KICS The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 8 (2022) 313–321

www.elsevier.com/locate/icte

Uniting cyber security and machine learning: Advantages, challenges and future research

Mohammad Wazid[a], Ashok Kumar Das[b,*], Vinay Chamola[c], Youngho Park[d,*]

[a] *Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun 248 002, India*
[b] *Center for Security, Theory and Algorithmic Research, International Institute of Information Technology, Hyderabad 500 032, India*
[c] *Department of Electrical and Electronics Engineering & APPCAIR, BITS-Pilani, Pilani Campus, 333 031, India*
[d] *School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, Republic of Korea*

Received 20 January 2022; received in revised form 31 March 2022; accepted 13 April 2022
Available online 21 April 2022

Abstract

Machine learning (ML) is a subset of Artificial Intelligence (AI), which focuses on the implementation of some systems that can learn from the historical data, identify patterns and make logical decisions with little to no human interventions. Cyber security is the practice of protecting digital systems, such as computers, servers, mobile devices, networks and associated data from malicious attacks. Uniting cyber security and ML has two major aspects, namely accounting for cyber security where the machine learning is applied, and the use of machine learning for enabling cyber security. This uniting can help us in various ways, like it provides enhanced security to the machine learning models, improves the performance of the cyber security methods, and supports effective detection of zero day attacks with less human intervention. In this survey paper, we discuss about two different concepts by uniting cyber security and ML. We also discuss the advantages, issues and challenges of uniting cyber security and ML. Furthermore, we discuss the various attacks and provide a comprehensive comparative study of various techniques in two different considered categories. Finally, we provide some future research directions.

Keywords: Cyber security; Machine learning; Internet of Things (ioT); Privacy; Security; Intrusion detection

1. Introduction

In the current era of computing devices, most of the devices that we use are connected to the Internet in an Internet of Things (IoT) environment. These devices share and transmit their data through the insecure (open) communication medium, also called as the Internet. Most of the time this data is sensitive in nature (i.e., healthcare data, banking data, insurance data, other finance related data, and social security numbers). The malicious entities, such as the online attackers (hackers) are always in search of that, where they play with the things (for example, they can launch attacks, like replay, man-in-the-middle, impersonation, credential guessing, session key computation, malware injection and data modification) [1,2]. Therefore, from time-to-time several researchers propose different security protocols to mitigate these attacks. The security protocols or cyber security protocols can be divided into different categories: "authentication protocols", "access control protocols", "intrusion detection protocols", "key management protocols", and "blockchain enabled security protocols". The summary of these protocols is given below.

- **Authentication protocols:** Authentication is a process of checking the genuineness (authenticity) of someone of some device. It can be performed through some credentials or factors (i.e., username, password, smartcard, biometrics), which are closely associated with the users or device. We can have user to user authentication, user to device authentication or device to authentication. On the basis of available factors, user authentication protocols can be again divided into three categories, i.e., one-factor user authentication protocol, two-factor user authentication protocol and three-factor user authentication protocol.
- **Access control protocols:** Access control is a process of putting restrictions on the unauthorized access of someone or some device(s). Users or devices can access the other users or devices in a secure way after the completion of all steps of a user/device access control protocol. Access control protocols

* Corresponding authors.
E-mail addresses: wazidkec2005@gmail.com (M. Wazid), ashok.das@iiit.ac.in (A.K. Das), vinay.chamola@pilani.bits-pilani.ac.in (V. Chamola), parkyh@knu.ac.kr (Y. Park).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2022.04.007

can be divided into two categories: (1) user access control and (2) device access control. User access control protocol can be used for the access control of the unauthorized users, whereas device access control protocol can be used for the access control of the unauthorized devices. Access control can be of certificate based or certificate less. Authorization is considered as a process through which an authority (i.e., a server) determines if an entity (i.e., a client) has permission to use the resource. It is usually performed in collaboration with authentication so that the server can know who the client is who's requesting access. It determines who has permission to access a resource and who does not.

- **Intrusion detection protocols:** An intrusion is something or somebody with the malicious intention. This may be some malicious programming script or some Internet attacker system, which is under the control of some hacker. Usually hackers try to inject some malware in the online devices to affect their performance or to breach the security of these devices (systems). For the detection and mitigation of intrusion, we need a specific category of protocols that come under "intrusion detection protocols". The intrusion detection can be performed in different ways i.e., signature based intrusion detection, anomaly based intrusion detection or hybrid intrusion detection, which is combination of both signature based and anomaly based schemes. The machine learning based or deep learning based intrusion detection (i.e., malware detection) is going very famous these days.
- **Key management protocols:** Key management protocols are used for secure key management among the various entities, such as some devices (for example, smart Internet of Things (IoT) devices and smart vehicles) and some users (smart home user, doctor, traffic inspector). Usually, a trusted registration authority does the registration of all entities of the communication system and then stores the secret credentials (i.e., secret keys) in their memory. We need a key management process for the fresh keys generation and their storing in the devices, key establishment and key revocation purposes. The devices/users can exchange their information in a secure way after the establishment of shared secret key (i.e., a session key), which may happen through the essential steps of an authenticated key agreement protocol.
- **Blockchain enabled security protocols:** Blockchain is one of the emerging technology of the era. Blockchain maintains data in the form of certain blocks, which are chained together with some hash values. In blockchain data is maintained in the form of distributed ledger, which is named as distributed ledger technology (DLT). All the genuine parties (sometimes miner) of the network have access to the DLT. The data that we store over the blockchain safe and secured against the various possible cyber attacks. Thus, the blockchain enabled security protocols are capable of defending the various cyber attacks [3].

Machine learning (ML) is a process in which computing systems learn from data and use algorithms to execute tasks without being explicitly programmed. Deep learning (DL) is a subset of artificial intelligence (AI) that is a sort of ML. DL is based on a complicated set of algorithms, which are modeled on the human brain. This allows unstructured data, such as documents, photos, and text, to be processed. ML refers to a computer's ability to think and behave without the need for human involvement. However, DL typically needs less ongoing human intervention. Due to this, it can analyze images, videos, and unstructured data in a better way than the traditional ML algorithms [4,5].

The uniting of cyber security and machine learning can help us in various ways. For example, enhanced security to the machine learning models, improved performance of the cyber security methods, effective detection of zero day attacks with less human intervention. However, it may suffer from various issues and security challenging, which should be handled carefully. Therefore, in this particular domain, we need some review study related to the "uniting of cyber security and machine learning" i.e., issues and challenges, various attacks, different protection schemes with their comparative study and some future research directions on which other researchers should work. Hence, we tried to conduct such studies in the proposed work [4,5].

2. Uniting cyber security and machine learning

2.1. Machine learning in cyber security

The systems, which are connected in the cyber space, are prone to various kind of attacks i.e., replay, man-in-the-middle (MiTM), impersonation, credentials leakage, password guessing, session key leakage, unauthorized data update, malware injection, flooding, denial of service (DoS) and distributed denial of service (DDoS) and many more. Therefore, we need some security protocol to detect and mitigate these attacks. The machine learning models (machine learning ML algorithms) can learn about various cyber attacks in the offline/online mode through the provided pre-processed dataset. The ML algorithms detect any sign of intrusion (some cyber attack) in the real time i.e., online mode. The scenario of "machine learning in cyber security" is depicted in Fig. 1. Here, we have a Internet connected system (i.e., laptops, desktops, smartphones, IoT devices), which can be used to perform various online tasks i.e., online financial transactions, online access of healthcare data, social security numbers, etc. Hackers are always in search of some vulnerabilities in such systems and if they get anything like that then they start their attacks. For the detection and mitigation of cyber attacks, different kinds of ML techniques i.e., supervised learning, unsupervised learning, reinforcement learning and deep learning can be used as per the situation. It is up-to the communication environment and available resources of the systems, which technique (i.e., i.e., supervised learning, unsupervised learning, reinforcement learning and deep learning) suites them in the best way. The learning (training) and prediction (testing) of cyber attacks can be done through the cloud servers as they have good computation and storage resources.

2.2. Cyber security in machine learning

The scenario of "cyber security in machine learning" is given in Fig. 2, which is also referred to machine learning

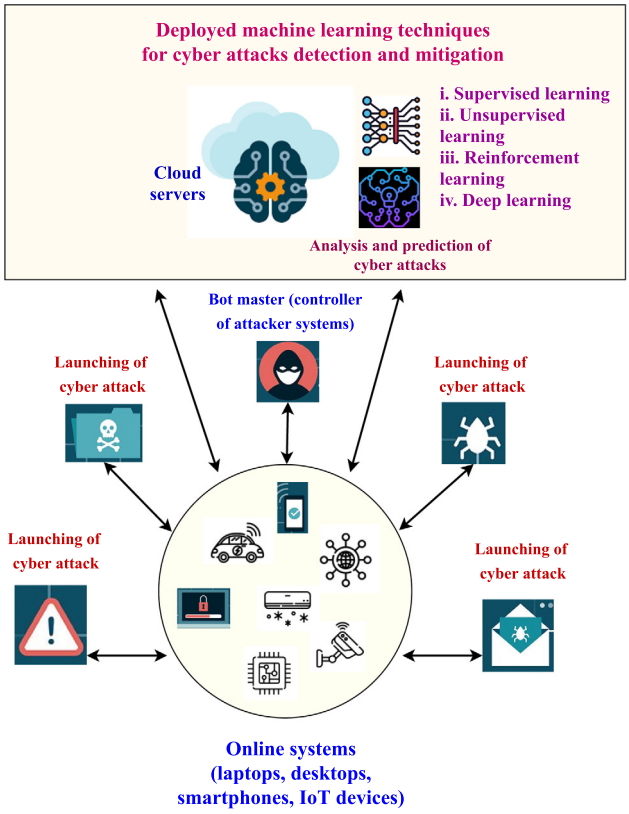

Fig. 1. Scenario of machine learning in cyber security.

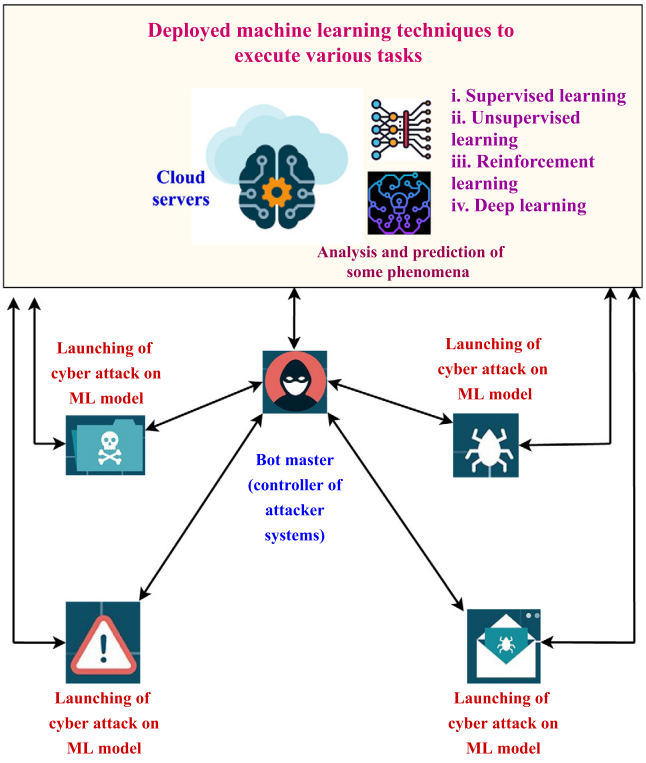

Fig. 2. Scenario of cyber security in machine learning.

(ML) security. The ML models are used for the analysis and prediction of various phenomena. However, the performance of ML models can be affected through the launching of some attacks i.e., dataset poisoning attack, model poisoning attack, privacy breach attack, membership inference attacks, runtime disruption attack, etc., [6]. Under the influence of these attacks ML models may give wrong prediction about the associated phenomena. In the "dataset poisoning attack", an attacker inserts adversarial examples (updated values) in dataset that causes the ML model to produce wrong predictions. Further in the "model poisoning attack" the attacker's focus is on corrupting the models by interfering with their internal workings and modifying the parameters. In the "privacy breach attack" the attacker works on exposing the sensitive data and he /she also tries for retrieving the valuable information of the model. Membership inference attack is the part of privacy breach. Furthermore, in the "runtime disruption attack" the attacker compromises the workflow of ML to affect the accurate prediction results by attacking the execution process of the model. Hence, there is a need for some cyber security mechanisms (i.e., encryption techniques, signature generation and verification techniques, hashing mechanisms) to protect against these attacks. Under the deployment of these cyber security mechanisms the ML models and the associated datasets become secure and we get the correct outcomes and prediction.

3. Advantages of uniting cyber security and machine learning

Both cyber security and machine learning are essential for each other and can improve their mutual performances. Some of the advantages of their uniting are as follows.

- **Full proof security of ML models:** As discussed earlier, the ML models are vulnerable to various attacks. The occurrence of these attacks may affect the working, performance and predictions of the ML models. However, these unwanted incidences can be secured through the deployment of certain cyber security mechanisms. Under the deployment of cyber security mechanisms the working and performance and the input datasets of the ML models become secured and we get the correct predictions and results [7].
- **Improved performance of cyber security techniques:** When we use the ML algorithms in the cyber security schemes (i.e., intrusion detection systems) that improve their performances (i.e., improved accuracy and detection rate with less false positive rate). ML techniques, like supervised learning, unsupervised learning, reinforcement learning and deep learning algorithms can be used as per the communication environment and the associated systems.
- **Effective detection of zero day attacks:** The cyber security methods, which detect the intrusion through the ML models seem very effective for the detect of zero day attacks (i.e., unknown malware attacks). It happens because they perform the detection with the help of some deployed ML models. The ML models work through collection and matching of certain features, if the features of a program matches with the malicious program's features then that can be considered as

the malicious program. This detection task can be performed by the ML models automatically. Thus, detection of zero day attacks can be performed effectively with the uniting of cyber security and machine learning.

• **Limited requirements of human intervention:** Most of the task in the ML based systems happen through the deployed ML models. When we unit cyber security with ML, most of the tasks for which these systems are deployed, they do that without any human involvement or with very less human intervention.

• **Quick scanning and mitigation:** The ML based intrusion detection systems work very efficiently to detect the presence of the attacks because they work through certain ML algorithms. Therefore, uniting of machine learning with the cyber security systems performs the scanning of intrusions very fast and also provide fast response in case of any sign of intrusion. The only thing that we need to take care is the suitable ML algorithm selection.

4. Overview of various threats and attacks

In this section, we provide the details of the following various attacks, which may occur in different computing environments.

• *Eavesdropping:* This attack is passive in nature which is also known as sniffing or snooping attack. In this attack, an adversary tries to listen the secret conversation of the communicating parties.

• *Traffic analysis:* This attack is passive in nature. In this attack, an adversary $\mathcal{A}$ intercepts the ongoing conversation and then examines the messages to get information like type of conversation, its pattern and behavior, location tracking and timing of information. The intercepted data further helps $\mathcal{A}$ to launch other associated attacks.

• *Replay attack:* In this attack, $\mathcal{A}$ intentionally does the retransmission of the captured messages, which were exchanged in the past. $\mathcal{A}$ does this to fool or misdirect the recipient and makes the legitimate users to act as per $\mathcal{A}$'s desires.

• *Man-in-the-middle attack (MiTM):* In this active attack, $\mathcal{A}$ makes independent connections with communicating entities and relays the messages to the both ends. Under such situations, the two communicating entities think that they are directly communicating with each other. Thus, $\mathcal{A}$ may intercept, delete, modify or insert a new information for transmission without any recognition [8].

• *Impersonation attack:* This attack is also active in nature, wherein $\mathcal{A}$ imitates one of the legitimate party of the network by deducing its identity and then transmits the modified or some fresh messages on the behalf that party to the other legitimate party.

• *Denial-of-Service (DoS) attack:* In DoS attack, $\mathcal{A}$ sends multiple fake requests (i.e., HTTP flood messages) to flood the victim's computing resources. Therefore, the service request of the legitimate user cannot be processed. Under such situation, the legitimate user cannot get the service of the network. There is another form of DoS attack, which is known as distributed denial-of-service (DDoS) attack in which $\mathcal{A}$ uses multiple machines (i.e., botnet) to send multiple request simultaneously to the victim's machine that consumes all computing resources of the system and that happens very fast. DoS or DDoS attacks can be performed through various flooding attacks i.e., SYN flood, HTTP flood, UDP flood, etc.

• *Malware attack:* These attacks are performed through the execution of malicious script at the victim's machine. The injected/installed malware is a file or a code, which performs unauthorized activities in the systems, such as stealing of data, illegal encryption of the drive or the stored data, modification of data, or deletion of data. Some malware types include keylogger, spyware, virus, ransomware, worms, trojan horse, etc., [6].

• *Scripting attack:* These attacks refer to the disclosure of information from some online database, which are maintained with some web server (i.e., online banking database). For example, "password cracking, structured query language (SQL) injection attack and cross-site scripting (XSS) attack" can be used to get the secret information from the system, like passwords, credit and debit card details.

• *Privileged insider attack:* This attack is performed by any privileged user of the system, who has access to the registration information of various users and devices. Since privileged insider has access to the sensitive information, this attack becomes a lot harder to defend and also has more adverse impact.

• *Physical stealing of smart devices:* These days most of the computing environments are operated through the use of smart devices, such as smart home appliances, smart healthcare devices, smart manufacturing devices. The smart devices are deployed without any physical security. If these smart devices are physically stolen by an adversary $\mathcal{A}$, they can be used for the extraction of sensitive information by making the use of power analysis attacks. After the extraction of sensitive information, the unauthorized tasks like illegal session key computation can be performed [9].

• *Birthday attack:* A birthday attack is a type of cryptographic attacks that takes an advantage of the mathematics behind the birthday problem, which may be found in a probability theory. The birthday attacks can be used for the malevolent purposes, such as guessing credentials (passwords). As described in the birthday paradox, this attack is based on a fixed degree of permutations and the higher possibility of collisions identified between random attack attempts. The birthday paradox (birthday problem) addresses the likelihood that some paired people in a group of n randomly selected people will share a birth date. The mathematics behind this problem inspired the birthday attack, a well-known cryptographic attack, that uses this probabilistic strategy to reduce the difficulty of cracking a hash function [10].

• *Dictionary attack:* A dictionary attack is a type of brute-force attack on a cryptographic system that is carried out maliciously. By systematically typing every word in a dictionary as a password, the attacker attempts to breach the system's security. A dictionary attack can also be used to find out what key is needed to decrypt a communication or document that has been encrypted. The attacker tries to break

the encryption or get access through a library of phrases or keywords that has been kept up to date. The words from a dictionary or numeric sequences can be utilized for automatic insertion into the target. The dictionary attacks are made easier by poor password utilization, such as upgrading passwords with sequential numbers, symbols, or characters. It works because some people use common words as passwords. These attacks are typically unsuccessful against systems that use multi-word passwords. Moreover, the passwords composed of uppercase and lowercase letters, and numbers in random combinations are also difficult for an attacker to break [10].

• *Stolen verifier attack:* In this malicious act, an attacker first tries to steal some devices (i.e., smart IoT devices) and then performs a power analysis attack on the memory units of these devices to extract sensitive information (i.e., secret credentials and keys) from their memory. The attacker eavesdrops some of the exchanged messages and then uses the extracted information to launch other potential attacks in the network, like unauthorized session key computation, password guessing, MiTM and impersonation attacks.

• *Unauthorized session key computation attack:* In this malicious act, an attacker tries to compute the session key, which is established between the legitimate entities of the network. To perform this task, the attacker tries various methods, such as physical device stolen attack, privileged insider attack and stolen verifier attack. It is always recommended to use the long term secrets (i.e., pseudo identities, secret keys) and short term secrets (i.e., random secret nonce values) for the computation of the session keys. This mechanism gives distinct keys in different sessions among different entities. Unfortunately, if a session key is revealed to the attacker, other session keys will be in safe hand, and it will provide the security to the remaining part of the communication.

• *Attacks on machine learning models:* We can broadly classified the attacks on ML model into four categories: (a) dataset poisoning attack, (b) model poisoning attack, (c) privacy breach attack and (d) runtime disruption attack [11].

• *Dataset poising attack:* In this attack, $\mathcal{A}$ uses the different methods to invade the training and testing data to affect the normal functioning of the ML task. $\mathcal{A}$ can use adversarial examples to attack the data server from where raw data has to be extracted. The compromising of the data sources helps to inserts the erroneous data, which possibly alters the functioning of the ML model. This further changes the output of the ML based system [12].

• *Model poisoning attack:* In model poisoning attack, $\mathcal{A}$ does parameter alteration through which $\mathcal{A}$ generates faulty output via interfering with the classifier. The parameters through which the classifier prepares ML model get altered. $\mathcal{A}$ can change sensitivity limits, rate of accession and cause under-fitting or over-fitting that further affects the normal execution of ML task [13].

• *Privacy breach:* The user's sensitive data and model's internal working mechanism can be compromised via various methods. The unprotected files and absence of encryption mechanism in the training and deployment phases of the ML task can cause the leaking of data. That further enables the unauthorized user to interfere with the model. It increases the privacy risks associated with the data as the privacy of the sensitive data may be breached [14]. Papernot et al. [15,16] discussed the different privacy preserving schemes to protect the privacy of model. They also discussed about the usage of noise generation to provide differential privacy to the data and ML model by "randomizing model's behavior" [17].

• *Runtime disruption attack:* An adversary $\mathcal{A}$ uses this task to delay or end the ongoing ML task. $\mathcal{A}$ usually targets the server at the time of deployment phase. Then, $\mathcal{A}$ tries to remotely disrupt the ongoing ML process. As a result, the normal functioning of the ML task gets disturbed, which results in wastage of the time and resources. $\mathcal{A}$ identifies the weak points (vulnerabilities) and penetrates the run time server through various attacks, like phishing, denial-of-service (DoS) attack and SQL injection attack. This attack can be mitigated through the decentralizing of ML work space. The blockchain based mechanisms can be deployed to further bifurcate and implement "distributed machine learning", which help to protect the integrity and privacy of the user's data and the associated ML models.

5. Issues and challenges of uniting of cyber security and machine learning

Though uniting of cyber security and machine learning provides enormous number of advantages. At the same time it has some issues and challenges, which need to be handled very carefully. Some of them are discussed below.

• **Compatibility issues:** The uniting of cyber security and machine learning contains different types of security techniques (i.e., encryption algorithms, signature generation and verification algorithms, hashing algorithms) and machine learning algorithms (clustering, classification, convolutional neural networks (CNNs)). Moreover, the data, which is the main input for analysis process comes from the different sources i.e., IoT devices. These IoT devices are operated through different communication techniques. During the amalgamation of these many algorithms, there may be the issues related to the compatibility. Therefore, we have to very selective, which algorithm works well with which algorithm and scheme. Hence compatibility related issues should be handled very carefully [18].

• **Overloading:** In uniting cyber security and machine learning, we use various algorithms as discussed earlier. For the execution of such algorithms, we need the resources in extra amount. Otherwise, the system will not work properly. Therefore, the amalgamation and use of various algorithms may cause the overloading to the system that may further affect the actual working of the system. For example, we cannot allocate entire resources of the system for the security related processes. We also need some resources for the execution of ML-related tasks. Hence, we should select the algorithms wisely and as per the resources of the communication environment. For example, for an encryption purpose, we would prefer to use the symmetric-key based encryption, known as the Advanced Encryption Standard (AES) algorithm

in place of any public key cryptographic algorithm for the secure communication of IoT, since AES requires less computation, communication and storage costs as compared to public key cryptographic algorithms. In that situation, we can also allocate the resources of the system for the execution of important tasks.

• **Accuracy:** In the uniting of cyber security and machine learning, we use various ML mechanisms i.e., machine learning (ML) models to predict about some physical phenomena (i.e., chances of roadside accident in the intelligent transportation system). The ML models work with the help of certain datasets, if we have some error in the dataset or in the settings of the ML model then this can give big trouble. For example, the obtained accuracy is not fully correct [19].

• **Flaws in security mechanisms:** In the uniting of cyber security and ML, we may use various cyber security mechanisms. If these mechanisms have some flaws, it may then cause the trouble to the security to the system. Most of the time, the hackers try to search for the zero-day vulnerabilities and then exploit them. In such situations, the sensitive data of the system may be revealed, changed or it may become unavailable. Therefore, the designers of the security protocols should have to be very careful while they design a new security protocol. The security of the newly designed protocol can be tested through certain mechanisms, like the Automated Validation of Internet Security Protocols and Applications (AVISPA) [20], which checks the security of the protocol against the replay and man-in-the-middle attacks through the formal security verification. Moreover, we can also go for the "Burrows–Abadi–Needham (BAN) logic test [21], which identifies the possibility of "secure mutual authentication among the communicating entities". Apart from these, we can also analysis the formal security of a security protocol through the Real-or-Random (ROR) model [22] implementation, which identifies the possibility of unauthorized session key computation attack on the designed authentication or access control or key management protocol. The security of the designed protocol can be evaluated and analyzed in this way.

6. Comparative study

In this section, we have done the comparison of various techniques in the categories of "machine learning in cyber security"and "cyber security in machine learning". The details are given below.

6.1. Performance comparison of machine learning in cyber security protocols

We perform a comparative analysis on the performance of various ML based intrusion (i.e., malware) detection schemes.

6.1.1. Summary of considered schemes

Kumar et al. [23] proposed a framework, which combined the advantages of ML models and blockchain to improve the intrusion detection for the IoT devices. They have implemented it via a sequential approach, i.e., through "clustering, classification, and blockchain". ML has automatically extracted the malware information using clustering and classification algorithms after that this information was stored over the blockchain network. Lei et al. [11] proposed a security scheme called as "EveDroid, a scalable and event-aware malware detection system". Unlike existing schemes, their scheme directly uses event group to describe apps' behaviors, which could capture higher level of semantics for the detection works. Nguyen et al. [24] proposed a technique for the Linux IoT botnet detection. The detection was based on the combination of "PSI graph and CNN classifier". Dinakarrao et al. [25] proposed a two-pronged method to detect the intrusions, where they had a runtime malware detector (HaRM) by employing "Hardware Performance Counter (HPC)" values to detect the malware and benign applications. Su et al. [26] proposed a light-weight technique to detect the DDoS malware in the IoT environments. They have extracted the malware images and utilized a light-weight convolutional neural network for their classification. The comparison of these schemes is given in Table 1.

For the comparative study part, we have only considered the recent schemes of the domain. The different performance parameters, like precision, recall, accuracy and F1-score are used. We compute these parameters through parameters, such as true positive (TP), false positive (FP), true negative (TN) and false negative (FN). If a "normal program" is detected as a "normal program" by the intrusion detection scheme, then it is called as "true negative (TN)"; whereas if a "normal program" is detected as a "malicious program" by the intrusion detection scheme, then it is called as "false positive (FP)". Similarly, if a "malicious program" is detected as a "malicious program" by intrusion detection scheme, it is called as "true positive (TP)"; whereas if a "malicious program" is detected as a "normal program" by intrusion detection scheme, it is called as "false negative (FN)" [23].

• **Precision:** It is known as "positive predicted value" which is the fraction of the correctly identified intrusion cases to all the predicted positive cases of intrusions, where Precision $= \frac{TP}{TP+FP}$.

• **Recall:** It is also known as "true positive rate or detection rate or sensitivity" which is treated as a fraction of correctly identified intrusion cases to the all real positive cases of intrusions, where Recall $= \frac{TP}{TP+FN}$.

• **Accuracy:** It is one of the most important parameters measured as the all correctly identified cases, which is formulated as Accuracy $= \frac{TP+TN}{TP+FP+TN+FN}$.

• **F1-score:** It is also known as F1-measure that is calculated through the harmonic mean of precision and recall. It gives the accurate estimate of the incorrectly classified cases than the accuracy and is formulated as F1-score $= \frac{2(\text{Precision}\times\text{Recall})}{\text{Precision + Recall}}$.

6.1.2. Comparison of various schemes

The comparison of the different schemes i.e., scheme of Kumar et al. [23], Lei et al. [11], Nguyen et al. [24], Dinakarrao et al. [25] and Su et al. [26] is given in Table 1. The schemes of Kumar et al. [23], Nguyen et al. [24], Dinakarrao et al. [25] and Su et al. [26] provide the accuracy value of

Table 1
Comparison of different ML-based intrusion detection schemes.

Scheme	Method used	Accuracy	F1-measure
Kumar et al. [23]	"Blockchain enabled ML driven detection"	98.00%	98.00%
Lei et al. [11]	"EveDroid"	N/A	99.00%
Nguyen et al. [24]	"Graph-based convolution neural network (CNN)"	92.00%	94.00%
Dinakarrao et al. [25]	"HaRM malware detector"	92.21%	N/A
Su et al. [26]	"Light-weight convolutional neural network (CNN)"	94.00%	N/A

Note: N/A: not available.

98%, 92%, 92.21% and 94%, respectively. Contrary to that the schemes of Kumar et al. [23], Lei et al. [11] and Nguyen et al. [24] provide the F1-measure of 98%, 99% and 94%, respectively. From the comparison, it has been observed that Kumar et al.'s scheme [23] provides better accuracy. However, the scheme of Lei et al. [11] provides maximum F1-measure.

6.2. Performance comparison of cyber security in machine learning protocols

We provide the details of different schemes that can secure the ML models.

6.2.1. Summary of considered schemes

Jagielski et al. [27] proposed a defense mechanism, which was resilient against the different poisoning attacks. They also provided formal guarantees about its convergence and an upper bound on the effect of poisoning attacks. Peri et al. [28] proposed a Deep k-NN defense mechanism against the "collision and convex polytope clean-label attacks" on "CIFAR-10 dataset". Chen et al. [29] presented De-Pois, an attack-agnostic defense mechanism for poisoning attacks. The key idea of their scheme was to train a mimic model. They have done this to imitate the behavior of the target model. Phong et al. [30] proposed a deep learning based mechanism to protect the gradients over the "honest-but-curious cloud server" via "additively homomorphic encryption". The various gradients were encrypted and stored on the cloud server. Payman and Zhang [31] proposed a mechanism to support secure arithmetic operations on shared decimal numbers, and propose MPC-friendly alternatives to non-linear functions such as "sigmoid and softmax", which were superior to the other existing schemes. Chen et al. [32] proposed a scheme for the detection and removal of backdoor for the neural networks. They demonstrated the effectiveness of proposed scheme for the "neural networks classifying text and images". As per their cliam it was the first scheme to capable of detecting poisonous data. Liu et al. [33] proposed the effective defense mechanism for the prevention of backdoor attacks. They implemented three backdoor attacks and used them to investigate two promising defense approaches, "pruning and fine-tuning". They then evaluated the "fine-pruning", which was the combination of "pruning and fine-tuning". Weber et al. [34] provided the unified framework via "randomized smoothing techniques". It had shown how it could be instantiated to certify the robustness against the "evasion and backdoor attacks". After that they presented the "robust training process (in short RAB) to smooth the trained model and to certify its resilience against the backdoor attacks. They derived the robustness bound for ML models trained with RAB. The comparison of these schemes is given in Table 2.

Table 2
Comparison of different cyber security based ML schemes.

Scheme	Type of attack	Method used	Accuracy
Jagielski et al. [27]	Poisoning attack	A reduced loss function to isolate points that deemed poisoned	76.80%
Peri et al. [28]	Poisoning attack	Isolation of poisoned point via comparison with nearest neighbors	91.80%
Chen et al. [29]	Poisoning attack	Generative adversarial networks to reconstruct a clean model	93.10%
Phong et al. [30]	Privacy attack	Preserving privacy via additive homomorphic encryption	97.00%
Payman and Zhang [31]	Privacy attack	Privacy preserving method	98.62%
Chen et al. [32]	Accessing attack	Activation cluster based scheme for backdoors removal	99.97%
Liu et al. [33]	Accessing attack	Combining fine tuning and pruning defense for efficient protection	98.6%
Weber et al. [34]	Accessing attack	Model deterministic test-time augmentation method to detect backdoor's existence	97.70%

6.2.2. Comparison of various schemes

From Table 2, it is clear that the scheme of Chen et al. [29] provides high accuracy i.e., 93.10% in the detection of poisoning attack. Moreover, the scheme of Payman and Zhang [31] provides high accuracy i.e., 98.62% accuracy in the detection of privacy breach attacks. Furthermore, the scheme of Chen et al. [32] provides high accuracy i.e., 99.97% accuracy in the detection of accessing attacks.

7. Future research

In this section, we discuss some of the future research directions of the "uniting of cyber security and machine learning", which should be considered by the researchers working in the same domain. S

• **Secrecy of exchanged and stored data:** Secrecy of the exchanged and stored data matters a lot. To maintain the secrecy of the data different types of security protocols have been proposed. However, these protocols fail in case of any flaw in the design or due to the happening of some zero day attack. Therefore, there is some scope of improvements as

online attackers (hackers) are going advance and use advance tools to break the security of the system. Hence, there is a requirement of new security protocols with additional security and functionality features, which can resist the zero day vulnerabilities as well.

• **Compatibility of different mechanisms and tools:** The "uniting of cyber security and ML" uses various mechanisms and tools (i.e., different types of security techniques. like encryption algorithms, signature generation and verification algorithms, hashing algorithms and machine learning algorithms, such as clustering, classification, CNNs). They also require different kind of hardware and configurations. Under such circumstances, there may be some issues related to the compatibility of these mechanisms and tools.

• **Overloading and performance:** In the uniting of cyber security and ML, we use various algorithms as discussed earlier. For the execution of these many algorithms we need some extra resources. Otherwise, the tasks will not be executed properly. Therefore, amalgamation and use of various algorithms may cause the overloading to the system that may further affect the actual working of the system. Hence, we should select the algorithms wisely and try to invent new lightweight algorithms may be in ML or in the security, which consume less resources of the systems.

• **Improvement in accuracy of the system:** The ML models work with the help of certain datasets, if we have some error in the dataset or in the settings of the ML model then this can cause problems. For example, the obtained accuracy is not fully correct or the system may make wrong prediction about something. Therefore, the researchers should work to overcome from such situations, new methods can be invented to detect the errors in the datasets or to improve the accuracy of the systems.

8. Lesson learned

We discussed about two different concepts by uniting cyber security and ML. We then discussed the advantages, issues and challenges of uniting cyber security and ML. Some of the advantages are as follows: "full proof security of ML models", "improved performance of cyber security techniques", "effective detection of zero day attacks" and "quick scanning and mitigation". However, this uniting also has some issues and challenges, like "compatibility issues", "overloading", "accuracy", etc. Furthermore, we discussed various attacks of the domain (i.e., eavesdropping, traffic analysis, replay, MiTM, impersonation, DoS, malware insertion, scripting, birthday, physical stealing of smart devices, dictionary, dataset poising, model poisoning and runtime disruption attacks. After that, we provided a comprehensive comparative study of various techniques in two different considered categories. For example, the scheme of Kumar et al. [23] performed better under the category of "machine learning in cyber security", whereas the scheme of Chen et al. [32] performed better under the category of "cyber security in ML". Some future research directions (i.e., "secrecy of exchanged and stored data", "compatibility of different mechanisms and tools", "overloading and performance" and "improvement in accuracy of the system") were also given so that other researchers could provide some solutions for those. Thus, there is a trade-off between the learning cost and performance. For example, DL is costlier than ML, however, it attains good predictive scores. Moreover, if we want to put more security, we need to invest more on the system resources.

9. Conclusion

We presented the details of two different concepts by uniting of cyber security and machine leaning: "machine learning in cyber security" and "cyber security in machine learning". We then discussed the advantages, issues and challenges of uniting of cyber security and ML. Further, we highlighted the different attacks and also provided a comparative study of various techniques in two different considered categories. Finally, some future research directions are provided.

Declaration of competing interest

The authors declare that there is no conflict of interest in this paper.

Acknowledgments

This research was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2020R1I1A3058605, and in part by the BK21 FOUR project funded by the Ministry of Education, Korea under Grant 4199990113966. The authors would like to thank the anonymous reviewers and the associate editor for their valuable suggestions and feedback on the paper.

References

[1] I. Butun, P. Osterberg, H. Song, Security of the internet of things: Vulnerabilities, attacks, and countermeasures, IEEE Commun. Surv. Tutor. 22 (1) (2020) 616–644, http://dx.doi.org/10.1109/COMST.2019.2953364.

[2] Z. Lv, L. Qiao, J. Li, H. Song, Deep-learning-enabled security issues in the internet of things, IEEE Internet Things J. 8 (12) (2021) 9531–9538.

[3] Y. Wang, J. Yu, B. Yan, G. Wang, Z. Shan, BSV-PAGS: Blockchain-based special vehicles priority access guarantee scheme, Comput. Commun. 161 (2020) 28–40.

[4] N. Magaia, R. Fonseca, K. Muhammad, A.H.F.N. Segundo, A.V. Lira Neto, V.H.C. de Albuquerque, Industrial internet-of-things security enhanced with deep learning approaches for smart cities, IEEE Internet Things J. 8 (8) (2021) 6393–6405.

[5] S.A. Parah, J.A. Kaw, P. Bellavista, N.A. Loan, G.M. Bhat, K. Muhammad, V.H.C. de Albuquerque, Efficient security and authentication for edge-based internet of medical things, IEEE Internet Things J. 8 (21) (2021) 15652–15662.

[6] Y. Sun, A.K. Bashir, U. Tariq, F. Xiao, Effective malware detection scheme based on classified behavior graph in IIoT, Ad Hoc Netw. 120 (2021) 102558.

[7] J. Yang, Z. Bian, J. Liu, B. Jiang, W. Lu, X. Gao, H. Song, No-reference quality assessment for screen content images using visual edge model and AdaBoosting neural network, IEEE Trans. Image Process. 30 (2021) 6801–6814.

[8] Y. Zhao, J. Yang, Y. Bao, H. Song, Trustworthy authorization method for security in industrial internet of things, Ad Hoc Netw. 121 (C) (2021).

[9] T.S. Messerges, E.A. Dabbish, R.H. Sloan, Examining smart-card security under the threat of power analysis attacks, IEEE Trans. Comput. 51 (5) (2002) 541–552.

[10] M.R.K. Soltanian, I.S. Amiri, Chapter 3 - problem solving, investigating ideas, and solutions, in: M.R.K. Soltanian, I.S. Amiri (Eds.), Theoretical and Experimental Methods for Defending Against DDOS Attacks, Syngress, 2016, pp. 33–45.

[11] T. Lei, Z. Qin, Z. Wang, Q. Li, D. Ye, EveDroid: Event-aware android malware detection against model degrading for IoT devices, IEEE Internet Things J. 6 (4) (2019) 6668–6680.

[12] J. Steinhardt, P.W. Koh, P. Liang, Certified defenses for data poisoning attacks, in: 31st International Conference on Neural Information Processing Systems, in: NIPS'17, Curran Associates Inc. Long Beach, California, USA, 2017, pp. 3520–3532.

[13] M. Aladag, F.O. Catak, E. Gul, Preventing data poisoning attacks by using generative models, in: 1st International Informatics and Software Engineering Conference, UBMYK, Ankara, Turkey, 2019, pp. 1–5, http://dx.doi.org/10.1109/UBMYK48245.2019.8965459.

[14] C. Huang, S. Chen, Y. Zhang, W. Zhou, J.J.P.C. Rodrigues, V.H.C. de Albuquerque, A robust approach for privacy data protection: IoT security assurance using generative adversarial imitation learning, IEEE Internet Things J. (2021) 1, http://dx.doi.org/10.1109/JIOT.2021.3128531.

[15] N. Papernot, P. McDaniel, X. Wu, S. Jha, A. Swami, Distillation as a defense to adversarial perturbations against deep neural networks, in: 2016 IEEE Symposium on Security and Privacy, 2016, pp. 582–597, http://dx.doi.org/10.1109/SP.2016.41.

[16] N. Papernot, A marauder's map of security and privacy in machine learning, in: 11th ACM Workshop on Artificial Intelligence and Security, Toronto, Canada, 2018.

[17] S. Pirbhulal, W. Wu, K. Muhammad, I. Mehmood, G. Li, V.H.C. de Albuquerque, Mobility enabled security for optimizing IoT based intelligent applications, IEEE Netw. 34 (2) (2020) 72–77.

[18] J. Yang, Y. Han, Y. Wang, B. Jiang, Z. Lv, H. Song, Optimization of real-time traffic network assignment based on IoT data using DBN and clustering model in smart city, Future Gener. Comput. Syst. 108 (2020) 976–986.

[19] R.R. Guimaraes, L.A. Passos, R.H. Filho, V.H.C.d. Albuquerque, J.J.P.C. Rodrigues, M.M. Komarov, J.P. Papa, Intelligent network security monitoring based on optimum-path forest clustering, IEEE Netw. 33 (2) (2019) 126–131.

[20] A. Armando, D. Basin, Y. Boichut, Y. Chevalier, L. Compagna, J. Cuellar, P.H. Drielsma, P.C. Heám, O. Kouchnarenko, J. Mantovani, S. Mödersheim, D. von Oheimb, M. Rusinowitch, J. Santiago, M. Turuani, L. Viganò, L. Vigneron, The AVISPA tool for the automated validation of internet security protocols and applications, in: K. Etessami, S.K. Rajamani (Eds.), Computer Aided Verification, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 281–285.

[21] M. Burrows, M. Abadi, R. Needham, A logic of authentication, ACM Trans. Comput. Syst. 8 (1) (1990) 18–36, http://dx.doi.org/10.1145/77648.77649.

[22] M. Abdalla, P.A. Fouque, D. Pointcheval, Password-based authenticated key exchange in the three-party setting, in: 8th International Workshop on Theory and Practice in Public Key Cryptography, PKC'05, in: Lecture Notes in Computer Science, vol. 3386, Les Diablerets, Switzerland, 2005, pp. 65–84.

[23] R. Kumar, X. Zhang, W. Wang, R.U. Khan, J. Kumar, A. Sharif, A multimodal malware detection technique for android IoT devices using various features, IEEE Access 7 (2019) 64411–64430.

[24] H.-T. Nguyen, Q.-D. Ngo, V.-H. Le, IoT botnet detection approach based on PSI graph and DGCNN classifier, in: 2018 IEEE International Conference on Information Communication and Signal Processing, ICICSP, Singapore, Singapore, 2018, pp. 118–122.

[25] S.M. Pudukotai Dinakarrao, H. Sayadi, H.M. Makrani, C. Nowzari, S. Rafatirad, H. Homayoun, Lightweight node-level malware detection and network-level malware confinement in IoT networks, in: Design, Automation Test in Europe Conference Exhibition, DATE, Florence, Italy, 2019, pp. 776–781.

[26] J. Su, D.V. Vasconcellos, S. Prasad, D. Sgandurra, Y. Feng, K. Sakurai, Lightweight classification of IoT malware based on image recognition, in: IEEE 42nd Annual Computer Software and Applications Conference, Vol. 02, COMPSAC, Tokyo, Japan, 2018, pp. 664–669.

[27] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, B. Li, Manipulating machine learning: Poisoning attacks and countermeasures for regression learning, in: IEEE Symposium on Security and Privacy, SP, San Francisco, CA, USA, 2018, pp. 19–35.

[28] N. Peri, N. Gupta, W.R. Huang, L. Fowl, C. Zhu, S. Feizi, T. Goldstein, J.P. Dickerson, Strong baseline defenses against clean-label poisoning attacks, in: ECCV Workshop, 2020, pp. 55–70.

[29] J. Chen, X. Zhang, R. Zhang, C. Wang, L. Liu, De-pois: An attack-agnostic defense against data poisoning attacks, 2021, CoRR, arXiv:2105.03592.

[30] L.T. Phong, Y. Aono, T. Hayashi, L. Wang, S. Moriai, Privacy-preserving deep learning via additively homomorphic encryption, IEEE Trans. Inf. Forensics Secur. 13 (2018) 1333–1345, http://dx.doi.org/10.1109/TIFS.2017.2787987.

[31] P. Mohassel, Y. Zhang, SecureML: A system for scalable privacy-preserving machine learning, in: IEEE Symposium on Security and Privacy, S&P, San Jose, USA, 2017, pp. 19–38, http://dx.doi.org/10.1109/SP.2017.12.

[32] B. Chen, W. Carvalho, N. Baracaldo, H. Ludwig, B. Edwards, T. Lee, I. Molloy, B. Srivastava, Detecting backdoor attacks on deep neural networks by activation clustering, in: SafeAI@AAAI, Honolulu, USA, 2019.

[33] K. Liu, B. Dolan-Gavitt, S. Garg, Fine-pruning: Defending against backdooring attacks on deep neural networks, 2018, CoRR, arXiv:1805.12185.

[34] M. Weber, X. Xu, B. Karlas, C. Zhang, B. Li, RAB: Provable robustness against backdoor attacks, 2020, arXiv, arXiv:2003.08904.