IMPROVING CYBERSECURITY USING AI

[1]Mrs. Anitha J, [2]Mr. Jagadish S, [3]Darshan K M

[1]*Asst. Professor of Dr. Ambedkar Institute of Technology, Dept. of MCA, Bangalore – 560060, Karnataka, India*
[2]*Student of Dr. Ambedkar Institute of Technology, Dept. of MCA, Bangalore – 560060, Karnataka, India*
[3]*Student of Dr. Ambedkar Institute of Technology, Dept. of MCA, Bangalore – 560060, Karnataka, India*

ABSTRACT

Because of the growing number of potential dangers and the persistent efforts of cybercriminals to stay one step ahead of law enforcement, cybersecurity is a rapidly developing field that has been a constant topic of discussion in the media over the past decade. Although the primary reasons for carrying out cyberattacks have remained largely unchanged over the years, the methods used in cybercrime have become increasingly sophisticated. Traditional cybersecurity solutions are becoming increasingly insufficient when it comes to detecting and mitigating new forms of cyberattack. Recent developments in cryptographic and artificial intelligence (AI) techniques, in particular machine learning and deep learning, hold the potential to equip cybersecurity professionals with the tools necessary to combat the ever-evolving threat posed by cybercriminals. In this article, we explore the potential of artificial intelligence to improve cybersecurity solutions by identifying both the benefits and drawbacks of AI. In addition, we discuss potential future research opportunities that are associated with the development of AI techniques in the field of cybersecurity across a variety of application domains.

KEYWORDS: ***Artificial intelligence, cybersecurity, cyberattacks, machine learning.***

I. INTRODUCTION

The term "cybersecurity" refers to a collection of practises, human behaviours, and computerised systems that work together to keep electronic resources secure. In a manner that is analogous to Moore's law, which predicts the doubling of components on an integrated circuit every two years (along with decreasing costs associated with chip manufacturing), cyber criminals are increasingly doubling the effectiveness of their attack tools for half the cost every few months [1]. It is anticipated that spending on cybersecurity will exceed $1 trillion worldwide between 2017 and 2021 [2,] and that spending on cybersecurity has already increased by almost 40 percent since 2013 to reach $66 billion [3]. Artificial Intelligence (AI)-based methods for enhancing cybersecurity have become an area of focus for researchers in the field of cybersecurity over the course of the past few years. In a similar vein, cybercriminals are employing AI in order to launch increasingly complex cyberattacks while simultaneously covering their digital footprints. This is being accomplished through the use of AI. However, the primary focus of this investigation is on how AI-based cybersecurity solutions might be able to deter attackers more successfully, thereby lowering the risk of data breaches or eliminating them entirely. The development These advancements are a direct result of the development of artificial intelligence. since its inception in the 1950s, led to a great deal of fascinating research that was conducted and systems that were developed. Following subsequent developments, machine learning and deep learning came into existence [4]. [Citation needed] Today, AI is being used in a wide variety of application areas, including healthcare, agriculture, space, the legal system, and manufacturing [5]–[9]. The continuous performance improvements in computer hardware and software (along with their decreasing costs), combined with new paradigms such as big data and cloud computing, have led to the development and deployment of a wide range of AI systems with varying capabilities. These AI systems have been made possible by the continuous performance improvements in computer hardware and software. Many of these AI systems can now perform a wide variety of difficult tasks, including learning, planning, problem solving, decision making, and the recognition of faces and voices. Since the 1980s, one more significant advancement in the field of artificial intelligence has been the emergence of technologies known as machine learning. These technologies assist computer systems in learning and adapting to different conditions by making use of their previous experiences, patterns, and knowledge. Deep learning is a subfield of machine learning that emerged about ten years ago. This subfield enables machines to discover hidden relationships in their input data, which in turn generates more accurate results for planning and predicting. Deep learning is also known as neural networks. Over the past little while, we have seen a rising interest in the application of of AI and machine learning techniques to fight cyberattacks. A strong motivation for the use of these techniques stems from the large amounts of data that are constantly being produced today, which requires significant resources and time to analyze and detect any patterns, anomalies, or intrusions in traffic data.

In a recent report by Juniper research, the authors predict that the cost of cybersecurity incidents will increase from

$3 trillion each year to more than $5 trillion in 2024, an aver- age yearly growth of 11 percent [10]. The key sources of cyberthreats include [11]:

1) *Script kiddies*: These are novices who have trained to create cyberattack tools to hack into vulnerable com- puting systems and to make a quick buck or boost their ego through such activities.
2) *Criminal organizations*: These include those involved in illegal operations, who launch cyberattacks that can cause a Denial of Service (DoS), steal data or state secrets as a result of data breaches, seek payments through ransomware, and so on.
3) *Nation states*: This involves state-sponsored cyber- criminal activities perpetrated against enemy nations with the intent of crippling the victim nation's economy or critical infrastructures, causing fatalities, disruption of state-sponsored programs, or to ultimately topple the government.
4) *Terrorists*: They attempt to cause nationwide losses and major disruptions to society's critical infrastructures, such as causing massive power outages in a victim country through cyberattacks.
5) *Spies (including business rivals)*: They steal trade secrets to gain an unfair market advantage.
6) *Disgruntled employees*: Employees who are stressed and unhappy with their jobs, rifts with management, or other factors may attempt to cause financial or rep- utation losses to the organization by carrying out a cyberattack against corporate resources.
7) *External attackers and insider threats*: Experts with a strong knowledge about the operation of computing resources as well as human behavior, who attempt to exploit vulnerable systems and gain (mainly finan- cially) through such acts or simply cause major disrup- tions to the organization's normal operations.

One type of threat that's becoming more prevalent and continuously evolving in complexity over the years is the zero-day threat which has not been previously seen by cybersecurity or software/hardware development staff. Con- sequently, the attacker exploits the computing resources' security vulnerability (software or hardware) the same day it becomes known. When a zero-day attack targets a soft- ware vulnerability, the patching of the security hole must be initiated from the software developer or vendor as quickly as possible. Such security patches take time to be created and rolled out on a global scale. During this interim period, all non-patched systems are exposed to the cyberthreat of the zero-day vulnerability. An example of such a threat is zero-day malware that can easily penetrate a target system while bypassing malware detection software such as anti- virus. Cybercriminals are using advanced techniques for code obfuscation, defined as concealment of malicious code within ‘‘legitimate-appearing code’’ that can be delivered to a vic- tims' system in the form of an email attachment. Naïve users may open these attachments or click an embedded link to a malicious website, leading to system compromise and more severe consequences—including data held for ransom, compromise, and even sensitive data disclosure. Hidden mal- ware within ads that appear on legitimate websites are also a clever technique for compromising end-user systems through zero-day exploits. Even the most up-to-date security software will not be able to detect obfuscated code embedded within such adware [12].

The German AV-TEST GmbH research institute for IT security registers more than 350,000 new malware programs and potentially unwanted applications every day. In fact, in 2019, the institute identified more than 140 million new malware programs, which translates to an equivalent of 266 types of malware every minute [13].

As the sophistication of cyberthreats increases, the key drivers pushing for increased cybersecurity at the corporate level include:

1) *Lack of cyber governance skills at the C-level.* Exec- utives such as the Chief Information Security Offi- cer (CISO) and the Chief Information Officer (CIO), do not easily make changes in security strategy at the corporate level. Such changes would safeguard corpo- rate resources against the ever-evolving and dynamic nature of cyber threats of contemporary times. The aggravating factor is the fact that cyber criminals are not privy to C-Level culture of organizations, and there- fore cybersecurity is increasingly posing a concern at executive meetings [1].
2) *Opportunities to harness state-of-the-art cybersecu- rity detection techniques.* Current computing systems become more efficient in data crunching, while at the same time the data required for cybersecurity analysis has become available. This trend has advanced cyber- security analysis techniques such as machine learning, data mining, and knowledge discovery. Data mining is a subcomponent of knowledge discovery, where a spe- cific sequence of steps is applied to data with the intent of extracting patterns. In addition, knowledge discov- ery also comprises data cleaning, selection, and the application of prior knowledge and established tech- niques for interpreting the results extracted. Machine learning and data mining significantly overlap, as they employ similar methods and processes. Whilst machine learning focuses on classification of data samples and prediction of events or behaviors, data mining focuses on the discovery of previously unseen patterns in data (very much similar to detection of zero-day cyber-attacks). The advancement of these techniques has become one of the key drivers for organizations to

achieve their goals, including their cybersecurity vig- ilance [14]–[16].

3) *Fragmented cybersecurity frameworks.* Despite hav- ing a plethora of frameworks for securing an orga- nization's resources against cyberthreats, the choice remains a largely difficult question for an organiza- tion's cybersecurity decision makers. Some industries such as the insurance sector do not have a proper reference model to follow to ensure the requisite cyber- security. This is attributed mainly to the lack of con- sumer data to build legitimate and illicit profiles, upon which machine learning or AI techniques can be applied; definitions of fraud differ between the insurance sector and the banking sector [17]. In the for- mer case, insurers mainly worry about policies being opened without a priori customer knowledge, and they operate in a fragmented regulatory environment. For instance, unlike banking, the insurance industry is not tightly regulated in the US, consequently encumbering the adoption of silver-bullet cyberprevention strategies because they invariably depend upon regulation. There- fore, the industry-specific cybersecurity framework, or lack thereof, hinders the realization of cybersecu- rity goals in a wide range of industries [18].A similar concern arises in Supervised Control and Data Acquisi- tion (SCADA) systems that comprise a range of com- mercial off-the-shelf hardware and software and rely upon standardized communication protocols. While integrity and availability are important cybersecurity concerns for SCADA systems, confidentiality is sec- ondary [19]. Precedence is typically given to safety, reliability, robustness, and maintainability of such sys- tems, and therefore security takes a backseat [20].

Research contributions of this work

We summarize the main contributions of this work as follows:

- We present an overview of the cybersecurity threat land- scape and discuss traditional security solutions (i.e., non-AI based solutions) that have been used to protect from the various threats.
- We discuss the weaknesses of traditional cybersecurity solutions and describe how emerging AI solutions can improve cybersecurity.
- Finally, we present some key challenges faced by the cybersecurity community that must be addressed in the future.

II. CYBERSECURITY THREATS AND LEGACY CYBERSECURITY SOLUTIONS

Over the last decade, many types of cyberthreats have emerged. Next, we briefly review those threats. According to a recent report [21], the top 10 cyberthreats we face today include:

1) *Denial of Service (DoS) attacks*: These attempt to overwhelm a victim system's computing resources by sending an overwhelming number of requests for it to process within a short period of time. Such attacks can be carried out in one of several ways: a single attacker machine can launch a DoS attack against a victim machine by transmitting a large number of network traffic packets that appear to be legitimate, to bypass security controls along the way; multiple attacker machines can participate in a distributed-style DoS attack, i.e., a Distributed Denial of Service (DDoS) attack, resulting in a similar outcome at the victim machine. DoS attacks are increasingly becoming more sophisticated and harder to detect, because of the ready availability of attacker tools, as well as the proliferation of the CyberCrime as a Service (CCaaS) market [22].

2) *Man-in-The-Middle (MiTM) attacks*: These are legacy cyberattacks carried out through the process of inter- ception of transmitted data on a communication line between two legitimate communicating parties. The attacker places itself either physically or virtually between two communicating parties, A and B, posing as A to communicate with B through the interception of A B messages and replacing these with malicious or tampered messages, and repeating the same process on the BA communication line, i.e., posing as party B and speaking to party A. Variant implementations of such an attack include IP address spoofing, wherein the malicious actor convinces legitimate systems that it is a trusted entity, enabling system access for the actor. A message replay attack involves the repeat transmission of a previously stored, stale message on the communication line, perpetrated by the malicious actor.

3) *Phishing and spear-phishing attacks*: These are car- ried out by crafting emails that appear legitimate and transmitting them to legitimate systems, with the intent of having the naïve end users click a link and divulge personal information. Such attacks exploit social engi- neering principles, wherein emails are made to appear legitimate to end users, luring them to trust them. Spear phishing is defined as a carefully designed attack that involves a thorough background search carried out by the malicious actor on susceptible victims, for subse- quent drafting of emails that appear to be very legit- imate, with the ''from'' field often containing trusted email addresses.

4) *Drive-by attacks*: These are carried out by malicious actors who skim through the web and search for vulner- able websites, so that they can implant malware scripts into the webservers. End users who visit the website are eventually infected with the malware, leading to system compromise, disclosure of sensitive data, and other damage.

5) *Password attacks*: These can be carried out by shoul- der surfing user keyboard activity, brute force into a system using common passwords, and crafting sophis- ticated passwords through the application of AI tech- niques [23], [24].

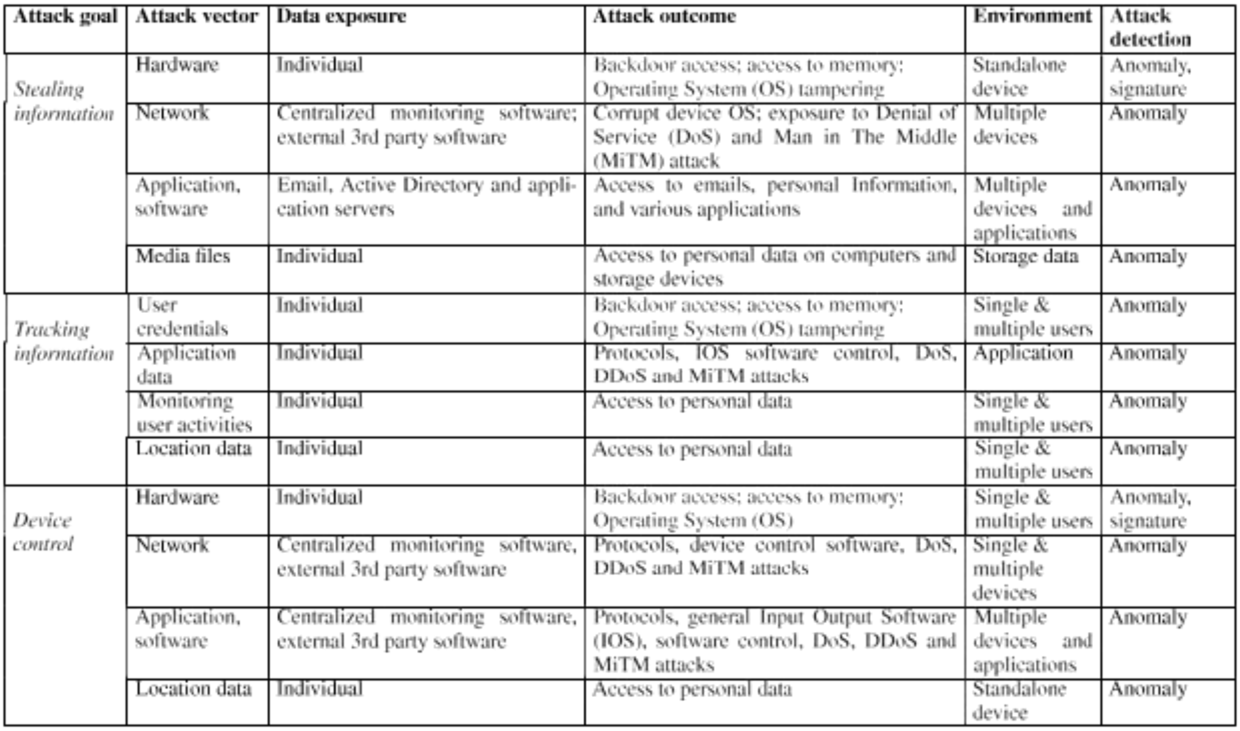
6) *Structured Query Language (SQL) injection attacks*: These are legacy cyberattacks that exploit vulnerabil- ities in the SQL language by injecting a webpage with input fields with SQL query code, that when executed at the webserver, would disclose some or all of the stored content on a backend database server, possibly including usernames and passwords.
7) *Cross-site scripting attacks*: These are carried out by injecting malicious code in a vulnerable web- server. Subsequent retrieval of the hosted webpages by naïve end-users would infect the victim's machine with malware. Such malware may transmit user data from the victim's machine to the malicious actor's servers, and may lead to the subsequent hijacking of web ses- sions, theft of credentials, installation of key stroke loggers, capture screenshots, and even taking control of the victim's machine remotely.
8) *Eavesdropping attacks*: These can be carried out by sniffing out the network communication line and mis- using obtained data. Malicious actors may either pas- sively sniff the line and obtain user data or actively attack the line, replacing messages with fictitious mes- sages, and masquerade as legitimate users.
9) *Birthday attacks*: This hash of a message, also known as a message digest, which can be computed using a standard algorithm such as the Secure Hash Algorithm- 1 (SHA-1). When this algorithm is applied to a message of arbitrary length, the output is a hash value of fixed length. The birthday attack refers to the attempt by a malicious actor to find two different messages that produce the same hash value. Consequently, the orig- inal message can be replaced with the other message that produces the same hash value, causing system and service disruption and data loss. Such attacks apply AI techniques to discover random messages that produce the same hash value as a legitimate message [25]
10) *Malware attacks*: One of the main difficulties to web-hosting organizations is that their websites can become the source of malware spread. According to Symantec's 2016 threat report, 78 percent of websites contain a critical vulnerability that can be exploited by the adversary to allow malicious code to run with- out any user interaction [26]. Strengthening a web- site's defenses involves deploying appropriate security controls such as web proxies, firewalls, and intrusion detection systems. A major issue here is the tradeoff between the right level of security controls and usabil- ity of websites being hosted. The higher the level of a website's usability, the greater the area of vulnerability for the website.

Network attacks are launched on the environment to disrupt services, steal individual/corporate data, and gain network intelligence. Malicious users exploit the Operating System's (OS's) weakness to gain access and tamper with the OS to achieve their malicious objectives. Some of these attacks are used to steal individual information, which can be used to gain access to individual/corporate data. In Table 1, we classified various network attacks based on their attack objectives, expected targeted device or application, data/ information exposed when specific attack is underway, type of environment affected when certain attacks occur, and how these attacks are detected.

Next, we briefly discuss traditional (non-AI) cybersecurity techniques for detecting cyberattacks:

1) *Game theory:* This has been previously applied to cybersecurity [27]–[29]. The malicious actor is con- sidered as one player in a game, and the victim's machine is the other player. Each player attempts to maximize his/her incentive through strategic move- ment, in which the player rationally justifies that the goal would be reached by the move. Each player's behaviors either can be known beforehand or remain concealed. An example of a game could be a smart grid environment where the attacker attempts to disrupt communication between a power system and a home, whereas the defender attempts to maintain connectivity between these various entities [30]–[32]. At each step of the game, the attacker and the defender would adopt strategies to be successful in their respective goals [33].
2) *Rate control:* Attacks against the availability of systems include DoS and DDoS. Rate-control techniques can minimize the impact on such systems' operation when they are under attack by reducing the volume of incom- ing network traffic, through basic traffic throttling and redefining permission lists [34].
3) *Heuristics:* Firewalls and intrusion detection systems commonly rely on heuristics to identify the most apt rule for classifying network traffic as legitimate or anomalous. One such technique [35], performs a sequence of steps comprising substring matching in order to identify suspicious website addresses. The second phase of the presented scheme comprises the scanning of the web address through the VirusTotal application (i.e. a website where one can supply a web address and gets a scored analysis about the degree of maliciousness of the input website), with the low- est score of the two scans considered for deciding on whether to let the data packets into the network or not.
4) *Signature-based intrusion detection:* A signature-based intrusion detection system makes use of a database that may store legitimate signatures corresponding to normal traffic or attack signatures corresponding to malicious traffic. The intrusion detection system matches the contents of incoming network packets with the stored signatures in real time [36]. This technique's drawback is that in the absence of relevant signatures, intrusion detection systems are limited in their capa- bilities to accurately detect malicious traffic entering a network.
5) *Anomaly-based intrusion detection:* This technique creates a model of what can be perceived as the norm. The models can be in terms of rule-based

TABLE 1. Various types of attacks, their impact, and approaches to detect them.



Attack goal	Attack vector	Data exposure	Attack outcome	Environment	Attack detection
Stealing information	Hardware	Individual	Backdoor access; access to memory; Operating System (OS) tampering	Standalone device	Anomaly, signature
	Network	Centralized monitoring software; external 3rd party software	Corrupt device OS; exposure to Denial of Service (DoS) and Man in The Middle (MiTM) attack	Multiple devices	Anomaly
	Application, software	Email, Active Directory and application servers	Access to emails, personal Information, and various applications	Multiple devices and applications	Anomaly
	Media files	Individual	Access to personal data on computers and storage devices	Storage data	Anomaly
Tracking information	User credentials	Individual	Backdoor access; access to memory; Operating System (OS) tampering	Single & multiple users	Anomaly
	Application data	Individual	Protocols, IOS software control, DoS, DDoS and MiTM attacks	Application	Anomaly
	Monitoring user activities	Individual	Access to personal data	Single & multiple users	Anomaly
	Location data	Individual	Access to personal data	Single & multiple users	Anomaly
Device control	Hardware	Individual	Backdoor access; access to memory; Operating System (OS)	Single & multiple users	Anomaly, signature
	Network	Centralized monitoring software, external 3rd party software	Protocols, device control software, DoS, DDoS and MiTM attacks	Single & multiple devices	Anomaly
	Application, software	Centralized monitoring software, external 3rd party software	Protocols, general Input Output Software (IOS), software control, DoS, DDoS and MiTM attacks	Multiple devices and applications	Anomaly
	Location data	Individual	Access to personal data	Standalone device	Anomaly

policies [37], mathematical models [38], and statistical techniques [39]. Deviations from the norm are regarded as attacks. When compared to the signature-based detection, such techniques have the advantage of being relieved from depending on signature patterns, thereby removing them from administrative efforts to collect signatures.

6) *Autonomous systems:* These have the capability to self-protect and self-heal, and to ensure reliability and availability [40], as in the case of the Bionic Autonomic Nervous System (BANS). This system is comprised of four different modules, namely, Cyber Neuron, Cyber Axon, Peripheral Nerve and Central Nerve. Cyber Neuron is used to protect against spy- ware and malware. Cyber Axon is an intelligent tool to recover from damage caused by spyware and malware. Similarly, Peripheral Nerve provides a robust defense against DoS/DDoS attacks by establishing a commu- nication path between multiple cyberneurons deployed on different devices. Last, Central Nerve serves as a knowledge base against new attacks and to dissemi- nate information to other security devices. Collabora- tive defense by peripheral nerves is proposed to block DoS and DDoS attacks through cooperation between devices within the network.
7) *End user security controls:* Current end-user devices such as mobile phones, smart portable devices (iPads), and personal computers require in-built security rather than add-ons [41]. End users might not update their devices with the latest security patches, with some vendors attempting to push automatic updates, in order to install security patches. The Wannacry ran- somware [42], [43] attack is an example of an attack wherein the latest security patches provided by the vendor were not applied on all the end-user devices. Most of the time users are not aware of the impli- cations of not applying the patches. In some cases, although some users may be aware of this fact, they do not either take the requisite action for securing their devices or they carry out incorrect procedures, expos- ing the devices through other vulnerabilities. A sug- gested control [41] is to perform ''out of sight'' secu- rity, where automatic updates are pushed by vendors directly to end-user devices without the user's involve- ment. However, the challenge would be that software vendors must ensure that the security updates guard against new attacks (also known as zero-day attacks) and work seamlessly with all pre- existing software on the end-user device.

III. ARTIFICIAL INTELLIGENCE

AI is concerned with how machines can think or act correctly, given what they know [44]. This universal definition includes how closely machines can think or act like humans (Fig. 1). At one end of the spectrum, machines are deemed to be intelligent if they can maximize the outcome on every state of the process. At the other end of the spectrum, the Turing Test [45] sets the standard on machine intelligence. Under

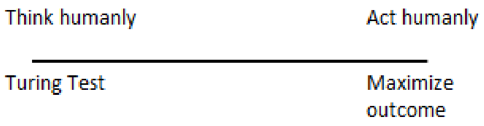


FIGURE 1. Spectrum on intelligent measures from thinking humanly through the Turing Test, to acting humanly to maximize the outcome.this test, a computer communicating with a human is said to have intelligence when the human cannot distinguish whether the responses come from a computer or a human. At both sides of the spectrum, AI embodies computing areas such as natural language processing, knowledge representation, logic, automated reasoning, machine learning, mathematics, and game theory. Early AI applications gave rise to thinking machines that solved puzzles such as geometry [46], checker games [47], and a family of blocks-world problems.

After the proliferation of the Internet in the late 1990s, software that behaved like humans gained popularity in terms of agent-based AI, commonly called bots. Ethical bots were made to spider the Internet for the benefit of search engines, yellow pages, and recommendation lists. They pro- vide protection against vandalism in Wikipedia articles where anybody can contribute as authors [48]. In contrast, mali- cious bots also emerged to cheat in online games [49], post spams [50], [51] and spread malware [52]. In mimicking online games, bot programmers analyzed the traffic flow between the game console and server to reverse engineer the game code [49], [53]. In posting spams, the bots mimicked the behavior of human when online, such as surfing the pages before posting a message in a forum, rather than continuously posting messages [51]. Malicious bots discourage cyber ser- vices to function properly, costing the service providers to have disheartened online visitors. As a result, some of the cybersecurity research investigated solutions that can detect and protect again malicious bots. Studies found that game bots were active longer, were less social e.g. exchanging items or participating in an auction, and have less variations in their sequence of actions when compared to human [49]. Furthermore, game bots are more interested to collect items, while human players seek to collaborate with other players to complete challenges/quests [48]. Similarly, spambots and malware bots can be detected from their behaviors being different than human, that can be detected through some distinctive communication patterns [50], [52].

The most relevant AI applications to the cybersecurity area are in intrusion detection systems [54]. Cybersecurity solutions often perform traffic analysis, where the Inter- net traffic is classified as either legitimate or malicious. At the dawn of the Internet, cyberattacks were identified with rule-based systems, where attacks could be detected based on their signatures. Over the years, as the number of Internet-connected devices and their applications increased, observing the huge amounts of network traffic being generated in real-time and creating rules which analyze this traffic have become time-consuming and make security protection systems behave defensively rather than proac- tively. Coupled with this trend, technological advances are also benefiting attackers who are developing new sophisti- cated attack strategies that can avoid detection by current security systems [4]. As the cyberthreat landscape contin- ues to rise, we need advanced tools and technologies which can help detect, investigate, and make decisions faster for emerging threats. AI has the potential to intelligently analyze and automatically classify large amounts of Internet traffic. Today, cybersecurity solutions, based on ML tech- nologies, are being used to automate the detection of attacks and to evolve and improve their capabilities over time. ML-based solutions are being used in intrusion detection systems [55]–[57] as they can handle large volumes of data and a wide range of data attributes (e.g. a large number of table columns) used for classification [54], [55]. Machine learning techniques learn from the collected Internet traf- fic to distinguish the malicious from the legitimate traffic class. It is worthwhile pointing out that due to the pervasive- ness of machine learning in addressing cybersecurity issues, the adoption of the ‘‘machine learning’’ terminology has become interchangeable with ‘‘Artificial Intelligence’’ in the cybersecurity field.

A. *MACHINE LEARNING*

Conventionally, machine learning methods can be classified into two categories: supervised and unsupervised learning. In supervised learning, data samples are labeled according to their class (e.g., malicious or legitimate). Training data, or data labeling is usually performed manually, requiring humans to detect data patterns with their classes. The trained data is input to an algorithm to create a mathematical model, which can output the predefined classes given new data sam- ples. In unsupervised learning, no data labeling or training is required. Instead, the algorithms determine the degree of coherence/dispersion among data samples, systematically creating classes, and then classifying these samples according to the quality of data coherence within the class and data modularity between the classes.

However, discussions in machine learning blur the dis- tinction between supervised/unsupervised machine learning algorithms. Mathematical, statistical, and probabilistic meth- ods are used by machine learning techniques, allowing unsu- pervised algorithms to label the data used by supervised algorithms [58]. This shows that taxonomy perspectives are converging, making it less essential to define machine learn- ing algorithms based on whether they are supervised or unsu- pervised [59]. Henceforth, we present an in-depth discussion of machine learning algorithms from a taxonomy perspective as described in [60], but in this section, we discuss the pre- dominant machine learning techniques that are effective for cybersecurity solutions.

Machine learning algorithms process data samples based on their determining factors, commonly called features. The data input is processed as a table of rows and columns, with rows serving as data samples and the columns representing

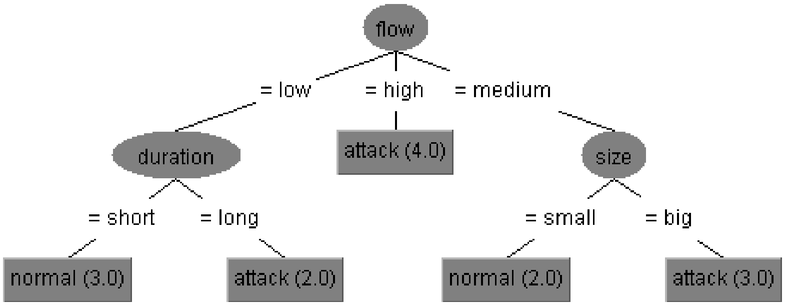


FIGURE 2. An example of a decision tree that classifies network traffic into attack and normal traffic type.

their features. Naïve Bayes is a machine learning technique used to classify data based on the Bayesian theorem [61] where the features are assumed to originate from independent events. The technique uses the computed probability of each class over all instances as the basis to find the probabil- ity of new data samples belonging to the class. Although the performance of Naïve Bayes classifiers degrades when more features come from dependent events, it is widely adopted [62]–[65], because it can inherently accept such a naïve assumption (that each feature comes from independent events) while still yielding acceptable results [66].

B. DECISION TREES

A decision tree is a technique used to create a set of rules from the training data samples. The algorithm iteratively finds a feature that best categorizes data samples. The iterative division creates a sequence of rules for every side of the categories, resulting in a tree-like structure, until data samples with only one class are found after a division. Fig. 2 shows a decision tree example that classifies network traffic using rules that lead to normal or attack traffic classifications. The tree shows that, for example, if the flow of the traffic is low, but the duration of the traffic pattern is long, then it is classified as an attack. The technique provides an intuitive method for detecting cybersecurity issues, because it shows the result of a decision according to the feature values, as what is required by classifying observed events in cybersecurity as either legitimate or an attack. For example, the flow rate, size, and duration were used by decision trees to detect DoS attacks in addition to source/destination error rates [67]. Fur- thermore, in detecting command injection attacks to robotic vehicles, decision trees were employed to categorize values from CPU consumption, network flow, and the amount of data written [68]. This technique's benefit is that once the effective series of rules has been found, intrusion detection systems can classify Internet traffic in real time. The quality of generated real-time alerts is one of the most important attributes in detecting cyberattacks.

A different approach is the Rule-Learning technique [69], which seeks to find a set of feature values for each itera- tion while maximizing a score that defines the classifica- tion result's quality—for example, the number of incorrectly classified data samples. Such an approach is similar to decision trees in that it generates a set of rules for clas- sification. While decision trees find the best feature val- ues that lead to a class, a rule-learning technique finds a set of rules that can describe a class. The advantage of a rule-learning technique is that it can factor human expert advice in generating rules. Consider a study that employed 28 features to detect DoS attacks in cloud networks [70]. The features consisted of computer and network indicators, such as Input/Output (IO) reads, memory used, TCP flags detected, and the number of system resources opened. It gen- erated a set consisting of rules derived from the features (e.g. IO_reads greater IO_reads(average)), and employed feature-ranking algorithms to discern the most relevant rules in finding the class. Afterward, the study employed human experts to optimize the rules, such as removing redundancies. Thus, the technique is suitable for intrusion detection systems where the configurations are mainly rule-based. Furthermore, the technique was generally employed as a performance benchmark to other machine learning techniques in detecting network intrusions [71], [72].

C. K-NEAREST NEIGHBORS

The k-Nearest Neighbor (k-NN) technique learns from data samples to create classes or clusters. It was first proposed as a non-parametric pattern analysis [73] to find the proportion of data samples in a neighborhood that yields a consistent esti- mate of a probability. The neighborhood was set as k-number of data samples according to a distance metric, usually the Euclidian distance to create clusters. The votes from all k neighbors decide how new data samples can be assigned to one of the clusters.

Fig. 3 illustrates the above technique. A new data sample (the red dot) was added to the data. In this example, the win- ning vote came from the highest number of data samples from one neighboring cluster. Hence, when k 3, the sample was put into Class 2. When k 9, the sample was put into Class 1. This technique is computationally complex even for small values of k. However, it is attractive for intrusion-detection systems because it can learn from new traffic patterns to reveal zero-day attacks as its unseen classes. Active research in this area thus seeks to find how k-NN can be used for

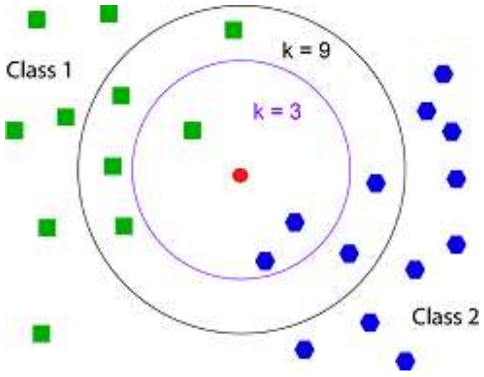


FIGURE 3. The k-Nearest Neighbor (k-NN) algorithm classifies data in class 1 and class 2, based on the k nearest data samples in the neighborhood from the new data sample.

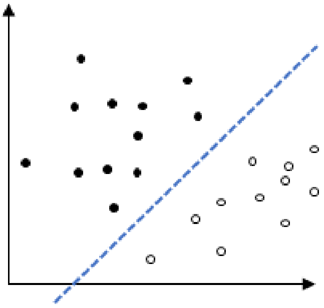


FIGURE 4. Support Vector Machines (SVMs) find a plane that separates data samples.

real-time detections of cyberattacks [74]. Recently, the tech- nique was employed to detect attacks such as data tampering and false data injection against industrial control systems [75] and smart grids [76]. It performs well when the data can be represented through a model that allows the measurement of their distance to other data–for example, in terms of a Gaussian distribution [75] or a vector [76].

D. SUPPORT VECTOR MACHINES

The Support Vector Machines (SVMs) [77] technique extends linear regression models. While classifying data samples, SVMs find a plane that separates data samples into two classes (as shown in Fig. 4).

The separating plane can be shaped to form linear, non- linear, polynomial, Gaussian, Radial, sigmoid, and so on depending on the function employed (called a kernel) [78]. SVMs can also separate multiclass data (that is, not only data to be classified into two classes such as legitimate versus attack class as what the previous examples showed, but rather data to be classified into more than two classes) by employing more than one plane. This makes SVMs an attractive tech- nique that can be used to analyze Internet traffic patterns, which often consist of several classes such as HyperText Transfer Protocol (HTTP), File Transfer Protocol (FTP), Post Office Protocol 3 (POP3), and Simple Mail Transfer Proto- col (SMTP) [79].

SVM is a supervised machine learning technique, which requires training data to create a classification model. There- fore, it is used in applications where attacks can be simu- lated [80]. For example, network traffic generated from the penetration testing conducted on a network system was used as the training data. SVM was employed to create a mathe- matical model to find a plane the penetration test traffic from normal traffic. A variation on its use creates a 1-class model for the normal traffic, while the model can be employed to detect anomalies when attack traffic was introduced [81]. From these perspectives, the benefit of SMVs enables the development of attack detection models through simulations.

Traditionally, privacy has been addressed through secure authentication mechanisms, such as encryption and security certificates. These mechanisms shift in the IoT, as devices are mobile, with data stored in the cloud. AI techniques can be used to maintain private communications when routing paths dynamically change, and when a third party stores the data. For example, learning automata was adopted to distribute secure certificates to moving vehicles [144], and artificial immune system algorithms were adopted to securely self-organize Wireless Sensor Network (WSN) ad hoc con- nections to serve mobile gadgets [145]. In WSNs, different IoT devices such as mobile gadgets dynamically join and leave the network. This causes traditional security measures such as port security (i.e., restricting traffic only to a known Media Access Control (MAC) address) inapplicable. Thus, in [145], the authors proposed features such as packet receiv- ing rate, packet mismatch rate, and energy consumption per packet received from a device to describe a device's behav- ior. They used artificial immune system algorithms to clas- sify a device's behavior as normal/abnormal. Upon detecting

IV. FUTURE CHALLENGES AND RESEARCH OPPORTUNITIES

A. THE RACE BETWEEN DEFENSE, OFFENSE, AND HUMANITY

Recent AI research advances in cybersecurity have fueled the race between the white hat (defenders) and black hat (offend- ers) hackers. Attackers can employ AI to mimic human behavior to achieve personal pride, power, or financial advan- tage. AI has led to the creation of intelligent agents that automatically click advertisements, play online games, and buy and resell best-seller seats for concerts [172]. AI has also manipulated public opinion in Venezuela by retweeting political content [173] and has affected the US presidential election by spreading tailored news [107]. Future research opportunities in cybersecurity are determined by how divid- ing lines can be drawn between developments and basic needs.

AI's use in cybersecurity impacts three major stakeholders: white hat hackers, black hat hackers, and end users (human- ity). The white hat and black hat hackers are the cohorts who promote the development of AI techniques. However, it is difficult to find the dividing line between the two groups to regulate technological deployment, because one's advance- ment follows the other's advances. Hence, it is imperative to investigate how AI can be employed for human basic needs and for developing cybersecurity controls.

B. INFRASTRUCTURE

The use of AI in cybersecurity is viewed as a race between law enforcement and cyberattackers. The leader in the race will be determined by his/her access to technical knowledge and the supporting computing infrastructure. AI algorithms are computationally expensive, because they are evolutionary by nature. Therefore, developing fast algorithms for the AI solutions shown in Table 2 should be an active research area. For example, to detect malware, hashing algorithms have been developed to input to the k-means clustering algorithms, to enable fast clustering of common data samples [174]. Developing relevant algorithms has become part of the recent race, but hardware development is another crucial part.

C. HARDWARE AND PLATFORM

Having access to state-of-the-art computing infrastructure will help solve AI problems efficiently and with efficacy. As the number of computing devices increases, the volume of traffic will also increase, thereby making it necessary to perform data analysis quickly. Consequently, analyzing data by using AI techniques requires high-end computing plat- forms. To address this challenge, cluster computing solutions such as Apache Spark and Hadoop have been employed to analyze cyber traffic [175], [176]. At the high end, quantum computing will be the breakthrough technology that helps solve complex computing problems. NASA's quantum computer [177] has been able to solve complex problems in a fraction of time–it is 100 million times faster [178] than traditional computers.

D. RESOURCES

Having easy access to the required resources when needed is crucial in implementing workable computing solutions. Currently, energy is seen as the scarce resource for many computing needs. For instance, Bitcoin blockchain consumes an equivalent energy of 29 average Australian households for a full day, only to commit one block [179].

When intelligent computers start to consume a signif- icantly larger chunk of resources which are shared with human beings, ethical issues regarding the use of AI will arise. One issue would be if intelligent machines have their own rights. In one way, the issue may seem irrel- evant because computers are viewed as having no con- sciousness [180]. In another way, researchers have started to debate whether intelligent computers should have rights regardless of the definition of consciousness [181]. The adop- tion of AI in cybersecurity extends the arguments on how to share scarce resources between intelligent computers and human. This will in turn motivate regulators to go back to the drawing board to justify what serves as development and basic needs. Ethical issues will also remain a future challenge when it comes to how AI can be employed for cybersecurity.

V.CONCLUSION

As the speed and sophistication of attacks increase, AI has become an indispensable technology in the cybersecurity area. This article showed how cyberthreats have increased, evolved in their complexities, and broadened their scope. We underscored how past cyberthreats remain relevant to future risks. We presented a comprehensive review of cyberthreats and solutions. In particular, we described how cyberattacks can be launched on different network stacks and applications, along with their impact. Cyberthreats will con- tinue to rise, even as the community identifies cyberthreats and develops solutions using a wide range of technologies and techniques.

In contemporary research, AI techniques have demon- strated their promise in combating future cybersecurity threats. The techniques propose a range of intelligent behaviors—from how machines can think to act humanly. Recently proposed AI-based cybersecurity solutions largely focused on machine learning techniques that involve the use of intelligent agents to distinguish between attack traffic and legitimate traffic. In this case, intelligent agents act as humans whose task is to find the most efficient classifica- tion rules. However, the cyberattack landscape today morphs from disrupting computers to sowing disorder in society and disturbing human wellbeing. We discussed this phenomenon in terms of how advances in technologies are transforming the ways cyberattacks can be launched, detected, and miti- gated. Through such advances, AI's role in cybersecurity will increase continuously. Novel AI techniques must be devel- oped to quickly detect and mitigate threats that impend upon societal and human wellbeing. In all likelihood, cybersecurity solutions will expand from intelligent agents acting humanly to thinking humanly.

Although AI's role in solving cybersecurity issues con- tinues to be investigated, some fundamental concerns exist surrounding where AI deployment can become regulated. For instance, as intelligent machines become more integral solu- tions for humanity, these machines increasingly will consume fundamental resources for life. When humans and machines compete for scarce resources, a new form of governance will promulgate. This in turn will engender a new research avenue.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions which helped us improve the content, quality, and presentation of this paper.

REFERENCES

[1] D. Venable. 2017. *Cybersecurity in 2017: When Moore's Law Attacks*. Accessed: Jun. 5, 2019. [Online]. Available: https://www.channelpartners online.com/blog/cybersecurity-in-2017-when-moore-s-law-attacks/

[2] S. Morgan. (Jun. 2019). Global Cybersecurity Spending Predicted to Exceed $1 Trillion From 2017–2021. Cybercrime Magazine. Accessed: Dec. 22, 2019. [Online]. Available: https://cybersecurityventures.com/ cybersecurity-market-report/

[3] Statista Research Department. (Aug. 2019). *Spending on Cybersecurity in the United States From 2010 to 2018*. Accessed: Dec. 22, 2019. [Online]. Available: https://www.statista.com/statistics/615450/cybersecurity-spending-in-the-us/

[4] Wall Street. (Aug. 2018). *How Artificial Intelligence and Machine Learning Will Impact Cyber Security*. Accessed: Jan. 5, 2020. [Online]. Available: https://wall-street.com/how-artificial-intelligence- and-machine-learning-will-impact-cyber-security/

[5] D. Yuhas. (Oct. 2017). Doctors Have Trouble Diagnosing Alzheimer's. AI Doesn't. NBC News. Accessed: Dec. 25, 2019. [Online]. Available: https://www.nbcnews.com/mach/science/doctors-have-trouble-diagnosing-alzheimer-s-ai-doesn-t-ncna815561

[6] M. McFarland. (Dec. 2017). Farmers Spot Diseased Crops Faster With Artificial Intelligence. CNN Business. Accessed: Dec. 25, 2019. [Online]. Available: https://money.cnn.com/2017/12/14/technology/corn-soybean-ai-farming/index.html

[7] C. Geib. (Jan. 2018). Nasa-Funded Research Will Let Unmanned Spacecraft Think' Using AI and Blockchain. Futurism. Accessed: Dec. 20, 2019. [Online]. Available: https://futurism.com/nasa-funds- autonomous-unmanned-

spacecraft

[8] E. Winick. (Dec. 2017). Lawyer-Bots are Shaking up Jobs. MIT Technol- ogy Review. Accessed: Dec. 25, 2019. [Online]. Available: https://www. technologyreview.com/s/609556/lawyer-bots-are-shaking-up-jobs/

[9] B. Morey. (Jun. 2019). Manufacturing and AI: Promises and Pitfalls. SME. Accessed: Dec. 25, 2019. [Online]. Available: https://www.sme. org/technologies/articles/2019/june/manufacturing-and-ai-promises- and-pitfalls/

[10] S. Morrow and T. Crabtree. (Aug. 2019). The Future of Cybercrime & Security. Juniper Research. Accessed: Dec. 25, 2019. [Online]. Available: https://www.juniperresearch.com/researchstore/ innovation-disruption/cybercrime-security?utm_source=juniperpr&utm_ campaign=pr1_thefutureofcybercrime_technology_aug19

[11] H. Taylor. (Sep. 2018). *What are Cyber Threats: How They Affect you and What to do About Them*. Accessed: Jun. 5, 2019. [Online]. Available: https://preyproject.com/blog/en/what-are-cyber-threats-how-they-affect-you-what-to-do-about-them/

[12] M. Cohen. *Zero-Day Attacks are Difficult but not Impossible to Defend Against*. Accessed: Jun. 5, 2019. [Online]. Available: https://eccit solutions.com/zero-day-attacks-difficult-not-impossible-defend/

[13] German AV-TEST GmbH Research Institute. *Malware*. Accessed: Dec. 22, 2019. [Online]. Available: https://www.av-test.org/en/statistics/ malware/

[14] W. Hall and J. Pesenti. (Oct. 2017). *Growing the Artificial Intelligence Industry in the UK*. Accessed: Dec. 30, 2019. [Online]. Available: http://ftp.shujuju.cn/platform/file/2017-10- 18/782c432045784854a04e458976aef0bf.pdf

[15] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, and L. Floridi, ‘‘Artificial intelligence and the ‘good society’: The US, EU, and UK approach,’’ *Sci. Eng. Ethics*, vol. 24, no. 2, pp. 505–528, 2018.

[16] E. Brynjolfsson, D. Rock, and C. Syverson, ‘‘Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics,’’ Nat. Bureau Econ. Res., Cambridge, MA, USA, Tech. Rep. w24001, 2017.

[17] E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, ‘‘The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,’’ *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, Feb. 2011.

[18] J. Borenstein. (Dec. 2018). *The Challenges of Adopting a Consistent Cybersecurity Framework in the Insurance Industry*. Accessed: Jun. 5, 2019. [Online]. Available: https://www.microsoft.com/security/ blog/2018/12/20/the-challenges-of-adopting-a-consistent-cybersecurity-framework-in-the-insurance-industry/

[19] C. Alcaraz and S. Zeadally, ‘‘Critical infrastructure protection: Require- ments and challenges for the 21st century,’’ *Int. J. Crit. Infrastruct. Protection*, vol. 8, pp. 53–66, Jan. 2015.