AI and machine learning: A mixed blessing for cybersecurity

Faouzi Kamoun
School of Engineering
ESPRIT
Tunis, Tunisia
faouzi.kammoun@esprit.tn

Farkhund Iqbal
College of Technological Innovation
Zayed University
Abu Dhabi, UAE
Farkhund.Iqbal@zu.ac.ae

Mohamed Amir Esseghir
School of Engineering
ESPRIT
Tunis, Tunisia
amiresseghir@gmail.com

Thar Baker
Dept of Computer Science
Liverpool John Moores University
Liverpool, UK
T.M.Shamsa@ljmu.ac.uk

***Abstract*— While the usage of Artificial Intelligence and Machine Learning Software (AI/MLS) in defensive cybersecurity has received considerable attention, there remains a noticeable research gap on their offensive use. This paper reviews the defensive usage of AI/MLS in cybersecurity and then presents a survey of its offensive use. Inspired by the System-Fault-Risk (SFR) framework, we categorize AI/MLS-powered cyberattacks by their actions into seven categories. We cover a wide spectrum of attack vectors, discuss their practical implications and provide some recommendations for future research.**

Keywords—Security, Cybersecurity, AI, machine learning, deep learning, neural networks, adversarial techniques

I. INTRODUCTION

In the era of ubiquitous Internet, cloud services, 5G mobile technology, and IoT, protecting organizational assets from harm and operations from disruptions has become more arduous than ever. In 2018, Cisco alone blocked seven trillion threats, or 20 billion threats a day, on behalf of their customers [1]. The FBI reported that Email Account Compromise (EAC), a scam aimed towards businesses and individuals performing wire transfer payment, resulted in losses estimated at more than $12.5 billion during the period January 2014-May 2018 [1]. In front of these ever-increasing threats, organizations need help. Some organizations are turning to Artificial Intelligence and Machine Learning Software (AI/MLS) to boosts their cybersecurity defenses for better automation, management, and effectiveness. In the sequel, the abbreviation AI/MLS will be used in a broader sense to encompass machine learning, pattern recognition, deep learning, and reinforcement techniques and technologies.

Historically, the InfoSec community has used AI/MLS defensively [2], for example in enhancing intrusion detection systems such as classifying malicious binaries or identifying anomalies in network traffic [3]. AI/MLS have the inherent capability to learn from past attacks and assist cybersecurity professionals to improve security solutions, empower digital forensic investigations, and curb cyberattacks. A survey reported in [4] indicated that 74% of businesses across the U.S and Japan have already begun using some form of AI/MLS to protect their organizational assets. However, AI/MLS tend to become a double-edged sword, as there have been growing concerns that they might be exploited by hackers to launch more complex attacks. To this regard, the same survey [4] indicated that 84% of security professionals are concerned that AI/MLS can be potentially weaponized to launch a new breed of sophisticated cyber-attacks that have strong potential to evade "traditional" as well as AI-powered cybersecurity defense layers. This might eventually lead to a potential AI/MLS vs AI/MLS war in the realm of cybersecurity. Today, adversarial AI/MLS models are poised to take hacking to a new level giving rise to a new breed of intelligent systems that can learn from experience and self-improve without explicit programming. It is therefore important that the InfoSec community understands the mechanisms by which cybercriminals can turn AI/MLS into weapons for malicious use so that they can put the proper defense mechanisms. Today, the potential misuse of AI/MLS models to launch cyberattacks remains an underexplored research topic [5-6]. Current literature has addressed different AI/MLS attacks in isolation with no comprehensive review or classification of these attacks. The aim of this paper is two folds: (1) shed light on the two facets (defensive/adversarial) of AI/MLS usage in the context of cybersecurity, and (2) outline a classification approach for AI/MLS-powered attacks. Such a classification can facilitate the identification of these attacks and establish the relationship among them, which may not be obvious when we look at them as a whole. To the best of our knowledge, this is the first contribution that aimed at (1) bringing the two facets of AI/MLS together and (2) proposing a classification of the AI/MLS-powered cybersecurity attacks.

The remaining of this paper is structured as follows: Section II presents a summary of the usage of AI/MLS models in cybersecurity defense. Section III discusses the key features of AI/MLS-powered cyberattacks and presents a classification of these attacks, based on their activities or actions. Section IV highlights the practical implications of weaponized AI/MLS models on the future of cybersecurity, whereas section V provides a summary of the paper and some suggestions for future research.

II. AI/MLS FOR CYBERSECURITY DEFENSE

When used as standalone tools or in conjunction with traditional defense methods, AI/MLS models offer powerful defensive tools to protect against cyberattacks and to assist in digital forensic investigations. In this section, we provide a summary of the key applications of AI/MLS in cybersecurity defense. For additional details, we refer the reader to the previous work of Garcia-Teodoro et al [7] Wu and Banzhaf [8], Buczak et al [9], Torres et al [10], and Berman et al [11].

978-1-7281-5628-6/20/$31.00 ©2020 IEEE

A. AI/MLS for Malware Detection and Classification

Several studies investigated the usage of AI/MLS for malware detection by exploiting the static and dynamic features of applications. Malware classification involves tagging a class of malware to a given sample to determine the malware type, which can help identify the motive of the attack. Anderson et al [12] noted that AI/MLS techniques offer a common approach to signatureless malware detection because they can generalize to never-seen-before malware families and polymorphic strains.

Kang et al [13] showed how a Long Short-Term Memory (LSTM) neural network can be used to estimate the effect of malware by analyzing the opcodes in its executable files and classifying the malware accordingly. Using a similar approach, Charan et al [14] showed how an LSTM can analyze large amounts of system event logs to detect Advanced Persistent Threat (APT) malware. Gupta et al [15] proposed a machine learning model to detect malware in Android-based smartphones through system calls. Other related contributions include the work of McLaughlin et al [16], Milosevic et al [17], and Yuan et al [18], among many others.

B. AI/MLS for Network Intrusion Detection

Several AI/MLS models have been proposed to support Network Intrusion Detection Systems (NIDS). Using the KDD-1999 dataset, Alom and Taha [19] performed K-means clustering to achieve a detection accuracy of 91.86%. Kim and Kim [20] applied Recurrent Neural Networks (RNNs) to intrusion detection, achieving a 100% detection rate and a 2.3% false alarm rate. Ding et al [21] proposed a real-time anomaly detection algorithm based on LSTM and Gaussian Mixture Model (GMM). Catak and Mustacoglu [22], and Chen et al [23] demonstrated the usage of Deep Convolutional Neural Networks (DCNNs) in the detection of DDoS attacks. Other related contributions include the work of Xia et al [24], Clements et al [25], Biswas [26], and Mirsky et al [27].

C. AI/MLS for Traffic Identification and Classification

Using Intranet TCP flow data, Wang [28] used a deep learning model to classify traffic flow types into 25 protocols, with precision between 91.74% and 100%, depending on the protocol type. Lotfollahi et al [29] used a CNN model to classify the type of network traffic as well as to recognize the type of application. Wang et al [30] proposed a CNN model that is capable of (1) distinguishing between VPN and non-VPN encrypted traffic streams and (2) classifying each traffic type into different levels. Other AI/MLS models have been proposed to classify encrypted network traffic, which can be used by firewalls and NIDS (e.g. Rezaei and Liu [31]; Aceto et al [32]).

D. AI/MLS for DGA, Botnet and Spam Detection

Domain Generation Algorithms (DGAs) can generate a large number of varying malicious domain names that can evade standard blacklisting and sink-holing defense methods [11]. DGAs are often associated with spam campaigns, malware communication with Command and Control (C2) servers, phishing, and DDoS attacks. Woodbridge et al [33] used RNNs to identify malicious domain names generated by DGAs. Other related contributions include the work of Tran et al [34], and Lison and Mavroeidis [35].

To address the problem of traditional rule-based behavioral models in detecting botnets, Torres et al [10] used an LSTM model to implement a Botnet detector. McDermott et al [36] used a trained LSTM model to identify four attack vectors of the Mirai Botnet [37], namely UDP, ACK, DNS and SYN floods. Hoang and Nguyen [38] proposed a two-phase detection model, based on machine learning and DNS query data, to increase the possibilities of detecting botnets.

Mi et al [39] proposed an auto-encoder followed by a classifier to identify spam emails with accuracy above 95%. Tzortzis and Lika [40] proposed a Deep Belief Network (DBN) for Spam Filtering. Alauthman et al [41] demonstrated the usage of machine learning in detecting phishing emails. Other contributions that applied AI/MLS models to detect email and web phishing attacks include the work of Benavides et al [42], Yi et al [43] and Aksu et al [44].

E. AI/MLS for Insider Threat Detection

Tuor et al [45] demonstrated how a DNN or an RNN model can effectively be used to analyze system logs of end-users and detect anomalies that might signal an insider threat event.

F. AI/MLS for Drive-by-download Attack Detection

Detecting drive-by-download is an active research area that has traditionally relied on anomaly detection methods. Deep learning neural networks provide more powerful approaches to detect and prevent this type of attack while reducing false alerts (e.g. Shibahara et al [46], Yamanishi [47]).

G. AI/MLS for Digital Forensic

Ariu et al [48] and Mitchell [49] discussed the important role AI/MLS can play in digital forensics. Building on these contributions, Karie et al [50] proposed a framework to embed deep learning cognitive computing techniques into cybersecurity forensics.

Recognizing that PDF documents are among the major vectors used to carry malware, Maiorca and Biggio [51] discussed the usage of AI/MLS as powerful PDF malware analysis tools that can support digital forensic investigations.

Traditional machine learning methods (e.g. Calhoun and Coles [52], Axelsson [53], Beebe et al [54]) have been applied to classify file fragments, a task that plays an important role in digital forensics. More recently, Chen et al [55] proposed a file fragment classification method based on grayscale image conversion and Deep learning. Hoon et al. [56] discussed the application of AI/MLS to Big data analytics in the context of DDoS digital forensics. Wang et al [57] applied machine learning tools to perform memory forensic analysis for the purpose of detecting kernel rootkits in Virtual Machines (VMs).

III. AI/MLS AS CYBERSECURITY THREAT

Rapid democratization of artificial intelligence has made AI/MLS-powered attacks a looming threat [58]. Access to open-source AI/MLS models, tools, libraries frameworks, and pre-trained deep learning models (e.g. TensorFlow, Keras,

Torch, Caffe, AlexNet, GoogleNet, ShuffleNet) make it easier for hackers to adapt AI models and tools in order to arm their exploits with more intelligence and efficiency.

What distinguishes adversarial AI/MLS-powered attacks from other traditional cyberattacks is the combinational set of speed, depth, automation, scale, and sophistication that these models tend to offer. In fact, AI/MLS models can bring three changes to the way threats are orchestrated and executed:

- Amplification in terms of the number of actors participating in an attack, the occurrence rate of these attacks and the number of attacked targets [6]
- Introduction of new threat vectors that would be impractical for humans to craft using traditional (preset, instruction-based) algorithms, and
- Injection of intelligence into traditional attack vectors, bringing new attributes and behavior to these threats, such as opportunism and polymorphism.

Inspired by the System-Fault-Risk (SFR) framework [59], we categorize AI/MLS-powered cyber-attacks by their activities (actions) into seven categories, as shown in Table 1.

TABLE I. AI-MLS POWERED ATTACKS CLASSIFICATION

AI/MLS models can be maliciously used to *(action)*			
Probe	Scan	Spoof	Flood
Misdirect	Execute	Bypass	

A. AI/MLS for Probing

In our context, we define probing as the capability of using AI/MLS to access an organizational asset to determine its characteristics. AI/MLS can be used to automate network probing [60]. More precisely, it can be used to intelligibly mine a large amount of public domain and social network data related to organizations and individuals, and which can be spread across multiple information sources, such as social media, news, blogs, forums, forums, and code repositories. This can enrich probing activities by maximizing the amount of gathered information. Some of this information can be used by hackers to launch more powerful and personalized social engineering attacks such as phishing, pretexting, baiting and quid pro quo attacks.

B. AI/MLS for Scanning

This refers to using AI/MLS to access a set of organizational assets sequentially to detect which assets have a specific characteristic. In operating system fingerprinting, Song et al [61] showcased how an Artificial Neural Network correctly identified operating systems with a 94% success rate, which is higher than the accuracy of conventional rule-based methods.

C. AI/MLS for Spoofing

This refers to using AI/MLS as a masquerade tool to disguise the identity of an entity. Earlier research has shown that AI/MLS models can be used for adversarial machine learning by poisoning AI/MLS engines that were supposed to protect against malware in the first place. For instance, Gu et al [62] highlighted the potential danger of the increasingly common practice of outsourcing pre-trained Convolutional Neural Networks (CNNs), publicly available online, to build security defenses. They showcased how a CNN can be maliciously trained by an adversary to create a stealthy backdoor neural network (BadNet) that behaves as expected on the user's training and validation samples, but misbehaves otherwise on carefully crafted attacker-chosen inputs.

D. AI/MLS for Flooding

This refers to using AI/MLS from single or multiple sources to overload an organizational asset's capacity. Fortinet [63] predicts that cybercriminals will replace botnets with self-learning "hivenets" and "swarmbots", a set of intelligent clusters of compromised devices, to create attack vectors at unprecedented scales. Hivenets share local intelligence and multiply as swarms, hence amplifying the scale of the attack.

Sagduyu et al [64] applied adversarial machine learning techniques in IoT systems to three types of over-the-air (OTA) wireless attacks, namely jamming, spectrum poisoning, and priority violation (evasion) attacks. AI/MLS models were also applied to trigger jamming attacks on wireless data transmission (Shi et al [65]; Erpek et al [66]).

CAPTCHA is used to restrict computer-automated submissions, hence reducing spam and frauds and preventing automated bots from conducting malicious activities. It ensures that the submission is being done by a human being. Various research initiatives have demonstrated the capability of AI/MLS in breaking CAPTCHA and Google reCAPTCHA with varying degrees of success. Cruz-Perez et al [67] proposed a reCAPTCHA breaker based on a Support Vector Machines (SVM) classifier that reported a segmentation success rate up to 82%. Using deep learning techniques for the semantic annotation of images, Sivakorn et al [68] developed an AI/MLS tag classifier that can guess the content of a reCAPTCHA image with an accuracy of 83.5% for Facebook image CAPTCHA. Other research contributions that aimed to crack CAPTCHA include the work of Yu and Darling [69].

E. AI/MLS for Misdirection

This refers to the capability of using AI/MLS to deliberately lie to a target and provoke an action based on deception; such is the case of cross-site scripting and email scams. This application has been the subject of considerable interest among the cybersecurity community.

AI/MLS can be used to generate malicious domain names, which can feed several types of cyberattacks, including spam campaigns, phishing emails and distributed DoS attacks [11]. Among the most prominent contributions, we cite the work of Anderson et al [70] that demonstrated the potential of Generative Adversarial Networks (GANs) to act as a malware tool by producing malicious domain names that can infiltrate current Domain Generation Algorithms (DGA) classifiers.

A Generative Adversarial Network (GAN) is a class of deep learning neural network architectures, introduced by Goodfellow et al [71], which is used in a wide spectrum of applications, especially in data augmentation, computer vision and image processing [11]. When provided with a training set,

a GAN can learn to generate new data that has very similar statistics to the training set.

As shown in Fig.1, a GAN consists of two competing neural networks trying to outsmart each other, often in a zero-sum game.

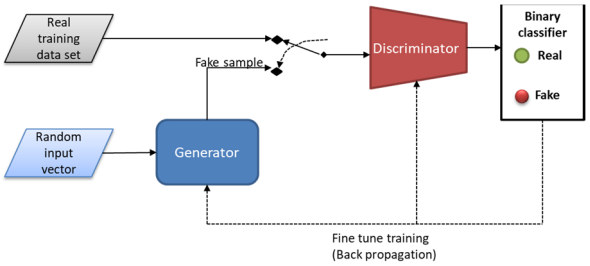

Fig. 1. Generative Adversarial Network (GAN) architecture

The first network acts as a generator that takes input data and generates new plausible data instances that have similar characteristics to the real data.

The second network is a discriminator that takes-in real data and data produced by the generator and decides whether the input is real (from the domain) or fake (produced by the generator network). While the discriminator learns to better discriminate between real and fake samples and penalizes the generator for producing implausible results, the generator learns (from the discriminatory power of the discriminator) to create new plausible samples that are even harder for the discriminator to classify as fake.

The two networks are trained together until the discriminator model is fooled about 50% of the time, implying that the generator network is capable of producing new data that is indistinguishable from the original training dataset. At this final stage, the discriminator model is discarded as the interest shifts towards the trained generator.

The power of GANs resides in their capability to produce malicious, yet genuine looking, data (e.g. domain names, URLs, email addresses, IP addresses) that hackers can use to infiltrate most NIDS.

Bahnsen et al [5] proposed DeepPhish, a deep neural network algorithm based on Long Short-Term Memory Networks that can learn from previous effective attacks to create new synthetic phishing URLs that have better chances of bypassing fraud defense mechanisms. By training DeepPhish on two separate threat actors, the authors reported an increase in the effective rate of the attack from 0.69% to 20.9% and from 4.91% to 36.28%, respectively.

By applying a fuzzing technique to email content, Palka and McCoy [72] showed how to launch an AI-powered phishing attack by crafting an email to evade conventional filters over the course of several simulations, regardless of the type of countermeasures being deployed.

Singh and Thaware [73] investigated how spammers can use AI/MLS to improve the success rate of their Business Email Compromise (BEC) scam phishing attacks. They showcased how AI/MLS models can be used (1) for reconnaissance by mining massive quantities of information availed from publicly available data (social media profiles, company website, current affairs, organizational chart, etc..), (2) for target profiling by identifying high-profile/weakest targets, and (3) for crafting customized genuine-looking emails. An AI/MLS model can be trained on genuine emails and self-learn how to create new contextualized emails that look convincing.

Seymour and Tully [3] demonstrated the usage of a recurrent neural network that has been pre-trained on generating tweets using a combination of spear phish pen-testing data, Reddit submissions, and tweets. The model was dynamically seeded with topics extracted from timeline posts of both the target and the users they retweet or follow. This allows the AI/MLS model to craft its own phishing bait. The model, named SNAP_R, uses clustering to identify high-value targets based on their level of social engagement. It was capable of sending simulated spear-phishing tweets to more than 800 users at an average rate of 6.75 tweets per minute. Experimental tests involving 90 twitter subscribers showed a success rate between 30% and 66%: a noticeable improvement over manual/bulk spear-phishing results.

Giaretta and Dragoni [74] observed that the Natural Language Generation (NLG) techniques can potentially enable attackers to target large community audiences with machine-tailored emails. For this purpose, they proposed the Community Targeted Phishing (CTP) technique to automatically craft such emails.

F. AI/MLS for Execution

This refers to the capability of using AI/MLS to execute a malicious process on a system process; such is the case of viruses and Trojans. For instance, with AI/MLS it might be possible to create a new breed of malware that can evade the best existing defenses. IBM DeepLocker [58], a hacked version of a video-conferencing software, embeds an evasive attack program that activates only when it detects the face of a target individual. This level of targeted stealth is achieved through a deep convolutional neural network that hides its attack payload in benign carrier applications and activates it when a given target is identified through several features such as geolocation, facial and voice recognition.

Jung et al [75] showcased the capability of AVPASS, an open-source AI-aided software, in mutating Android malware to bypass anti-virus (AV) solutions. This was done by inferring AV features and detection rules as well as by obfuscating the Android binary (APK) while minimizing information leaking by sending fake malware.

Anderson et al [12] proposed a black box attack using a deep Reinforcement Learning (RL) agent that is equipped with a set of functionality-preserving operations that it may perform on Windows Portable Executable (PE) files. The RL agent can evade ML-based anti-malware PE engine. This suggests that AI/MLS can be used to detect what other ML-based malware detection mechanisms are "looking" for and hence create a malware that can evade detection by detecting blind spots in the AI/MLS model.

Petro & Morris [76] demonstrated the capability of DeepHack, a proof of concept open-source AI/MLS-based hacking tool that uses ML algorithms to break into a Web application or to perform a penetration test in full autonomy and with no prior knowledge of apps and databases. DeepHack learns how to exploit multiple kinds of vulnerabilities simply by itself through trial and errors, and reward mechanisms.

In another example demonstrating the usage of AI/MLS to create evasive malware, Hu and Tan [77] proposed a GAN-based algorithm to generate adversarial malware samples that was able to bypass black-box ML-based detection models. The algorithm uses a trained substitute detector to fit the black-box malware detection algorithm, and a generative network to transform malware samples into adversarial examples.

G. AI/MLS for Bypassing

This refers to using AI/MLS to create an alternative method to access an organizational asset or to elevate access privilege to a given asset.

AI/MLS can be used by hackers to optimize the process of cracking admin passwords by reducing the number of probable passwords based on collected data about the end-users or their organization. Hitaj et al [78] showcased how a deep learning-based approach, named PassGAN, uses a GAN to autonomously learn the distribution of real passwords from actual password leaks and to generate high-quality password guesses. Experimental results showed that PassGAN was able to surpass rule-based password guessing tools.

Zhou et al [79] proposed a deep learning model that works with a face recognition library to launch an attack on the social authentication system of Facebook which requires users to identify the correct names of friends tagged in photos.

Das et al [80] proposed a deep neural network solution to perform a cross-device power Side-Channel Analysis (SCA) attack aiming at breaking the secret key of an embedded device by exploiting the side-channel leakage emanating from the physical implementation of an AES-128 target encryption engine. Other contributions that aimed to apply machine learning techniques to perform profiling power SCA cross-device attacks include the work of Golder et al [81], Hospodar [82] and Lerman et al [83].

IV. IMPLICATIONS

This research suggests that, soon, organizations will be compelled to incorporate AI/MLS into their cybersecurity strategies and move swiftly towards building capacity in AI/MLS technologies. There is also a need to raise awareness among AI/MLS researchers, cybersecurity academic and professional communities, policymakers and legislators about the interplay between AI/MLS models and cybersecurity and highlight the imminent dangers that weaponized AI/MLS models can pose to cybersecurity. Integrating the offensive usage of AI/MLS into the cyber warfare strategies of nations, particularly from the perspective of developing and deploying weaponized malware is a research direction that is worth pursuing [2].

One implication of this study points to the need of researchers and cybersecurity professionals to:

- Examine the potential flaws of existing AI-based defense layers in thwarting AI/MLS adversarial threats, and
- Develop new AI/MLS approaches for cybersecurity that take adversary into account.

V. CONCLUSION

AI/MLS models have already proven to be both a blessing and a curse on the cybersecurity front. This suggests that current cybersecurity defenses will most probably become obsolete and that new defense mechanisms will be required.

We are currently learning about adversarial AI/MLS applications through reports and demos reported by "white hat" hackers and few high-tech companies whose goals are to increase awareness among cybersecurity professionals. How long it will take to see AI/MLS weaponized attacks in action remains to be seen, though this might have already happened, as it is difficult to ascertain that a cyberattack was powered by AI/MLS.

It would be interesting to explore the usage of AI/MLS to implement next-generation IDS systems with intelligent autonomous response capabilities that can quickly detect and also stop in-progress cyberattacks. The application of AI/MLS in cyber threat response remains an underexplored research topic that is worth pursuing.

It is also recommended to proactively anticipate the potential use-cases of misusing AI/MLS models and share the corresponding countermeasures with the InfoSec community. We hope that this research will stimulate new contributions in the emerging field of weaponized AI/MLS models and their implications on the future of cybersecurity profession, education, and research.

REFERENCES

[1] S.Morgan, "Cybersecurity almanac: 100 facts, figures, predictions and statistics," Cisco and Cybersecurity Ventures Press Release, https://cybersecurityventures.com/cybersecurity-almanac-2019/, accessed December 5, 2019.

[2] C. Easttom, "A methodological approach to weaponizing machine learning," In proceedings of AIAM' 19: 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, pp. 1-5, 2010.

[3] J. Seymour, and P. Tully, "Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter," Black Hat USA 2016. https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter-wp.pdf, accessed January 2019.

[4] "Knowledge gaps: AI and machine learning in cybersecurity: Perspectives from U.S. and Japanese IT professionals," Webroot Cybersecurity Report, 2019, https://www-cdn.webroot.com/6015/4999/4566/Webroot_AI_ML_Survey_US-2019.pdf , accessed November 10, 2019.

[5] A.C.Bahnsen, I. Torroledo, L.D. Camacho, and S. Villegas, S, "DeepPhish: Simulating Malicious AI," In Proceedings of the APWG Symposium on Electronic Crime Research (eCrime), pp. 1-9. 2018.

[6] M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, et al, "The malicious use of artificial intelligence: Forecasting, prevention, and Mitigation", Technical Report, 2018, https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/1c6q2kc4v_50335.pdf. Accessed Janaury 3, 2019.

[7] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez," Anomaly-based network intrusion detection: Techniques, systems and challenges," Comput. Secur, vol. 28, pp. 18–28, 2009.

[8] S.X. Wu, and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," Appl. Soft Comput. Vol. 10, pp. 1–35, 2010.

[9] L. Buczak, and E. Guven, E, "A Survey of data mining and machine learning methods for cybersecurity," IEEE Commun. Surv. Tutor, 18, pp. 1153–1176, 2016.

[10] J.M. Torres, C.I. Comesaña, and P.J. García-Nieto, "Machine learning techniques applied to cybersecurity," Int. J. Mach. Learn. Cybern, pp. 1–14, 2019.

[11] D.S. Berman, A.L.S. Buczak, J.S. Chavis, and C.L. Corbett, "A Survey of deep learning methods for cybersecurity, Information, vol. 10, no.122, pp. 1-35, 2019.

[12] H.S. Anderson, A. Kharkar, B. Filar, B. Roth, "Evading machine learning malware detection," Black Hat USA 2017, July 22-27, 2017. https://www.blackhat.com/docs/us-17/thursday/us-17-Anderson-Bot-Vs-Bot-Evading-Machine-Learning-Malware-Detection-wp.pdf , accessed November 6, 2018.

[13] J. Kang, S. Jang, S. Li, Y.S. Jeong, and Y. Sung, "Long short-term memory-based Malware classification method for information security," Computers & Electrical Engineering, vol. 77, pp. 366-375, 2019.

[14] P.V.S. Charan, T.G.Kumar, and M.P.M. Anand, "Advance persistent threat detection using Long Short Term Memory (LSTM) neural networks, Communications in Computer and Information Science, vol. 985, pp. 45-54, 2019.

[15] B.B. Gupta, S.Gupta, S. Goel, N. Bhardwaj, and J. SinghIn., "A prototype method to discover malwares in Android-based smartphones through system calls," In: Machine Learning for Computer and Cyber Security, B. Gupta, B.B. & Sheng, Q.Z. (eds), Chapter 7, pp. 1-25, 2019. Taylor & Francis

[16] N. McLaughlin, J.M. Del Rincon, B. Kang, S. Yerima, et al, "Deep android malware detection," In Proceedings of the 7th ACM on Conference on Data and Application Security and Privacy, Scottsdale, AZ, USA, pp. 301–308, 2017.

[17] N. Milosevic, A.Dehghantanha, and K.K.R. Choo, "Machine learning aided Android malware classification," Computers and Electrical Engineering, vol. 61, pp.266-274, 2017.

[18] Z. Yuan, Y. Lu, and Y. Xue, "Droiddetector: Android malware characterization and detection using deep learning,' Tsinghua Sci. Technol,vol. 21, pp.114–123, 2016.

[19] M.Z. Alom, and T.M. Taha, "Network intrusion detection for cybersecurity using unsupervised deep learning approaches," In Proceedings of the 2017 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, pp. 63–69, 2017.

[20] J. Kim, and H. Kim, "Applying recurrent neural network to intrusion detection with hessian free optimization," In proceedings of the International Conference on Information Security Applications, Jeju Island, Korea, pp. 357-369, 2015.

[21] N. Ding, H. Ma, H. Gao, Y. Ma, and G.Tan, "Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model," Computers & Electrical Engineering, vol. 79, pp. 1-11, 2019.

[22] F.O. Catak, and A.F. Mustacoglu, "Distributed denial of service attack detection using autoencoder and deep neural networks," Journal of Intelligent & Fuzzy Systems, vol. 37, no. 3, pp. 3969-3979, 2019.

[23] J. Chen, Y. Yang, K. Hu, H. Zheng, and Z. Wang, "DAD-MCNN: DDoS attack detection via multi-channel CNN," In Proceedings of the 11th International Conference on Machine Learning and Computing: ICMLC '19, pp. 484-488, 2019.

[24] S. Xia, M. Qiu, M. Liu, M. Zhong, and H. Zhao, "AI-enhanced automatic response system for resisting network threats," In M. Qiu (Ed.): SmartCom 2019, LNCS 11910, pp. 221–230, 2019.

[25] J. Clements, Y. Yangy, A.A. Sharma, H. Huy, and Y. Lao, "Rallying adversarial techniques against deep learning for network security, arXiv preprint arXiv:1903.11688v1, pp. 1-8, 2019

[26] S.K. Biswas, S. K, "Intrusion detection using machine learning: A comparison study," International Journal of Pure and Applied Mathematics, vol. 118, no. 19, pp. 101-114, 2018.

[27] Y. Mirsky, T. Doitshman, Y. Elovici, A. Shabtai, and A. Kitsune, "An ensemble of autoencoders for online network intrusion detection," arXiv preprint arXiv:1802.09089, pp. 1-15, 2018.

[28] Z. Wang, "The Applications of Deep Learning on Traffic Identification", BlackHat, 2015, https://www.blackhat.com/docs/us-15/materials/us-15-Wang-The-Applications-Of-Deep-Learning-On-Traffic-Identification-wp.pdf , accessed March 23, 2019.

[29] M. Lotfollahi, R. Shirali, M.J. Siavoshani, and M. Saberian, "Deep packet: A novel approach for encrypted traffic classification using deep learning," arXiv preprint arXiv:1709.02656, pp. 1-13, 2017.

[30] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang," End-to-end encrypted traffic classification with one-dimensional convolution neural networks, In Proceedings of the 2017 IEEE International Conference Intelligence and Security Informatics, Beijing, China, pp. 43–48, 2017.

[31] S. Rezaei, and X. Liu,"Deep learning for encrypted traffic classification: An overview, IEEE Comm Mag, vol. 57 (5), no. 5, pp. 76 – 81, 2019.

[32] G. Aceto. C. Domenico, M. Antonio, and P. Antonio, "Mobile encrypted traffic classification using deep learning," In Proc of the Network Traffic Measurement and Analysis Conference, pp. 1-8, 2018.

[33] J. Woodbridge, H.S. Anderson, A. Ahuja, and D. Grant," Predicting domain generation algorithms with long short-term memory networks," arXiv preprint arXiv:1611.00791, pp.1-13, 2016.

[34] D. Tran, H. Mac, V. Tong, H.A. Tran, and L.G. Nguyen, L.G. "A LSTM based framework for handling multiclass imbalance in DGA botnet detection," Neurocomputing, vol. 275, pp. 2401–2413, 2018.

[35] P. Lison, and V. Mavroeidis, "Automatic detection of malware-generated domains with recurrent neural models," arXiv preprint arXiv:1709.07102, pp. 1-12, 2017.

[36] C.D. McDermott, F. Majdani, and A. Petrovski, "Botnet detection in the internet of things using deep learning approaches," In Proceedings of IJCNN' 2018, pp. 1-8, 2018.

[37] C. Kolias, G. Kambourakis, A. Stavrou, and V. Voas, "DDoS in the IoT: Mirai and botnets," Computer, vol. 50, pp. 80-84, 2017.

[38] X.D. Hoang, and Q.C. Nguyen, "Botnet detection based on machine learning techniques using DNS query data," Future Internet, vol. 10, no. 5, pp. 1-11, 2018.

[39] G. Mi, Y. Gao, and Y. Tan, "Apply stacked auto-encoder to spam detection," In Proceedings of the International Conference in Swarm Intelligence, Beijing, China, pp. 3–15, 2015.

[40] G. Tzortzis, and A. Likas, "Deep belief networks for spam filtering," In Proceedings of the 19th IEEE International Conference on ICTAI, Patras, Greece, pp. 306–309, 2007.

[41] M. Alauthman, M. Almomani, M. Alweshah, W. Omoush, and K. Alieyan, " Machine learning for phishing detection and mitigation," In: Machine Learning for Computer and Cyber Security, B. Gupta, and Q.Z. Sheng, (eds), pp. 1-27, Taylor & Francis, 2019.

[42] E. Benavides, W. Fuertes, S. Sanchez, and M. Sanchez, M." Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review," in Á. Rocha and R. P. Pereira (eds.), Developments and Advances in Defense and Security, Smart Innovation, Systems and Technologies vol. 152, pp. 51-64, 2020.

[43] P. Yi, Y. Guan, F. Zou, Y. Yao, W. Wang, and T. Zhu, "Web phishing detection using a deep learning framework," Wirel. Commun. Mob. Comput, pp. 1–9, 2018.

[44] D. Aksu, Z. Turgut, S. Üstebay, and M.A. Aydin, "Phishing analysis of websites using classification techniques," pp. 251–258. Springer, Singapore, 2019.

[45] A. Tuor, S. Kaplan, B. Hutchinson, N. Nicholsand, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," arXiv preprint arXiv:1710.00811, pp. 1-9, 2017.

[46] T. Shibahara, K. Yamanishi, Y. Takata, D. Chiba, M. Akiyama, T. Yagi, Y. Ohsita, and M. Murata, "Malicious URL sequence detection using event de-noising convolutional neural network," In Proceedings of the IEEE ICC Conference, Paris, France, pp. 1–7, 2017.

[47] K. Yamanishi, "Detecting Drive-By Download Attacks from Proxy Log Information Using Convolutional Neural Network," Master Thesis, Osaka University: Osaka, Japan, pp. 1-32, 2017

[48] D. Ariu, G. Giacinto, and F. Roli, "Machine learning in computer forensics," In Proceedings of the 4th ACM workshop on Security and artificial intelligence, AISec 11, pages 99–103, 2011.

[49] F. Mitchell, "The use of artificial intelligence in digital forensics: an introduction," Digital Evidence and Electronic Signature Law Review, vol 7, pp. 35-41, 2010.

[50] N.M. Karie , V.R. Kebande, and H.S. Venter, "Diverging deep learning cognitive computing techniques into cyber forensics," Forensic Science International: Synergy vol.1, pp 61- 67, 2019.

[51] D. Maiorca, and B. Biggio, B. "Digital investigation of PDF files: Unveiling traces of embedded malware," IEEE Security and Privacy Magazine, vol. 17, no.1, pp. 63 – 71, 2019.

[52] W.C. Calhoun, and D. Coles, "Predicting the types of file fragments," Digital Investigation, vol. 5, pp. S14–S20, 2008.

[53] S. Axelsson, "The normalised compression distance as a file fragment classifier," Digital Investigation, vol. 7, no. 8, pp. S24–S31, 2010.

[54] N.L. Beebe, L.A. Maddox, L. Liu, and M. Sun, "Sceadan: Using concatenated n-gram vectors for improved file and data type classification," IEEE Transactions on Information Forensics and Security, vol. 8, no. 9, pp. 1519–1530, 2013.

[55] Q. Chen, Q. Liao, Z. Jiang, J. Fang, S. Yiu, G. Xi, et al, "File fragment classification using grayscale image conversion and deep learning," In Proceedings of the IEEE Symposium on Security and Privacy Workshops, pp. 140-147, 2018.

[56] K.S. Hoon K.C. Yeo, S. Azam, B. Shanmugam, and F. De Boer, "Critical review of machine learning approaches to apply big data analytics in DDoS forensics," In Proceedings of ICCCI'2018, Coimbatore, India, pp. 1-5, 2018.

[57] X. Wang, J. Zhang, A. Zhang, and J. Ren, J. "TKRD: Trusted kernel rootkit detection for cybersecurity of VMs based on machine learning and memory forensic analysis," Mathematical Biosciences and Engineering, vol. 16, no.4, pp. 2650–2667, 2019.

[58] D. Kirat, J. Jang, and M.P. Stoecklin, "DeepLocker Concealing Targeted Attacks with AI Locksmithing," IBM Presentation, BlackHat USA 2018, https://i.blackhat.com/us-18/Thu-August-9/us-18-Kirat-DeepLocker-Concealing-Targeted-Attacks-with-AI-Locksmithing.pdf, accessed May 8, 2019.

[59] N. Ye, C. Newman, and T. Farley, "A System-fault-risk framework for cyber-attack classification," Information Knowledge Systems Management vol.5, pp. 135–151, 2005.

[60] "2018 Cybersecurity Guide: Hackers and defenders harness design and machine learning," HP 2018 Report, pp. 1-22. https://www8.hp.com/h20195/v2/GetPDF.aspx/4AA7-2519ENW.pdf, accessed March 4, 2019.

[61] J. Song, C, Cho, and Y. Won, "Analysis of operating system identification via fingerprinting and machine learning," Computers and Electrical Engineering, vol. 78, pp. 1-10, 2019.

[62] T. Gu, B. Dolan-Gavitt, and S. Garg, "Badnets: Identifying vulnerabilities in the machine learning model supply chain," arXiv preprint arXiv:1708.06733, pp. 1-13, 2017.

[63] "Fortinet predicts highly destructive and self-learning "Swarm" cyberattacks in 2018," Fortinet Press release, https://www.fortinet.com/corporate/about-us/newsroom/press-releases/2017/predicts-self-learning-swarm-cyberattacks-2018.html, accessed March 5, 2019.

[64] Y.E. Sagduyu, Y. Shi, and T. Erpek, "IoT network security from the perspective of adversarial deep learning," arXiv preprint arXiv:1906.00076, pp. 1-9, 2019.

[65] Y. Shi, Y.E. Sagduyu, T. Erpek, K. Davaslioglu, Z. Lu, and J. Li, "Adversarial deep learning for cognitive radio security: Jamming attack and defense strategies," IEEE ICC Workshop, pp. 1-6, 2018.

[66] T. Erpek, Y.E. Sagduyu, and Y. Shi "Deep learning for launching and mitigating wireless jamming attacks," IEEE Trans. Cogn. Comm. & Networking vol.5, no.1, pp. 2-14, 2019.

[67] C. Cruz-Perez, O. Starostenko, F. Uceda-Ponga, V. Alarcon-Aquino, and L. Reyes-Cabrera, "Breaking reCAPTCHAs with unpredictable collapse: Heuristic character sSegmentation and recognition," Proceedings of MCPR'12, Huatulco, Mexico, pp. 155—165, 2012.

[68] S.Sivakorn, J. Polakis, J, and A.D. Keromytis, "I'm not a human: Breaking the Google reCAPTCHA," Black Hat USA 2016. https://www.blackhat.com/docs/asia-16/materials/asia-16-Sivakorn-Im-Not-a-Human-Breaking-the-Google-reCAPTCHA-wp.pdf, accessed April 6, 2019.

[69] N. Yu, and K. Darling, "A low-cost approach to crack python CAPTCHAs using AI-based chosen-plaintext attack," Applied Sciences, vol. 9, pp. 1-17, 2019.

[70] H.S. Anderson, J. Woodbridge, and B. Filar, "DeepDGA: Adversarially-tuned domain generation and detection,' In Proceedings of ACM Workshop on Artificial Intelligence and Security, Vienna, Austria, pp. 13–21, 2016.

[71] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al, "Generative adversarial nets," In Advances in Neural Information Processing Systems; MIT Press: Cambridge, MA, pp. 2672–2680, 2014.

[72] S. Palka, and D. McCoy, "Fuzzing E-mail Filters with Generative Grammars and N-Gram Analysis," Usenix WOOT, pp. 1-10, 2015.

[73] A. Singh, and V. Thaware, "Wire me through machine learning," Black Hat USA 2017, Las Vegas, https://www.blackhat.com/docs/us-17/wednesday/us-17-Singh-Wire-Me-Through-Machine-Learning.pdf, accessed July 16, 2018.

[74] A. Giaretta, and N. Dragoni, "Community targeted spam: A middle ground between general spam and spear phishing through natural language generation," arXiv preprint arXiv:1708.07342v2, pp.1-8, 2018.

[75] J. Jung, C. Jeon, M. Wolotsky, I. Yun, and T. Kim,"AVPASS: automatically bypassing android malware detection system," Black Hat USA 2017, https://taesoo.kim/pubs/2017/jung:avpass-slides.pdf, accessed July 16, 2018.

[76] D. Petro, D, and B. Morris, "Weaponizing machine learning: Humanity was overrated anyway," Presentation at DEF CON 25, 2017, Las Vegas http://hwcdn.libsyn.com/p/8/8/1/881758917f6d6a03/DEFCON-25-Report.pdf?c_id=16503562&cs_id=16503562&expiration=1573591467&hwt=54e7017ba652d406843948d4cd0aca7a, accessed July 16, 2018.

[77] W. Hu, and Y. Tan, "Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN," arXiv preprint arXiv: 1702.05983, pp. 1-7, 2017.

[78] B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, "PassGAN - a deep learning approach for password guessing," arXiv preprint arXiv: 1709.00440, pp. 1-12, 2018.

[79] W. Zhou, W. Chai, and H. Ma, "Deep learning based attack on social authentication system," In proc of ITNEC'2019, pp. 982-986, 2019.

[80] D. Das, A. Golder, J. Danial, S. Ghosh, A. Raychowdhury, and S. Sen, "X-DeepSCA: Cross-device deep learning side channel attack," Proceedings of the 56th ACM/IEEE Design Automation Conference, Las Vegas, NV, pp. 1-6, 2019.

[81] A. Golder, D. Das, J. Danial, S. Ghosh, S. Sen, and A. Raychowdhury, "Practical approaches toward deep-learning-based cross-device power side-channel attack," IEEE Transactions on Very Large-Scale Integration Systems, vol. 27, no. 12, pp. 2720 – 2733, 2019.

[82] G. Hospodar, B. Gierlichs, and E. De Mulder, "Machine learning in side-channel analysis: A first study," Journal of Cryptographic Engineering, vol. 1, no. 4, pp. 293–302, 2011.

[83] L. Lerman, R. Poussier, O. Markowitch, and F.X. Standaert, "Template attacks versus machine learning revisited and the curse of dimensionality in side-channel analysis: Extended version," Journal of Cryptographic Engineering, vol. 8, no.4, pp. 301–313, 2018.