaches and technical solutions that can exploit Big Data and analytics in management solutions, as well as platform improvements and specific designs to speed up the processing of large amounts of data that a network typically generates.

We have accepted 10 papers from 36 papers submitted to the open call in this special issue that address the underlying challenges of *Big Data Analytics for Management* and present novel theoretical and experimentation results.

II. Special Issue Overview

The ten accepted papers in this special issue cover three important areas of *Big Data Analytics for Management*: analytics for network management, anomaly detection and security, and analytics platforms and applications.

A. Analytics for Network Management

Four papers in this special issue focus on Analytics for Network management, considering data centers and SDN networks.

In "AWESoME: Big Data for Automatic Web Service Management in SDN", Trevisan *et al.* [item 1) in the Appendix] present a comprehensive web traffic management approach based on a "per service" management concept to identify and prioritize all traffic of important web services. It leverages big data algorithms to automatically build models describing the traffic of web services, and uses the models to install rules in SDN switches to steer all flows related to the originating services. Extensive experimental results are shown using various available traffic traces.

In "Big Data Analysis-based Secure Cluster Management for Optimized Control Plane in Software-Defined Networks", Wu *et al.* [item 2) in the Appendix] explore a big data analysis-based security cluster management architecture for an optimized control plane. To achieve this, the authors propose an ant colony based optimization scheme and a security authentication scheme. Simulations and evaluations show the feasibility and the efficiency of the proposed cluster management architecture.

In "Spatial-Temporal Prediction Models for Active Ticket Managing in Data Centers", Xue *et al.* [item 3) in the Appendix] present a methodology for active ticket managing to achieve efficient time series prediction and capacity planning. In doing so, their goal is to reduce virtual machine and box usage tickets that are issued in production data centers. They evaluate their methodology on a large number of data center production traces as well as on a cluster running MediaWiki.

In "Mining Causality of Network Events in Log Data", Kobayashi *et al.* [item 4) in the Appendix] propose a method to mine causality of network events from large size of heterogeneous network log messages. The proposal leverages a causal inference algorithm that reconstructs causal structures from a set of time series of events. The authors evaluate their method on 15 months of network syslog data obtained in a nation-wide academic network.

B. Anomaly Detection and Security

Three papers in this special issue focus on Anomaly Detection and Security, using Big Data approaches.

In "Detecting Botclouds at Large Scale: a Decentralized and Robust Detection Method for Multi-Tenant Virtualized Environments", Cogranne *et al.* [item 5) in the Appendix] propose a two-step approach based on Principle Component Analysis and statistical hypothesis theory for the detection of malicious activities perpetrated by virtual hosts infected by botnets. They validate their approach on a large scale dataset providing real container traces as well as simulated ones. This enables them to reach to public cloud context in their evaluations.

In "Anomaly Detection in Complex Real World Application Systems", Gow *et al.* [item 6) in the Appendix] present a Whole of Service Anomaly Detection (WoSAD) methodology that is based on a mix of data driven and model based methods for service profiling and measurement. The authors examine anomaly detection efficiency in two case studies using six detection models in a large Australian Financial Services Organisation. The proposed WoSAD methodology is demonstrated to be more efficient than alternative individual transactions and service models.

In "Social Plane for Recommenders in Network Performance Expectation Management", Zhang *et al.*[item 7) in the Appendix] propose a "social plane" that relies on recommended measurements based on "content-based filtering" to run similarity analysis and subscribe to useful measurements, and "collaborative filtering" to share knowledge on anomaly symptoms. The authors show the effectiveness of the social plane approach within a SoyKB Big Data application case study using social network creation and mingling of experts.

C. Analytics Platforms and Applications

Three papers in this special issue focus on Analytics Platforms and Applications, leveraging network properties to improve performance.

In "Efficient Deep Neural Network Serving: Fast and Furious", Yan *et al.* [item 8) in the Appendix] present a dynamic scheduling framework powered by an interference-aware queueing-based analytical model. It identifies and switches to the optimal parallel configuration of the serving system to minimize response latency for deep neural network serving. The benchmark based evaluation demonstrates its good latency prediction accuracy and efficiency as well as its abilities to identify optimal parallel configurations and to adapt to changing load conditions.

In "Group Mobility Detection and User Connectivity Models for Evaluation of Mobile Network Functions", Suzuki *et al.* [item 9) in the Appendix] present a group mobility detection method based solely on signaling data. Then, using this method, the authors build connected/idle duration models for users to characterize network utilization. The simulation results show that the group mobility detection and the connection/idle duration models based on control plane data analytics are useful for the development of mobility-aware functions in base stations.

In "Automatic Generation of Workload Profiles using Unsupervised Learning Pipelines", Prats *et al.* [item 10) in the Appendix] introduce a methodology for discovery of resource consumption phases in data center applications. The authors propose to combine Conditional Restricted Boltzmann Machines and Hidden Markov Models to deliver an unsupervised phase detection method. The technique is shown to be effective in identifying workload phases in Apache Hadoop and Spark workloads, among others.

ACKNOWLEDGMENT

APPENDIX: RELATED WORK

1) M. Trevisan, I. Drago, M. Mellia, H. H. Son, and M. Baldi, "AWESoME: Big Data for Automatic Web Service Management in SDN," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
2) J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "Big Data Analysis-based Secure Cluster Management for Optimized Control Plane in Software-Defined Networks," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
3) J. Xue, R. Birke, L. Y. Chen, and E. Smirni, "Spatial-Temporal Prediction Models for Active Ticket Managing in Data Centers," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
4) S. Kobayashi, K. Otomo, K. Fukuda, and H. Esaki, "Mining Causality of Network Events in Log Data," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
5) R. Cogranne, G. Doyen, N. Ghadban, and B. Hammi, "Detecting Botclouds at Large Scale: a Decentralized and Robust Detection Method for Multi-Tenant Virtualized Environments," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
6) R. Gow, F. A. Rabhi, and S. Venugopal, "Anomaly Detection in Complex Real World Application Systems," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
7) Y. Zhang, P. Calyam, S. Debroy, and S. S. Nguri, "Social Plane for Recommenders in Network Performance Expectation Management," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
8) F. Yan, Y. He, O. Ruwase, and E. Smirni, "Efficient Deep Neural Network Serving: Fast and Furious," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
9) M. Suzuki, T. Kitahara, S. Ano, and M. Tsuru, "Group Mobility Detection and User Connectivity Models for Evaluation of Mobile Network Functions," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.
10) D. B. Prats, J. L. Berral, and D. Carrera, "Automatic Generation of Workload Profiles using Unsupervised Learning Pipelines," IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 1, Mar. 2018.

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.

BIOGRAPHIES

Giuliano Casale received a Ph.D. degree in Computer Engineering from Politecnico di Milano, Italy, in 2006. In 2010 he joined the Department of Computing at Imperial College London, UK, where 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, cloud computing, and Big data, topics on which he has published more than 120 refereed papers. He has served on the technical program committee of over 80 conferences and workshops and as co-chair for conferences in the area of performance engineering such as ACM SIGMETRICS/Performance. He is a member of the IFIP WG 7.3 group on Computer Performance Analysis and since 2015 serves in the ACM SIGMETRICS Board of Directors.

Rajiv Ranjan received the Ph.D. degree (2009) in computer science and software engineering from University of Melbourne, Australia. He has been a Reader (Associate Professor) of Computing Science at Newcastle University since 1 September 2015. Rajiv is an internationally renowned researcher in the areas of cloud computing, Internet of Things (IoT), and big data. He has authorship of about 225 peer-reviewed scientific publications, including publications in IEEE Trans. on Parallel and Distributed Systems, IEEE Trans. on Computers, Journal of Computer and System Sciences, ACM Computing Surveys, and IEEE/ACM World Wide Web conference, PVLDB. He is the inventor of CloudSim (https://goo.gl/SvB0m5) which is the worlds most adopted and cited (2800+ Google citations since 2011) distributed systems simulation and benchmarking framework. Along with researchers from Chinese Academy of Sciences, he invented G-Hadoop (https://goo.gl/1f7oow) - the first-ever framework to support big data processing across distributed datacentres. G-Hadoop successfully integrated, managed, and processed 2.36 PB of remote sensing big data across 8 satellite datacentres, and holds the current world record for remote sensing big data management and processing. He is one of the world's most cited computer scientists (top 0.09% of 2 million researchers) with 9,950+ citations, an H-index of 42 and G-index of 94 (ref: Google Scholar).His Web of Science (H-index of 21 and 2550+ citations) and Scopus (H-index of 25 and 4600+ citations) indices are also internationally leading.
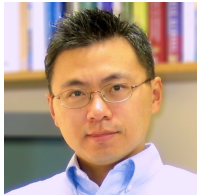
Yixin Diao received the Ph.D. degree in electrical engineering from Ohio State University, Columbus, OH, USA. He is currently a Research Staff Member with IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. He is the author of more than 80 papers in systems and services management and is the coauthor of the book *Feedback Control of Computing Systems*. He has received several Best Paper Awards from IEEE/IFIP Network Operations and Management Symposium, IFAC Engineering Applications of Artificial Intelligence, and IEEE International Conference on Services Computing. He is an Associate Editor for *IEEE Transactions on Network and Service Management*, and *Journal of Network and Systems Management*.

Nur Zincir-Heywood received the PhD degree in Computer Science and Engineering in 1998 from Ege University, Turkey. She is a Full Professor of Computer Science at Dalhousie University, Canada. She is on the editorial board of the IEEE Transactions on Network and Service Management and is the Technical Program Co-chair of IFIP/IEEE Traffic Measurement and Analysis Conference 2018. She has been a co-organizer for the IEEE/IFIP International Workshop on Analytics for Network and Service Management since 2016. Her research interests include data driven techniques for cyber security and network management, topics on which she has published over 200 fully reviewed papers. She has several best paper awards as well as the supervisor for recipient of the IFIP/IEEE IM 2013 Best PhD Dissertation Award in Network Management. She is a recipient of the 2017 Women Leaders in the Digital Economy Award and a member of the IEEE and the ACM.

Marco Mellia received the Ph.D. degree in electrical engineering from Politecnico di Torino, Italy. His current research interests are in the in the area of traffic monitoring and analysis, in cyber monitoring, and Big Data analytics. Marco Mellia has co-authored over 250 papers published in international journals and presented in leading international conferences. He won the IRTF ANR Prize at IETF-88, and best paper award at IEEE P2P'12, ACM CoNEXT'13, IEEE ICDCS'15. He is part of the editorial board of ACM/IEEE Transactions on Networking, IEEE Transactions on Network and Service Management, and ACM Computer Communication Review. He holds a position as Associate Professor at Politecnico di Torino, Italy.