IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON SECURITY AND PRIVACY IN EMERGING DECENTRALIZED COMMUNICATION ENVIRONMENTS

Received September 14, 2019, accepted October 5, 2019, date of publication October 22, 2019, date of current version November 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2948912

Machine Learning for Security and the Internet of Things: The Good, the Bad, and the Ugly

FAN LIANG, WILLIAM GRANT HATCHER, WEIXIAN LIAO, WEICHAO GAO, AND WEI YU
Department of Computer and Information Sciences, Towson University, Towson, MD 21252, USA

Corresponding author: Wei Yu (wyu@towson.edu)

This work was supported in part by the US National Science Foundation (NSF) under Grant CNS 1350145, and in part by the University System of Maryland through the Wilson H. Elkins Professorship Award.

ABSTRACT The advancement of the Internet of Things (IoT) has allowed for unprecedented data collection, automation, and remote sensing and actuation, transforming autonomous systems and bringing smart command and control into numerous cyber physical systems (CPS) that our daily lives depend on. Simultaneously, dramatic improvements in machine learning and deep neural network architectures have enabled unprecedented analytical capabilities, which we see in increasingly common applications and production technologies, such as self-driving vehicles and intelligent mobile applications. Predictably, these technologies have seen rapid adoption, which has left many implementations vulnerable to threats unforeseen or undefended against. Moreover, such technologies can be used by malicious actors, and the potential for cyber threats, attacks, intrusions, and obfuscation that are only just being considered, applied, and countered. In this paper, we consider the good, the bad, and the ugly use of machine learning for cybersecurity and CPS/IoT. In detail, we consider the numerous benefits (good use) that machine learning has brought, both in general, and specifically for security and CPS/IoT, such as the improvement of intrusion detection mechanisms and decision accuracy in CPS/IoT. More pressing, we consider the vulnerabilities of machine learning (bad use) from the perspectives of security and CPS/IoT, including the ways in which machine learning systems can be compromised, misled, and subverted at all stages of the machine learning life-cycle (data collection, pre-processing, training, validation, implementation, etc.). Finally, the most concerning, a growing trend has been the utilization of machine learning in the execution of cyberattacks and intrusions (ugly use). Thus, we consider existing mechanisms with the potential to improve target acquisition and existing threat patterns, as well as those that can enable novel attacks yet to be seen.

INDEX TERMS Security, machine learning, cyber physical systems, Internet of Things, applications, distributed environments.

I. INTRODUCTION

The development of the Internet has had an extraordinary influence on the past few decades, globally interconnecting networked devices such as computers, switches, routers, etc. Over the rapid development of the Internet, the scope of Internet connectivity has enabled what is the predominant method of communication and interaction among humans. Internet-based applications provide various content and services, not only improving the efficiency of commerce, manufacturing, and education, among others, but also redefining aspects of human life and work. These dramatic changes to the commercial and living habits of humans depend heavily on the development of the Internet and the various kinds of devices it comprises, most especially, embedded devices that serve as the backbone of Internet service and remote data collection. These embedded devices are network connected to communicate with each other, and have their own characteristics that distinguish them from other traditional computers and network devices, such as limited computation capacity and power supply. Therefore, it is necessary to extend the Internet to allow the connection and interaction of heterogeneous devices.

The Internet of Things (IoT) is clearly the next evolution and extension of the Internet, massively integrating IoT devices (sensors, cameras, smartphones, etc.) to communicate across network infrastructures [1], [2]. The fundamental purpose of IoT is providing a networking platform so that

The associate editor coordinating the review of this manuscript and approving it for publication was Shui Yu.

data from the physical environment can be captured, shared, and analyzed to create precise digital models for a variety of things [3], [4]. These digital models can then be shared and leveraged to analyze the status of target environments and predict or simulate real-world events. IoT technologies and devices are being deployed ever more widely, in many different industries, and the volume of collected data has increased prodigiously. Based on such technology and data, vertical architectures of application, networking, and physical layers are deployed, denoted as Cyber-Physical Systems (CPS), and are widely leveraged in different application domains, including energy, transportation, city infrastructure, healthcare, manufacturing, and home automation, among others [5]–[15]. It is worth noting that CPS are closed-loop systems that utilize IoT devices to gather data, which is then analyzed by CPS applications, and the results of which are used to control the CPS themselves. Specifically, enabled by IoT, CPS are critical infrastructure systems comprised of a vertical hierarchy of command and control, networking, and sensing and actuation nodes. The complex networking and sensing/actuation systems are enabled and achieved via IoT [16], while the particular command and control systems and device architectures are often unique and domain-specific.

Big data analysis is one of the important keys to enabling CPS [3], [4], [17], [18]. Because of the development of IoT, the volume of data is massively increasing, yielding significant benefits and challenges. In particular, the data analysis results can be more informative based on the comprehensive and large volume of collected data. Nonetheless, the huge data collection poses formidable pressure in the data training process, where not all data is reliable, timely, or valuable. Therefore, machine learning, as a powerful yet complex data analysis tool, is critical to CPS. Furthermore, big data analysis plays a key role in achieving automation for CPS, enabling CPS to automatically adapt to new situations according to the results of big data analysis that is based on machine learning.

However, utilizing machine learning in CPS creates uncertainties, especially in the realms of security and privacy. As a persistent concern in all computing systems, security is critical for continued reliability, confidentiality, integrity, and availability [19]. Most pressing is the critical, unforeseen, and unheeded vulnerabilities in machine learning and CPS/IoT systems in use today. For instance, in machine learning, a variety of adversarial techniques have been demonstrated repeatedly for thwarting the inference of trained models, such as attacks that modify input [20] and attacks that learn to produce edge cases and induce incorrect evaluations [21]. Likewise, in IoT/CPS, weak and exploitable software implementation coupled with limited device resources make IoT nodes easy targets that can be compromised on a global scale [22]–[24].

Focusing on the benefits and challenges of introducing machine learning techniques into cybersecurity and CPS, we categorize the use of machine learning in cybersecurity and CPS by its attributes and roles, namely as: the Good, the Bad, and the Ugly. Using CPS as an example, the ''Good'' represents the improvements and benefits brought by leveraging machine learning to assist CPS. The ''Bad'' represents the adversarial attacks against machine learning itself to maliciously deceive the training process and induce erroneous results. Finally, the ''Ugly'' represents the utilization of machine learning as tools to attack generic systems, such as CPS. Based on the three defined roles, we discuss the impacts of introducing machine learning into cybersecurity and CPS. In this work, we survey the spectrum of research at the intersection of machine learning, CPS, and cybersecurity, and attempt to elicit potential solutions and best practices.

Specifically, our contributions are summarized as follows:

- We consider the application of machine learning in cybersecurity and CPS contexts. We assess good, bad, and ugly uses of machine learning. Concretely, we assess the use of machine learning to achieve positive goals (good), attacks against machine learning systems and life-cycles (bad), and the use of machine learning to conduct or aid in attacks (ugly). Note that the last two cases are of particular importance as numerous machine learning systems are already deployed in production systems and applications, and defending against intelligent machine learning based attacks is unprecedented.
- We identify areas in need of further research in each category (good, bad, and ugly) of use, particularly toward the improvement of CPS and cybersecurity. Necessary developments include the improvement of mature machine learning to be more efficient, transferable, and auditable; the improvement of defensive techniques against adversarial attacks and the ex-filtration of trained data; and techniques to segment and containerize machine learning processes to resist attacks. The maturity of research and real-world use of machine learning systems have not mitigated the concerning weaknesses of machine learning systems to attack, nor generated defenses against the novel and emerging use of machine learning in aiding attacks.

The remainder of this paper is as follows. In Section II, we outline our analytical framework for the study of machine learning in cybersecurity and CPS. In Section III, we consider machine learning for good in cybersecurity and CPS, as well as providing a brief overview of machine learning techniques. In Section IV, we present flaws and vulnerabilities in machine learning in the context of cybersecurity and CPS. Furthermore, we survey the existing adversarial learning techniques and provide a vision for protecting the machine learning process. In Section V, we study how adversaries utilize machine learning as tools to launch and achieve attacks. Finally, we provide some concluding remarks in Section VI.

II. PROBLEM SPACE AND OVERVIEW OF MACHINE LEARNING TECHNOLOGIES

In this section, we first outline the problem space of applying machine learning techniques for cybersecurity and CPS.

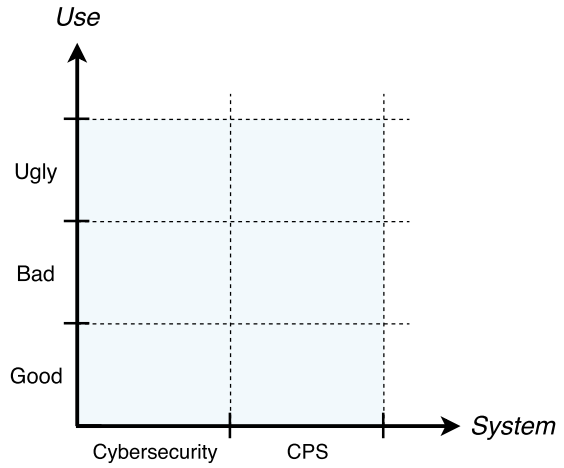


FIGURE 1. A Framework for analyzing the interrelationships of Machine Learning use in Cybersecurity and CPS.

We then provide an overview of the dimensions of machine learning schemes.

A. PROBLEM SPACE

Considering the diverse applications of machine learning paradigm, this paper primarily focuses on the ramifications for two areas, namely Cybersecurity and CPS. Here, we consider systems to be any generic ubiquitous networks or Internet connected systems (smart home, smart watch, smartphone, etc.), as well as CPS enabled by IoT technologies. In the case of the Cybersecurity domain, we seek to broadly address machine learning applications for security issues as applied in the Internet and intranet scenarios. As shown in Fig. 1, we assess the use of machine learning schemes in the three main categories: good, bad, and ugly.

To be specific, the ''good'' use of machine learning indicates that the benign or positive application of machine learning for the improvement of task-specific performance or the implementation of beneficial applications. Example cases include improving medical diagnostics, progressing the state-of-the-art in computer vision and speech services, and others [25]–[29]. Likewise, the ''bad'' use of machine learning is deemed as the potential risks in the process of machine learning, such as data collection, preprocessing, training, implementation, and decision-making, as well as training processes (i.e., training, testing, validation, and implementation) that are vulnerable to subversion and exploitation, leading the output of machine learning models and the resulting decision-making to be incorrect or wholly unreliable. Finally, we define the ''ugly'' use of machine learning as malicious uses of machine learning for negative or detrimental applications or software, such as in improving the efficacy of malware, and aiding in the subversion of automated computer systems and security detection mechanisms, among others.

Recent works have highlighted the vulnerabilities of overall machine learning systems, as well as machine learning models at all levels. Of particular note, the implementation of machine learning models provides users with the ability to apply input and retrieve output, enabling users to potentially infer model operation given enough time and effort. Such mechanisms allow users to determine, through trial and error, where edge cases exist and for what cases is the model not well suited. Nonetheless, models can be duplicated, extracted, and attacked during the processes of all levels. While significant efforts have been devoted for ''good'' uses of machine learning, the ''bad'' and ''ugly'' cases are emerging, becoming more sophisticated and applicable to models in production software systems that deserve more attention.

B. DIMENSIONS OF MACHINE LEARNING SCHEMES

As a number of existing works have discussed in thorough detail the various machine learning schemes, we only present a brief categorical review. The typical learning methods are supervised learning, unsupervised learning, and reinforcement learning schemes. Based on the outcome of learning tasks, we categorize different dimensions of machine learning algorithms by classification, regression, dimensionality reduction, clustering, and density estimation, among others [30].

1) SUPERVISED LEARNING

Typical supervised learning algorithms learn a function/model that maps inputs to outputs based on labeled training input-output pairs. They analyze the training data and infer a function that can be used for mapping new data samples. Based on whether target labels are discrete or numeric, the learning process is defined as classification and regression, respectively.

- **Classification:** The output of the classification is a finite set of categorical classes. These classes can be binary (two classes), such as anomaly detection, or multitudinous, such as handwritten number recognition and others.
- **Regression:** The output of regression tasks is continuous values for the examined instances. Example results might include a 97% probability that the object is malware, and a 3% probability it is not.

2) UNSUPERVISED LEARNING

The main difference between supervised learning and unsupervised learning is the availability of class labels. In unsupervised learning, all input samples are unlabeled, and the evaluation of the trained model will not necessarily rely on the accuracy of mapping input to output classes, but on achieving some broader goal. This category includes the learning tasks of dimensionality reduction, clustering, and density estimation, among others.

- **Dimensionality reduction:** In this case, the target is discriminant analysis. Typically, dimensionality reduction can be utilized in auto-encoders and reduce the dimensionality of the input data. In addition, they can be used to reduce noise or redundant data in video [31].

- **Clustering:** This is generally utilized to group data using mathematical, probabilistic, or statistical means. This is performed through alternatively selecting cluster centroids and cluster membership. Examples include real-time image registration using Self-Orienting Feature Maps (SOFMS) [32], and the combination of TSK-DBN fuzzy learning (i.e., Takagi-Sugeno-Kang(TSK) system with Deep Belief Network (DBN)) [33].
- **Density estimation:** This is basically the statistical extraction or approximation of a target data distribution. Some examples of density estimation include noise reduction of binaural assisted listening devices [34] and utilizing CNNs to estimate the traffic density of intersections through analyzing heterogeneous distributed video [35].

3) REINFORCEMENT LEARNING

Reinforcement learning is considered to be somewhere between supervised and unsupervised learning, as the input of reinforcement learning has no label information, but instead is associated with a reward value, such that each execution improves the decision-making of the overall model, and typically by maximizing the rewards. This can be represented by the perception-action-learning loop. The two major reinforcement methods are policy search and value function approximation.

- **Policy search:** This is the search for an optimal policy using gradient-based or gradient-free methods. For example, Google's Alpha Go is based on policy search, and can learn without any human intervention or interaction and still achieves superiority [36].
- **Value function approximation:** This method estimates the expected rewards of actions and attempts to reach an optimized learning process and results. The key component of the value function is the state-action value function, known as the quality function [37].

III. POSITIVE MACHINE LEARNING: GOOD USE

Modern machine learning, and deep learning techniques, powered by deep, convolutional, and recursive neural networks, have developed rapidly with the availability of increased computational power and large collections of data, and have been leveraged in a variety of predictive and analytic applications for commercial and consumer use, such as self-driving vehicle systems, natural language assistants and interpreters, and others [38], [39]. These applications are transformative, enabling truly smart systems and advancing the state-of-the-art in a variety of areas. As discussed in Section I, the development of IoT and CPS enable smart applications through the collection, storage, and analysis of massively large, distributed datasets. In this section, we focus on the leveraging of machine learning technologies for positive uses in Cybersecurity and CPS. We first study the approaches of utilizing machine learning to detect network attacks and improve network security. Then, we illustrate the benefits of utilizing machine learning techniques in CPS.

A. MACHINE LEARNING IN CYBERSECURITY

Machine learning models, while requiring significant time and efforts to train and test, can be easily deployed to conduct basic inference. Moreover, machine learning models are significantly improved by training on larger, more representative datasets. As an area of particular need for active analysis and discovery, intrusion detection is required to maintain secure systems and to notify system users and administrators of unintended access so that further actions (attribution, mitigation and removal) can be carried out. In the following, we investigate the details of using different machine learning techniques to detect intrusion attacks.

The Intrusion Detection System (IDS) is a typical system designed to monitor protected networks and systems for malicious activities, and is an important approach to protecting cyber infrastructures and enforcing system security [40]–[42]. With the increasing development of machine learning, the integration of machine learning has improved the performance of IDS and enabled the detection of unforeseen intrusions. We now discuss the utilization of various machine learning techniques in both misuse detection and anomaly detection.

1) MISUSE DETECTION

In the conventional IDS, misuse detection compares the current activities against large databases of attack signatures. Essentially, misuse detection utilizes existing attack records to detect anomalies. There have been a number machine learning techniques developed for this process [43]–[46]. For example, Hodo *et al.* [44] integrated Artificial Neural Networks (ANNs) for misuse detection, primarily using supervised and unsupervised procedures to improve the performance of IDS. In the supervised learning procedure, the learning model is trained based on a labeled training dataset. In the unsupervised learning procedure, an unlabeled dataset is used to train the learning model, and the model utilizes the Self-Organizing Map (SOM) to finalize the results without human intervention. Furthermore, an ANN model with three layers and two learning algorithms were proposed. Their evaluation results showed that, based on the ANN model, their IDS performance reached 99.4% (as compared with 96.3% without ANN). In addition, Lin *et al.* [43] proposed a cluster center and nearest neighbor (CANN) approach to combine multiple learning techniques and improve detection performance. Specifically, their developed approach focused on extracting representative features of the attack by computing two metrics. The first metric is the distance from the data sample to its cluster center. The second metric is the distance from the data to the nearest neighbor within the cluster. Furthermore, the approach rebuilds the data features by using the distances, and formats the data features as a k-Nearest Neighbor (k-NN) classifier. Based on the experimental results, the CANN classifier performed better than the k-NN classifier on the KDD-Cup 99 dataset.

In addition, there are a number of studies that have focused on reducing the learning cost to optimize the performance of misuse detection [47]–[49]. For instance, Subba *et al.* [47] proposed an ANN-based misuse detection technique that integrates an intelligent agent in the system. The intelligent agent can identify the underlying patterns in both abnormal and normal data, and simplifies the learning process into new and unseen audit records. The proposed approach could achieve better performance than other intrusion detection models based on SVM, Naive Bayes, and C4.5 algorithms. Likewise, Alheeti *et al.* [49] proposed an ANN-based misuse detection approach to protect vehicular ad hoc networking (VANET). Since the VANET is dynamic, open, wireless, and has no fixed security infrastructure, it is more vulnerable to attacks. The proposed misuse detection approach identifies abnormal system behaviors by tracking key features in various files and by training on a dataset of abnormal features. Moreover, the extracted features from the trace files are auditable data, and the method can increase the detection speed.

2) ANOMALY DETECTION

In the case of anomaly detection, a number of existing research efforts have demonstrated the successful use of a variety of machine learning techniques [50]–[53]. For example, Jiang *et al.* [50] leveraged a deep learning technique to carry out the detection of virtual MAC spoofing attacks. Bontemps *et al.* [51] investigated a Long-Short Term Memory (LSTM) model to detect abnormal activities in the cyber environment. Specifically, while a number of previous studies carried out anomaly detection by training on normal and anomalous behaviors, their study focused on designing a mechanism that considers or recalls recent detection results in continuous data. As a solution to enabling continuous anomaly detection, the LSTM RNN-based detection method was designed to use time series data to train the model and adopt a circular array to detect collective anomalies. The method leverages the average error of the circular array as a threshold. If the predicted error is greater than the threshold and increases continuously, this indicates a collective anomaly.

In addition, Bivens *et al.* [54] proposed a comprehensive IDS which includes a preprocessing stage, clustering of normal traffic, normalization, an ANN training stage, and an ANN decision stage. In the training stage, it utilizes the Self-Organizing Map (SOM), in order to train the model by time period. After the learning process, the model can be used to identify abnormal activities. Meanwhile, a new SOM starts to learn a new traffic pattern and train a new multilayer feed-forward perceptron (MLP) attack classifier. The evaluation results reported 100% successful identification of normal behaviors and 76% successful identification of abnormal behaviors. Likewise, Yin *et al* [52] proposed a recurrent neural network IDS (RNN-IDS) to increase the identification rate. They deeply studied the performance of using binary classification and multi-class classification for the RNN model. By identifying the number of RNN layers and neurons, the proposed RNN used binary classification to increase the identification rate of the anomaly detection. Also, Tian *et al.* [53] designed a distributed deep learning-based detection scheme that can be deployed on edge devices to deal with web attack detection.

Other studies have focused on designing efficient and flexible learning approaches to handle different situations. For instance, Javaid *et al.* [55] proposed a Self-Taught Learning (STL)-based deep learning network to handle unforeseen and unpredictable attacks. Based on the experimental results, the investigated approach could outperform Support Vector Machines, Naive-Bayes, Random Forests, and Self-Organized Maps. Our prior work [56] focused on identifying the correlation and dependency in time series data, aiming at carrying out anomaly detection. Also, we developed a joint optimization framework, which aimed at optimizing the Variational Auto Encoder, the Deep Belief Network, and the Gaussian mixture model together, leading to detection performance improvement [57].

3) MALWARE DETECTION

Furthermore, smartphones are a critical source of data associated with users and businesses, making up the bulk of current contributions, yet security and privacy issues are dramatically dire in mobile devices [58], [59]. To combat such threats, a number of research efforts have been carried out. For example, Booz *et al.* [60] designed a mechanism to optimize deep learning model so that the accurate detection Android malware can be realized. Also, Comar *et al.* [61] combined both supervised and unsupervised learning schemes to detect malware. Additionally, McGiff *et al.* [62] designed a multi-input multimodal detection scheme based, leveraging both permission and hardware feature data to improve malware detection accuracy. Similarly, Yuan *et al.* [63] developed a deep learning framework based on DBN to detect malware in Android apps, and achieved approximately 96.5% detection accuracy.

Considering other types of devices, Ding *et al.* [64] proposed a scheme to extract opcode sequences from Windows PE files in order to detect malware via DBN, illustrating the ability of DBNs to conduct classification tasks. Uwagbole *et al.* [65] designed a system that utilizes static and dynamic deep learning mechanisms to deal with SQL injection attacks. Their system utilized logistic regression to train the learning model and obtain the attack features. Meanwhile, Kim *et al.* [66] focused on leveraging the LSTM neural network model as a component in an IDS, which achieved better performance in network attack detection than other comparable methods, as applied to DoS attacks. As evidenced by these diverse research methods and techniques, enabling the detection of ongoing attacks such that response and mitigation are possible in real time is critical and challenging.

B. MACHINE LEARNING IN CPS

Besides being applied to cybersecurity, machine learning can be applied to address control, networking, and computing

challenges in CPS. In this subsection, we first review existing studies that leverage machine learning for CPS, and then introduce challenges and future research directions.

1) SOLVING SECURITY ISSUES IN CPS

Security and privacy are major concerns in CPS systems and applications. Industry 4.0, smart homes, smart grids, and many other smart-world applications necessarily require capable and resilient security mechanisms, given the nature of the critical infrastructures that they monitor and control [23]. Moreover, cyber-attacks that target CPS typically display common features, such as false data injection attacks, also known as data integrity attacks, which aim to mislead the targeted CPS, such as smart grid, and smart transportation, among others, by manipulating or injecting erroneous data [67]–[77]. Using the smart grid as an example, false data injection attacks have been shown to greatly impact key functional components in the power grid, such as state estimation [68], [69], optimal power flow [71], and energy price [72], [78], along others. Given the explosion in scale of data collection and distribution in CPS, malicious data samples will inevitably be collected. Machine learning technologies, with unprecedented analytical abilities, have the potential to analyze large input datasets and improve security in CPS.

Specifically, we can leverage malicious data or identified features as inputs to train machine learning models to detect cyberattacks. For instance, data integrity attacks are critical threats in power grid systems, which are typical energy CPS. As a mechanism to detect data integrity attacks in the AC power grid, An *et al.* [79] proposed a deep reinforcement learning-based scheme. Their study focused on leveraging deep-Q-network detection (DQND) in the main network and a target network so that the model can be adjusted to optimize the defensive strategy. To seek efficiency of learning process, a quantification of the observation space sliding window was utilized. Their experimental results demonstrated that the accuracy of the proposed scheme in several IEEE bus systems was better than some existing detection schemes. Similarly, He *et al.* [80] focused on detecting data integrity attacks. They proposed an optimized model to capture the behavior of one type of false data injection attack in the power grid system. In detail, they leveraged the Conditional Deep Belief Network (CDBN) to learn the historical behavior patterns of false data injection attacks. Distinguishing their work from other applications of CDBN, their proposed mechanism was designed as a classifier, while other CDBN mechanisms were designed for time-series data. Moreover, to reduce the complexity of training, the proposed CDBN utilized the Conditional Gaussian-Bernoulli RBM (CGBRBM) technique applied to the first hidden layer. Finally, the experimental results demonstrated the superiority of their scheme in the IEEE 118-bus test system.

Additionally, machine learning technologies have their own privacy issues, especially in distributed machine learning systems [81], [82]. The transmission of data and parameters between distributed computing nodes and central server poses serious privacy implications. Related to this issue, Shokri and Shmatikov [82] proposed a solution to deal with issues of privacy protection in distributed learning models. Their solution aims to enable the transmission of parameters while maintaining privacy and accuracy. Zhang *et al.* [81] developed a privacy-preserving scheme that enables multiple users to conduct collaborative deep model training. Also, Yu *et al.* [83] proposed a differentially private approach that optimized both privacy loss and model accuracy.

2) CHALLENGES OF LEVERAGING MACHINE LEARNING FOR CPS SECURITY

Despite the solutions described above, challenges still persist in regard to security issues in CPS. We now discuss the challenges of leveraging machine learning to solve such security issues.

- In general, there are several prerequisites for establishing an effective machine learning model. It is necessary to balance high detection capabilities against resource consumption. Additionally, machine learning models need to be supplied with relevant and timely data of high quality in order to obtain accurate results. Moreover, data streams supply continuous input to enable real-time detection, but are difficult to verify on-the-fly.
- The diversity and rapid evolution of malware and malicious code increase the difficulties of identification and detection. Malware exist that can be copied to 3 million new samples in an hour, and some new attacks are able to bypass end-point detection and can be launched at variable rates [84]. At the same time, the training process of machine learning is generally tedious, and while it is possible to detect new malware in the wild, this is only possible with a trained model. Moreover, such detection is obviously unforeseen, and often adversaries use malware detection systems to test their malware against. Simply put, the evolution of malware is continuous, and how to develop and update well-trained machine learning models to enable effective detection is a challenging problem.

Towards solving some of the aforementioned issues, potential solutions include using high-dimensional data and incremental learning for non-stationary data. Using high-dimensional data can increase model complexity, accuracy, and diversity of the features for malware fingerprinting. Nonetheless, processing the high-dimensional data is computationally expensive. Related to this issue, Chen *et al.* [85] proposed marginalized stacked denoising autoencoders (mSDAs) for this purpose. Their mSDAs have two key parameters with which to control the dimensionality of the data, namely the amount of noise and the number of layers, which simplify the training process and reduce training time.

Meanwhile, to increase training speed and handle incremental learning, online learning strategies have been

proposed [86], [87]. For example, Liang *et al.* [86] investigated an online learning strategy that utilized slices of continuous data to update machine learning models in order to fit the new datasets dynamically. Likewise, Zhou *et al.* [87] proposed an online learning algorithm that implements the denoising autoencoder, and proved that the proposed algorithm could increase the convergence rate by using incremental datasets. In addition, distributed learning platforms can extend computation capabilities, offering another potential solution.

3) IMPROVING THE PERFORMANCE OF CPS

Recall that machine learning has demonstrated an impressive capacity for data analysis and for obtaining unprecedented insights and intelligence. Clearly, machine learning has the potential to be utilized in a variety of systems for a diverse array of applications, including natural language analysis, image and video recognition, smart transportation, smart grid, wireless networks, and others [28], [39], [88]. In addition, machine learning can be leveraged for control and network systems to promote system automation [39], [89].

A number of existing studies have sought to utilize machine learning in CPS [90]–[94]. One particular focus has been on optimizing network scheduling. To this end, Jiang *et al.* [90] investigated existing machine learning techniques and leveraged a variety of learning techniques to improve network performance in IoT systems. Specifically, they utilized machine learning to optimize scheduling for the IoT network. Also, Zhu *et al.* [91] proposed a Q-learning based network scheduling algorithm to select appropriate operations and leverage multi-channel network resources to optimize system throughput. In addition, Lopez-Martin *et al.* [92] proposed a scheme to classify network traffics, leveraging machine learning to analyze package headers, and then classify traffic into different groups so that high priority groups can prioritize network resources.

Another research direction is the optimization of computing resource management [95]. One way to optimize computing resources is to design and deploy distributed platforms and algorithms [96]. For example, Liu *et al.* [97] utilized deep reinforcement learning to solve the resource allocation problem adaptively in the cloud computing system. Their mechanism considers the problem of minimizing power consumption management through effective dynamic power management via machine learning implementation. He *et al.* [98] focused on improving Quality of Experience (QoE) by optimizing computing resource allocation. Specifically, they denoted the optimized computing resources as ''green resources'', achieving their goal through the proposed deep learning based green resource selector, which selects best-fit resources.

In addition, there are a number of more machine learning applications that assist CPS operations, such as the co-design or integration of networking and control components in industrial IoT, in the field of energy generation and monitoring, and traffic congestion prevention, among others [99]–[107]. For instance, Xu *et al.* [99] designed reinforcement learning based technique that could enable the automatic configuration of control and networking components in manufacturing CPS. Liang *et al.* [101] investigated an edge computing based machine learning technique that enables the automatic recognition of components in the manufacturing process so that industry automation can be supported.

A number of distinct deep learning models have been proposed to predict peak electricity usage and related environmental factors, and provide reasonable recommendations for the power suppliers [100], [102], [103], [105], [108]. For example, Yu *et al.* [102] designed a statistics-based technique to understand the statistical distribution of energy usage. The authors also leveraged machine learning techniques to carry out energy usage forecasting. Also, in the field of smart transportation, Lv *et al.* [106] leveraged machine learning to predict traffic congestion, while Ma *et al.* [107] proposed a CNN model to analyze traffic congestion using images of real-time traffic conditions.

4) CHALLENGES OF LEVERAGING MACHINE LEARNING FOR CPS

Similar to applying machine learning for cybersecurity, issues persist in leveraging machine learning for applications in CPS. First, a single machine learning model alone may not fit all tasks in one or many situations that need to be addressed. Generally speaking, one particular machine learning model is trained for a specific problem, or at most can be retrained to another similar task. Moreover, CPS are themselves diverse [109], making it difficult to generalize one machine learning model to cover all situations. Thus, a variety of models and diverse data are necessary for system-wide solutions.

Second, the deep neural network training process is generally a black box process [110], through which we obtain output from input without knowing how exactly the model obtains the results, though we do know the parameters and weights manipulated in the training process. Indeed, for distinctly complex models, the tracking of each input's contribution to each neuron's updates is generally infeasible, as the parameter transmission from one layer to another layer is complex. In CPS, data is often in the form of data streams, dynamically generated in real-time to be stored, examined, analyzed, and eventually trained upon by learning models. Thus, how to dynamically adjust parameters in neural networks is a challenge.

Third, edge computing technologies provide flexibility and redundancy for CPS, and enable critical real-time decision-making and actuation [111]. Nonetheless, edge computing nodes may not be able to support the deep learning training process due to their limited computation power in comparison with cloud infrastructures. Thus, how to optimize deep learning models and systems to reduce computation requirements and increase efficiency in distributed learning are key challenges that are currently being investigated.

Furthermore, the stability of deep learning models in edge computing infrastructures is particularly important, and thus, it is necessary to develop error recovery schemes for distributed machine learning.

Finally, machine learning models have strict requirements for the sizes, shapes, and types of input data. Even though CPS collect massive data, the quality of such data may not be guaranteed, especially as the lifetime of newly created IoT hardware will be unverified. The input data for machine learning must be transformed from raw data into some particular data format, a process that incurs massive computation costs, or the machine learning systems must be able to cope with such raw data and noise inherently. As discussed, data in CPS systems is continuously collected by a variety of heterogeneous sensors, and how to handle the raw data is an enormous challenge.

C. SUMMARY AND FUTURE DIRECTIONS

Based on the study and discussion in this section, machine learning techniques can significantly improve the performance of cybersecurity and CPS. Specifically, utilizing machine learning techniques can improve intrusion detection performance (detection speed, accuracy, flexibility, etc.) by training on historical intrusion datasets. Likewise, in CPS, machine learning techniques are not only able to improve the security of the system, but are also able to assist the system to achieve automation.

Nonetheless, there are a number of unresolved challenges.

- First, the training process is time and computationally expensive and traditional machine learning cannot handle dynamic systems. For example, intrusion detection systems are dynamic systems, in which new training data are continuously generated. Training on the new data continuously takes significant time, is mostly inefficient, and may degrade in performance. Similarly, CPS are also dynamic systems, and while some studies have leveraged machine learning techniques to address problems in CPS, the effectiveness and efficiency of machine learning in dynamic CPS environments are still unknown. Especially, given the rapid speed of data updates and low latency response requirements, the overhead of retraining on new datasets cannot be spared by the system.
- Second, applying one well-trained machine learning model to multiple scenarios is a challenging issue. Generally speaking, one particular machine learning model is trained for a specific problem and cannot be easily applied to another case without retraining and reconfiguration. Thus, the retraining process cannot be avoided when applying a model to other datasets. Considering that cybersystems and CPS are designed for specific scenarios, and that the systems are unique and diverse, applying one well-trained machine learning model for all the different systems is a critical problem.
- Third, machine learning is a black-box process, and backtracking through specific training steps is difficult, if not impossible. Potentially, the machine learning process in action cannot be fully explained and audited. Finally, since cybersystems and CPS are dynamic systems, responding to rapidly changing situations is a challenging issue, because models cannot dynamically adjust parameters to correct for changes to a system.

Considering the challenges outlined above, we now discuss possible future research directions, both from the perspective of advances to machine learning practices and mechanism, as well as the perspective of cybersecurity/CPS scenarios.

- First, from the machine learning perspective, the optimization of machine learning algorithms to better fit cybersecurity and CPS scenarios is key. In order to more feasibly leverage machine learning for cybersecurity and CPS scenarios, the reduction of training time and the increase in reusability of models are critical, and should be addressed in immediate future research. Some studies have illustrated the ability of online learning and distributed learning to increase training speeds and mitigate problems in response times. Thus, online and distributed learning approaches are feasible, particularly for dynamic systems. Specifically, in dynamic systems, online learning can update a model by retraining on slices of new data, instead of on the entire dataset, clearly increasing the training speed. Additionally, distributed learning deploys machine learning algorithms to cloud or edge computing nodes, which dramatically improves computation capabilities and, in the case of edge nodes, reduces network latency. In addition, machine learning is a black-box process, and explaining how results are achieved through the training process is an open problem. Explainable machine learning is indeed an important topic, as it enables auditable machine learning, and the evaluation of models to determine whether they were properly or improperly trained and fit to data. Moreover, the development of explainable machine learning could further benefit machine learning development and implementation, providing unforeseen insights into optimal machine learning implementations and guiding machine learning practitioners. Thus, the development of explainable machine learning theories, mathematical models, and tools, can enable further advances in the field and improve usability. Most especially, determining appropriate model shapes and layers can be optimized automatically, but this process is extremely time consuming, a problem that compounds the difficulty of long training times. Being able to rapidly define optimal model parameters is an important goal that could increase usability and reduce development time.
- Second, considering cybersecurity and CPS scenarios, how to deploy machine learning algorithms and allocate computation resources efficiently and in a scalable way are critical issues that need additional research. Clearly, as cyberspace and CPS are dynamic heterogeneous systems, distributing machine learning to make use of the

available capacity for computing is an ideal solution to reduce training time and expand system capabilities. Thus, how to deploy machine learning to heterogeneous and remote computation nodes, and how to select the most efficient computation resources, are critical barriers for system optimization and improved capacity. While many resource allocation algorithms exist, they are not necessarily designed for cybersecurity and CPS scenarios and may not be effective or even viable in this case, and thus, further work is needed. Additionally, in CPS, networking systems transmit data and control signals, which often have requirements of high-speed communication for real time or near real time execution. In this case, transmitting large amounts of data is expensive and can overload the network infrastructure, posing additional demands on machine learning training in terms of preprocessing and local processing to avoid data transmission. Data normalization, discretization, and sampling could be appropriate solutions, but these solutions are not satisfactory in computation-limited devices in collaboration for time-sensitive CPS requirements. Therefore, mechanisms to reduce the size of collected and stored data in constrained CPS devices are another possible research direction.

IV. ADVERSARIAL WEAKNESS IN MACHINE LEARNING: BAD USE

As a particularly novel attack paradigm, the use of machine learning mechanisms to subvert benign machine learning based software and products is a particularly virulent threat, especially considering the widespread nature of machine learning software and development [112], [113]. While some comfort rests in the potential to use these same adversarial methods to improve the original models against subversion, unfortunately, some of these attacks have no satisfactory solution as of yet [114], [115]. Typical examples include the use of machine learning to subvert ReCaptcha human-user verification systems [116], [117], GANs for the subversion of benign deep learning systems [118], the ability to extract information related to training and testing datasets based on the trained models [119], and the duplication of an existing model [120], among others. This is a critical problem for both academic and industry groups that spend untold hours of work and money to develop production-ready machine learning models.

Clearly, attacks against machine learning models have the potential to degrade system performance when such models are used as critical components, inducing serious malfunctions and errors in cybersystems and CPS, since the attacks significantly reduce the accuracy of learning results. Toward a thorough consideration of such attacks, Wei *et al.* [121] provided a general formulation and categorization of adversarial examples. Additionally, the authors provided the basic principles for the design adversarial attack algorithms. Thus, thoughtful consideration for attacks on machine learning is necessary, and should include evaluation and defense against attacks on training (black box attack) and on the model itself (white box attack) to mitigate risks. In the following, we first provide an overview of adversarial learning, and then introduce the impacts of launching adversarial machine learning in cybersecurity and CPS contexts.

A. OVERVIEW OF ADVERSARIAL LEARNING

Adversarial learning can be applied in different learning stages, and we classify adversarial attacks based on these learning phases.

1) ATTACKS ON DATA COLLECTION

Generally speaking, we can consider generic attacks that could tamper with the system during the data collection and pre-processing stages, as follows:

- **Evasion attacks:** In an evasion attack [122], an adversary can attempt to escape inspection in the system test process by manipulating test samples. This kind of attack can affect the machine learning model at the creation phase and induces the model to output incorrect results. Because evasion attacks do not touch any training data and do not participate in the training process, it is easy to deceive the system and escape inspection.
- **Poisoning attacks:** In this case, an adversary attempts to inject carefully designed training data samples that cannot be distinguished by domain experts and that affect the machine learning model. The manipulated data is injected into the training phase in order to contaminate the training process. Obviously, the contaminated data negatively affect the accuracy of the results. As an adversarial example, Demontis *et al.* [123] discussed the two major factors of such an attack, the first being the adversarial vulnerability of the target model, and the second being the complexity of the surrogate model. Utilizing these two major factors, the authors then proposed a model that could successfully attack the target machine learning model.
- **Exploratory attacks:** In exploratory attack scenarios [124], an adversary focuses on detecting the learning algorithms and structures of the target machine learning models. In this way, the adversary can manipulate the parameter passing in the models to achieve their attack goals.

2) ATTACKS ON THE TRAINING PHASE

These attacks directly target the training process by either injecting carefully modified data into the training datasets or by manipulating the training logics so that the training results can be affected. This is the most straightforward attack on the machine learning process. In detail, the following two attack strategies can be considered to target the training phase.

- **Data modification and injection:** Assuming an adversary cannot obtain the machine learning algorithms and configurations of the target model, the training data could instead be assessed by the adversary.

Thus, modification of the data and false data injection prior to the training process (during the data pre-processing phase) will affect the final results as desired.

- **Logic manipulation:** In this case, an adversary attacks the machine learning model by controlling the logic of the model to tamper with the learning results. This is considered one of the highest threats to the machine learning process.

3) ATTACKS ON THE TESTING PHASE

These attacks are distinguished from attacks in the training phase, in that attacks in the training phase attempt to utilize modified or altered datasets to impact the machine learning training process, while attacks in the testing phase attempt to induce the model to output incorrect results after training has been completed. The attacks on the testing phase can be done in either a white-box or black-box manner. We consider the formal definition of the training process of a learning model as proposed by Chakraborty *et al.* [125], which is denoted by $\theta \leftarrow f(\mathbf{X}, y, r)$, where f represents the machine learning model and θ is the parameter learned in the training process. Also, r is the randomness of the training process (random initial weights, dropout, etc.) and $(\mathbf{X}, y)$'s are input pairs, which are assumed to be independently and identically sampled from a data distribution μ. We discuss the details of both white-box and black-box attack in the following.

4) WHITE-BOX ATTACKS

In white-box attack scenarios, an adversary has perfect knowledge of model parameters θ and the structure of the target model f, such as the types of neural networks, the number of the layers and neurons, and the size of the input tensors, among others. Furthermore, the adversary obtains the learning algorithms (linear regression, classification, etc.) in the training process and is able to access the training datasets μ. Having complete information of the training procedure and trained model, an adversary can inspect the vulnerabilities of the target model and launch attacks that modify the parameters of the model or the internal weights and biases directly. Therefore, the white-box attack is considered a strong and targeted adversarial method.

- **Model inversion attacks:** The model inversion attack has generally been applied to facial recognition models. For example, Fredrikson *et al.* [126] focused on a case of linear classifiers in facial recognition and proposed a white-box attack by deeply exploring the decision trees of trained models. Experimental results demonstrated that an adversary could recover recognizable images by obtaining labels and learning models. Furthermore, Wu *et al.* [127] extended those studies and proposed a game-based methodology inspired by the ''two worlds'' concept of cryptographic definitions to precisely detect this attack (i.e., when an adversary has obtained the structure of the model). The game-based methodology can also be used in a decision process to evaluate the robustness of the machine learning model before the model is released. Additionally, their methodology can help strengthen models against attacks without degrading performance by reducing model invertibility and utilizing less noise. Likewise, Hidano *et al.* [128] proposed a general inversion framework (GMI) aiming at cases that have non-sensitive attributes for the adversary to compromise.
- **Privacy inference attacks:** Another typical white-box attack is the privacy inference attack. For example, Nasr *et al.* [129] extended membership inference attacks on machine learning models to white-box attack scenarios. They analyzed the outputs of the activation functions, and proposed a white-box privacy inference attack which adjusts the parameters of the learned model to measure privacy leakage. In greater detail, they utilized the privacy vulnerabilities of the gradient descent algorithm. The results illustrated that even well-trained models can be affected by the proposed white-box privacy inference attack.

5) BLACK-BOX ATTACKS

Black-box attacks assume little knowledge of internal machine learning model. Instead, an adversary must carefully choose the input datasets and analyze prior input and output information to assess the vulnerabilities of the machine learning model [130]–[132]. For example, Dong *et al.* [130] proposed a black-box attack that leverages momentum-based iterative algorithms to increase the effectiveness of the attack. Specifically, in the iteration process, the methods can automatically switch directions and escape detection.

- **Non-adaptive attacks:** Considering that an adversary can obtain training datasets (input datasets) $\mathbf{X}$ and the results (output) y. Such an adversary could then create a machine learning model f' and train the model by utilizing the original training data. By modifying the parameters and structures of the model, the adversary forces the output y' of the model to match the output y of the target model. The process generally creates a mirror of the target model and the adversary can launch white-box attacks on their newly created model to obtain vulnerabilities that apply to both their copy model and the original. The discovered vulnerabilities could then be used in attacks.
- **Membership inference attacks:** In membership inference attacks, also known as adaptive attacks, an adversary cannot obtain any information related to the training input and output of the target model. Nonetheless, they can obtain the well-trained model that has been published for consumer use. The adversary then utilizes a marked training dataset to train the model. Similar to launching a non-adaptive attack, the adversary creates a mirror model and trains the model with dataset $\mathbf{X}$, resulting in output y' equaling y. Finally, the adversary obtains

the details of the target model and can launch attacks. Related to this attack category, Shokri *et al.* [119] investigated the issue of training an inference model so that prediction differences between the inference and target models can be recognized, such as whether particular inputs were trained on or not. Their results indicate that classical machine learning models are vulnerable to membership inference attacks.

Based on the above discussions, we generalize several key points. In black-box attack scenarios, the adversary does not obtain the details of the target model, such as the model structure, parameters θ, and others. Instead, the adversary attacks the model by utilizing and analyzing the relationship between input and output datasets, and either creates an agent learning model to practice the attack action, or directly manipulates the input datasets to force the target model to obtain incorrect results.

B. ADVERSARIAL LEARNING IN CYBERSECURITY

As we discussed in Section III-A, the integration of machine learning with IDS offers an effective approach to improving cybersecurity, through implementations such as Hidden Markov Models (HMM) [133], Support Vector Machines (SVM) [134], Decision Trees [135], N-grams [136], and Artificial Neural Networks [137], among others. Nonetheless, some existing research has indicated that machine learning based IDS have vulnerabilities to adversarial attacks [138]–[142]. Specifically, adversaries are able to launch malicious attacks against the machine learning components, in order to inject or manipulate the learning results and cause the failure of the IDS.

For example, Rubinstein *et al.* [143] evaluated poisoning techniques against the training phase of the machine learning in IDS and developed a defensive scheme. They illustrated three different poisoning attack approaches, which can substantially increase the chances of evading IDS detection. Furthermore, they demonstrated that the poisoning attacks can reduce the efficiency of the target IDS, and proposed a strong and robust antidote that can detect poisoning attacks against machine learning in IDS and maintain the efficiency of system. One particular problem with the majority of protection and prevention strategies against data poisoning attacks is that they are generally only effective when the ratio of poisoned data is small. Thus, there remains significant vulnerability to highly poisoned data, even though it may be quite difficult to add the necessary degree of poisoned data into the original data. Moreover, while honeypots can be utilized to recognize suspicious network traffic, adversaries are able to imitate honeypots to inject poisoned data to mislead machine learning algorithms and attack the IDS [144], [145].

In response to these and other vulnerabilities, some solutions have been developed to deal with adversarial attacks. For instance, Nelson *et al.* [146] developed a technique using the Reject On Negative Impact (RONI) mechanism. In detail, the proposed approach pre-processes the original data prior to the training stage, generating different datasets for training

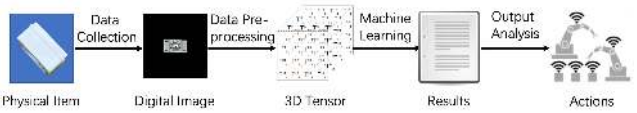


FIGURE 2. An example of the machine learning pipeline.

that add testing data samples. In addition, in the training stage, a detector is trained simultaneously, which can reject test samples if some malicious data is detected. Nonetheless, one limitation of this approach is that it is computationally expensive, especially if the training dataset is large. An additional technique against data poisoning attacks is robust statistics and boosting [147]. In this approach, the statistics framework evaluates the median and median absolute deviation (MAD) in order to monitor the robustness of machine learning algorithms. Furthermore, boosting can smooth the impact of training data samples in the machine learning model, and mitigate the impacts of the poisoned data.

C. ADVERSARIAL LEARNING IN CPS

As the basis of IoT, sensors are able to collect and transmit data to computing nodes over a variety of network connections. With the increasing deployment of smart devices, more and more data can be collected, which further benefits the data analysis process and enables more comprehensive results. To handle the massive volume of data and datasets, machine learning is imperative in CPS for analyzing data, enabling rapid and appropriate decisions, and achieving automation. Nonetheless, this integration also increases the risks of attacks against machine learning models and processes. The increasing capabilities of adversaries to attack the machine learning process could directly lead to critical failures, disruption, and damage. In the following, we discuss the impacts of using adversarial learning in CPS.

The machine learning process can be simplified and generalized as a data flow process [148]. This process generally consists of four stages from input to output: (i) the physical objectives are captured by sensors, generating datasets that are then stored locally or remotely, (ii) pre-processing is performed to transform raw data for input into the specific machine learning models, (iii) the machine learning models process input data and generate output based on their internal algorithms, logic, and training, and (iv) the system takes actions according to the output from the machine learning models. As a representative example, we consider the system pipeline of industrial IoT, as shown in Fig. 2.

In CPS, a sensor captures data (such as a camera capturing an image of a component). Then, the data (digital images) are pre-processed by computation nodes and transformed for machine learning input (tensor). The chosen machine learning model, such as a Convolutional Neural Network (CNN) for image processing, takes the pre-processed data and calculates generates output based on prior training and design. Finally, the CPS decides upon some action according to the results obtained from the model. From this example, attack

surfaces can be defined based on the different stages of the pipeline. These surfaces include the data collection and data transmission stage, the data storage stage (both cloud and local), the machine learning training and verification stage, and system decision stage. Since machine learning participates in the entire loop of CPS, launching attacks against machine learning itself can cause massive losses for CPS operations. Also, as we discussed before, adversaries are able to launch attacks that target machine learning in each of the different stages.

1) ADVERSARIAL LEARNING IN THE DATA COLLECTION PHASE

Machine learning mechanisms are data-driven techniques, and the training results highly depend on the quality of the raw data. Bad source data can cause unpredictable training results. In CPS, the data collection phase is the initial stage of the entire CPS. Additionally, some adversaries are able to inject fake data to poison the raw datasets, which can obviously impact the machine learning results. Furthermore, CPS require multiple datasets collected by different subsystems and organizations, increasing the complexity of protecting the source data. For instance, in the smart transportation system, autopilot systems navigate vehicles automatically based on traffic and GPS data. Meanwhile, the system also receives other related information (parking information, gas station information, etc.) for additional, potentially external sources. In this phase, malicious adversaries could inject false data into the original datasets to compromise machine learning results and further impact the decision-making process of CPS. Thus, it is important to have detection, protection, and prevention methods to ensure the correctness of source data. In this direction, to prevent false data attacks, Golle *et al.* [149] designed a scheme to evaluate the validity of vehicular ad hoc network (VANET) data. Specifically, their proposed approach evaluates the network nodes to generate a score, and then categorizes the nodes into two groups: safe and risky. By creating a VANET model with complete information of the nodes in a real VANET, managers can evaluate the location and physical properties of nodes, and validate the correctness of the data from a particular node.

As another example, in the smart city, a number of different kinds of sensors are deployed by different organizations and owners to provide information through wired and wireless networks. Nonetheless, the complexity of the network structure and its management pose high risks for data integrity. To combat this, Ghafouri *et al.* [150] proposed an effective detector for inspecting sensor and communications failures to ensure data integrity in the smart city. The detector utilizes the Gaussian processes for optimization, as well as an approach for computing optimal parameters. Evaluating the proposed detector on the OpenTripPlanner platform, the results showed that the detector could reliably increase data integrity in the smart city. Furthermore, Ghafouri *et al.* [151] proposed a general framework that considered attacks on a subset of sensors in CPS, with specific emphasis on overcoming the limitations of their prior work [150]. This framework consists of a general anomaly detection module that predicts a measurement for each sensor and leverages three different regression models: linear, neural network, and combined. A Stackelberg game was designed to evaluate the thresholds for each sensor to balance false alarms. The implemented framework was demonstrated to be effective without increasing the false alarm rate.

2) ADVERSARIAL LEARNING IN THE TRAINING AND TESTING PHASE

Adversarial learning in the training and testing phase can be used to attack machine learning mechanisms directly by manipulating learning algorithms or injecting malicious datasets into both training and testing processes. For instance, in a smart grid system, machine learning techniques are widely utilized for predicting electrical power usage in order to balance demand and supply. Thus, attacking the smart grid by injecting fake electricity consumption data is one potential approach to attack the smart grid. Nonetheless, the smart grid is a dynamic system, and compromising the machine learning results of such a system requires the injection of fake data continuously, which is not efficient from an attack perspective. Instead, attacking machine learning modules directly (e.g., black-box or white-box attacks) can have the maximum impact. Thus, attacks directly on the training and testing phases pose serious risks to CPS.

To consider and counter such attacks, Hu and Tan [118] studied black-box attacks in CPS and proposed a generative adversarial network (GAN)-based algorithm to thwart machine learning based attack detection modules. The proposed MalGAN utilizes a substitute detector to fit the target black-box attack detection module. The substitute detector can simulate the original attack detection module and is used to evaluate and modify malware samples to minimize the detection rate. The implementation results show that MalGAN is capable of reducing the attack detection rate to nearly zero and negates the original machine learning based attack detection module. Similarly, Eykholt *et al.* [152] proposed an attack mechanism, called Robust Physical Perturbations (RP2), to maximize the attack impact against CPS. The proposed RP2 utilizes Deep Neural Networks and attacks the testing process, fooling the testing process with good fit results when the fit is actually poor. To detect attacks utilizing machine learning models, Jones *et al.* [153] designed an unsupervised learning algorithm to detect the attacks against machine learning models. They utilized the signal temporal logic (STL) formula, which is typically used for early detection via online monitoring. A simulate train brake system was used as the attack target in their experiments, and the results demonstrated that the STL can detect attacks with machine learning models.

D. SUMMARY AND FUTURE DIRECTIONS

In this section, we have reviewed the majority of existing adversarial learning attacks in different machine learning

phases for both cybersecurity and CPS. Machine learning techniques have shown several glaring vulnerabilities that can be leveraged to launch adversarial attacks. In particular, adversaries are able to maliciously inject false data or noise into original datasets to disturb the training results. For example, adversaries can tamper with IDS training data, modifying categorical labels or add inaccurate information, inducing the machine learning algorithms to output incorrect results. Although there exist some detection and protection approaches, most approaches have limitations, such as being too computationally expensive, as well as only being effective when malicious data make up a small percentage of the entire dataset. In addition, the complexity of the structure and management of CPS causes additional risks, which malicious adversaries can easily manipulate, specifically in training data and learning processes. For example, data collected by different sensors may belong to various geo-distributed organizations. It thus is difficult to ensure the full protection of all the sensors at all locations and organizations. Thus, there are still numerous open challenges in detecting adversarial attacks.

Since machine learning techniques are widely deployed in cybersecurity and CPS applications, the risks of attacks against machine learning are massive and increasing. Thus, as immanently needed research, the discovery of vulnerabilities in existing systems, the development of defensive approaches, and the discovery of vulnerabilities in defensive strategies are critical.

- First, to detect adversarial learning, it is necessary to understand and leverage correlations and distinctions between false and real data [69], as well as to understand the prominent and state-of-the-art techniques of machine learning [3], as mechanisms to develop and carry out attacks. Typically, we can consider adversaries to launch attacks in unpredictable ways with evolutionary methods, iterating upon detectable attacks to improve subversion. For example, adversaries can hide their attack methods by slowly poisoning data over time, such as by manipulating sensors and modifying parameters over a long period to evade detection and remain within a local statistical range. These adversarial attacks are quite difficult to detect [132]. To combat such attacks, proposed methods include nonparametric cumulative sum schemes, as well as distributed detection mechanism. Nonetheless, there exists no comprehensive strategy for adversarial detection in either cybersecurity or CPS scenarios.
- Second, as we have identified, machine learning mechanisms themselves have vulnerabilities and are easy targets. Indeed, existing research has demonstrated that adversarial learning can compromise many different machine learning models, including DNN, CNN, and RNN, among others. Thus, strengthening the security of machine learning algorithms and models is another critical research direction.
- Third, based on our discussion, machine learning techniques are widely used in cybersecurity and CPS scenarios, and adversaries will increasingly attack machine learning models. Thus, a comprehensive defensive strategy is necessary, not only to secure machine learning techniques, but also to develop appropriate management and policies to prevent bad practices in machine learning development. A variety of sensors are geographically dispersed in complex CPS utilizing distinct protocols and managed by various organizations and stakeholders. In this case, it is difficult to unify defense strategies in deployment, and thus it is necessary to deploy robust data verification strategies to ensure the correctness of transmitted data. In addition, to protect machine learning in the training and testing phases, protection and redundancy mechanisms need to be added to the learning algorithms, as well as the underlying architectures.

V. MALICIOUS MACHINE LEARNING: UGLY USE

As outlined in the prior sections, machine learning techniques have become the most popular data analysis methods, and are being applied in both cybersecurity and CPS to analyze complex sensor data. As we discussed, machine learning can not only increase the detection rate of cybersecurity, but also improve the performance of CPS to achieve automation and artificial intelligence. Moreover, since machine learning techniques embody critical roles in such fields, attacking machine learning models and algorithms can achieve critical damage across entire systems. On the other hand, because the machine learning techniques have powerful data analysis capabilities, they can also be used by malicious adversaries to analyze the vulnerabilities of cybersystems and CPS, as well as assist in successfully delivering attacks. In this section, we discuss some cases related to utilizing machine learning as attacking tools against cybersecurity and CPS.

A. OVERVIEW

Machine learning techniques are just starting to be used by adversaries to improve the effectiveness of their attacks. Especially against the cybersystems and CPS, adversaries are able to leverage machine learning to analyze the vulnerabilities of cybersystems and CPS, since the cybersystems and CPS widely utilize the data analysis to improve performance. In the following, we provide an overview of leveraging machine learning to attack cybersystems and CPS.

In the cybersystems, machine learning and neural networks can be utilized by malicious adversaries to amplify and enhance some types of existing attacks [154]–[157]. For example, adversaries are able to utilize machine learning algorithms to replicate and imitate the regular actions of users and hide attacking actions. In addition, by applying data analysis, machine learning and neural networks are able to aid adversaries to find and detect network system vulnerabilities. Moreover, adversaries can use existing external benign machine learning systems to improve their attack methodologies, such as in the use of malware detection systems

to improve detection avoidance, as well as to improve their understanding of a target. For example, adversaries are able to use machine learning to obtain which attack method is the most efficient approach against a specific network structure.

Machine learning can also be utilized to attack verification code systems and gain unauthorized access [158]–[160]. For example, Rosebrock [160] proposed a CNN model to recognize verification codes and obtain access authentication. Furthermore, network attacks have a well-known history of causing site outages and damage to systems. Just as system administrators have used machine learning to analyze network traffic for anomalies and intrusions, so too can adversaries use machine learning models to analyze network traffic and predict peaks in network loads. They can then launch precisely timed strong attacks to damage systems at critical usage peaks, or apply stealthy attacks that mimic typical network traffic to remain anonymous or increase network loads discreetly and cause network failures.

In the following, after a brief overview, we consider malicious machine learning used against generalized cybersystems and CPS. To automatically monitor and control systems in CPS, distributed sensors are deployed to necessary locations to enable communication across the CPS via network facilities. The complexity of the system and the breadth of devices deployed increase the vulnerabilities and risk for the system. In particular, there are three components in CPS that can be targeted by adversaries: computing, control, and network. Attacks on the network component have already been mentioned above. We illustrate the attacks which focus on computing and control components below.

Targeting the computing component, adversaries are able to use machine learning to analyze computing task loads and predict the next computing load peaks. With this information, adversaries could inject additional computing tasks during peak periods to increase computing time and cause system errors, especially for time-sensitive systems. For example, in a distributed computing platform, since the computation nodes have limited computing power, a computation task is sent to different computation nodes. Depending on the specific computation resource selection algorithms, computing loads may not be balanced in the target time period, meaning that some computation nodes have higher computing loads than others. In this case, adversaries are able to use historical task load datasets for each computation node as training datasets and train the machine learning models to predict upcoming computation load peaks. Injecting additional computation tasks during computation load peaks could then maximize the effect of the attack and cause system errors and failures.

In the control component, machine learning can create models that themselves improve the function or performance of attacks, in order to analyze the effect of an attack. Moreover, it is possible to analyze the control process to obtain key values to damage the control process itself. For example, in a typical manufacturing process, a temperature control system precisely maintains the temperature of industrial processes, and the control process is precise and time-sensitive. There are several key actions in this process, including temperature variations, control signal intervention, and temperature recovery time. Based on these key values, adversaries are able to create multiple features for machine learning model training, and utilizing the trained model, an adversary could obtain the relationships of those parameters. Furthermore, by manipulating the control signal intervention time or the control process, the adversary could then disturb the system control process and cause system anomalies.

B. CYBERSECURITY

In an era where the malicious use of machine learning is commonplace, rapidly developing machine learning techniques offer huge benefits to cybercriminals [161]. For example, existing research has shown the ability of Artificial Intelligence to power malware [162]. Utilizing machine learning techniques, adversaries can configure an agent machine to automatically coordinate different malware and launch attacks targeting various vulnerabilities of the network system, affording attacks that easily overload a victim's defensive strategies. For example, Falco *et al.* [163] utilized artificial intelligence planning techniques that focused on the defensive strategies of network systems and that identified vulnerabilities automatically. Moreover, they proposed an automated attack generation scheme that can output detailed attack trees. In the following, we discuss different machine learning attacks in different malicious uses.

1) MALICIOUS ACCESS

One of the most intuitive malicious uses of machine learning is to obtain access permission for unauthorized users, which we call malicious access. We know that machine learning techniques have shown impressive achievements in machine vision. These mechanisms have the potential to be used by malicious adversaries to deceive network authentication systems. For example, Agarwal *et al.* [164] proposed a machine learning based social media analysis framework that utilizes machine learning to capture and analyze the captcha during the login process. By comparing and training the huge captured datasets, machine learning can bypass captcha based verification systems that are designed to exclude machine users. Likewise, Stark *et al.* [165] investigated a CNN based attacking neural network, which recognizes captcha photos and subvert the human/machine distinguishing process. In detail, to increase the correctness of the model, they utilized an active learning approach, which trains a comparably small slice of training data in the initial phase, and adds new training data continuously in subsequent training processed. The evaluation results showed the proposed attacking neural network achieved a success rate of over 83% in bypassing the captcha systems.

Furthermore, utilizing machine learning techniques, malicious adversaries are able to effectively leverage leaked user information, analyzing historical user name and password data, in order to guess the current user names

and passwords. Based on the huge size of data and related information, machine learning can significantly increase the speed of malicious password cracking [166]. For example, Lyastani *et al.* [167] carried out a large-scale investigation on how password managers could influence the real-world passwords of users. In addition, Hitaj *et al.* [168] proposed an enhanced password analysis tool named passGAN that utilizes Generative Adversarial Network (GAN) architecture. Their passGan can autonomously train on leaked passwords from actual systems and then provide high-quality password guesses. Their evaluation results demonstrated that passGAN could achieve better performance than some traditional password analysis tools, such as ''HashCat'' and ''John the Ripper''. Other studies focused on leveraging machine learning to generate fake human voices, fingerprints, and human faces to deceive authentication systems. These can be used against network authentication systems that utilize biometric authorization mechanisms, such as facial recognition and fingerprint identification. Machine learning techniques can easily create fake information to bypass such authentication systems. An example application can clone voices [169] and is able to deceive authentication systems.

2) EVASIVE ATTACKS

In network security systems, generally speaking, the process of creating malicious programs requires particular programming tools, whose features are recorded by security systems such as IDS. The IDS can then detect malicious features by checking for known malware signatures [170]. Nonetheless, since machine learning techniques can be utilized by malicious adversaries, machine learning can be used to evolve malicious programs by learning the vulnerabilities of IDS [171]. Moreover, machine learning techniques are able to generate computer code and programs automatically by learning from existing programs [172]. Thus, without supervision and control, machine learning can make existing malicious programs more effective and evade detection or tracking, even by human investigators.

3) PHISHING AND RANSOMWARE

Machine learning can additionally be used to drive ransomware attacks [173]. At present, most criminal organizations leverage machine learning to modify well-known ransomware programs and generate various samples [174]. These cyberattacks are autonomous in selecting targets, infiltration, evading detection, and sabotaging the target. Yet, there still exists significant space for leveraging machine learning to further improve the efficiency of attacks. For example, traditional cyberattacks, such as distributed denial of service (DDoS) attacks, ransomware, and backdoor attacks, all attack the targets autonomously. Nonetheless, these rely on predefined configurations. Thus, integrating machine learning systems to improve the attack process could complete the entire attacking loop and supervise malicious programs to carry out complex tasks, possibly evolving malicious programs automatically.

C. CPS

As we discussed before, CPS have three key components, which are computing, control, and networking. CPS involve a variety of technologies, such as distributed computing, machine learning, and artificial intelligence, among others. The distributed structure and complex techniques raise potential risks and vulnerabilities for adversaries to exploit. In addition, increasingly, CPS utilize data collection and analysis to achieve automation, which also enables a variety of negative effects. For instance, adversaries are able to utilize machine learning to analyze data to improve attack effects. As we have shown above, the application of machine learning for attacks against human users and traditional systems, while nascent, demonstrates significant power for disruption. Indeed, even in CPS, traditional weaknesses remain in a number of cases (i.e., human users), and limitations of the technologies may exacerbate or generate new weaknesses (i.e., system scale and capabilities).

We now discuss some existing research on CPS attacks. For example, Gerdes *et al.* [175] described an attack focusing on degrading the performance of automated vehicular transportation systems. In the target system, vehicle platooning strategies for motorcades and adaptive cruise control for each vehicle are key components. Those strategies manage and control automated vehicles efficiently and safely. In their study, the authors leveraged a maliciously controlled vehicle to interfere with the system and force the system to reduce the speed of vehicles to avoid accidents. Moreover, to achieve the best attack effect, it is possible to utilize machine learning models to analyze vehicle platooning datasets and determine which action causes the worst effects. Then, adversaries can control malicious vehicles to achieve the desired attack effects. Likewise, Chen *et al.* [176] investigated an attack designed to change the CPS state from the current to a target state, reducing the probability of being detected by the defense system, while ensuring the attack effect. Since CPS typically run in different states, they designed a system model and formulated a cost function to compute the minimal probability of being detected by the defense system. In fact, there is also an approach that utilizes machine learning to analyze the minimal probability. It remains to be determined if machine learning systems can subvert these systems in practice. In the case of traditionally programmed software systems, it likewise remains to be seen whether machine learning systems can circumvent or overcome specifically designed logic. Especially in CPS, these questions are paramount, as the number of reachable devices in complex CPS is unprecedented, and the resulting disruption may be costly on a scale yet unforeseen. In the case of Industrial IoT, the damage could result in the collapse of a business, notwithstanding the potential for physical harm that could result from subverted machinery.

In addition, various critical infrastructures have internal interdependence which is physical, cyber, geographical, and logical [177]. Because of these complex relationships, attacks on each component can propagate through different domains

and cause secondary and cascading damage [77]. For example, adversaries are able to attack a data trading system to increase the data price, which then prevents a CPS from obtaining a necessary dataset to execute the computing process. Furthermore, leveraging machine learning technologies, adversaries are able to counterfeit credentials to obtain access permission. For instance, machine learning can analyze verification codes and adversaries can then bypass the verification code to access the system. Moreover, machine learning can synthesize false face identification data to deceive systems.

D. OTHER MALICIOUS USES OF MACHINE LEARNING

As stated above, we consider the malicious use of machine learning as the application of machine learning models in the attack of some target. Understanding that machine learning provides unprecedented analytical power, it can be utilized by an adversary as easily as a benign actor. Nonetheless, we also observe that this particular strategy is quite new, and few examples exist of this mechanism in research or in practice. As subsets of machine learning for malicious use, we first consider targets that are traditional and generic in scope. These can be human users over Internet or intranet, as well as autonomous or machine-type systems.

1) ATTACKS ON GAMES AND GAMBLING

Game scenarios are probably the most obvious targets for machine learning based attacks. Particularly, the optimization of strategy and the autonomy available in computing systems make game scenarios ideal targets. Indeed, recent significant media attention has focused on tangible competitions and exhibitions that have demonstrated the power of artificial intelligence in learning and winning complex strategic competitions, such as the games Go, Shogi, and Chess mastered by AlphaZero [178], the AIIDE Starcraft Contest [179], in which Facebook's entry took second place [180], one-on-one and five-on-five Dota2 matches played by OpenAI's bot iterations [181], [182], and the online collectible card game (CCG) Hearthstone [183], [184], among others.

The application of machine learning in this context is particularly problematic, as it specifically subverts the competition for undeserved monetary gain. In its extreme, one could imagine user-programmed machine learning models for use in online gambling and betting to gain a competitive edge. Moreover, while this may be banned by the laws or policies governing such sites, it may be difficult to police, especially if sophisticated machine learning models are designed and trained to imitate typical user behavior (e.g., improvement over time, and skill levels and win frequency within normal bounds) to subvert detection. Further still, the use of machine learning directly with software application programmable interfaces (APIs) provides direct advantages that human competitors simply cannot compete with.

2) AUCTIONS AND COMPETITIONS

In a similar way, we consider the use of machine learning to learn optimal strategies for auction scenarios that are sufficiently complex, as well as when human user patterns can be learned and subverted. For instance, Zhang *et al.* [185] used n-gram and LSTM models to predict the end of bidding in online penny auctions, such as DealDash, with high accuracy, and clustered bidders into groups. Assuming that other such auctions operate at the same time, it is feasible that these platforms could be subverted by machine learning models. In a tangentially related work, Chen and Qiu [186] developed a Q-learning algorithm to be used by secondary users in a cognitive radio spectrum allocation auction. In a competitive setting, all tested nodes used the algorithm to learn from competitors and achieved optimal results for their own needs. In this case, all users have the same learning capacity and mechanism, which may be infeasible in reality, and could be vulnerable to adversarial systems. Moreover, from this example, we can consider a similar mechanism as applied against human users who may be severely outmatched in ability to analyze user patterns and achieve optimal rewards.

3) HUMAN-CENTRIC ATTACKS

Furthermore, human-centric attacks have the capacity to accelerate and optimize social engineering techniques, which can fool the general population into incorrect action through deliberate manipulations. A particularly powerful demonstration was offered by Seymour and Tully [187], which automated and improved the subversion capabilities of spear phishing attacks via the use of recurrent neural networks. Their mechanism specifically targeted high-valued users. Through the analysis of successful attacks, attack efficacy can be improved. Additionally, Melicher *et al.* [188] implemented a password strength assessment through machine learning algorithms, which ultimately over-estimated password strength. Applied in the opposite direction (i.e., against users), such a password assessment system could be employed to mitigate massive volumes of leaked and subverted passwords and improve intrusion systems.

4) BLACKMAIL AND TARGET DELIVERY

In terms of blackmail and reputation damage, the fabrication of images, video, and audio recordings [189] can be used as leverage for adversaries to extort value and actions from human victims. While these mechanisms are most successful when they are trained on high-quality images and videos of the target (most specifically celebrities), this type of data is becoming more available via purchase or theft from insecure social media platforms. In the most direct attack yet using machine learning, a recent attack developed by Kirat [190] utilizes embedded machine learning software to ensure payload delivery to the target, through methods such as facial recognition.

E. SUMMARY AND FUTURE DIRECTIONS

In this section, we have illustrated some malicious uses of machine learning in the cybersecurity and CPS areas. By utilizing machine learning techniques, the malicious adversaries are able to configure a malicious agent that automatically

and continuously selects, infiltrates, evades, and even sabotages the target systems. Specifically, in the selection stage, the malicious agent automatically evaluates the probability of compromising targets. In the infiltration stage, the malicious agent can leverage machine learning algorithms to enhance the effects of attacks by analyzing the attack results. From the attack results, machine learning can obtain the best means of attack for a specific system. In the evasion stage, the malicious agent can hide the attack actions by using machine learning schemes, such as manipulating the system defense strategy or duplicating legitimate activities.

After the prior three stages, the system may have been infected by the malicious agent, leading to the sabotage stage. In this stage, machine learning can deploy viruses, and control systems and ransomware. Based on the discussions, leveraging machine learning techniques as a potential malicious attack tool will raise significant risks for cybersecurity and CPS. Although there are many protection and detection strategies, these generally focus on traditional attacks, such as Denial of Service, SSL attacks, and Backdoor attacks, among others. In addition, malicious use of machine learning is dangerous, since machine learning techniques are able to produce falsified photos and video, and there are minimal detection and defense approaches.

Because of the risks detailed above, we now outline a number of potential solutions and research directions.

- First, one interesting approach to preventing or limiting machine learning based attacks is by limiting the abilities of production machine learning systems. For example, a recent study proposed the use of virtual machines (VMs) with boundaries for running machine learning processes to mitigate attack damage and limit the reach of adversaries [191]. This type of mechanism can also be used to separate machine learning processes, and in concert with other techniques such as transfer learning [192], could be used to separate training and test data and prevent data theft. In reducing the effectiveness of attacks, we consider systems that integrate machine learning with human users, making the final critical decision. In these cases, examples of which include smart transportation and smart grids, machine learning systems can feed information and decision analytics to human users to implement. In such cases, machine learning systems must be reliable and trustworthy. Thus, how to combine the machine learning smoothly with human decisions is a necessary research direction. In addition, how to prevent such integrated systems from being subverted is critical. Often, human users are easier targets to subvert, and thus developing and improving implementation strategies to mitigate human error are also necessary.
- Second, another possible approach to thwarting the use of machine learning based attacks is to accelerate the testing and deployment of machine learning models and attacks in simulated environments, or sandboxes, and systematically studying the attacks, which leverage machine learning techniques to extract critical fingerprint and process information. By doing so, researchers can achieve a better understanding and increase the confidence of detecting and protecting systems against real-world attacks. In addition, protecting against the leakage of sensitive data is critical. Machine learning based attacks often rely on analyzing datasets relevant to the target, avoiding machine learning attacks requires comprehensive data protection and secured segregation of storage. Moreover, enhanced detection methods are necessary, such as AI against AI. Specifically, since machine learning is now a tool for malicious adversaries as well as positive actors, defenders must employ advanced machine learning techniques as defensive weapons too. This includes machine learning to detect traditional and non-traditional attacks, machine learning for advanced authorization, and machine learning in combination with tested security techniques for multi-level security and defense.
- Finally, while advanced legal structures are potential deterrents, we know that these will not prevent bad actors from developing advanced attack techniques, yet these legal frameworks should be strengthened nonetheless. We need to formulate and optimize related laws to limit the malicious use of machine learning wherever possible, and enable recourse for those systems targeted. As well, it is imperative that government entities and communities are aware of the deficiencies in technologies that have been widely deployed, as well as threats that utilize such technologies to achieve more devastating results. This awareness should lead to actions to enable regulatory agencies, oversight bodies, and security agencies to investigate and develop appropriate responses to such threats in the form of policy positions and the development of further technologies for public and private use.

VI. FINAL REMARKS

In this paper, we have developed a broad understanding of machine learning for positive and negative uses, and have extolled the vulnerabilities of machine learning systems against traditional and machine learning based attacks. In the context of cybersecurity and CPS, we have considered the good use of machine learning, especially toward improving system performance and achieving automation. We have likewise presented the bad use of machine learning. That is, how the widespread use of machine learning raises new and unresolved vulnerabilities in a variety of systems, and the significant lack of defensive capabilities. Finally, we have addressed in detail the ugly use of machine learning, or the weaponization of machine learning toward the subversion of user confidentiality, system reliability, and service, and the improvement of intrusion and obfuscation mechanisms. Of particular concern, the vulnerabilities of existing machine learning systems provide unprotected attack surfaces, ripe for

exploitation. At the same time, the use of machine learning to improve attack success, efficacy, and strength should raise alarms across all industries and research, as the lack of defenses against machine learning based attacks make us all vulnerable. Critical research is necessary to strengthen detection and defenses against such machine learning based attacks, especially in critical infrastructure systems with the potential for massive disruption, destruction, and loss of life.

ACKNOWLEDGMENT

Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

REFERENCES

[1] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017.

[2] J. A. Stankovic, "Research directions for the Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 3–9, Feb. 2014.

[3] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2923–2960, 4th Quart., 2018.

[4] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, and W. Zhao, "A survey on big data market: Pricing, trading and protection," *IEEE Access*, vol. 6, pp. 15132–15154, 2018.

[5] G. Xu, W. Yu, D. Griffith, N. Golmie, and P. Moulema, "Toward integrating distributed energy resources and storage devices in smart grid," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 192–204, Feb. 2017.

[6] J. Lin, W. Yu, X. Yang, Q. Yang, X. Fu, and W. Zhao, "A novel dynamic en-route decision real-time route guidance scheme in intelligent transportation systems," in *Proc. IEEE 35th Int. Conf. Distrib. Comput. Syst.*, Jun. 2015, pp. 61–72.

[7] P. Moulema, W. Yu, D. Griffith, and N. Golmie, "On effectiveness of smart grid applications using co-simulation," in *Proc. 24th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2015, pp. 1–8.

[8] D. Li, Q. Yang, W. Yu, D. An, Y. Zhang, and W. Zhao, "Towards differential privacy-based online double auction for smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 971–986, Aug. 2019.

[9] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibañez, "Internet of vehicles: Architecture, protocols, and security," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3701–3709, Oct. 2018.

[10] S. B. Baker, W. Xiang, and I. Atkinson, "Internet of Things for smart healthcare: Technologies, challenges, and opportunities," *IEEE Access*, vol. 5, pp. 26521–26544, 2017.

[11] J. Xu, H. Guo, and S. Wu, "Indoor multi-sensory self-supervised autonomous mobile robotic navigation," in *Proc. IEEE Int. Conf. Ind. Internet (ICII)*, Oct. 2018, pp. 119–128.

[12] Z. Guan, Z. Lv, X. Du, L. Wu, and M. Guizani, "Achieving data utility-privacy tradeoff in Internet of medical things: A machine learning approach," *Future Gener. Comput. Syst.*, vol. 98, pp. 60–68, Sep. 2019.

[13] A. Kusiak, "Smart manufacturing," *Int. J. Prod. Res.*, vol. 56, nos. 1–2, pp. 508–517, 2018.

[14] S. Wu, J. B. Rendall, M. J. Smith, S. Zhu, J. Xu, H. Wang, Q. Yang, and P. Qin, "Survey on prediction algorithms in smart homes," *IEEE Internet Things J.*, vol. 4, no. 3, pp. 636–644, Jun. 2017.

[15] H. Wang, M. S. Mahmud, H. Fang, and C. Wang, *Wireless Health*. Cham, Switzerland: Springer, 2016. [Online]. Available: https://link.springer.com/content/pdf/bfm%3A978-3-319-47946-0%2F1.pdf

[16] X. Yang, J. Lin, W. Yu, P.-M. Moulema, X. Fu, and W. Zhao, "A novel en-route filtering scheme against false data injection attacks in cyber-physical networked systems," *IEEE Trans. Comput.*, vol. 64, no. 1, pp. 4–18, Jan. 2015.

[17] F. Liang, C. Qian, W. G. Hatcher, and W. Yu, "Search engine for the Internet of Things: Lessons from Web search, vision, and opportunities," *IEEE Access*, vol. 7, pp. 104673–104691, 2019.

[18] S. Yu, M. Liu, W. Dou, X. Liu, and S. Zhou, "Networking for big data: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 531–549, 1st Quart., 2016.

[19] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A.-R. Sadeghi, and S. Tarkoma, "IoT SENTINEL: Automated device-type identification for security enforcement in IoT," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 2177–2184.

[20] B. Biggio, B. Nelson, and P. Laskov, "Poisoning attacks against support vector machines," 2012, *arXiv:1206.6389*. [Online]. Available: https://arxiv.org/abs/1206.6389

[21] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 7167–7176.

[22] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other Botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.

[23] X. Liu, C. Qian, W. G. Hatcher, H. Xu, W. Liao, and W. Yu, "Secure Internet of Things (IoT)-based smart-world critical infrastructures: Survey, case study and research opportunities," *IEEE Access*, vol. 7, pp. 79523–79544, 2019.

[24] Z. Ling, J. Luo, Y. Xu, C. Gao, K. Wu, and X. Fu, "Security vulnerabilities of Internet of Things: A case study of the smart plug system," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1899–1909, Dec. 2017.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[26] J. Lemley, S. Bazrafkan, and P. Corcoran, "Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision," *IEEE Consum. Electron. Mag.*, vol. 6, no. 2, pp. 48–56, Apr. 2017.

[27] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[28] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[29] N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 146–153, Jun. 2017.

[30] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, no. 1, pp. 1–47, 2002.

[31] B. Su, X. Ding, H. Wang, and Y. Wu, "Discriminative dimensionality reduction for multi-dimensional sequences," *IEEE Trans. Pattern Anal. Mag. Intell.*, vol. 40, no. 1, pp. 77–91, Jan. 2018.

[32] L.-L. Ge, Y.-H. Wu, B. Hua, Z.-M. Chen, and L. Chen, "Image registration based on SOFM neural network clustering," in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 6016–6020.

[33] X. Zhang, X. Pan, and S. Wang, "Fuzzy DBN with rule-based knowledge representation and high interpretability," in *Proc. 12th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, Nov. 2017, pp. 1–7.

[34] D. Marquardt and S. Doclo, "Noise power spectral density estimation for binaural noise reduction exploiting direction of arrival estimates," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2017, pp. 234–238.

[35] C. Yeshwanth, P. S. A. Sooraj, V. Sudhakaran, and V. Raveendran, "Estimation of intersection traffic density on decentralized architectures with deep networks," in *Proc. Int. Smart Cities Conf. (ISC2)*, Sep. 2017, pp. 1–6.

[36] E. Gibney, "Google AI algorithm masters ancient game of go," *Nature News*, vol. 529, no. 7587, p. 445, 2016.

[37] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.

[38] Y. Wu, L. Liu, C. Pu, W. Cao, S. Sahin, W. Wei, and Q. Zhang, "A comparative measurement study of deep learning as a service framework," *IEEE Trans. Services Comput.*, to be published.

[39] W. G. Hatcher and W. Yu, "A survey of deep learning: Platforms, applications and emerging research trends," *IEEE Access*, vol. 6, pp. 24411–24432, 2018.

[40] Z. Chen, G. Xu, V. Mahalingam, L. Ge, J. Nguyen, W. Yu, and C. Lu, "A cloud computing based network monitoring and threat detection system for critical infrastructures," *Big Data Res.*, vol. 3, pp. 10–23, Apr. 2016, doi: 10.1016/j.bdr.2015.11.002.

[41] W. Yu, G. Xu, Z. Chen, and P. Moulema, "A cloud computing based architecture for cyber security situation awareness," in *Proc. IEEE Conf. Commun. Netw. Security (CNS)*, Oct. 2013, pp. 488–492.

[42] D. Zhang, L. Ge, R. Hardy, W. Yu, H. Zhang, and R. Reschly, "On effective data aggregation techniques in host-based intrusion detection in MANET," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2013, pp. 85–90.

[43] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowl.-Based Syst.*, vol. 78, pp. 13–21, Apr. 2015.

[44] E. Hodo, X. Bellekens, A. Hamilton, P.-L. Dubouilh, E. Iorkyase, C. Tachtatzis, and R. Atkinson, "Threat analysis of iot networks using artificial neural network intrusion detection system," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, May 2016, pp. 1–6.

[45] L.-P. Yuan, W. Hu, T. Yu, P. Liu, and S. Zhu, "Towards large-scale hunting for Android negative-day malware," in *Proc. 22nd Int. Symp. Res. Attacks, Intrusions Defenses (RAID)*. Chaoyang District, Beijing: USENIX Association, Sep. 2019, pp. 533–545. [Online]. Available: https://www.usenix.org/conference/raid2019/presentation/yuan

[46] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, "CT-GAN: Malicious tampering of 3D medical imagery using deep learning," in *Proc. 28th USENIX Secur. Symp. (USENIX Secur.)*. Santa Clara, CA, USA: USENIX Association, Aug. 2019, pp. 461–478. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/mirsky

[47] B. Subba, S. Biswas, and S. Karmakar, "A neural network based system for intrusion detection and attack classification," in *Proc. 22nd Nat. Conf. Commun. (NCC)*, Mar. 2016, pp. 1–6.

[48] E. Bertino and E. Ferrari, "Big data security and privacy," in *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*. Cham, Switzerland: Springer, 2018, pp. 425–439.

[49] K. M. A. Alheeti, A. Gruebler, and K. D. McDonald-Maier, "An intrusion detection system against malicious attacks on the communication network of driverless cars," in *Proc. 12th Annu. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2015, pp. 916–921.

[50] P. Jiang, H. Wu, C. Wang, and C. Xin, "Virtual MAC spoofing detection through deep learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[51] L. Bontemps, V. L. Cao, J. McDermott, and N.-A. Le-Khac, "Collective anomaly detection based on long short-term memory recurrent neural networks," in *Proc. Int. Conf. Future Data Secur. Eng.* Wiesbaden, Germany: Springer, 2016, pp. 141–152.

[52] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[53] Z. Tian, C. Luo, J. Qiu, X. Du, and M. Guizani, "A distributed deep learning system for Web attack detection on edge devices," *IEEE Trans. Ind. Informat.*, to be published.

[54] A. Bivens, C. Palagiri, R. Smith, B. Szymanski, and M. Embrechts, "Network-based intrusion detection using neural networks," *Intell. Eng. Syst. Artif. Neural Netw.*, vol. 12, no. 1, pp. 579–584, 2002.

[55] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proc. 9th EAI Int. Conf. Bio-Inspired Inf. Commun. Technol. (BIONETICS)*, 2016, pp. 21–26.

[56] Y. Guo, W. Liao, Q. Wang, L. Yu, T. Ji, and P. Li, "Multidimensional time series anomaly detection: A GRU-based Gaussian mixture variational autoencoder approach," in *Proc. 10th Asian Conf. Mach. Learn.*, 2018, pp. 97–112.

[57] W. Liao, Y. Guo, X. Chen, and P. Li, "A unified unsupervised Gaussian mixture variational autoencoder for high dimensional outlier detection," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 1208–1217.

[58] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 925–941, 2nd Quart., 2014.

[59] W. Yu, H. Zhang, L. Ge, and R. Hardy, "On behavior-based detection of malware on Android platform," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 814–819.

[60] J. Booz, J. McGiff, W. G. Hatcher, W. Yu, J. Nguyen, and C. Lu, "Tuning deep learning performance for Android malware detection," in *Proc. 19th IEEE/ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput. (SNPD)*, Jun. 2018, pp. 140–145.

[61] P. M. Comar, L. Liu, S. Saha, P.-N. Tan, and A. Nucci, "Combining supervised and unsupervised learning for zero-day malware detection," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 2022–2030.

[62] J. McGiff, W. G. Hatcher, J. Nguyen, W. Yu, E. Blasch, and C. Lu, "Towards multimodal learning for Android malware detection," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2019, pp. 432–436.

[63] Z. Yuan, Y. Lu, Z. Wang, and Y. Xue, "Droid-Sec: Deep learning in Android malware detection," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 371–372, Aug. 2014, doi: 10.1145/2740070.2631434.

[64] Y. Ding, S. Chen, and J. Xu, "Application of deep belief networks for opcode based malware detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3901–3908.

[65] S. O. Uwagbole, W. J. Buchanan, and L. Fan, "Numerical encoding to tame SQL injection attacks," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS)*, Apr. 2016, pp. 1253–1256.

[66] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long short term memory recurrent neural network classifier for intrusion detection," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Feb. 2016, pp. 1–5.

[67] W. Yu, D. Griffith, L. Ge, S. Bhattarai, and N. Golmie, "An integrated detection system against false data injection attacks in the smart grid," *Secur. Commun. Netw.*, vol. 8, no. 2, pp. 91–109, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sec.957

[68] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Secur. (CCS)*, New York, NY, USA, 2009, pp. 21–32, doi: 10.1145/1952982.1952995.

[69] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On false data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 717–729, Mar. 2014.

[70] J. Lin, W. Yu, X. Yang, G. Xu, and W. Zhao, "On false data injection attacks against distributed energy routing in smart grid," in *Proc. IEEE/ACM 3rd Int. Conf. Cyber-Phys. Syst. (ICCPS)*, Washington, DC, USA, 2012, pp. 183–192, doi: 10.1109/ICCPS.2012.26.

[71] Q. Yang, Y. Liu, W. Yu, D. An, X. Yang, and J. Lin, "On data integrity attacks against optimal power flow in power grid systems," in *Proc. 14th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2017, pp. 1008–1009.

[72] J. Lin, W. Yu, and X. Yang, "Towards multistep electricity prices in smart grid electricity markets," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 1, pp. 286–302, Jan. 2016.

[73] J. Lin, W. Yu, N. Zhang, X. Yang, and L. Ge, "Data integrity attacks against dynamic route guidance in transportation-based cyber-physical systems: Modeling, analysis, and defense," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8738–8753, Sep. 2018.

[74] Q. Yang, L. Chang, and W. Yu, "On false data injection attacks against Kalman filtering in power system dynamic state estimation," *Secur. Commun. Netw.*, vol. 9, no. 9, pp. 833–849, 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sec.835

[75] X. Yang, X. Zhang, J. Lin, W. Yu, and P. Zhao, "A Gaussian-mixture model based detection scheme against data integrity attacks in the smart grid," in *Proc. 25th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2016, pp. 1–9.

[76] Q. Yang, D. An, R. Min, W. Yu, X. Yang, and W. Zhao, "On optimal PMU placement-based defense against data integrity attacks in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1735–1750, Jul. 2017.

[77] W. Liao, S. Salinas, M. Li, P. Li, and K. A. Loparo, "Cascading failure attacks in the power system: A stochastic game perspective," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2247–2259, Dec. 2017.

[78] X. Zhang, X. Yang, J. Lin, G. Xu, and W. Yu, "On data integrity attacks against real-time pricing in energy-based cyber-physical systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 1, pp. 170–187, Jan. 2017.

[79] D. An, Q. Yang, W. Liu, and Y. Zhang, "Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach," *IEEE Access*, vol. 7, pp. 110835–110845, 2019.

[80] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Trans. Smart Grid.*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[81] Q. Zhang, C. Wang, H. Wu, C. Xin, and T. Phuong, "GELU-net: A globally encrypted, locally unencrypted deep neural network for privacy-preserved learning," in *Proc. IJCAI*, Jul. 2018, pp. 3933–3939.

[82] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, pp. 1310–1321.

[83] L. Yu, L. Liu, C. Pu, M. E. Gursoy, and S. Truex, "Differentially private model publishing for deep learning," 2019, *arXiv:1904.02200*. [Online]. Available: https://arxiv.org/abs/1904.02200

[84] T. Liggett, "Evolution of endpoint detection and response platforms," Ph.D. dissertation, School Arts, Utica College, Utica, NY, USA, 2018.

[85] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," 2012, *arXiv:1206.4683*. [Online]. Available: https://arxiv.org/abs/1206.4683

[86] F. Liang, W. G. Hatcher, G. Xu, J. Nguyen, W. Liao, and W. Yu, "Towards online deep learning-based energy forecasting," in *Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN)*, 2019, pp. 1–9.

[87] G. Zhou, K. Sohn, and H. Lee, "Online incremental feature learning with denoising autoencoders," in *Proc. Artif. Intell. Statist.*, 2012, pp. 1453–1461.

[88] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, to be published.

[89] H. Xu, W. Yu, D. Griffith, and N. Golmie, "A survey on industrial Internet of Things: A cyber-physical systems perspective," *IEEE Access*, vol. 6, pp. 78238–78259, 2018.

[90] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[91] J. Zhu, Y. Song, D. Jiang, and H. Song, "A new deep-Q-learning-based transmission scheduling mechanism for the cognitive Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2375–2385, Aug. 2017.

[92] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network traffic classifier with convolutional and recurrent neural networks for Internet of Things," *IEEE Access*, vol. 5, pp. 18042–18050, 2017.

[93] S. Sen, J. Koo, and S. Bagchi, "TRIFECTA: Security, energy efficiency, and communication capacity comparison for wireless IoT devices," *IEEE Internet Comput.*, vol. 22, no. 1, pp. 74–81, Jan./Feb. 2018.

[94] C. Fachkha, E. Bou-Harb, A. Keliris, N. D. Memon, and M. Ahamad, "Internet-scale probing of CPS: Inference, characterization and orchestration analysis," in *Proc. NDSS*, 2017, pp. 1–15.

[95] W. Liao, C. Luo, S. Salinas, and P. Li, "Efficient secure outsourcing of large-scale convex separable programming for big data," *IEEE Trans. Big Data*, vol. 5, no. 3, pp. 368–378, Sep. 2019.

[96] Y. Cui, W. He, C. Ni, C. Guo, and Z. Liu, "Energy-efficient resource allocation for cache-assisted mobile edge computing," in *Proc. IEEE 42nd Conf. Local Comput. Netw. (LCN)*, Oct. 2017, pp. 640–648.

[97] N. Liu, Z. Li, J. Xu, Z. Xu, S. Lin, Q. Qiu, J. Tang, and Y. Wang, "A hierarchical framework of cloud resource allocation and power management using deep reinforcement learning," in *Proc. IEEE 37th Int. Conf. Distrib. Computing Syst. (ICDCS)*, Jun. 2017, pp. 372–382.

[98] X. He, K. Wang, H. Huang, T. Miyazaki, Y. Wang, and S. Guo, "Green resource allocation based on deep reinforcement learning in content-centric IoT," *IEEE Trans. Emerg. Topics Comput.*, to be published.

[99] H. Xu, X. Liu, W. Yu, D. Griffith, and N. Golmie, "Reinforcement learning-based control and networking co-design for industrial Internet of Thing SV," Towson Univ., Towson, MD, USA, Tech. Rep., 2019.

[100] D. C. Mocanu, E. Mocanu, P. H. Nguyen, M. Gibescu, and A. Liotta, "Big IoT data mining for real-time energy disaggregation in buildings," in *Proc. IEEE Int. Conf. Syst., Man, (SMC)*, Oct. 2016, pp. 3765–3769.

[101] F. Liang, W. Yu, D. Griffith, and N. Golmie, "Towards edge-based deep learning in industrial Internet of Things," Towson Univ., Towson, MD, USA, Tech. Rep., 2019.

[102] W. Yu, D. An, D. Griffith, Q. Yang, and G. Xu, "On statistical modeling and forecasting of energy usage in smart grid," in *Proc. Conf. Res. Adapt. Convergent Syst. (RACS)*, New York, NY, USA, 2014, pp. 12–17, doi: 10.1145/2663761.2663768.

[103] Y. Huang, X. Ma, X. Fan, J. Liu, and W. Gong, "When deep learning meets edge computing," in *Proc. IEEE 25th Int. Conf. Netw. Protocols (ICNP)*, Oct. 2017, pp. 1–2.

[104] P. Zhao, D. Quan, W. Yu, X. Yang, and X. Fu, "Towards deep learning-based detection scheme with raw ECG signal for wearable telehealth systems," in *Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2019, pp. 1–9, doi: 10.1109/ICCCN.2019.8847069.

[105] Y. Wang, B. Song, P. Zhang, N. Xin, and G. Cao, "A fast feature fusion algorithm in image classification for cyber physical systems," *IEEE Access*, vol. 5, pp. 9089–9098, 2017.

[106] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[107] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[108] J. Booz, W. Yu, G. Xu, D. Griffith, and N. Golmie, "A deep learning-based weather forecast system for data volume and recency analysis," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2019, pp. 697–701.

[109] D. Ding, Q.-L. Han, Y. Xiang, C. Ge, and X.-M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, Jan. 2018.

[110] B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," *Pattern Recognit.*, vol. 84, pp. 317–331, Dec. 2018.

[111] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang, "A survey on the edge computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2018.

[112] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2805–2824, Sep. 2019.

[113] E. Quiring, A. Maier, and K. Rieck, "Misleading authorship attribution of source code using adversarial learning," in *Proc. 28th USENIX Secur. Symp. (USENIX Secur.)*. Santa Clara, CA, USA: USENIX Association, Aug. 2019, pp. 479–496. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/quiring

[114] H. L. J. Bijmans, T. M. Booij, and C. Doerr, "Inadvertently making cyber criminals rich: A comprehensive study of cryptojacking campaigns at Internet scale," in *Proc. 28th USENIX Secur. Symp. (USENIX Secur.)*. Santa Clara, CA, USA: USENIX Association, Aug. 2019, pp. 1627–1644. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/bijmans

[115] S. Islam, A. Moghimi, I. Bruhns, M. Krebbel, B. Gulmezoglu, T. Eisenbarth, and B. Sunar, "SPOILER: Speculative load hazards boost Rowhammer and cache attacks," in *Proc. 28th USENIX Secur. Symp. (USENIX Secur.)*. Santa Clara, CA, USA: USENIX Association, Aug. 2019, pp. 621–637. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/islam

[116] S. Sivakorn, J. Polakis, and A. D. Keromytis, "I'm not a human: Breaking the Google recaptcha," Black Hat, Singapore, Tech. Rep., Apr. 2016.

[117] C. Cruz-Perez, O. Starostenko, F. Uceda-Ponga, V. Alarcon-Aquino, and L. Reyes-Cabrera, "Breaking reCAPTCHAs with unpredictable collapse: Heuristic character segmentation and recognition," in *Proc. Mex. Conf. Pattern Recognit.* Berlin, Germany: Springer, 2012, pp. 155–165.

[118] W. Hu and Y. Tan, "Generating adversarial malware examples for black-box attacks based on GAN," 2017, *arXiv:1702.05983*. [Online]. Available: https://arxiv.org/abs/1702.05983

[119] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 3–18.

[120] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction apis," in *Proc. 25th USENIX Secur. Symp. (USENIX Secur.)*, 2016, pp. 601–618.

[121] W. Wei, L. Liu, M. Loper, S. Truex, L. Yu, M. E. Gursoy, and Y. Wu, "Adversarial examples in deep learning: Characterization and divergence," Jun. 2018, *arXiv:1807.00051*. [Online]. Available: https://arxiv.org/abs/1807.00051

[122] B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Šrndić, P. Laskov, G. Giacinto, and F. Roli, "Evasion attacks against machine learning at test time," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Berlin, Germany: Springer, 2013, pp. 387–402.

[123] A. Demontis, M. Melis, M. Pintor, M. Jagielski, B. Biggio, A. Oprea, C. Nita-Rotaru, and F. Roli, "Why do adversarial attacks transfer? Explaining transferability of evasion and poisoning attacks," in *Proc. 28th USENIX Secur. Symp. (USENIX Secur.)*. Santa Clara, CA, USA: USENIX Association, Aug. 2019, pp. 321–338. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/demontis

[124] D. Yu and L. Deng, "Deep learning and its applications to signal and information processing [exploratory DSP]," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 145–154, Jan. 2011.

[125] A. Chakraborty, M. Alam, V. Dey, A. Chattopadhyay, and D. Mukhopadhyay, "Adversarial attacks and defences: A survey," 2018, *arXiv:1810.00069*. [Online]. Available: https://arxiv.org/abs/1810.00069

[126] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, pp. 1322–1333.

[127] X. Wu, M. Fredrikson, S. Jha, and J. F. Naughton, "A methodology for formalizing model-inversion attacks," in *Proc. IEEE 29th Comput. Secur. Found. Symp. (CSF)*, Jun./Jul. 2016, pp. 355–370.

[128] S. Hidano, T. Murakami, S. Katsumata, S. Kiyomoto, and G. Hanaoka, "Model inversion attacks for online prediction systems: Without knowledge of non-sensitive attributes," *IEICE Trans. Inf. Syst.*, vol. E101.D, pp. 2665–2676, Nov. 2018.

[129] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Stand-alone and federated learning under passive and active white-box inference attacks," 2018, *arXiv:1812.00910*. [Online]. Available: https://arxiv.org/abs/1812.00910

[130] Y. Dong, F. Liao, T. Pang, H. Su, J. Zhu, X. Hu, and J. Li, "Boosting adversarial attacks with momentum," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9185–9193.

[131] S. Baluja and I. Fischer, "Learning to attack: Adversarial transformation networks," in *Proc. AAAI*, 2018, pp. 2687–2695.

[132] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14410–14430, 2018.

[133] I. Corona, D. Ariu, and G. Giacinto, "HMM-Web: A framework for the detection of attacks against Web applications," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009, pp. 1–6.

[134] R. Perdisci, G. Gu, and W. Lee, "Using an ensemble of one-class SVM classifiers to harden payload-based anomaly detection systems," in *Proc. ICDM*, vol. 6, 2006, pp. 488–498.

[135] D. Maiorca, G. Giacinto, and I. Corona, "A pattern recognition system for malicious PDF files detection," in *Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit.* Berlin, Germany: Springer, 2012, pp. 510–524.

[136] K. Wang, G. Cretu, and S. J. Stolfo, "Anomalous payload-based worm detection and signature generation," in *Proc. Int. Workshop Recent Adv. Intrusion Detection*. Berlin, Germany: Springer, 2005, pp. 227–246.

[137] D. Fisch, A. Hofmann, and B. Sick, "On the versatility of radial basis function neural networks: A case study in the field of intrusion detection," *Inf. Sci.*, vol. 180, no. 12, pp. 2421–2439, 2010.

[138] S. P. Chung and A. K. Mok, "Advanced allergy attacks: Does a corpus really help?" in *Proc. Int. Workshop Recent Adv. Intrusion Detection*. Berlin, Germany: Springer, 2007, pp. 236–255.

[139] M. Barreno, B. Nelson, A. D. Joseph, and J. D. Tygar, "The security of machine learning," *Mach. Learn.*, vol. 81, no. 2, pp. 121–148, 2010.

[140] H. Aljifri, M. Smets, and A. Pons, "IP traceback using header compression," *Comput. Secur.*, vol. 22, no. 2, pp. 136–151, 2003.

[141] J. Newsome, B. Karp, and D. Song, "Paragraph: Thwarting signature learning by training maliciously," in *Proc. Int. Workshop Recent Adv. Intrusion Detection*. Berlin, Germany: Springer, 2006, pp. 81–105.

[142] R. Perdisci, D. Dagon, W. Lee, P. Fogla, and M. Sharif, "Misleading worm signature generators using deliberate noise injection," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2006, p. 15.

[143] B. I. P. Rubinstein, B. Nelson, L. Huang, A. D. Joseph, S.-H. Lau, S. Rao, N. Taft, and J. D. Tygar, "ANTIDOTE: Understanding and defending against poisoning of anomaly detectors," in *Proc. 9th ACM SIGCOMM Conf. Internet Meas.*, 2009, pp. 1–14.

[144] M. Van Gundy, H. Chen, Z. Su, and G. Vigna, "Feature omission vulnerabilities: Thwarting signature generation for polymorphic worms," in *Proc. 23rd Annu. Comput. Secur. Appl. Conf. (ACSAC)*, Dec. 2007, pp. 74–85.

[145] Z. Li, M. Sanghi, Y. Chen, M.-Y. Kao, and B. Chavez, "Hamsa: Fast signature generation for zero-day polymorphic worms with provable attack resilience," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2006, p. 15.

[146] B. Nelson, M. Barreno, F. J. Chi, A. D. Joseph, B. I. Rubinstein, U. Saini, C. A. Sutton, J. D. Tygar, and K. Xia, "Exploiting machine learning to subvert your spam filter," in *Proc. LEET*, vol. 8, 2008, pp. 1–9.

[147] R. A. Servedio, "Smooth boosting and learning with malicious noise," in *Proc. Int. Conf. Comput. Learn. Theory*. Berlin, Germany: Springer, 2001, pp. 473–489.

[148] B. Bagheri, S. Yang, H.-A. Kao, and J. Lee, "Cyber-physical systems architecture for self-aware machines in industry 4.0 environment," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 1622–1627, 2015.

[149] P. Golle, D. Greene, and J. Staddon, "Detecting and correcting malicious data in VANETs," in *Proc. 1st ACM Int. Workshop Veh. Ad Hoc Netw.*, 2004, pp. 29–37.

[150] A. Ghafouri, A. Laszka, A. Dubey, and X. Koutsoukos, "Optimal detection of faulty traffic sensors used in route planning," in *Proc. 2nd Int. Workshop Sci. Smart City Oper. Platforms Eng.*, 2017, pp. 1–6.

[151] A. Ghafouri, Y. Vorobeychik, and X. Koutsoukos, "Adversarial regression for detecting attacks in cyber-physical systems," 2018, *arXiv:1804.11022*. [Online]. Available: https://arxiv.org/abs/1804.11022

[152] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning models," 2017, *arXiv:1707.08945*. [Online]. Available: https://arxiv.org/abs/1707.08945

[153] A. Jones, Z. Kong, and C. Belta, "Anomaly detection in cyber-physical systems: A formal methods approach," in *Proc. 53rd IEEE Conf. Decis. Control*, Dec. 2014, pp. 848–853.

[154] B. Dickson. (2019). *The Security Threats of Neural Networks and Deep Learning Algorithms*. [Online]. Available: https://bdtechtalks.com/2018/12/27/deep-learning-adversarial-attacks-ai-malware/

[155] Y. Liang, Z. Cai, J. Yu, Q. Han, and Y. Li, "Deep learning based inference of private information using embedded sensors in smart devices," *IEEE Netw.*, vol. 32, no. 4, pp. 8–14, Jul./Aug. 2018.

[156] A. J. Williamson, "Exploring the dark side and the downside of entrepreneurship with machine learning, sentiment analysis and experience sampling methodologies," Ph.D. dissertation, Dept. Manage. School, Univ. Sheffield, Sheffield, U.K., 2019.

[157] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[158] W. Wen, "Security analysis of a color image encryption scheme based on skew tent map and hyper chaotic system of 6th-order CNN against chosen-plaintext attack," *Multimedia Tools Appl.*, vol. 75, no. 6, pp. 3553–3560, Mar. 2016.

[159] B. Mei, Y. Xiao, R. Li, H. Li, X. Cheng, and Y. Sun, "Image and attribute based convolutional neural network inference attacks in social networks," *IEEE Trans. Netw. Sci. Eng.*, to be published.

[160] A. Rosebrock, *Deep Learning for Computer Vision With Python: ImageNet Bundle*. PyImageSearch, 2017.

[161] M. Kassner. (2017). *Using AI-Enhanced Malware, Researchers Disrupt Algorithms Used in Antimalware*. [Online]. Available: https://www.techrepublic.com/article/using-ai-enhanced-malware-researchers-disrupt-algorithms-used-in/antimalware/

[162] O. Kubovic and P. Kosinar. (2018). *Can Artificial Intelligence Power Future Malware?* [Online]. Available: https://www.welivesecurity.com/wp-content/uploads/2018/08/Can_AI_Power_Future_Malware.pdf

[163] G. Falco, A. Viswanathan, C. Caldera, and H. Shrobe, "A master attack methodology for an AI-based automated attack planner for smart cities," *IEEE Access*, vol. 6, pp. 48360–48373, 2018.

[164] S. Agarwal, A. Sureka, and V. Goyal, "Open source social media analytics for intelligence and security informatics applications," in *Proc. Int. Conf. Big Data Anal.* Berlin, Germany: Springer, 2015, pp. 21–37.

[165] F. Stark, C. Hazırbas, R. Triebel, and D. Cremers, "Captcha recognition with active deep learning," in *Proc. Workshop New Challenges Neural Comput.*, 2015, p. 94.

[166] I. Thomson. (2017). *AI Slurps, Learns Millions of Passwords to Work Out Which Ones You May Use Next*. [Online]. Available: https://www.theregister.co.uk/2017/09/20/researchers_train_ai_bots_to_crack_passwords/

[167] S. G. Lyastani, M. Schilling, S. Fahl, M. Backes, and S. Bugiel, "Better managed than memorized? Studying the impact of managers on password strength and reuse," in *Proc. 27th USENIX Secur. Symp. (USENIX Secur.)*, 2018, pp. 203–220.

[168] B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, "PassGAN: A deep learning approach for password guessing," in *Proc. Int. Conf. Appl. Cryptogr. Netw. Secur.* Berlin, Germany: Springer, 2019, pp. 217–237.

[169] (2017). *Ultra-Realistic Voice Cloning and Text-to-Speech*. [Online]. Available: https://lyrebird.ai/

[170] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 16–24, 2013.

[171] P.-Y. Chen, H. Zhang, Y. Sharma, J. Yi, and C.-J. Hsieh, "ZOO: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models," in *Proc. 10th ACM Workshop Artif. Intell. Secur.*, 2017, pp. 15–26.

[172] D. Gershgorn. (2017). *Microsoft's AI is Learning to Write Code By Itself, Not Steal It*. [Online]. Available: https://qz.com/920468/artificial-intelligence-created-by-microsoft-and-university-of-cambridge/is-learning-to-write-code-by-itself-not-steal-it/

[173] P. Arntz. (2018). *How Artificial Intelligence and Machine Learning Will Impact Cybersecurity*. [Online]. Available: https://blog.malwarebytes.com/security-world/2018/03/how-artificial-intelligence-and-machine-learning-will-impact-cybersecurity/

[174] M. Beltov. (2017). *Artificial Intelligence Can Drive Ransomware Attacks*. [Online]. Available: https://www.informationsecuritybuzz.com/articles/artificial-intelligence-can-drive-ransomware-attacks/

[175] R. M. Gerdes, C. Winstead, and K. Heaslip, "CPS: An efficiency-motivated attack against autonomous vehicular transportation," in *Proc. 29th Annu. Comput. Secur. Appl. Conf.*, 2013, pp. 99–108.

[176] Y. Chen, S. Kar, and J. M. F. Moura, "Optimal attack strategies subject to detection constraints against cyber-physical systems," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1157–1168, Sep. 2018.

[177] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *IEEE Control Syst.*, vol. 21, no. 6, pp. 11–25, Dec. 2001.

[178] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, and T. A. Lillicrap, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, Dec. 2018.

[179] D. Churchill. (2018). *A History of Starcraft AI Competitions StarCraft AI Competition*. [Online]. Available: https://www.cs.mun.ca/~dchurchill/starcraftaicomp/history.shtml

[180] (2018). *An Old-Fashioned AI Has Won a Starcraft Shootout*. [Online]. Available: https://www.technologyreview.com/f/612438/an-old-fashioned-ai-has-won-a-starcraft-shootout/

[181] J. Vincent. (2018). *Did Elon Musk's AI Champ Destroy Humans at Video Games? It's Complicated*. [Online]. Available: https://www.theverge.com/2017/8/14/16143392/dota-ai-openai-bot-win-elon-musk

[182] (2018). *OpenAI Five*. [Online]. Available: https://openai.com/five/

[183] E. Bursztein. (2018). *Elie*. [Online]. Available: https://elie.net/blog/hearthstone/i-am-a-legend-hacking-hearthstone-with-machine-learning-defcon-talk/-wrap-up/

[184] E. Bursztein, "I am a legend: Hacking hearthstone using statistical learning methods," in *Proc. IEEE Conf. Comput. Intell. Games (CIG)*, Sep. 2016, pp. 1–8.

[185] X. Zhang, S. Shan, S. Tang, H. Zheng, and B. Y. Zhao, "Penny auctions are predictable: Predicting and profiling user behavior on dealdash," in *Proc. 29th Hypertext Social Media (HT)*, New York, NY, USA, 2018, pp. 123–127, doi: 10.1145/3209542.3209576.

[186] Z. Chen and R. C. Qiu, "Q-learning based bidding algorithm for spectrum auction in cognitive radio," in *Proc. IEEE Southeastcon*, Mar. 2011, pp. 409–412.

[187] J. Seymour and P. Tully, "Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter," Black Hat, Alexandria, VA, USA, Tech. Rep., 2016, vol. 37.

[188] W. Melicher, B. Ur, S. M. Segreti, S. Komanduri, L. Bauer, N. Christin, and L. F. Cranor, "Fast, lean, and accurate: Modeling password guessability using neural networks," in *Proc. 25th USENIX Secur. Symp. (USENIX Secur.)*, 2016, pp. 175–191.

[189] (2018). *Inside the Pentagon's Race Against Deepfake Videos*. [Online]. Available: https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

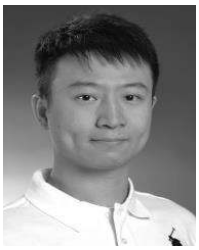
[190] D. Kirat. (2018). *DeepLocker Concealing Targeted Attacks With AI Locksmithing*. [Online]. Available: https://i.blackhat.com/us-18/Thu-August-9/us-18-Kirat-DeepLocker-Concealing-Targeted-Attacks-with-AI-/Locksmithing.pdf

[191] W. Brendel, J. Rauber, and M. Bethge, "Decision-based adversarial attacks: Reliable attacks against black-box machine learning models," 2017, *arXiv:1712.04248*. [Online]. Available: https://arxiv.org/abs/1712.04248

[192] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

FAN LIANG received the bachelor's degree in computer science from Northwestern Polytechnical University, China, in 2005, and the master's degree in computer engineering from the University of Massachusetts Dartmouth, in 2015. He is currently pursuing the Ph.D. degree in computer science with Towson University. His current research interests include big data, the Internet of Things, and security.

WILLIAM GRANT HATCHER received the B.Sc. degree in materials science and engineering from the University of Maryland, and the master's degree in computer science from Towson University, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include mobile computing and security, big data, and machine learning.



WEIXIAN LIAO received the B.S. degree in information engineering from Xidian University, Xi'an, China, in 2012, the M.S. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2015, and the Ph.D. degree in computer engineering from Case Western Reserve University, Cleveland, OH, USA, in 2018. He is currently an Assistant Professor with the Department of Computer and Information Sciences, Towson University. His current research interests include cybersecurity and optimization in big data applications, cyber physical systems, and machine learning.

WEICHAO GAO received the B.S. degree from Fudan University, Shanghai, China, in 2005, the M.B.A. degree from the University of Michigan, MI, USA, in 2011, and the M.S. degree in computer science and technology from Towson University, MD, USA, in 2017, where he is currently pursuing the Ph.D. degree. His current research interests include the Internet of Things, cyberspace security, and computer networks.

WEI YU received the B.S. degree in electrical engineering from the Nanjing University of Technology, Nanjing, China, in 1992, the M.S. degree in electrical engineering from Tongji University, Shanghai, China, in 1995, and the Ph.D. degree in computer engineering from Texas A&M University, in 2008. He was with Cisco Systems, Inc., for nine years. He is currently a Full Professor with the Department of Computer and Information Sciences, Towson University, MD, USA. His current research interests include cyberspace security and privacy, cyber-physical systems, the Internet of Things, and big data. He was a recipient of the 2014 NSF Faculty CAREER Award, the 2015 University System of Maryland (USM) Regents' Faculty Award for Excellence in Scholarship, Research, or Creative Activity, the University System of Maryland (USM)'s Wilson H. Elkins Professorship Award, in 2016, and the Best Paper Awards from the IEEE ICC 2008, ICC 2013, IEEE IPCCC 2016, and WASA 2017.