Guest Editorial Introduction to the Special Section on Artificial Intelligence Security: Adversarial Attack and Defense

ARTIFICIAL Intelligence (AI) has been widely adopted in various applications such as face detection, speech recognition, machine learning, etc. Due to the lack of theoretical explanation, recent works show that AI is vulnerable to adversarial attacks, especially deep neural networks could be easily fooled by adversarial examples that are in the form of subtle perturbations to the inputs. The intrinsic vulnerability of AI might incur severe security problems in areas like automatic driving, face payment, voice command control, etc. Adversarial learning is one typical defense method, which can migrate such security risk of AI by training with generated adversarial examples. However, this method cannot defend the growing number of adversarial attacks.

The special section of "Artificial Intelligence Security: Adversarial Attack and Defense" focused on the state-of-art adversarial attack and defense methods, and explored how these security problems could affect other areas such as cyberspace security, and internet-of-things. Thanks to the extensive efforts of the reviewers, the great support from the former Editor-in-Chief Dr. Dapeng Oliver Wu, and the current Editor-in-Chief Dr. Jianwei Huang, we were able to accept 17 contributed articles covering several important topics, from the adversarial attack and defense methods, applications in cyberspace security, blockchain, Cyber-Physical-Social Systems, and security issues in federated learning, reinforcement learning, and online learning. A brief review of the accepted articles are given as follows:

Choi *et al.* in "EEJE: Two-Step Input Transformation for Robust DNN against Adversarial Examples" consider a two-step input transformation architecture, which not only provided good robustness against various adversarial examples, but also worked well for the legitimate input. They also proposed a practical implementation of two-step input transformation architecture, called EEJE.

Gu *et al.* in "Gradient Shielding: Towards Understanding Vulnerability of Deep Neural Networks" propose iterative gradient shielding method to better understanding adversarial examples against deep neural networks (DNNs). They assumed adding perturbations to the key/sensitive regions of the image could fool image classification DNNs. The gradient shielding method was proposed to verify the assumption and the experimental results corroborated the perspective.

Guo *et al.* in "Coverage Guided Differential Adversarial Testing of Deep Learning Systems" propose DLFuzz, the coverage guided differential adversarial testing framework. It aimed to guide deep learning systems exposing incorrect behaviors, where differential testing was leveraged to avoid manually checking effort and the trouble of collecting similar functional deep learning systems for cross-referencing. They also designed multiple neuron selection strategies to improve neuron coverage efficiently during testing.

Ma and Shi in "AESMOTE: Adversarial Reinforcement Learning with SMOTE for Anomaly Detection" present a combined anomaly-detection framework, which considered the auto-learning ability from reinforcement learning joint with the potentials of class-imbalance techniques to improve detection performance. They introduced an adapted learning strategy on the behaviors of environment agent with techniques such as SMOTE, ROS, NearMiss1 and NearMiss2. The proposed model AESMOTE achieved an outstanding performance with F1 greater than 0.824.

Zhang *et al.* in "Spectral-based Directed Graph Network for Malware Detection" consider a graph-based deep learning method for malware detection, where the weighted graph matrix normalization methods can be performed to transform asymmetrical adjacency matrix so that the multi-aspect graph representations can be learned in a spectral way. They fused the multi-aspect and multi-layer graph representations for accurate classification with a combined loss function.

Pei *et al.* in "Detecting False Data Injection Attacks using Canonical Variate Analysis in Power Grid," proposed a canonical variate analysis-based detection method to rapidly identify false data injection attacks, which are undetected by traditional residual-based detection methods in smart grid state estimation. Both cross-correlation and auto-correlation of received meter measurements were considered among consecutive time slots, and the changes of state variables could be directly obtained to effectively monitor attacks.

Zhu *et al.* in "SEDMDroid: An enhanced stacking ensemble of deep learning framework for Android malware detection" developed a novel framework SEDMDroid to detect Android

Digital Object Identifier 10.1109/TNSE.2021.3073637

malware, based on an enhanced stacking ensemble of deep learning algorithm. This stacking ensemble of deep learning method is a two-tier architecture, which includes an ensemble of base Multi-Layer Perception (MLP) classifiers and a fusion Support Vector Machine (SVM) classifier.

Xia *et al.* in "IDS Intelligent Configuration Scheme Against Advanced Adaptive Attacks" analyze the advanced intelligent collaboration problem among attackers and the optimization problem of configuration between IDSs. Further, they defined an advanced adaptive attack based on intrusion-sharing incentive mechanism, and propose an IDS intelligent configuration scheme based on evolutionary game to detect our defined attack. Lastly, the experimental results showed that the proposed scheme is effective and feasible.

Bosri *et al.* in "Integrating Blockchain with Artificial Intelligence for Privacy-Preserving in Recommender Systems" solved the problem in the current recommender system, which was collecting customer's data without ensuring privacy. They have constructed a client-driven recommendation system using blockchain, where user data transaction records were stored on the blockchain. They have also proposed an incentive mechanism for the users. Companies would provide incentives to the user for using their data to calculate recommendations

Li *et al.* in "A Blockchain-based Traceable Self-tallying E-voting Protocol in AI Era" considered a trust-oriented e-voting application, which requires voters and initiators have strong mutual trust, still facing with over-centralized problem. To address this issue, they proposed a blockchain-based traceable self-tallying e-voting system by taking advantage of an event-oriented linkable group signature and a homomorphic time-lock puzzle, which balanced the anonymity and accountability, the voting scale and efficiency in a decentralized e-voting system.

Doku *et al.* in "On the Blockchain-Based Decentralized Data Sharing for Event Based Encryption to Combat Adversarial Attacks" propose a variant of Attribute-Based Encryption (ABE), called Event-Based Encryption (EBE), that would help avoid adversarial attacks. They introduced a decentralized data sharing network powered by the blockchain technology that ensured data undergoes a thorough vetting process before it was accepted to the network.

Patel *et al.* in "KiRTi: A Blockchain-based Credit Recommender System for Financial Institutions" propose KiRTi, a deep-learning- based credit-recommender scheme for public blockchain to facilitate smart lending operations between prospective borrowers (PB) and prospective lenders (PL) to eliminate the need of third party credit-rating agencies (CRAs) for credit-score (CS) generation, which guaranteed loan grants to PB from PL was secured, authorized, and automated so as to expedite the disbursement process.

Wang *et al.* in "Learning in the Air: Secure Federated Learning for UAV-Assisted Crowdsensing" present a blockchain-based federated learning framework for the trusted knowledge sharing, reliable contribution recording, and prevention of single point failure in UAV-assisted crowdsensing. In the proposed framework, a differential privacy-based data perturbation mechanism was designed for UAV's privacy preservation in learning, and a reinforcement learning-based incentive mechanism was devised to promote UAV's high-quality knowledge sharing.

Chen *et al.* in "Zero Knowledge Clustering Based Adversarial Mitigation in Heterogeneous Federated Learning" address the adversarial issue against federated learning in IoT systems and designed a zero-knowledge clustering based defensive approach to combat with adversarial update, as federated learning has great potential in Internet of Things (IoT)-based systems. The analysis confirmed the convergence and extensive experiments validated the efficacy of the designed approach in comparison with the existing schemes.

Yang *et al.* in "FedSteg: A Federated Transfer Learning Framework for Secure Image Steganalysis" consider a secure federated transfer learning framework, where information sharing can only be encrypted parameters so that image steganalysis can be performed without the disclosure of private data. They proposed a federated transfer learning framework that collaborated all the scattered data form different participants to train a general model, and performed transfer learning to achieve tailored classifier for each participants.

Zhang *et al.* in "Security-Aware Virtual Network Embedding Algorithm based on Reinforcement Learning" propose a secure virtual network embedding algorithm, which was essentially a secure network resource allocation algorithm. They set the security level for network nodes, utilized reinforcement learning agent training to ensure the security and reliability of resource allocation, and optimized the basic performance of virtual network embedding algorithm.

Yin *et al.* in "Online Learning Aided Adaptive Multiple Attribute-based Physical Layer Authentication in Dynamic Environments" consider an online learning aided adaptive PHY-layer authentication framework for enhanced authenticity provisioning. Particularly, the idea of exploring and exploiting multi-attributes was adopted, and artificial intelligence aided search algorithms were formulated to facilitate adaptive selection of the Most Effective PHY-layer Attributes (MEA) through learning their historical authenticity performance.

In summary, the collected articles not only offer fundamental attack and defense algorithms, but also provide innovative application scenarios in various fields, such as cyberspace security, blockchain, federated learning, etc. We hope that this timely special section will trigger more future work in the emerging area.

XIAOJIANG DU, *Guest Editor*
Department of Computer and Information Sciences
Temple University
Philadelphia, PA 19122 USA
e-mail: dxj@ieee.org

WILLY SUSILO, *Guest Editor*
School of Computing and Information Technology
Faculty of Engineering and Information Sciences
University of Wollongong
Wollongong, NSW 2522, Australia
e-mail: wsusilo@uow.edu.au

MOHSEN GUIZANI, *Guest Editor*
Computer Science and Engineering Department
Qatar University
Doha, Qatar
e-mail: mguizani@ieee.org

ZHIHONG TIAN, *Guest Editor*
Cyberspace Institute of Advanced Technology
Guangzhou University
Guangzhou 510006, China
e-mail: tianzhihong@gzhu.edu.cn

Xiaojiang Du (Fellow, IEEE) is currently a tenured Full Professor and the Director of the Security and Networking Lab, Department of Computer and Information Sciences, Temple University, Philadelphia, PA, USA. He has authored or coauthored more than 500 high quality papers. His research interests include security, wireless networks, and systems. He is a Life Member of ACM..

Willy Susilo (Fellow, IEEE) is currently a Professor with the School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, NSW, Australia. He is also the Director of the Institute of Cybersecurity and Cryptology, School of Computing and Information Technology, University of Wollongong.

Mohsen Guizani (Fellow, IEEE) is currently a Professor with the Computer Science and Engineering Department, Qatar University, Doha, Qatar. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is a Senior Member of ACM

Zhihong Tian (Member, IEEE) is currently a Professor and the Dean with the Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou, China. He is also a part-time Professor with Carlton University, Ottawa, ON, Canada. He is a Senior Member of the China Computer Federation. He is a Distinguished Professor with Guangdong Province Universities and Colleges Pearl River Scholar.