International Journal of Scientific Engineering and Science
Volume 5, Issue 2, pp. 35-40, 2021. ISSN (Online): 2456-7361

Defense by Artificial Intelligence in Cyber Attack

Ahood Hameed S. ALThobyti[1], Samah Mohammed S. ALHusayni[2], Sabah M. Alzahrani[3]
[1, 2, 3]College of Computers and Information Technology, Taif University, Taif21944, Saudi Arabia
Email address: ahood.a1 @ tvtc.gov.sa, sama7muhamad @ gmail.com, sabahalzahrani2018 @ gmail.com

Abstract— *In light of the technical developments and the amount of data used and the sources available in the network and the various devices that expose it to a cyber attack, it has become necessary to use advanced technologies that have the ability to search for any threat and discovered it and prevent it from causing damage. The attack, which means that it may be used negatively to achieve damage, which makes it imperative for us to update and develop the methods used and databases. In this paper, the study was focused on the most important information about artificial intelligence, cyber attack and the advantages of artificial intelligence AI-Based approach for defending against cyberspace attacks and Defense Strategy For Artificial Intelligence Adversarial attack, and finally AI applications of security threats in cyberattack. It is a narrative review summarizes the findings of some existing literature for readers from outside the field who require a rapid and general summary in the role of artificial intelligence in cyber attacks.*

I. Introduction

Latest rapid advances in the applications of artificial intelligence technology (AI) in different fields. Based on high accuracy, great availability , and expertise in the range like medicine component checking [1], mind cycle rebuilding [2], dynamic dynamometric records analyzes[3]and DNA gene effect study[4]artificially intelligent technology has been seen in picture analysis, object tracking, voice commands, translation software and more highly developed areas [5]. Whereas Szegedy et al. [6] indicated that neural nets are susceptible to antagonistic threats, studies have increasingly being a popular spot in artificial intelligence, and investigators have continuously introduced new techniques of the attacks and protection strategies[5]. Around one million malware codes exist daily and are applied for cyber attacks. So there is an increasing awareness that there is a basic limit to a way of reacting to such attacks at each stage (network, port) [7]. Nonetheless, artificial intelligence systems are susceptible to attacks that restrict the deployment in main cyber security sectors of artificial intelligence technology, So must be more evolving AI in enhancing the robustness of AI platform against adversarial attacks [5]. Offenses occur at smarter methods so different protection techniques be there used to counter these threats [7]. Nevertheless, more and more people believe that there is an important limit to the individual response to any smart violation attack [7]. The cyber-threat intelligence analysis technology, therefore, focuses on an evaluation of the attacking community, identification of the attack pattern and the gathering of information in the judgment call-making process through the collection, and study of a wide variety of cyber attack data [7]. Most AI solutions for identification issues are based on a method of "brute force" in finding solutions, that depend on computing resources for a wide variety of matching choices. Nonetheless, the brute-force approach applied in Watson and other profound learning platforms presently in use rely on sheer computing resources to create effective AIs with a wide variety of training stimuli to improve trust AI [8].

The range of this paper is Part 2 the advantages of artificial intelligence for defending against cyberspace attacks; Part 3 Protection strategies AI towards offenses; Part 4 Defense Strategy for Artificial Intelligence Adversarial attack; Finally, AI applications of security threats in a cyberattack.

II. Beneficial Serves of AI

AI is used to propel careful capabilities in the area of cyber security. Given its incredible robotization and information investigation abilities, AI can be utilized to dissect a lot of information with effectiveness, precision, plus rapidity. A framework of AI is able to exploit the knowledge that he define and comprehend the last dangers to distinguish comparative assaults later on, regardless of whether their examples change. Without a doubt, man-made consciousness has a few points of interest with regards to cyber security at the next perspectives:

1- AI finds fresh and advanced Modifies to the assault adaptability: traditional innovation is centered on last and depends vigorously at recognized assailants and assaults, exit space to vulnerable sides at identifying surprising occasions in new assaults. The restrictions of old resistance innovation are currently being tended to through canny innovation.

2- AI can deal with the volume of information: AI can upgrade organize security by creating self-ruling security frameworks to identify assaults and react to breaks. The volume of security alarms that show up day by day can be overpowering for security gatherings. Consequently distinguishing and reacting to dangers has assisted with decreasing crafted by organize security specialists and can help with recognizing dangers more viably than different strategies.

3-The protection framework of AI would adapt to react smarter at hazards after a while: AI distinguishes dangers dependent on application conduct and an entire system's action. After some time, AI security framework finds out passing and conduct for the standard system, and produces an ordinary pattern for that point, each and every perversion from the standard can be detected from detect attacks. Man-made intelligence procedures appear to be a best in the class region of research that improves the safety efforts for the internet[9].

4. Protection strategies AI towards offenses: In recent times scientists have suggested many strategies for detecting or categorizing ransomware, net interference, phishing, and spam attacks utilizing AI tools, combating Enhanced Persistence Threat (APT), and for defining domain creation algorithm

(DGA) domain. We categorize those writings under the fourth key categories for this segment: malicious identity; invasion detections for the net; phishing and Spam identification; and those that abused APT and DGAs identification. Fig. 1 demonstrates the key fields of protection AI use

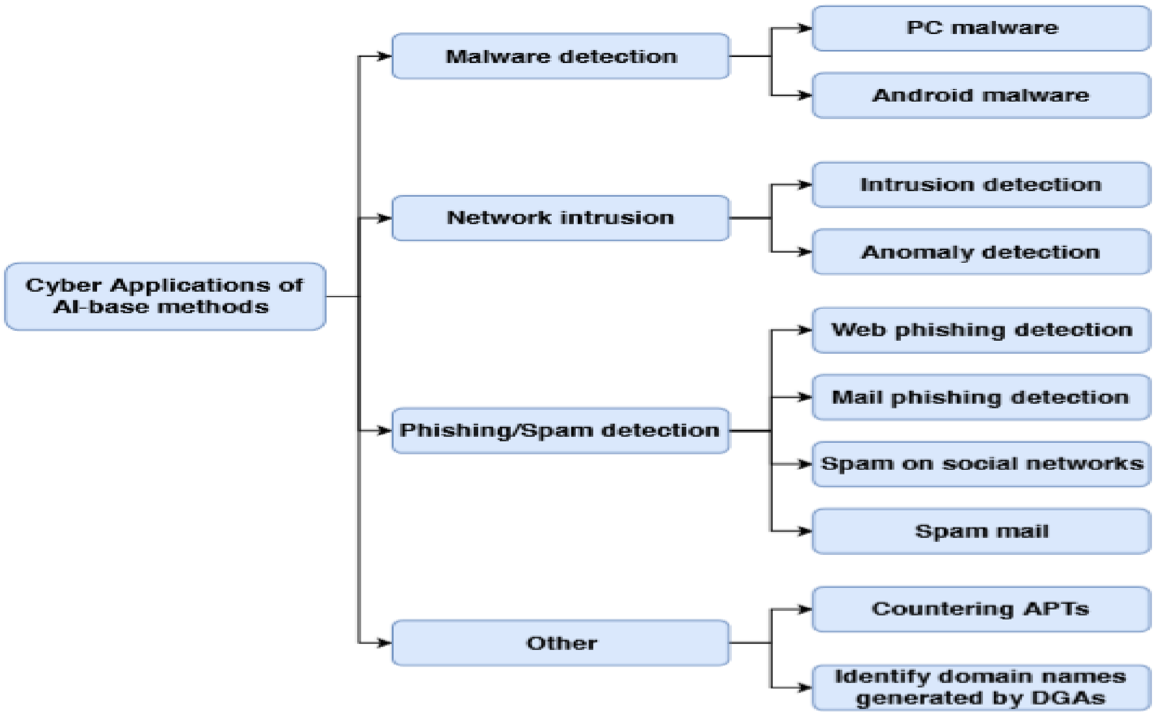

Figure 1. Key sectors in defense solutions use AI technologies [9]

A. Identify Malicious Programs

The malicious program is an aggregate concept to some kinds for pernicious programming, for example, infections, worms, trojan ponies, misuses, botnet, retroviruses, and today, malware is a famous technique for digital assault. Malicious program effect in an advanced community is colossal thus a lot of work of receiving AI systems was performed to forestall and relieve malicious programs. The latest & imperative commitments use insight for malware identification and anticipation—depicted so shadows. In [10]the creators received Machine learning to make an operational system for equipment helped Malicious programs recognition dependent on virtual memory get to designs. The proposed strategy utilized calculated relapse, a help vector machine, and an arbitrary woods classify and perform on the benchmark set for the investigations. The creators revealed that the system has a genuine confident pace of 99% with an under 5% bogus constructive degree. Then, the researchers in [11]introduced a structure for characterizing and distinguishing noxious programming utilizing information mining and Machine learning classify. In that effort, both marks built and inconsistency built highlights are broke down for recognition. The observational outcome demonstrated that the proposed classical is capacity with a little bogus alert degree and a great location degree. Afterward. [12] constructed a profound knowledge design for smart Malicious program identification. to identify obscure malware. The creator asserted that a heterogeneous profound learning system could improve the general execution in malware location contrasted and customary shallow learning techniques and profound learning strategies. An ongoing pattern of research in malware recognition concentrated on versatile malware when all is said in done and android malware specifically. AI, alongside profound learning, was a significant achievement around there. In [13], a profound serpentine nervous system (CNN) was embraced to recognize Malicious programs. The crude operation code grouping from a dismantled software is utilized to arrange Malicious programs. The creators in[14]used a help vector engine (SVM) and the greatest can't sign consents from the entirety of the authorization information to recognize considerate and vindictive applications. At [15] the creators introduced original Machine Learning calculations, in particular, turn timberland, for Malicious programs character. A technical communication system (ANN) and the crude groupings of API strategy song was used at [16] toward identifying Android Malicious programs[17] presented a crossbreed classical dependent on profound Auto coding (DAE) and a serpentine nervous system (CNN) to raise the precision and capacity of huge scope android malicious program recognition. Additional exploration course that pulled in the consideration of researchers was the utilization of biologically aroused strategies for the malicious program category. Those procedures was for the most part utilized to include advancement and enhancing the parameter for the classifiers[9].

B. Interruption Detection

An interruption recognition framework (IDS) is a framework that should shield the framework from conceivable incidents, violations, or up and coming dangers. Computers based-intelligence based-procedure are fitting for creating IDS, and beat different systems due to their flexibility, versatility, fast counts, and brisk learning. Subsequently, numerous specialists considered keen strategies to progress the exhibition of IDS. The emphasis was on creating advanced highlights and enlightening the class to lessen the bogus alerts. About ongoing prominent investigations were recorded as follows. [18] joined a help route engine (SVM) and an extraordinary LM with modifying kimplies as a model for IDS. In the meantime, Kabir et al. The proposed strategy was approved through the KDD'99 Cup dataset and acquired a sensible exhibition as far as precision and efficiency. The assessment was led by utilizing genuine system traffic from a college and acquired a precision of 96.53% and a bogus alert of 0.56%. In light of the outcomes acquired, the calculation demonstrated great execution, [19] presented a learn classical for a quick learning system (FLN) in light of PSO named PSO-FLN. In the ongoing investigation by Chen. [20], a staggered versatile joined interruption discovery technique consolidating white rundown innovation and AI was introduced. The white rundown was utilized to filter the correspondence, and the AI classical was utilized to distinguish anomalous correspondence. In object, the versatile PSO calculation and the swarm calculation was utilized to improve the constraints for the AI classical. The technique was tried on KDD'99 Cup, Gas Pipeline, &Recent data sets. The exact outcome indicated that the proposal classical is efficient with different assault kinds. In [21], in spite of the fact that the precision degree was high, 94.53[9].

C. SPAM Detection & Phishing

Phish assault was a digital assault that endeavors to take client's character or monetary certifications. New, phish assaults were one of the greatest threatening dangers on the network. Different tale savvy methods were utilized to adapt to those issues. The creators in [22] offered an enemy of phish technique, which used a few distinctive ML calculations and nineteen highlights to recognize phishing sites from authentic ones. The creators guaranteed classical accomplished a 99.39% genuine optimistic degree. Another methodology via Feng. [23]useful a neuronal system for identify the phish spots by embracing the Monte Carlo calculation and hazard minimize guideline. Observational outcomes demonstrated that their classical arrived at a 97.71% exact discovery degree and a 1.7% bogus alert rate. An ongoing report directed by [24] presented a continuous enemy of phishing framework, which used seven distinctive classification calculations and characteristic language handling (NLP) based highlights. As per the creators, their methodology got a promising outcome with a 97.98% precision rate. The creators announced that their methodology arrived at a 98% precision degree. The phrasing SPAM alludes to spontaneous mass E-mail (garbage E-mail). Spam E-mail can prompt safety matters and the wrong substance. To conquer the downsides of these digital dangers, as of late researchers applied different novel, smart procedures to manufacture spam refinery frameworks. [25] joined help engine path and Simple Bayes to build up a spam refinery framework. The proposed framework was assessed by the DATAMALL dataset and got an extraordinary spam-location exactness. The creators in [26] structured a spam order system utilizing modification research to upgrade the spam classifier. In their effort, the progression scan was used for highlight removal, & the SVM was utilized for classifying. The PSO calculation was received to include determination, and the SVM & choice sapling for classifying. The creators asserted that the propose framework was effective.[27] gave a half breed way to deal with distinguishing the spam profiles on Twitter utilizing web-based social networking investigation and bio-propelled figuring. particular, they used a modified k_implies incorporated duty calculation (LFA) through disorganized charts to recognize spam sender. A sum of 14,235 files were utilized to assess the presentation of the technique. As per the analyses, the proposed framework got astounding outcomes as far as exactness, accuracy, and review [9].

III. DEFENSE STRATEGY FOR ARTIFICIAL INTELLIGENCE ADVERSARIAL ATTACK

The researchers classified opponent attack defense tactics based on three categories: data, change, model alteration, and support resources.

A. Data Modification

These techniques include altering the training data set or adjusting data in the test process, involving adverse testing, concealed gradient rates, preventing transferability, compression of data, and the pseudo-random of data [5]. The adversarial specimens were initially injected and its tags adjusted to make the template robust against the opponent [6] and t is impractical to introduce in an opponent training all unidentified attack specimens leading to reduced adversary learning, Levels hiding: defenses for level-based attacks [28] and attacks by the opposing of approach crafting and this approach covers opponents' knowledge about template gradation[5],Preventing the transferability Although the transferability property continues even if the classifier has a diverse structure or is focused on a disjoint sample, the secret to avoiding a black-box intrusion is to block the transferability of oppositional samples and This approach has the benefit of labeling the disturbance data as a null item, instead of identifying it as the initial tag So this approach is probably one of the most powerful mechanisms of security[5], Data pressured. It is an effective method against some attacks, not strong ones Likewise, the compressed approach applied by Display Compression Technological advancements (DCT) in combating uniform interruption attacks [29] has also tended to be inadequate and the main drawback of these data compression-based protection mechanisms is that a huge amount of compression can result in a reduction in the performance of the main image description, whereas a tiny percentage of compression is always inadequate to eliminate the effect of the trouble [30], and Data Randomized Xie et al.

[31] have shown that the process of randomized resizing oppositional data sets can minimize oppositional dataset efficacy, Likewise, adding such randomized textures to the oppositional datasets may also decrease their frustration with the network structure[5].

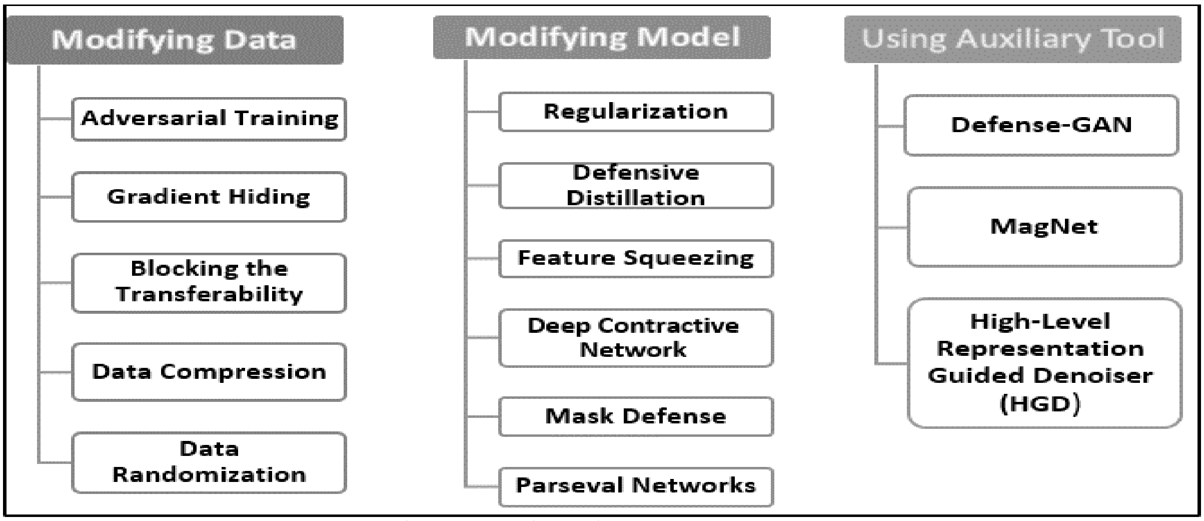

Figure 2. Defense Strategy for Artificial Intelligence Adversarial Attack

B. Update Model

The pattern recognition model, such as regularization, compression of functions, strong contractive networks, and mask safety, can be modified. Regularization This approach proposes at enhancing the generalization potential of the destination sample by inserting standard terms defined as penalty limits to the cost function and making the template well adaptable to predict attacks on an undefined sample and The papers [32] employed the technique of regularization to enhance the algorithm's robustness and obtained positive results in combating oppositional attacks [6], Defensive Distillation The initial distilling technology works at small-scale compressing and maintaining an initial precision of the big-scale model, thus defense distilling should not alter the model level and provides a model with an easier output surface and a lower susceptibility to disruption that boost model solidity and In black-box attacks, the successful security distillation cannot be ensured So, Papernot et al. [33] instead of it suggested extendable distillation protection technology [5], Function Squash, which is primarily designed to limit the difficulty and interference in the data representation by low sensitivity. While this strategy can avoid adverse attacks efficiently, it decreases the precision of the classification of actual samples as well [5], Wide Agreement Nets In.[34] implemented a sort on broad compressed nets that utilizes an automated codec for noise cancellation to minimize noise adversity and has been shown to have some protective impact from attacks as L-BGFS [6], Mask Defense: The cover surface educated the real photos and associated adverse reaction samples and encoded deviations from the existing network design structure's performance characteristics and the main mass in the extra layer is commonly believed to be equivalent to the network's most sensitive function. So these attributes are obscured in the final ranking by requiring the extra layers to weight zero primarily so It will avoid the variance in identification reports produced by opponent samples[5]. Parseval NetIn. [35] suggested the Parseval network like protection solution to adversary attacks. This network regularizes the hierarchy by regulating the worldwide Lipschitz network static. In the theory of the possibility of seeing the network as a blend of features on each layer, it is possible to maintain a tiny Lipschitz static for such roles with strong positive ions for minor input disruption [5].

C. Auxiliary Tool

This technique applies to the use of external methods for the neural network template as a supplementary method. While defense GAN has shown its effectiveness in the protection of attacks, it relies on GAN's expressiveness and relational efficiency Moreover, the practice of the GAN is difficult, i.e. the reliability of defense GANs obviously decreases without sufficient training. MagNetMeng et al. [36] presented a MagNet system that reading the final classifier results as a blackbox without accessing or changing any internal layer details and In order to distinguish permissible and harmful samples, MagNet deploys a detector which tests the range from the reference sample and refuse the pattern as the interval exceeds the level. It also requires a leader to turn the opponent's pattern into a clear legal sample using an auto codec [5], And High- Level Representation Guided Denoiser (HGD) This issue can be successfully solved using a failure function in contrast with the data from the destination model with the clear picture and the rusty picture, which is unique from the regular de noise detection system like the pixel level reconstructive loss function that has an issue with error amplification, HGD has been implemented by Liao et al. [37] to develop a reliable destination model for white and black box attacks. A further benefit of HGD practice can be training on a fairly limited dataset and used for protecting models other than the one which will lead it. The author has suggested three various HGD methodologies [5].

IV. AI Applications of Security Threats in Cyber Attack

Top U.S. security officials say artificial intelligence (AI) can change information security and cyber warfare. CSIS which is a politics study body has documented major cyber-attacks on governments, defense, and wide-technology companies, or financial crimes in large amounts of a Million dollars [38]. Whilst a lot of factors may trigger the escalating rise in cyber attacks, AI ultimately rules [38]. Also, AI establishes more efficient, readily available mechanisms for attacks [38]. While it is hard to gauge the technologies employed, the large availability of AI cyber attack and coding analysis already indicates that a malicious AI program is a very common occurrence. AI offers a clear advantage in web penetration and security [38]. Cyber-dependent countries are, as the data shows, exposed to political manipulation and these cyber-attacks are not today [38]. No wonder AI cyber security work is growing fast, one research article offers instructions for designing malicious AI applications, for example, The study reveals practical ways of creating malicious machine learning programs that damage people [38]. The key argument is, therefore, that an attacker needs no prior knowledge of his goal so that a network under attack can penetrate successfully where Cyberspace enables us to interact, exchange, study and share knowledge globally [38]. Nonetheless, every year cyber-attacks occur and are rising where Information violations have increased complaints from governments, private claimants, and the general opinion [38]. Research and data on a globally are must as protected as the algorithms and network systems by which they are secured [38]. There is therefore considerable uncertainty regarding protection at any level, the armament of cyberspace and subsequent cyber warfare are creating a world where defense models are useless where the capacity to hack an opponent vehicle, an aircraft, a nuclear reactor, and a communication infrastructure [38].

Closing with Outlook Direction:

For this article, we discuss the importance of AI in cyber attack and AI-based approaches for defending against cyberspace attacks which are malware identification, interruption detection, and phishing and SPAM detection .we also explore defense strategy for artificial intelligence adversarial attack which are data modification, update the model and using the auxiliary tool. Finally, we discuss AI applications of security threats in the cyber attack. We look forward to having thoughtful plans in applying artificial intelligence and enhancing the durability of its platforms in the face of attacks.

References

[1] Ma, J., et al., Deep neural nets as a method for quantitative structure–activity relationships. Journal of chemical information and modeling, 2015. 55(2): p. 263-274.

[2] Helmstaedter, M., et al., Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature, 2013. 500(7461): p. 168-174.

[3] Ciodaro, T., et al. Online particle detection with neural networks based on topological calorimetry information. in Journal of physics: conference series. 2012. IOP Publishing.

[4] Xiong, H.Y., et al., The human splicing code reveals new insights into the genetic determinants of disease. Science, 2015. 347(6218): p. 1254806.

[5] Qiu, S., et al., Review of artificial intelligence adversarial attack and defense technologies. Applied Sciences, 2019. 9(5): p. 909.

[6] Szegedy, C., et al., Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199, 2013.

[7] Son, K.-h., B.-i. Kim, and T.-j. Lee, Cyber-attack group analysis method based on association of cyber-attack information. KSII Transactions on Internet & Information Systems, 2020. 14(1).

[8] Kelley, T. and K. Dickerson, A Review of Artificial Intelligence (AI) Algorithms for Sound Classification: Implications for Human-Robot Interaction (HRI). 2020, CCDC Army Research Laboratory Adelphi United States.

[9] Truong, T.C., Q.B. Diep, and I. Zelinka, Artificial Intelligence in the Cyber Domain: Offense and Defense. Symmetry, 2020. 12(3): p. 410.

[10] Xu, Z., et al. Malware detection using machine learning based analysis of virtual memory access patterns. in Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017. 2017. IEEE.

[11] Chowdhury, M., A. Rahman, and R. Islam. Malware analysis and detection using data mining and machine learning classification. in International Conference on Applications and Techniques in Cyber Security and Intelligence. 2017. Springer.

[12] Ye, Y., et al., DeepAM: a heterogeneous deep learning framework for intelligent malware detection. Knowledge and Information Systems, 2018. 54(2): p. 265-285.

[13] McLaughlin, N., et al. Deep android malware detection. in Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy. 2017.

[14] Li, J., et al., Significant permission identification for machine-learning-based android malware detection. IEEE Transactions on Industrial Informatics, 2018. 14(7): p. 3216-3225.

[15] Zhu, H.-J., et al., DroidDet: effective and robust detection of android malware using static analysis along with rotation forest model. Neurocomputing, 2018. 272: p. 638-646.

[16] Karbab, E.B., et al., MalDozer: Automatic framework for android malware detection using deep learning. Digital Investigation, 2018. 24: p. S48-S59.

[17] Wang, W., M. Zhao, and J. Wang, Effective android malware detection with a hybrid model based on deep autoencoder and convolutional neural network. Journal of Ambient Intelligence and Humanized Computing, 2019. 10(8): p. 3035-3043.

[18] Al-Yaseen, W.L., Z.A. Othman, and M.Z.A. Nazri, Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Systems with Applications, 2017. 67: p. 296-303.

[19] Ali, M.H., et al., A new intrusion detection system based on fast learning network and particle swarm optimization. IEEE Access, 2018. 6: p. 20255-20261.

[20] Chen, W., et al., Multi-level adaptive coupled method for industrial control networks safety based on machine learning. Safety Science, 2019. 120: p. 268-275.

[21] Garg, S. and S. Batra, Fuzzified cuckoo based clustering technique for network anomaly detection. Computers & Electrical Engineering, 2018. 71: p. 798-817.

[22] Jain, A.K. and B.B. Gupta, Towards detection of phishing websites on client-side using machine learning based approach. Telecommunication Systems, 2018. 68(4): p. 687-700.

[23] Feng, F., et al., The application of a novel neural network in the detection of phishing websites. Journal of Ambient Intelligence and Humanized Computing, 2018: p. 1-15.

[24] Sahingoz, O.K., et al., Machine learning based phishing detection from URLs. Expert Systems with Applications, 2019. 117: p. 345-357.

[25] Feng, W., et al. A support vector machine based naive Bayes algorithm for spam filtering. in 2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC). 2016. IEEE.

[26] Kumaresan, T. and C. Palanisamy, E-mail spam classification using S-cuckoo search and support vector machine. International Journal of Bio-Inspired Computation, 2017. 9(3): p. 142-156.

[27] Aswani, R., A.K. Kar, and P.V. Ilavarasan, Detection of spammers in twitter marketing: a hybrid approach using social media analytics and bio inspired computing. Information Systems Frontiers, 2018. 20(3): p. 515-530.

[28] Tramèr, F., et al., Ensemble adversarial training: Attacks and defenses. arXiv preprint arXiv:1705.07204, 2017.

[29] Akhtar, N., J. Liu, and A. Mian, Defense against Universal Adversarial Perturbations. arXiv 2017. arXiv preprint arXiv:1711.05929.

[30] Das, N., et al., Keeping the bad guys out: Protecting and vaccinating deep learning with jpeg compression. arXiv preprint arXiv:1705.02900, 2017.

[31] Xie, C., et al., Adversarial examples for semantic segmentation and object detection. arXiv 2017. arXiv preprint arXiv:1703.08603.

[32] Rozsa, A., M. Gunther, and T.E. Boult, Towards robust deep neural networks with BANG. arXiv preprint arXiv:1612.00138, 2016.

[33] Papernot, N. and P. McDaniel, Extending defensive distillation. arXiv preprint arXiv:1705.05264, 2017.

[34] Gu, S. and L. Rigazio, Towards deep neural network architectures robust to adversarial examples. arXiv preprint arXiv:1412.5068, 2014.

[35] Cisse, M., et al., Houdini: Fooling deep structured prediction models. arXiv preprint arXiv:1707.05373, 2017.

[36] Meng, D. and H. Chen. Magnet: a two-pronged defense against adversarial examples. in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017.

[37] Liao, F., et al., Defense against Adversarial Attacks Using High-Level Representation Guided Denoiser. arXiv e-prints, page. arXiv preprint arXiv:1712.02976, 2017.

[38] Haney, B.S., Applied Artificial Intelligence in Modern Warfare and National Security Policy. Hastings Sci. & Tech. LJ, 2020. 11: p. 61.