Hindawi
Security and Communication Networks
Volume 2019, Article ID 4109836, 2 pages
https://doi.org/10.1155/2019/4109836

Editorial

Big Data Analytics for Cyber Security

Pelin Angin,[1] Bharat Bhargava,[2] and Rohit Ranchal[3]

[1]*Department of Computer Engineering, Middle East Technical University, Ankara, Turkey*
[2]*Department of Computer Science, Purdue University, West Lafayette, IN, USA*
[3]*IBM Watson Health Cloud, Cambridge, MA, USA*

Correspondence should be addressed to Pelin Angin; pangin@ceng.metu.edu.tr

Received 4 August 2019; Accepted 8 August 2019; Published 8 September 2019

The era of Internet of Things with billions of connected devices has created an ever larger surface for cyber attackers to exploit, which has resulted in the need for fast and accurate detection of those attacks. The developments in mobile computing, communications, and mass storage architectures in the past decade have brought about the phenomenon of big data, which involves unprecedented amounts of valuable data generated in various forms at a high speed. The ability to process these massive amounts of data in real time using big data analytics tools brings along many benefits that could be utilized in cyber threat analysis systems. By making use of big data collected from networks, computers, sensors, and cloud systems, cyber threat analysts and intrusion detection/prevention systems can discover useful information in real time. This information can help detect system vulnerabilities and attacks that are becoming prevalent and develop security solutions accordingly.

Big data analytics will be a must-have component of any effective cyber security solution due to the need of fast processing of the high-velocity, high-volume data from various sources to discover anomalies and/or attack patterns as fast as possible to limit the vulnerability of the systems and increase their resilience. Even though many big data analytics tools have been developed in the past few years, their usage in the field of cyber security warrants new approaches considering many aspects including (a) unified data representation, (b) zero-day attack detection, (c) data sharing across threat detection systems, (d) real time analysis, (e) sampling and dimensionality reduction, (f) resource-constrained data processing, and (g) time series analysis for anomaly detection.

This special issue has attracted original contributions that utilize and build big data analytics solutions for cyber security in a variety of fields. All submissions underwent a meticulous review process, and nine papers were accepted for publication in this special issue. The following is a short summary of the findings of each of these papers.

Cyber Physical Power Systems (CPPS) are a critical infrastructure and therefore a favorable target of cyber-attacks. In "VHDRA: A Vertical and Horizontal Intelligent Dataset Reduction Approach for Cyber-Physical Power Aware Intrusion Detection Systems," the authors proposed the use of the Nonnested Generalized Exemplars (NNGE) algorithm and showed that it is among the most accurate and suitable classification methods for developing an intrusion detection system for CPPS because of its ability to classify multiclass scenarios and handle heterogeneous datasets. Furthermore, VHDRA proposed mechanisms to improve the classification accuracy and speed of the NNGE algorithm and reduce the computational resource consumption. It achieves this by vertical reduction of the dataset features by selecting only the most significant features and horizontally reduces the size of data while preserving original key events and patterns within the datasets using the State Tracking and Extraction Method approach.

In "Integrating Traffics with Network Device Logs for Anomaly Detection," the authors presented Traffic-Log Combined Detection (TLCD), which is a multistage intrusion analysis system that overcomes the inefficacy of existing anomaly detection systems that search logs or traffics alone for evidence of attacks but do not perform further analysis of attack processes. TLCD correlates log data with traffic characteristics to reflect the attack process and construct a federated detection platform. Specifically, it can discover the process steps of a cyberattack, reflect the current network status, and reveal the behaviors of normal users.

Experiments with different cyberattacks demonstrated that TLCD provides high accuracy and a low false positive rate.

Role-based access control (RBAC) is a predominant access control model and is widely used in both commercial and research settings. A key requirement of RBAC is to identify appropriate roles that capture business needs. Role mining is a common approach to discover user roles from existing datasets using data mining. The interdependent relationships between user permissions must be considered to prevent security vulnerabilities. In "RMMDI: A Novel Framework for Role Mining Based on the Multi-Domain Information," the authors proposed a role mining framework based on multi-domain information. It utilizes the information from multiple domains such as physical, network, and digital, to find the relationships and similarity between user permissions, and aggregates the interdependent permissions under the same role using multiview community detection methods.

Governments and enterprises are frequently exposed to coordinated cyberattacks such as advanced persistent threat (APT). Such attacks require exploiting multiple systems within an organization to gain unauthorized access to data for an extended period by staying undetected. Detection and prevention of such attacks requires classifying the disparate data from multiple systems based on its semantics and correlating it through a comprehensive analysis. The article titled "HeteMSD: A Big Data Analytics Framework for Targeted Cyber-Attacks Detection Using Heterogeneous Multisource Data" addresses these gaps by complementing the analysis using human security experts. It presents a multilayer design of the framework and discusses the identification of security related characteristics in the data, classification of data based on degree of security semantics, and different types of correlation analysis.

In "Optimizing Computer Worm Detection Using Ensembles," the authors addressed the problem of detecting computer worms in networks. They focused particularly on the problem of detecting sophisticated computer worms that use code obfuscation techniques and developed a behavioral machine learning model to detect computer worms. The achieved results are promising in terms of accuracy and generalization to new datasets.

In "Malware Detection on Byte Streams of PDF Files Using Convolutional Neural Networks," the authors designed a convolutional neural network to tackle malware detection on PDF files. They collected malicious and benign PDF files and manually labeled the byte sequences within them. The proposed network was designed to interpret high-level patterns among collectable spatial clues, predicting whether the given byte sequence has malicious actions or not. The experimental results showed that the proposed approach outperforms several machine learning models.

Due to the numerous benefits of cloud computing, it is becoming the go-to technology for hosting services and storing data. However, the utilization of cloud brings inherent risks and uncertainty due to lack of visibility into the cloud and loss of control over operations applied to shared data. A key requirement in cloud-based data storage and sharing is to ensure the integrity of shared data. In "Integrity Audit of Shared Cloud Data with Identity Tracking," the authors proposed a public auditing scheme for dynamic group-oriented data sharing in cloud environments. They introduced a new role called Rights Distribution Center (RDC) to track the membership and identity of users. The approach enables performing third party audits to verify data integrity while protecting the privacy of user identity.

Automated data mining can help in extracting important information from unstructured text for various cybersecurity use cases. However, lack of a high-quality large labeled dataset has been a hindrance for information security research. Crowdsourcing can be an effective way to quickly obtain a large labeled dataset at low cost, but the crowd annotations may be of lower quality than those of experts. In "Multifeature Named Entity Recognition in Information Security Based on Adversarial Learning," the authors proposed solutions by first identifying the common features in crowdsourced annotations using generative adversarial networks. Due to the diversity and specificity of the entity categories in cybersecurity, only the basic word and character features can be used, but these features alone are not sufficient for effective named entity recognition. To address this, the domain dictionary and sentence dependency features were used as additional features to again identify the entities and improve the quality of crowdsourcing annotations.

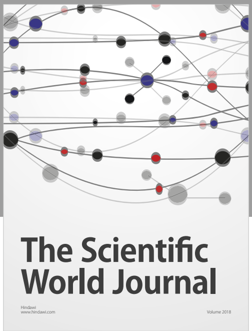
The rise of cloud computing has resulted in data storage and computation being delegated to the untrusted cloud, leading to a series of challenging security and privacy threats. While fully homomorphic encryption can be used to protect the privacy of cloud data and solve the trust problem of a third party, the key problem of achieving fully homomorphic encryption is reducing the increasing noise during the ciphertext evaluation. In "Generalized Bootstrapping Technique Based on Block Equality Test Algorithm," the authors investigated the bootstrapping procedure used to construct a fully homomorphic encryption scheme. They proposed a new block homomorphic equality test algorithm and gave an instance based on the FH-SIMD scheme. Both theoretical analysis and experiment simulation demonstrated the high performance of the proposed bootstrapping algorithm.

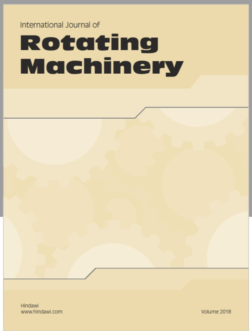
Conflicts of Interest

The Editors declare that there are no conflicts of interest.

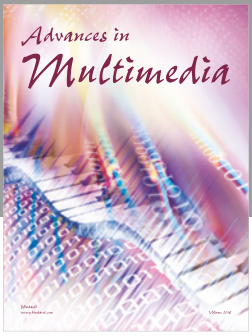
Pelin Angin
Bharat Bhargava
Rohit Ranchal

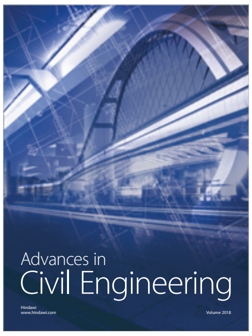
Journal of
Engineering

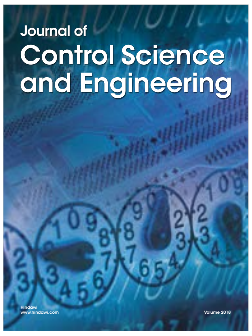

The Scientific
World Journal


International Journal of
Rotating
Machinery

Journal of
Sensors

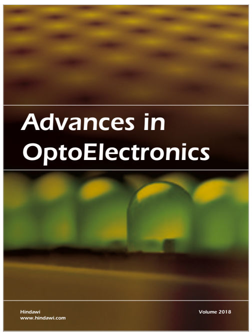

Advances in
Multimedia


Advances in
Civil Engineering


Journal of
Control Science
and Engineering

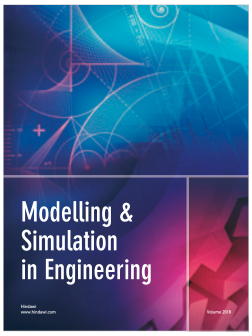
Journal of
Robotics

Journal of
Electrical and Computer
Engineering

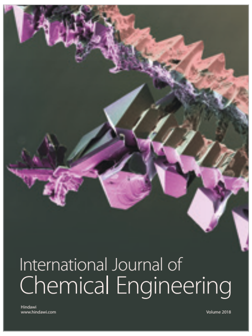

Advances in
OptoElectronics

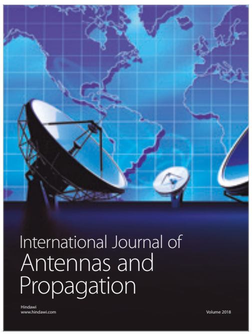
VLSI Design

International Journal of
Navigation and
Observation

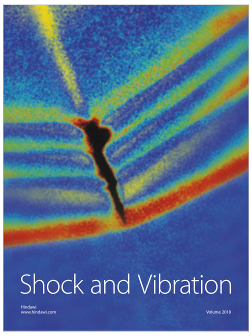

Modelling &
Simulation
in Engineering

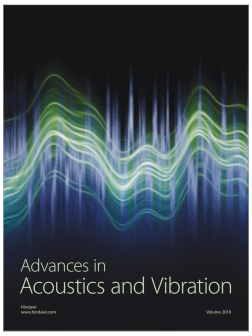
International Journal of
Aerospace
Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration