Special Issue on Big Data Applications in Cyber Security and Threat Intelligence – Part 1

Kim-Kwang Raymond Choo, *Senior Member, IEEE*, Mauro Conti, *Senior Member, IEEE*, and Ali Dehghantanha, *Senior Member, IEEE*

THIS last decade has witnessed a tremendous rapid increase in volume, veracity, velocity and variety of data (also commonly referred to as the four V's of big data in the literature[1]) generated by different cyber security solutions and as part of cyber investigation cases. When a significant amount of data is collected from or generated by different devices and sources, intelligent big-data analytical techniques are necessary to mine, interpret and visualize such data. To mitigate existing cyber security threats, it is important for big-data analytical techniques to keep pace.

Therefore, in special issue we focus on cutting-edge from both academia and industry, with a particular emphasis on novel techniques to mine, interpret and visualize big-data from a wide range of sources and can be applied in cyber security, cyber forensics and threat intelligence context. Every submission was reviewed by at least three independent subject matter experts. Of the 40 submissions received, only 20 submissions were selected for inclusion in this special section of *IEEE Transactions on Big Data* (i.e., acceptance rate of 50 percent).

We will now introduce the first 11 accepted submissions (hereafter referred to as papers).

In the paper entitled "Iterative Classifier Fusion System for the Detection of Android Malware", Jemal Abawajy and Andrei Kelarev presented the Iterative Classifier Fusion System (ICFS) for Android malware detection. Specifically, in the system, classifiers are applied iteratively with their proposed iterative feature selection (IFS) procedure. The authors then determined the optimal configurations for ICFS, as well as evaluating the effectiveness of ICFS with competing classifiers.

In the paper entitled "Spamdoop: A privacy-preserving Big Data platform for collaborative spam detection", Abdelrahman AlMahmoud, Ernesto Damiani, Hadi Otrok and Yousof Al-Hammadi presented a privacy-preserving collaborative spam detection platform. The latter leverages a standard Map Reduce facility, and uses a highly parallel encoding technique to facilitate the detection of spam campaigns. The authors evaluated the platforms performance using a synthetic spam dataset.

In the paper entitled "Detecting Anomalous Behavior in Cloud Servers by Nested Arc Hidden SEMI-Markov Model with State Summarization", Waqas Haider, Jiankun Hu, Yi Xie, Xinghuo Yu and Qianhong Wu presented a mathematical model that can be used to model dynamic usage behavior and facilitate the detection of anomalous used behavior. The authors then collected data using IXIA Perfect Storm and the commercial security-test hardware platform cyber range, and evaluated the proposed model's reliability.

In the paper entitled "Big Data for Cybersecurity: Vulnerability Disclosure Trends and Dependencies", MingJian Tang, Mamoun Alazab and Yuxiu Luo presented a framework to take as input historical vulnerability data and analyze and determine the (long-term) dependency (ies) between different vulnerability(ies) and exploit disclosure(s).

In the paper entitled "Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud", Hui Cui, Robert Deng, Yingjiu Li and Guowei Wu presented an attribute-based storage system with secure deduplication in a hybrid cloud setting. In their approach, a private cloud is tasked with duplicate detection and a public cloud manages the storage. The proposed approach allows one to share data with other users based on pre-specified access policies, at the standard notion of semantic security for data confidentiality.

In the paper entitled "Efficient Recommendation of De-identification Policies using MapReduce" Xiaofeng Ding, Li Wang, Zhiyuan Shao and Hai Jin proposed a parallel algorithm, based on MapReduce, to facilitate the computing of skylines over large scale de-identification policies. To achieve improved performance, the authors also designed a novel approximate skyline computation scheme.

In the paper entitled "A Secure Higher-Order Lanczos-Based Orthogonal Tensor SVD for Big Data Reduction",

1. See https://www.ibmbigdatahub.com/infographic/four-vs-big-data (last accessed July 25, 2019).

- *K.-K. R. Choo is with the Department of Information Systems and Cybernetic Security, University of Texas at San Antonio 1 UTSA Cir, San Antonio, TX 28249. E-mail: raymond.choo@fulbrightmail.org.*
- *M. Conti is with the Department of Mathematics, University of Padua, Via Trieste, Padua 63-35121, Italy. E-mail: conte@math.unipd.it.*
- *A. Dehghantanha is with the School of Computer Science, University of Guelph, Guelph, ON N1G 2W1, Canada. E-mail: adehghan@uoguelph.ca.*

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TBDATA.2019.2933039

2332-7790 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Jun Feng, Laurence T. Yang, Guohui Dai, Wei Wang and Deqing Zou designed an orthogonal tensor singular value decomposition (SVD) method, comprising a higher-order Lanczos-based orthogonal tensor SVD algorithm (to reduce higher-order data), and a secure orthogonal tensor SVD method (to outsource the orthogonal tensor SVD algorithm's computation tasks to the cloud).

In the paper entitled "Efficient Secure Outsourcing of Large-scale Convex Separable Programming for Big Data", Weixian Liao, Changqing Luo, Sergio Salinas and Pan Li designed an outsourcing scheme for convex separable programming problems (CSPs). Their approach allows the client to verify the correctness of returned results.

In the paper entitled "SEEN: A Selective Encryption Method to Ensure Confidentiality for Big Sensing Data Streams", Deepak Puthal, Xindong Wu, Surya Nepal, Rajiv Ranjan and Jinjun Chen proposed SEEN, a selective encryption method, to secure big sensing data streams, in order to achieve multiple levels of confidentiality and data integrity.

In the paper entitled "Heterogeneous Data Storage Management with Deduplication in Cloud Computing", Zheng Yan, Lifang Zhang, Wenxiu Ding and Qinghua Zheng presented their heterogeneous data storage management scheme, which can be deployed across multiple cloud providers.

In the paper entitled "Context-driven Granular Disclosure Control for Internet of Things Applications" Arezou Soltani Panah, Ali Yavari, Ron van Schyndel, Dimitrios Georgakopoulos and Xun Yi studied data obfuscation and designed a privacy preservation framework for *Internet of Things (IoT)* environment. Using smart vehicle system as a case study, they demonstrated how data stream watermarking can be utilized to preserve privacy.

In the next issue, we will introduce the remaining nine papers.

Kim-Kwang Raymond Choo
The University of Texas
San Antonio, USA

Mauro Conti
University of Padua
Italy

Ali Dehghantanha
University of Guelph
Canada
Guest Editors

Kim-Kwang Raymond Choo (SM'15) received the PhD degree in information security from the Queensland University of Technology, Australia, in 2006. He currently holds the Cloud Technology Endowed Professorship at the University of Texas at San Antonio (UTSA). In 2016, he was named the Cybersecurity Educator of the Year - APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and in 2015 he and his team won the Digital Forensics Research Challenge organized by the Germany's University of Erlangen-Nuremberg. He is the recipient of the 2019 *IEEE Technical Committee on Scalable Computing (TCSC)* Award for Excellence in Scalable Computing (Middle Career Researcher), 2018 the UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty, Outstanding associate editor of 2018 for *IEEE Access*, British Computer Society's 2019 Wilkes Award Runner-up, 2019 *EURASIP Journal on Wireless Communications and Networking (JWCN)* Best Paper Award, Korea Information Processing Society's *Journal of Information Processing Systems (JIPS)* Survey Paper Award (Gold) 2019, *IEEE Blockchain* 2019 Outstanding Paper Award, *IEEE TrustCom* 2018 Best Paper Award, ESORICS 2015 Best Research Paper Award, 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, Fulbright Scholarship, in 2009, 2008 Australia Day Achievement Medallion, and British Computer Society's Wilkes Award, in 2008. He is also a fellow of the Australian Computer Society and co-chair of IEEE Multimedia Communications Technical Committee's Digital Rights Management for Multimedia Interest Group, he is a senior member of the IEEE.

Mauro Conti (SM'14) received the PhD degree from the Sapienza University of Rome, Italy, in 2009. After the PhD degree he was a post-doc researcher at Vrije Universiteit Amsterdam, the Netherlands. He is full professor with the University of Padua, Italy, and affiliate professor with the University of Washington, Seattle, USA. In 2011 he joined as assistant professor the University of Padua, where he became associate professor, in 2015, and full professor, in 2018. He has been visiting researcher at GMU (2008, 2016), UCLA (2010), UCI (2012, 2013, 2014, 2017), TU Darmstadt (2013), UF (2015), and FIU (2015, 2016, 2018). He has been awarded with a Marie Curie Fellowship (2012) by the European Commission, and with a Fellowship by the German DAAD (2013). His research is also funded by companies, including Cisco, Intel, and Huawei. His main research interest is in the area of Security and Privacy. In this area, he published more than 280 papers in topmost international peer-reviewed journals and conference. He is area editor-in-chief for *IEEE Communications Surveys & Tutorials*, and associate editor for several journals, including *IEEE Communications Surveys & Tutorials*, *IEEE Transactions on Information Forensics and Security*, *IEEE Transactions on Dependable and Secure Computing*, and *IEEE Transactions on Network and Service Management*. He was program chair for TRUST 2015, ICISS 2016, WiSec 2017, and general chair for SecureComm 2012 and ACM SACMAT 2013, he is a senior member of the IEEE.

Ali Dehghantanha (SM'17) received the PhD degree in security in computing. He has served for more than a decade in a variety of industrial and academic positions with leading players in Cyber-Security and Artificial Intelligence. He is currently director of the Cyber Science Lab (http://cybersciencelab.org/) in the University of Guelph, Ontario, Canada. The Cyber Science Lab (CSL) is a research lab focused on advancing knowledge and practice in security and privacy of machine learning systems to build trustable ML agents for a variety of threat hunting, threat attribution and digital forensics tasks. He is eminently qualified in the field of cyber security; he has an EU Marie Curie fellowship in cyber forensics. He is a Certified Information Systems Security Professional (CISSP), a Certified Information Security manger (CISM), and a Certified Cyber Forensics Professional (CCFP). He is a fellow of the UK Higher Education Academy (HEA) and a senior member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.