Introduction to the Special Issue on Data Mining for Cybersecurity

Nathalie Japkowicz
American University

Yuval Elovici
Ben-Gurion University of the Negev

Computer and communication systems are subject to repeated security attacks. Given the variety of new vulnerabilities discovered every day, the introduction of new attack schemes, and the ever-expanding use of the Internet, it is not surprising that the field of computer and network security has grown and evolved significantly in recent years. Attacks are so pervasive nowadays that many firms, especially large financial institutions, spend over 10% of their total information and communication technology budget directly on computer and network security. Changes in the type of attacks and the identification of new vulnerabilities have resulted in a highly dynamic threat landscape that is unamenable to traditional security approaches.

Data mining techniques that explore data in order to discover hidden patterns and develop predictive models have proven to be effective in tackling the aforementioned information security challenges. In recent years, classification, anomaly detection, and temporal analysis have all been used among other techniques to discover and generalize attack patterns in order to develop powerful solutions for coping with the latest threats.

The articles presented in this special issue are quite representative of the field of data mining applied to cybersecurity—both in terms of the tasks and domains that they consider and in terms of the solutions that they propose. Specifically, the tasks represented in this issue include user authentication through biometrics, SCADA systems vulnerability assessment, user action identification in IoT encrypted traffic, and network anomaly and intrusion detection in large computer networks as well as in small ones such as car controller networks. In order to address all the issues surveyed in this volume, a plethora of approaches are presented including ensemble methods, one-class classification methods, text mining, transfer learning, data stream mining, and temporal analyses via neural networks. The principal problems tackled by these techniques are problems of reliability, the need to function in different environments, or adaptability to dynamic conditions either due to natural changes to the systems or to adversarial settings.

The paper by Pisani, Lorena, and de Carvalho entitled “Adaptive Biometric Systems using Ensembles” presents an ensemble of one-class classification algorithms designed to increase the robustness of adaptive user authentication methods that degrade due to changes in biometric features over time.

The paper by Samtani, Yu, Zhu, Patton, Matherly, and Chen entitled “Identifying SCADA Systems and Their Vulnerabilities on the Internet of Things: A Text-Mining Approach” proposes

Published by the IEEE Computer Society
1541-1672/18/$33.00 ©2018 IEEE

text and data mining approaches to identify and assess a large number of dangerous vulnerabilities in SCADA devices that control critical infrastructures.

The paper by Rege, Obradovic, Asadi, Parker, Pandit, Masceri, and Singer entitled "Predicting Adversarial Cyber Intrusion Stages Using Autoregressive Neural Networks" compares the performance of four autoregressive approaches to predict dynamic adversarial movement through a 12-step cyberintrusion chain model.

The paper by Noble and Adams entitled "Real-time Dynamic Network Anomaly Detection" proposes a framework for streaming network anomaly detection that detects anomalies on various edges of the network and combines this information in order to increase the reliability of its detection.

The paper by Grolman, Finkelstein, Puzis, Shabtai, Celniker, Katzir, and Rosenfeld entitled "Transfer Learning for User Action Identification in Mobile Apps via Encrypted Traffic Analysis" proposes a transfer learning method for determining user actions from encrypted traffic over a large range of app versions, mobile operating systems, and device models.

Finally, the paper by Taylor, Leblanc, and Japkowicz entitled "Probing the Limits of Anomaly Detectors for Automobiles with a Cyberattack Framework" proposes an attack framework for automotive cyberattacks and compares the performance of recurrent neural networks and multivariate Markov chains on the detection of attacks generated by this framework.

In summary, this special issue touches upon a very topical problem which will gain more and more prominence as time progresses. The articles published in this volume provide both a comprehensive introduction to the types of issues encountered in the field and present sophisticated solution to tackle them. It is an important read for anyone interested in the state-of-the-art in cybersecurity

ABOUT THE AUTHORS

Nathalie Japkowicz is a professor in the Department of Computer Science at American University, Washington, DC, USA. Contact her at nathalie.japkowicz@american.edu.

Yuval Elovici is a Professor in the Software and Information Systems Engineering as well as the Director of the Deutsche Telekom Laboratories at Ben-Gurion University of the Negev, Beer Sheva, Israel. Contact him at elovici@bgu.ac.il.