Application of Deep Machine Learning in Cybersecurity

Sarana Tse[1], Niharika Kakumani[2], Savannah Muniz[3]
Vahid Emamian[2], Senior IEEE Member
[1]Electrical Engineering Department
[2]Computer Science Department
[3]Computer Engineering Department
St. Mary's University, San Antonio, TX
smuniz5@mail.stmarytx.edu

ABSTRACT

The evolution of technology has brought in many changes across the globe. Technology has no limit in expanding its scope; there are smart gadgets like Alexa, Google Home and Apple Pod that act like a personal assistant; there are smart homes, with the control of home appliances from faraway places; so many more examples come to mind with the help of the Internet. The world is constantly looking for 'the next best thing' in terms of science and technology. As the world is becoming more digitalized and with increasing availability to technology, raising security challenges introduce an immediate need for robust techniques to combat various complex-cyber security-attacks. The attackers are improving their skills and coming up with new techniques to break through security infrastructures. The traditional cybersecurity tools are unable to defend, detect, and simply just keep up with the onslaught of today's attacks. Thus, cybersecurity is becoming an overwhelmingly complex and sophisticated field. However, there is a silver-lining in the implementation of cybersecurity with the help of deep machine learning to improve the attacks detection rates and response time to the attacks. This paper focuses on adoption of deep machine learning in cyber security. The first section gives an overview about various cyber threats, deep learning as a subset of machine learning, and artificial neural networks. The second section discusses various security goals to be achieved by the machine learning algorithms, followed by applications of machine learning in cybersecurity. The fourth section discusses three use cases to demonstrate how machine learning is applied in real-time scenarios to enhance the security. The final section summarizes the research paper.

Keywords - machine learning, cybersecurity, artificial neural networks,

1. INTRODUCTION

Cybercrime is one of the world's fastest growing threats to security. This trend unfortunately corresponds to the growing dependence on computer networks and information technology. Cybersecurity crimes and a greater reliance on these systems targeted correspond to an estimated $6 trillion global cost by 2021 under the 2020 Official Annual Cybercrime Report by Cybersecurity Ventures [1]. No longer do we live in the age where Artificial Intelligence is restricted to the film industry. In fact, we began to explore these uncharted waters in the mid-1950s with the introduction of the Logic Theorist [2], a problem-solving program presented at a Dartmouth conference. The fact that a computer, a machine that worked in binary operations, could mimic human thinking, fast-tracked the development and evolution of AI to include the branch of deep learning, or DL, around 2000. DL is characterized by the multiple layers of artificial neural networks, a model inspired by the neurons of the brain. The term *deep* refers to the number of layers involved, as compared to *shallow* learning. In today's world, applications of DL are not far-fetched. In fact, DL can be found in virtual assistants (i.e., Siri or Alexa), autonomous vehicles, facial/image recognition, and even personal Netflix and shopping recommendations.
And so, it is no secret that humans have entered a technology revolution that includes smart devices that populate homes and businesses, smarter communication systems, and significant advancements in AI. However, the growing trend of our lives becoming more technology-centric expose us to more cybersecurity vulnerabilities.

The majority of these vulnerabilities are associated with the Internet. In early days, when applications were developed separately on a user's system, there was a limited scope for vulnerabilities. The introduction of a common platform to share web-applications, the Internet, saw the quantity of attacks launched begin to rise. Morris worm was the first attack to be launched on the Internet in 1988, in which many systems fell victim. Several SQL injection attacks were launched to get into the company's database. Hence, there was and is an immediate necessity to make defense systems strong in an effort to minimize the impact of any attack. Cybersecurity plays a crucial role in securing not only the company's infrastructure but also the security of smart systems at homes, education system databases, and so much more. The basic security infrastructure of any business or system consists of a network and cybersecurity system. These systems, in turn, collect large amounts of data and analytics for future reference. It becomes clear that to study any kind of present attacks, data from previous attacks is needed, consequently forming huge datasets which are difficult to analyze manually. Machine learning systems are a subdomain of Artificial Intelligence, with its self-learning ability, it can manipulate

large sets of data in various fields. With the help of machine learning, cybersecurity can make our lives more secure.

This research paper describes a literature review of DL applications with the ever-growing cybersecurity threats. The overview of cyber threats and deep learning as a subset of machine learning (ML) will be covered in the first section. Following will be a brief introduction of cybersecurity goals, where the utmost purpose is protection of the computer system, network, or data involved. For the purposes of this paper, the Cybersecurity CIA Triad is covered; it consists of the cybersecurity goals of confidentiality (along with privacy), integrity, and availability. The next section defines applications of DL concerning cybersecurity. These applications include Intrusion Detection Systems and Prevention Systems (IDSs/IPSs), malware detection, spam/social engineering detection, network traffic analysis, and user behavior analysis. For each DL application listed under this section, references to important work in the area are provided. Examples of these DL applications are further studied and explained in the following section, with an emphasis on use cases concerning the Internet of Things (IoT), IDSs, and Android malware detection. The final section concludes the research carried out for the paper.

2. Overview of Cyber Threats and Machine Learning

Our society with its many infrastructures is mainly dependent on computer networks. The more our society depends on computer networks, the more we are exposed to cybercrimes. The Symantec cybercrime has released a report in April 2012 [3], stating that cyber-attacks cost $114 billion every year. According to their reports, every second at least 14 adults are exposed to cyber threats. Also, every individual faces at least one cyber-attack in their lifetime. Data breaches in organizations are frequent. According to 2020 cybersecurity statistics [4], and shown in Figure 1, companies spend $3.92 on average for data breaches. Hence, cybersecurity is required to protect our computer software, hardware, infrastructure, and network from malicious users. With the advent of new technologies there are many security risks arising. The most trending cyber threats, and the most dangerous cyber threats in recent times, are *Ransomware*, *Advanced Persistent Threat*, *Insider Threats*, *Malware*, *Botnets* and *Cyber Espionage*.

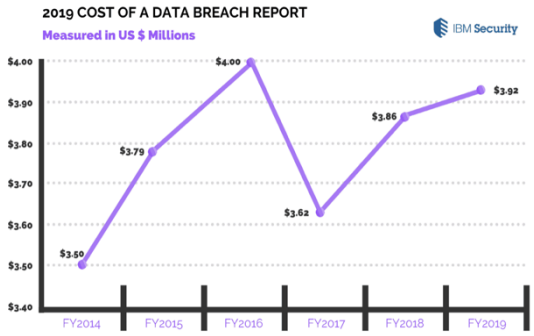

Figure 1: Average total cost of data breach spent by organizations globally

Ransomware is a malicious software that was identified in 1989 [5] and the attackers use it to gain unauthorized access over a company's database or systems or simply encrypt the entire sensitive information in the company. The attacker usually demands a huge ransom to give a decryption key or to return access to the users. A *malware attack* is of potential threat to the Internet users. Malware is a software designed by the attackers to infect the computer or network systems without the user's consent. Popular malwares are Trojan Horse, Rootkits, keyloggers and Spywares. *Botnets* also come under this category, but they launch denial of service attacks and send spams to the user's systems. *Cyber espionage* is a way of stealing confidential information from the users without their consent with the help of proxy servers or cracking techniques. Similarly, *Advanced Persistent* and *Insider threats* are used to access a company's server without their presence being known. Whatever the method used to launch an attack, the sole motive is to gain unauthorized access over a company's database or servers. This not only affects various companies, but also targets smart homes and smart cars. A car can be hacked with the help of simple Bluetooth. All these increasing attacks are making cyber security increasingly complex and sophisticated with traditional defense techniques and tools becoming outdated. With the outdated techniques and tools, it is taking nearly 240 days to detect an intrusion in the system. The outdated techniques are increasing the challenge of finding a tool or method that reacts quickly to cyber threats. One of the promising and rapidly growing fields that can help cyber security is machine learning.

Machine Learning is the study of computing algorithms which will continuously improve by the experience of the model. Usually, machine learning algorithms will build a mathematical model, based on the given sample data (training data) in order to make accurate predictions. Nowadays, machine learning is widely used in various industries – automobile, computer vision, and email filtering. The term machine learning was coined by Arthur Samuel [6]. Later, in the mid 90s, many people developed it and Tom M. Michael gave credence to the term. It is mainly related to the computation statistics which focus on predictions through the use of computers. Machine learning is often referred to as predictive analysis, where we can predict the results by giving certain output. Machine

learning approaches are mainly divided into three types which are supervised learning, unsupervised learning, and reinforcement learning. In *supervised learning*, the computer is given sample inputs and desired outputs. The mapping of inputs to outputs takes place and gives desired results. In *unsupervised learning*, the computer is given the input without labels. The algorithm then discovers hidden patterns in the data and generates output. In *reinforcement learning*, the computer program analyzes an adynamic environment in which it will analyze and perform certain goals.

To understand the meaning of deep learning, as a subset of ML, we turn to Geoffrey Everest Hinton, known to some as the "Godfather of Deep Learning" and the Conference on Neural Information Processing Systems (NIPS) of December 2012. Considered a revolutionary event in the history of DL, it was at the NIPS conference that Hinton and two graduate students displayed an ImageNet (dataset of 15 million labeled images of 22000 categories) classification error reduction of 20 percent through use of an eight-layered convolutional neural network (CNN) [7]. This was an application of DL. Prior ML algorithms could not boast the same and thus, the popularity and intrigue of DL flourished. It was concluded that improved results would stem from larger networks (constrained by GPU performance) and larger datasets.

In defining DL, there are two takeaways from that singular event – scalability and feature training. DL works by modeling the dataset in a neural network, or layered approach. The term "deep" refers to the number of layers involved. As the layers increase and the model grows, the performance improves; hence DL models are scalable. In fact, Hinton et al. began defining "deep" in this way in [8] and [9], in which he introduced a multilayer neural network of Boltzmann machines, or a deep Boltzmann Machine (DBM). Each layer consisted of a Boltzmann machine, where its output was fed into the next layer's input, learning hidden features of the layers below. The proposed model boasted of feedback and increasingly complex learning.

This layered approach is a consequence of modeling inspired by the brain. In technical terms, the algorithms used are based on Artificial Neural Networks. Artificial Neural Networks (ANN) are a type of artificial intelligence technology based on the innerworkings of neurons inside the brain. The ANN systems are fed data, and the system can adapt to changes depending on the information given. Algorithms are implemented and created based on their data experiences and seek to determine what may happen next, just as humans learn from their own experiences. ANNs consist of systems of nodes, similar to neuron connections in the brain. The nodes send data to each other depending on the algorithm. All the parts between input and output nodes are known as the hidden layers. "Each layer passes its output to the next layer and the last layer outputs the result" [10]. Though the structure of ANNs is pretty complex, here's a simple example on how it

can work. In determining what materials are going through the machine, the images of the materials will go through the first layer, where the algorithm of the ANN will examine a certain characteristic of the image, known as a feature. It will then proceed to go through several more layers, such as analyzing the textures, shapes, and so on. Finally, at the very end, the system concludes what materials have gone through [11]. What makes ANNs more popular now are their ability to generate nonlinear models, which allows for the model of EX-OR logic [10]. A disadvantage is that the more data the machine needs to go through or if the task becomes very complex, the amount of power and time needed increases as well [11].

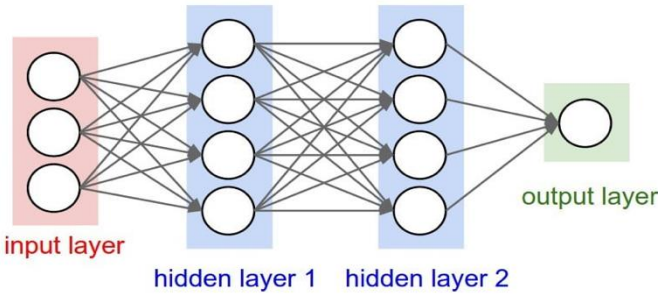

Figure 2: Artificial Neural Network Model Retrieved from https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network, 2019

Returning to the ImageNet demonstration, the second takeaway is that DL algorithms use feature learning to solve the AI representation problem of feature selection when defining. Feature training is the *automatic* extraction of these defining features. The result is to enable the computer to build complex concepts out of simpler concepts [12]. Another leader in DL and one of the masterminds behind Google Brain, Andrew Ng, likens DL algorithms to brain simulations that strive to improve the performance and ease of learning algorithms and make revolutionary advancements in ML and AI [13].

3. SECURITY GOALS

The advancements in deep learning have reiterated the importance of security now more than ever. To illustrate, InfoSec (information security) professionals use the CIA Triad (see Figure 3). Interestingly enough, the triad has nothing to do with the Central Intelligence Agency; instead, it models the interconnectedness of the cybersecurity principles of confidentiality, integrity, and privacy. The purpose of the triangle configuration suggests that all three components are needed to define a system as secure. They are all equally important.

Figure 3: CIA Triad. From *Spanning, A Kaseya Company*.

3.1. Confidentiality

Machine learning systems are used to identify security threats and breaches in other systems. Before that can happen, the security of the machine learning system has to be trusted. Therefore, the information within the ML system has to be confidential so that it can be trusted. According to CIA triad, confidentiality means protection of data or resources from unauthorized access. Data confidentiality is a big task to be achieved in machine learning systems as there is sensitive data within the training models of the system. There are certain extraction methods to extract data from the machine learning systems and this affects the health care system the most. The hackers usually perform two types of extractions attacks and they are *Model Extraction Attack* and *Model Inversion Attack*.

Model extraction attacks usually extract the model parameters by sending queries to the model. The attacker makes use of a prediction application programming interface to send queries and the model, in turn, returns extra information than required. This attack can be prevented by following countermeasures such as eliminating incomplete queries and not returning complete information. The second type of attack is the Model Inversion Attack, in which attacks are carried out by finding the inputs that provide sensitive information from the training datasets when given as input to the models. This kind of attack can be minimized by adding noise to the returned output, but it is not apt when the request made is legitimate. Also, the sensitive features are to be identified and certain restrictions must be placed to get access.

3.2. Privacy

In the age we are living in where the amount of information being shared is extensive, privacy threats are real and prevalent. Norton, a leader in cybersecurity, defines *privacy* as the rights to information, regarding its access and usage [14]. Throw

DL into the mix and the privacy definition is expanded to include protection of the model itself and its training data. To reinforce the significance of privacy in DL applications, [15] presents the scenario of a medical application, with patient health data, known to be protected under the Health Insurance Portability and Accountability Act (HIPAA). According to [16], a threat to privacy occurs when an adversary gains access to the system that hosts the model and the data, opening the potential for recovery and extraction of critical information.

3.3. Integrity

Maintaining integrity in ML is very crucial. Without it, attacks can easily change the outputs of what we want and cause harm. An example of compromised integrity in ML could be "attacks that attempt to induce false positives in a face recognition system" allowing anyone to have access to the system [17]. Thus, we can see how essential it is to have good integrity in ML. When people hear about a compromise in the integrity of ML, most will think it's from an adversary attack. However, the integrity of the ML system can be compromised from the very beginning in its learning phase. If the ML system is fed bad samples or data, it will give outputs we do not want and develop harmful algorithms. For example, the attacker can poison the data set so that during the learning phase, the ML system would learn to not tag the anomaly. In learning so, the attacks could be expected to go unflagged, or unnoticed, in the future.

3.4. Availability

Availability is the ability of authorized users to access the system, network, and data at will. An interruption of this freedom is an availability attack. The most common availability attack is a denial-of-service, or DoS attack. Such attacks can happen through the flooding of a network server or individual host with traffic until the attacked party is unable to access said system. In DL systems, availability attacks can be understood as resulting in false positives, or misclassifications. Regarding malicious activity detection, a false positive can mean misclassification of benign instances as malicious [18]. As a result, these errors can lead to doubt in the DL system’s efficacy.

4. APPLICATIONS OF DL IN CYBERSECURITY

Threats to cybersecurity can typically be characterized into two categories: causative and exploratory attacks. Causative attacks involve the attacker having control over the training data. These types of attacks can consist of input label manipulation through adversarial data injection or modification. On the other hand, exploratory attacks attempt to learn about the learning algorithms; they do not affect the training data.
Deep learning applications in response to these exploratory and causative attacks are explored in the following subsections.

4.1. Intrusion Detection and Prevention Systems

Intrusion detection and firewalls have been an integral part of any system's network infrastructure. As the attackers adopt new techniques and tools, it becomes increasingly difficult for the firewall to detect novel attacks. Intrusion Detection Systems, or IDSs, serve the purpose of continuous monitoring of networking traffic and generate alerts if any suspicious activity is found. The techniques used by intrusion detection systems are unable to handle dynamic and complexity of attacks on networks. Moreover, most of the systems using these techniques show high false positive and false negative detection rates and they lag to adapt to the changes in malicious behavior. Hence, deep machine learning techniques are proposed to improve the detection rates of the intrusion detection systems. Researchers have proposed two machine learning categories of approaches for intrusion detection systems. They are approaches based on *Artificial Intelligence* and *Computational Intelligence* techniques [19]. The artificial intelligence techniques include decision trees, k-nearest neighbor, multi-layer perceptron and support vector machines. Computational intelligence techniques include fuzzy logic, artificial neural networks and artificial immune systems. However, these two approaches share common features such as fault tolerance, high computation speed and error resilience.

Laskov et al proposed an artificial intelligence method to perform comparative analysis between supervised and unsupervised learning [20]. Here, supervised refers to classification techniques, while unsupervised refers to clustering techniques. They have framed two scenarios to study the identification ability of the intrusion detection system. First scenario consisted of training and test datasets from the familiar unknown pattern and the second scenario consisted of test data coming unseen attack patterns. These two scenarios helped them to study how the intrusion detection system reacts to new attack patterns. The results showed that supervised algorithms mainly decision tree algorithms have given 95% accurate results [20] compared to unsupervised algorithms for the first scenario. However, unsupervised exhibited better results in scenario two, i.e., were new attack patterns were considered. They showed that among both unsupervised algorithms provided more robust results. Figure 4 shows the average detection rates of supervised and unsupervised methods in both the scenarios.

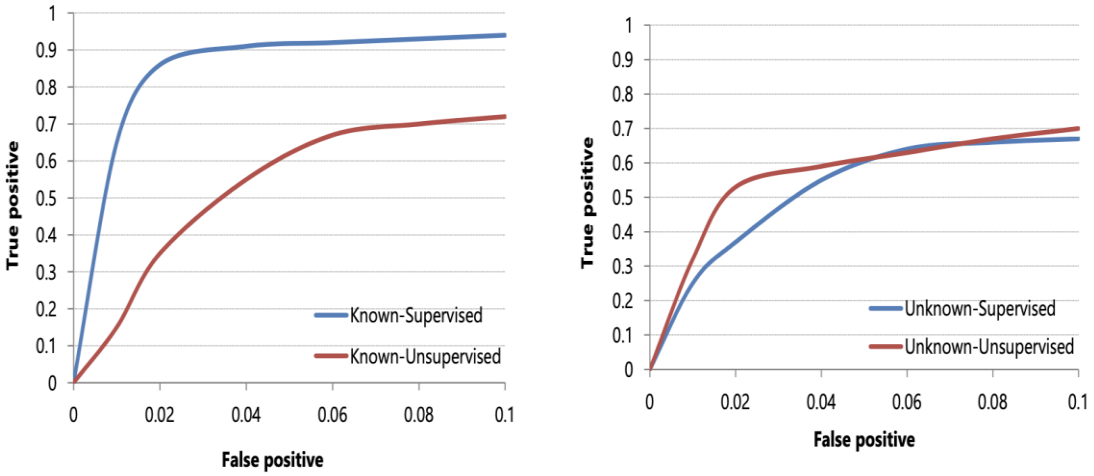

Figure 4: Detection rates of methods evaluated in the two scenarios.

Unlike artificial intelligence, computer intelligence approaches deal with complex problems which are not solved by traditional methods. Computer intelligence-based approaches are classified into four main techniques: genetic algorithms, artificial neural networks, fuzzy logic, and artificial immune systems. These techniques are resilient and have fast anomaly detection capacities. Hence, these techniques are explored by researchers in the field of intrusion detection.

Genetic algorithms are designed to give optimal solutions. Sinclair et al. has used this algorithm in combination with decision trees to design an intrusion detection system that automatically differentiates malicious traffic from normal traffic [21]. They have developed network traffic rules using genetic algorithms and each rule was represented by a genome. The system develops rules and decisions-based support for the given training data set. For instance, if a rule matches malicious behavior it is taken into consideration and if the rule matches normal behavior then it is disregarded. In this manner an algorithm is developed that is inclined towards the intrusive nature. Hence this rule can help the intrusion detection system to find the abnormal behavior. Similarly, other techniques have their own specialty. Artificial neural Networks are used to provide higher detection rates. Mukkamala et al has used a support vector machine to develop classifiers that could identify intrusions based on user behavior and they have achieved up to 99.25% accuracy [22]. Fuzzy logic techniques provide reduction in false alarm rates while detecting intrusive behavior. The nature of these machine learning techniques will be of great help to design and improve intrusion detection rates and practicality and all these algorithms are to be practically implemented for further improvisations.

4.2. Malware Detection

McAfee, another leader in cybersecurity, defines *malware* as "any type of malicious software designed to harm or exploit any programmable device, service, or network." The goal is the extraction of critical data. From there, the critical data can be used as a means of leverage. According to the 2019 Accenture Cybercrime Report [23], malware tops the list of the most expensive cybercrimes, growing 11% more expensive from 2018 to 2019 (see Figure 5).

DL handles malware by learning to distinguish suspicious activity from the benign in a system. Traditional ML malware detection algorithms distinguish the features of malware code through static and dynamic analysis. Static analysis captures data without running the malicious code. Dynamic analysis captures the data while executing the malware in a controlled environment. ML methods rely on feature learning and representation that once known, risk feature evasion by adversaries. DL outperforms ML methods, because of its ability to learn new malware patterns from existing. DL methods are not limited to characterizing malware through only a handful of patterns, instead adding upon those patterns [24].

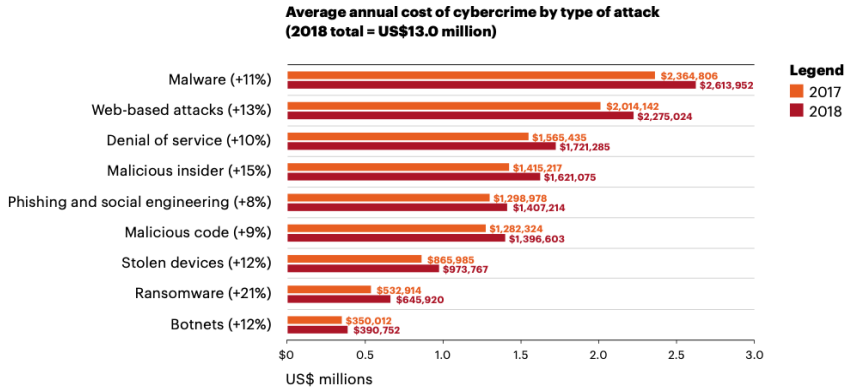

Figure 5: Average annual cost of cybercrime by type of attack, 2018-2019.
From *Accenture Ninth Annual Cost of Cybercrime Study*, 2019.

Li et al. proposes a deep learning framework based on a single autoencoder and a Deep Belief Network, or DBN [25]. An autoencoder is an unsupervised layered approach to learning through an encoder and decoder [12]. The single autoencoder was used for data dimensionality reduction. The goal of reducing data dimensionality is to uncover the main feature(s) in the data. DBNs are composed of multiple simpler unsupervised networks. In this case, restricted Boltzmann machines (RBMs) were used. RBMs are generative stochastic artificial neural networks that boast probability distribution learning. The model, with varying iterations, was tested on the KDDCUP'99 dataset and compared to the malware detection accuracy of a single DBN model (91.4%). The combined autoencoder

and DBN model with 10 pre-trained iterations and 10 fine-tuning iterations achieved an accuracy of 92.10%. Said model did not achieve a significant difference in CPU time, suggesting the need for further improvements.

In a similar approach that achieved greater accuracy, Hardy et al. proposes a deep learning framework for malware detection (DL4MD) based on *stacked* AutoEncoders, or SAEs [26]. SAEs make use of multiple autoencoders to train, with each autoencoder output leading to the input of the next [27]. The DL4MD model is based on Windows API calls and was trained and tested on a Comodo Cloud Security Center dataset. The DL4MD model achieved an accuracy of 95.6% in testing, outperforming Artificial Neural Networks, Support Vector Machines, Naïve Bayes, and Decision Tree malware detection techniques. Hardy suggests that the model would perform well in industry applications (a definite area of need), due to the model's detection efficiency, about 0.1 second per unknown sample.

4.3. Spam and Social Engineering Detection

Having a robust ML system can help "mitigate the integrity attacks," in particular the spam attacks [17]. A lot of systems have spam filters, but there are ways that an adversary can get around that. One way is to avoid using words that the ML system learned to flag during the training period. In doing so, the spam email can avoid being caught. Another way would be for the adversary to interfere in the ML training, so that the spam email gets passed.

Twitter, a social media platform, has had a long history of a spam problem. Wu et al. [28] describes how the ability to share messages on Twitter so quickly causes an increase of spam posts. Many which include "suspicious URLs to redirect users to phishing or malicious websites" [28]. Originally, to detect spam, the characteristics of the account were monitored, such as number of followers and age of account, if the account posted shortened URLs, and how much they posted. However, that method only had an 85% accuracy. Blacklisting was then implemented, but it was seen that the majority of uses would have clicked and visited the malicious web page before the webpage could be blacklisted.

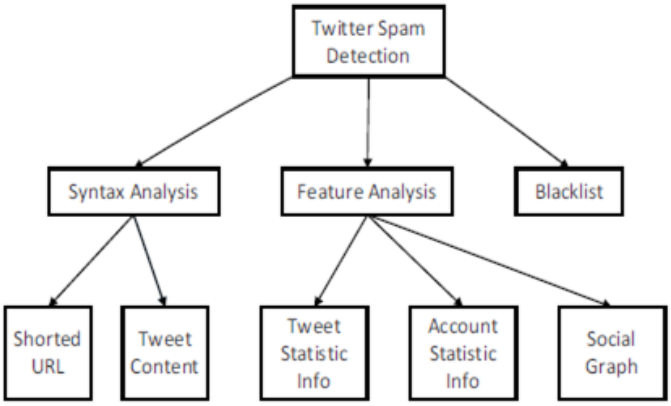

Figure 6: Twitter Spam Detection Model. From *Twitter Spam Detection Based on Deep Learning*, 2017

Wu et al. [28] then describes a new spam detection technique which can be implemented in Twitter. This new method utilizes multiple machine learning algorithms unlike the method described previously. The vector-based characteristic training and binary classifier, taught in deep learning, allows for better efficiency when processing the data. Word2Vec and Doc2Vec are applied so that each word or tweet can be mapped out in a dataset. In doing so, the system can learn multiple features or characteristics of a tweet. To test out their technique, nearly 2 million tweets were processed through and the system had to decide if the tweet was spam or not. The technique did well with around 95% precision and 92% to 99% accuracy [28].

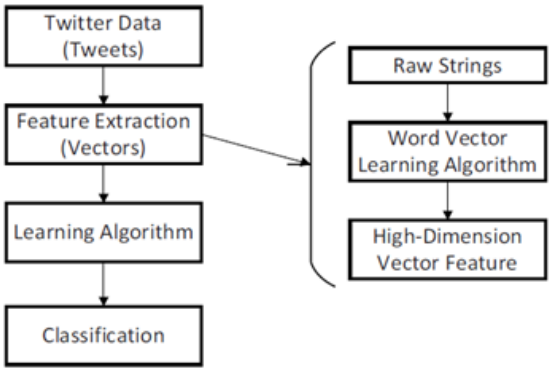

Figure 7: New Twitter Spam Detection Model. From *Twitter Spam Detection Based on Deep Learning*, 2017

4.4. Network Traffic Analysis

Network traffic analysis is one way of characterizing a user based on their "traffic activity patterns" [29]. Where they go, how they get there, and how long they stay there can be used to profile a person. User profiling is significant, because each device that is used to connect to a network has its own IP address. If a system notices that a certain device is coming in from an unusual area or has a certain traffic pattern similar to another attack, that user can be flagged and monitored to ensure that a malicious attack doesn't occur.

With the recent developments network traffic has increased and is becoming difficult to detect and analyze malware behavior, classify internet traffic, determine any intrusion, identify protocols and applications, network control and so on. In the network analysis it is very important to identify the types of protocols and applications in the network flow and the crucial part is segregating the traffic for analyzing the network flow and taking appropriate actions. In the context of accurately analyzing traffic and improving its quality of service many researchers have proposed methods using machine learning.

Soysal et al. proposed a technique to identify protocols and applications in the network flow with the help of Bayesian Networks, Decision Trees and Multilayer Perceptron techniques [30]. These techniques were tested on the National Academic Network of Turkey datasets. For efficiency in the results the data was extracted from the servers. With the help of these techniques' protocols, source and destination ports, packet size and applications were found. Among the three techniques Decision Tree method was proven to be cost effective and efficient. This technique provides good results based on the previous data extracted from the servers under an assumption that traffic comes from a known class of attacks. However, the majority of the network traffic flow consists of zero-day applications which are not taken care of. Zero-day traffic makes up to 60% of the network flows from the network traffic dataset [31].

Zang et al has proposed a framework known as Robust Traffic Classification to deal with zero-day network flow classification [32]. They proposed a three module-based framework and the three modules: 1) unknown discovery, 2) bag of flows-based traffic classification, and 3) system update. The purpose of the unknown discovery module is to spontaneously identify the zero-day traffic packets from the unlabeled traffic of the target network. The K-means clustering was adopted by them to find training datasets from the unknown application traffic. Random Forest method was used to label the known and unknown classes. They have used the bag of flows approach to develop a classifier for the robust traffic classification. This module develops a classifier using prelabeled training data and zero-day traffic data as an input. Finally, the system update module will analyze

the zero-day traffic flow and the results were accurate when tested.

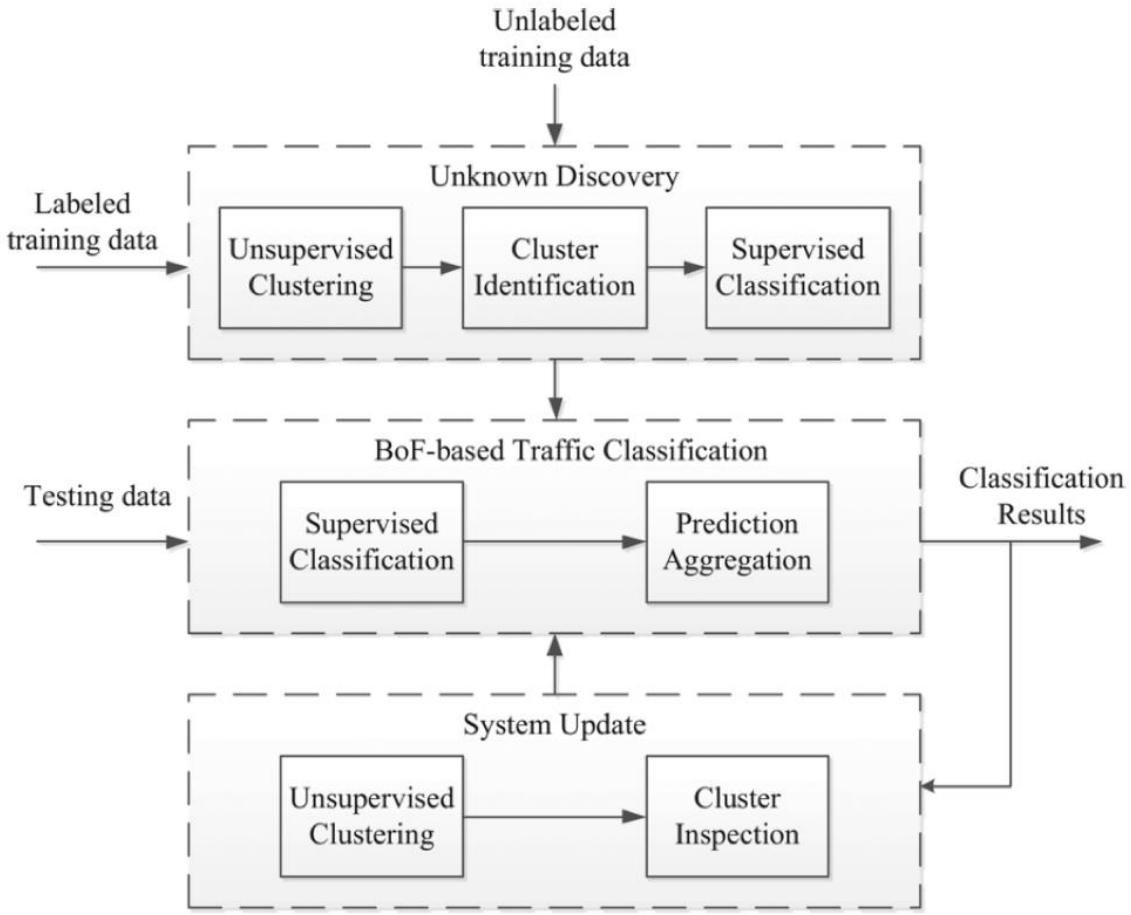

Figure 8: Robust Network Traffic Classification model

4.5. User Behavior Analysis

User behavior analysis is one way of characterizing a user by determining the time the user stays on something and how much time there is in between two activities [33]. Kind of like setting up a social behavior profile of that user. “User behavior analysis is based on user-generated data that largely reflect the characteristics of individual users” [34]. Other characteristics such as “location, posting time, and writing style” can be used to determine whether it is the same person with multiple accounts. This becomes important because sometimes users who create multiple accounts with malicious intent, can be tracked by their behavior throughout the multiple accounts.

An important concept in user behavior analysis is the potential to identify insider threats. An insider threat is a security risk that occurs within a targeted organization. The insider actor possesses authorized access to that organization’s network, data, and resources, posing a threat to confidentiality, integrity, and availability.

Tuor et al. proposes deep learning models that are targeted at anomaly detection [35]. The models are based on recurrent and deep neural networks. Recurrent neural networks, or RNNs, introduce feedback into previous layers, whereas other neural networks operate as strictly feedforward. The models rank anomaly scores, which are then assessed by analysts to determine if said rank is indicative of an insider threat. The models were compared to Principal Component Analysis, Support Vector Machine, and Isolation Forest models. The best-performing model resulted in a 93.5% data reduction for the data analysts, simultaneously reducing time and work for the analysts. This meant that the model was capable of significantly identifying and separating the possibly malignant data from the benign.

Another topic that falls under user behavior analysis is the ability to detect fake news (misinformation), by uncovering the source of the news, or more specifically, the *user* behind the news. Ruchansky et al. proposes the CSI (Capture, Score, and Integrate) model [36]. Like the Tuor model [35], CSI uses a recurrent neural network. The model breaks down a news sample into three components: text (language consistency and quality), response (reception, inclination to share, like, etc.), and source (authors, media source, etc.). The RNN is responsible for capture of the text and response features and the computation of the score per user. Integration of the results of those modules provides a more accurate prediction. Both TWITTER and WEIBO datasets were used in training and testing, because of the dataset's inclusion of text, response, and source characteristics. CSI model achieved 89.2% accuracy on the TWITTER dataset and 95.3% on the WEIBO dataset.

5. Use Cases: An Exploration into Machine Learning Applications for Cybersecurity

To better understand the use of DL in cybersecurity, we explore various use cases in the following subsections.

5.1. Internet of Things

The Internet of Things, or IoT, is used to connect billions of devices together throughout the world. It is estimated that 50 billion devices will be connected together by the end of year 2020 [37]. The implementation of IoT is enhancing real-time applications such as smart homes, smart cars, smart education and smart ecosystems. In IoT systems, various components are integrated together, and several devices are deployed together, this introduces new security challenges. Most of the IoT devices are working in unattended environments, like smart cameras installed in a house that are not monitored by any

security professionals. This allows any intruder to gain physical access or unauthorized access by launching denial of service attacks. For example, Mirai took down most of the IoT devices using botnets [37]. Much more complex attacks than Mirai can be expected and it is difficult to develop mechanisms for a specific attack. The Internet of Things is a mixture of radio frequency identifications, wireless sensor networks and cloud computing [38] and due to this the attack surface of the IoT devices is larger, also easily accessible to the Internet hackers. Present encryption techniques are not efficient to countermeasure various kinds of attacks launched on IoT devices. Due to the vulnerable nature of IoT devices many attackers are launching new types of attacks before researchers are finding a solution for the previous one. Hence, developing a robust IoT device is needed and this can be done by considering energy efficiency, security, IoT data analytics aspects during the developing stages. Monitoring of IoT devices regularly will help to identify the threats. Machine learning and Deep learning techniques can be employed to monitor the IoT data continuously. As illustrated in Figure 9, the ability of machine learning/deep learning algorithms to monitor the IoT devices continuously will minimize the attacks. Machine learning and Deep learning algorithms identify and differentiate abnormal behavior of the device from the normal behavior that makes threat detection much easier. The input data to IoT devices can be taken and fed to ML/DL algorithms as training data to study normal behavior patterns of a device when it interacts.

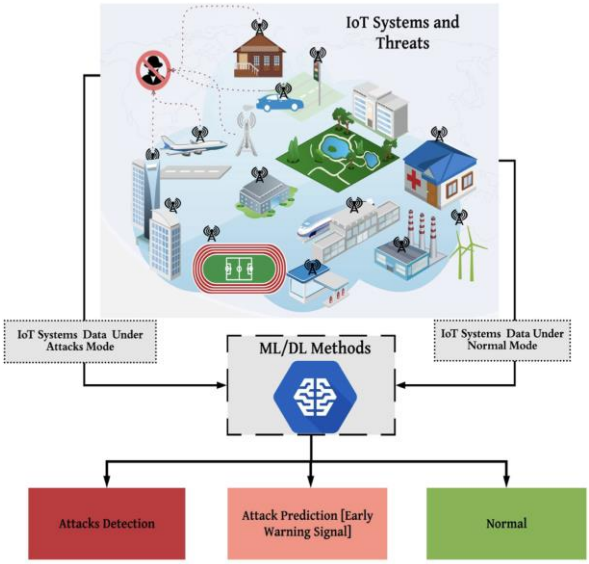

Figure 9: Model highlighting the role of ML/DL in IoT security.

Attacks such as spoofing try to gain unauthorized access to an IoT device by impersonating a node with medium access control address or RFID tags. Once the attacker gains access, he can launch distributed denial of service attacks. Not only that, attackers can send signals that are illegitimate to jam the IoT transmissions. Due to open source IoT devices the attackers can launch various attacks such as Sybil, Man-in-the-middle,

Eavesdropping and so on. To prevent unauthorized access authentication machine learning technique helps to distinguish between a legitimate source code and illegal source code. Traditional physical layer authentication techniques have limitations such as computational problems, less memory allocation and batter to identify attacks. Machine learning techniques such as supervised learning, reinforcement learning or unsupervised learning techniques can be used to provide secure authentication.

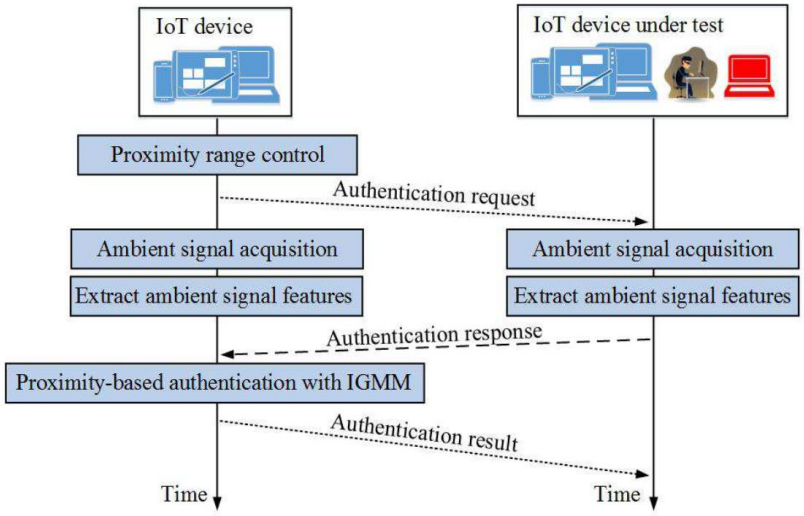

Figure 10: Illustration of implementing IGMM in authentication of IoT security

Proximity based authentication using unsupervised learning techniques like infinite gaussian mixture model as proposed in [38] can be used to provide secure authentication by preventing spoofing and without revealing location of the IoT devices. Infinite gaussian mixture model is a nonparametric Bayesian model which is used to determine receiver signal strength indicator of the transmitting radio signals [38]. As described in Figure 10 the IoT device sends requests to IoT devices under test requesting features of the signals such as receiver signal strength indicator, packet arrival time, MAC address and so on. The IoT device then sends the features to a legal receiver which applies the infinite gaussian mixture model to compare the received features with observed features in the proximity-based test. If the recorded features of the IoT device under test are the same, then the IoT device authenticates access to access its resources. When tested, this technique showed reduction in error detection rate from 20% to 5%.

Supervised learning techniques can be employed to help detect malicious software in IoT devices. Often malware gets induced in the devices without consent or user's knowledge. The presence of malware might give access to the user's sensitive information such as bank details, password and so on to the attackers. Malware detection techniques

using K-NN and random forest classifiers as proposed in [39] can be used to study the behavior of the device and identify any abnormal behavior. As in Figure 11, as and when the IoT device filters network traffic, it stores IP addresses, protocols, file type, frame length, frame number, and host post number in a database. The data stored in the database is given as input to the machine learning algorithm, the K-NN model assigns labels to the network traffic elements and the random forest builds a classifier to identify the malware based on labeled inputs.

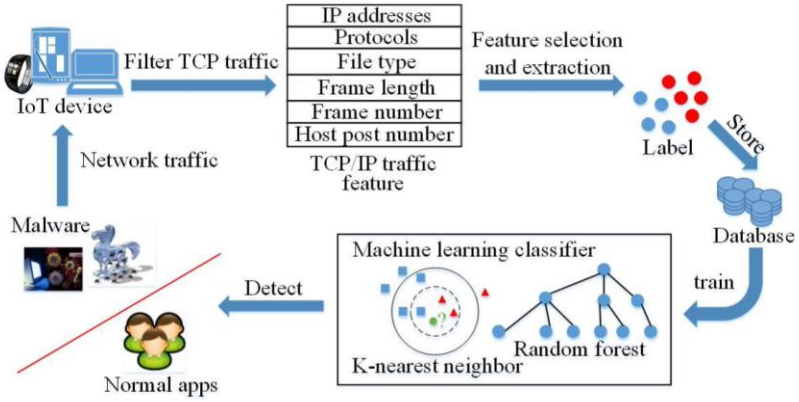

Figure 11: Implementation of machine learning algorithms in malware detection of IoT devices.

5.2. Intrusion Detection Systems

An intrusion detection system (IDS) has changed the way people live their lives. Phones became smart phones, and homes are now following suit, in becoming smart homes. All these smart devices make life easier, allowing us to do other things or give us more time in the day to work on other things. As we rely more and more on machines to monitor things, we need to make sure that these devices are secure.

Scenarios, such as in [40], the IDS was compromised. Ring devices have made headlines for adding a “peace of mind” in your life. This peace comes from the fact that you can always monitor who is where and that you don’t necessarily have to open the door to interact with a visitor. However, with a compromised IDS, we lose the sense of security we thought we had with these smart devices.

Deep learning comes in and allows us to help defend against these adversary attacks by learning to detect any anomalies. However, there was little research in the risks of using deep learning in IDS and the “vulnerability discovered in recent years greatly limit the application of deep neural networks in security-critical areas such as

self-driving, safety-critical voice-controllable systems, and IDS" [40]. To fix this, we use several layers of neural networks to create a deep neural network where many different attacks can be classified and then prevented.

5.3. Malware Detection in Android Devices

One of the most common and dangerous threats to a computer system, malware is also one of the most expensive to resolve [23]. As malicious software, malware can affect any computer system, meaning any operating system. Dubbed the best-selling operating system worldwide since 2011, even Android is susceptible to the threat.

Z. Yuan et al. explores Android malware detection through various machine learning techniques and one deep learning application in [41]. Typically, mobile systems use a risk communication technique to warn users of a potential for malware when downloading. The ineffectiveness of the technique is apparent when considering the inconspicuous nature of malware. That is why the authors of [41] turn to DL. Through static and dynamic analysis, a dataset of 202 features was compiled from multiple Android apps.

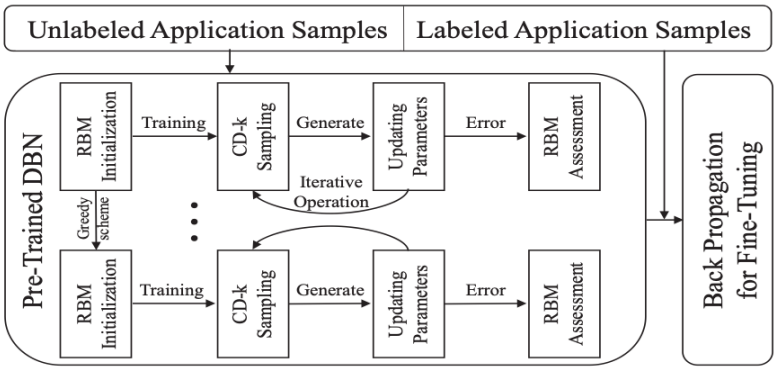

Figure 12: Droid-Sec Framework Model. From *Droid-Sec: Deep Learning in Android Malware Detection*, 2014.

The DL model employed consisted of a semi-supervised training algorithm, broken down into an unsupervised pre-training phase and a supervised back-propagation phase. The pre-training phases consisted of a DBN of RBMs. A similar structure is defined in [25]. In the next phase, the supervised propagation phase, the pre-trained data was labeled. Back-propagation, a common phase in DL approaches, is responsible for adjusting the weights of the neurons in an effort to produce more accurate results and a reduction in errors. Thus, the model was implemented (see Figure 12).

Using a public application set of both malware and benign apps, the deep learning model was validated. At 3 layers with 150 neurons per layer, the model was able to achieve

a maximum accuracy of 96.5% [41]. The DL model outperformed the following machine learning techniques: SVM, Naïve Bayes, C4.5, Logistic Regression, and Multi-layer Perceptron.

Because the authors tested their model on real world data, they alluded to the promising potential of such deep learning techniques in industry. The trend of increased malware threats and high resolution costs point to a research ripe field in deep learning malware detection solutions.

6. CONCLUSIONS

The theme of deep learning research is often that there is not enough. With the potential to continuously improve our lives through numerous industries, DL also has a significant stake in the cybersecurity field. More and more we recognize cybercrimes as a growing and prevalent threat that can cost millions of dollars to resolve. The conclusion, there has to be more research into what deep learning applications can offer to cybersecurity.

REFERENCES

[1]. Cybersecurity Ventures. (2019). "2019 Official Annual Cybercrime Report". Herjavec Group. https://www.herjavecgroup.com/wp-content/uploads/2018/12/CV-HG-2019-Official-Annual-Cybercrime-Report.pdf

[2]. R. Anyoha. "The History of Artificial Intelligence". Harvard University, August 28, 2017. Retrieved from: http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

[3]. J. Jang-Jaccard, S. Nepal. "A Survey of Emerging Threats in Cybersecurity." J*ournal of Computer and System Sciences, Academic Press,* 10 Feb. 2014, https://doi.org/10.1016/j.jcss.2014.02.005

[4]. R. Sobers. (Updated: 10/26/2020). "110 Must-Know Cybersecurity Statistics for 2020: Varonis." Inside Out Security. 26 Oct. 2020. Web. 20 Nov. 2020.

[5]. Devi, R. S. and D. Mohankumar. "AN EMPIRICAL STUDY ON CYBER SECURITY THREATS AND ATTACKS." (2019). https://www.semanticscholar.org/paper/AN-EMPIRICAL-STUDY-ON-CYBER-SECURITY-THREATS-AND-Devi-Mohankumar/3f197fcc0bd775fc2e970a71139736e606e1ec5b

[6]. J. Alzubi, A. Nayyar, A. Kumar. (2018) Machine Learning from Theory to Algorithms: An Overview. Journal of Physics: Conference Series. 1142. 012012. 10.1088/1742-6596/1142/1/012012.

[7]. A. Krizhevsky, I. Sutskever, G. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks", Communications of the ACM, May 2017. Retrieved from: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[8]. R. Salakhutdinov, G. Hinton. “Deep Boltzmann Machines”. University of Toronto, 2006. Retrieved from: http://proceedings.mlr.press/v5/salakhutdinov09a/salakhutdinov09a.pdf
[9]. G. Hinton. R. Salakhutdinov. “Reducing the Dimensionality of Data with Neural Networks”. Science, 28 July 2006. Retrieved from: https://www.cs.toronto.edu/~hinton/science.pdf
[10]. Buczak, Anna L., and Erhan Guven. “A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection.” IEEE COMMUNICATIONS SURVEYS &TUTORIALS, vol. 18, no. 2, 2016, pp. 1153–1175. SECOND QUARTER 2016.
[11]. L. Dormehl. “What Is an Artificial Neural Network? Here's Everything You Need to Know”. Digital Trends, Digital Trends, 6 Jan. 2019, www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/.
[12]. I. Goodfellow, Y. Bengio, A. Courville. “Deep Learning”. MIT Press, 2016. Retrieved from: https://www.deeplearningbook.org
[13]. J. Brownlee. “What is Deep Learning?”. Machine Learning Mastery, August 14, 2020. Retrieved from: https://machinelearningmastery.com/what-is-deep-learning/
[14]. S. Symanovich. “Privacy vs. Security: What’s the Difference?”. NortonLifeLock, 28 August 2019. Retrieved from: https://us.norton.com/internetsecurity-privacy-privacy-vs-security-whats-the-difference.html
[15]. F. Bastani, T. Tang. “Improving Security of Wireless Communication in Medical Devices”. Massachusetts Institute of Technology.
[16]. N.Papernot, P. McDaniel, A. Sinha, M. Wellman. “SoK: Towards the Science of Security and Privacy in Machine Learning”. [arXiv:1611.03814v1][cs.CR], 11 November 2016.
[17]. Kwon, Seoyun J. *Artificial Neural Networks*. Nova Science Publishers, Inc, 2011. *EBSCOhost*,search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=439593&site=eds-live&scope=site.
[18]. M. Barreno, B. Nelson, A. Joseph, J. Tygar. “The Security of Machine Learning”. SpringerLink.com, 20 May 2010.
[19]. M. Zamani.” Machine Learning Techniques for Intrusion Detection”, 2013. https://www.researchgate.net/publication/259212150_Machine_Learning_Techniques_for_Intrusion_Detection/citation/download
[20]. P. Laskov, P. Dssel, C. Schfer, and K. Rieck. Learning Intrusion Detection: Supervised or Unsupervised?. *In Image Analysis and Processing ICIAP 2005*, volume 3617 of Lecture Notes in Computer Science, pages 50–57. Springer Berlin Heidelberg, 2005.
[21]. C. Sinclair, L. Pierce, and S. Matzner. An Application of Machine Learning to Network Intrusion Detection. In *Proceedings of the 15th Annual Computer Security Applications Conference*, ACSAC ’99, Washington, DC, USA, 1999.
[22]. S. Mukkamala, G. Janoski, and A. Sung. Intrusion Detection Using Neural Networks and Support Vector Machines. In *Proceedings of the 2002 International Joint Conference on Neural Network (IJCNN),* volume 2, pages 1702–1707, 2002.
[23]. Accenture Security. “The Cost of Cybercrime: Ninth Annual Cost of Cybercrime Study”. Ponemon Institute, 2019. Retrieved from:

https://www.accenture.com/_acnmedia/pdf-96/accenture-2019-cost-of-cybercrime-study-final.pdf
[24]. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran and S. Venkatraman. "Robust Intelligent Malware Detection Using Deep Learning," in *IEEE Access*, vol. 7, pp. 46717-46738, 2019. Retrieved from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8681127&isnumber=8600701
[25]. Y. Li, R. Ma, R. Jiao. "A Hybrid Malicious Code Detection Method based on Deep Learning". *In International Journal of Security and Its Applications*, Vol 9(5), 205-216, 2015. Retrieved from: http://www.covert.io/research-papers/deep-learning-security/A%20Hybrid%20Malicious%20Code%20Detection%20Method%20based%20on%20Deep%20Learning.pdf
[26]. W. Hardy, L. Chen, S. Hou, Y. Ye, X. Li. "DL4MD: A Deep Learning Framework for Intelligent Malware Detection." *International Conference Data Mining*, West Virginia University [ISBN 1-60132-431-6, CSREA Press]. Retrieved from: https://www.covert.io/research-papers/deep-learning-security/DL4MD-%20A%20Deep%20Learning%20Framework%20for%20Intelligent%20Malware%20Detection.pdf
[27]. G. Liu, H. Bao, B. Han. "A Stacked Autoencoder-Based Deep Neural Network for Achieving Gearbox Fault Diagnosis" *Mathematical Problems in Engineering*, vol. 2018, Article ID 5105709, 10 pages, 2018. Retrieved from: https://doi.org/10.1155/2018/5105709
[28]. Wu, Tingmin, et al. "Twitter Spam Detection Based on Deep Learning." Proceedings of the Australasian Computer Science Week Multiconference on - ACSW '17, 2017, doi:10.1145/3014812.3014815.
[29]. Chowdhury, Rajarshi Roy, et al. Network Traffic Analysis Based IoT Device Identification. 2020. EBSCOhost, search.ebscohost.com/login.aspx?direct=true&db=edsarx&AN=edsarx.2009.04682&site=eds-live&scope=site.
[30]. M. Soysal, E.G. Schmidt, "Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison," Performance Evaluation, vol. 67, no. 6, pp. 451–467, 2010.
[31]. H. Kim et al., "Internet traffic classification demystified: myths, caveats, and the best practices". In *Proc. ACM CoNEXT Conf.*, 2008, pp. 1–12.
[32]. J. Zhang, X. Chen, Y. Xiang, W. Zhou, and J. Wu, "Robust network traffic classification," *IEEE/ACM Transactions on Networking (ToN)*, vol. 23, no. 4, pp. 1257–1270, 2015.
[33]. Yan, Qiang, et al. "Social Network Based Microblog User Behavior Analysis." Physica A:Statistical Mechanics and Its Applications, vol. 392, no. 7, 2013, pp. 1712–1723., doi:10.1016/j.physa.2012.12.008.
[34]. Deng, Kaikai, et al. "A User Identification Algorithm Based on User Behavior Analysis in Social Networks." IEEE Access, vol. 7, 2019, pp. 47114–47123., doi:10.1109/access.2019.2909089.
[35]. A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, S. Robinson. "Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams".

[arXiv:1710.00811v2] [cs.NE], 15 Dec 2017. Retrieved from: https://arxiv.org/pdf/1710.00811.pdf
[36]. N. Ruchansky, S. Seo, Y. Liu. “CSI: A Hybrid Deep Model for Fake News Detection.” *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Pages 797-806, November 2017. Retrieved from: https://dl.acm.org/doi/abs/10.1145/3132847.3132877
[37]. M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali and M. Guizani, "A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security," in *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1646-1685, thirdquarter 2020, doi: 10.1109/COMST.2020.2988293.
[38]. I. Andrea, C. Chrysostomou, and G. Hadjichristofi, “Internet of Things: Security vulnerabilities and challenges,” in *Proc. IEEE Symposium on Computers and Commun*, pp. 180–187, Larnaca, Cyprus, Feb. 2015.
[39]. L. Xiao, Q. Yan, W. Lou, G. Chen, and Y. T. Hou, “Proximity-based security techniques for mobile users in wireless networks,” *IEEE Trans. Information Forensics and Security*, vol. 8, no. 12, pp. 2089–2100, Oct. 2013.
[40]. Chiu, Allyson. “She Installed a Ring Camera in Her Children's Room for 'Peace of Mind.' A Hacker Accessed It and Harassed Her 8-Year-Old Daughter.” The Washington Post, WP Company, 13 Dec. 2019, www.washingtonpost.com/nation/2019/12/12/she-installed-ring-camera-her-childrens-room-peace-mind-hacker-accessed-it-harassed-her-year-old-daughter/.
[41]. Z. Yuan, Y. Lu, Z. Wang, Y. Xue. “Droid-Sec: Deep Learning in Android Malware Detection”. *Proceedings of the 2014 ACM Conference on SIGCOMM*, Pages 371-372, August 2014. Retrieved from: https://dl.acm.org/doi/abs/10.1145/2619239.2631434
[42]. F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, “Evaluation of machine learning classifiers for mobile malware detection,” *Soft Computing*, vol. 20, no. 1, pp. 343–357, Jan. 2016.
[43]. Wang, Zheng. “Deep Learning-Based Intrusion Detection With Adversaries.” IEEE Xplore, 9 July 2018, ieeexplore.ieee.org/abstract/document/8408779.