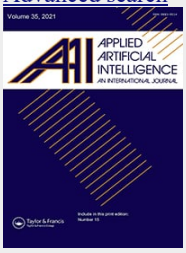

Applied Artificial Intelligence
An International Journal
Latest Articles
Review

The Emerging Threat of Ai-driven Cyber Attacks: A Review

Blessing Guembe
,
Ambrose Azeta
,
Sanjay Misra
,
Victor Chukwudi Osamor
,
Luis Fernandez-Sanz
&
Vera Pospelova

Received 05 Nov 2021, Accepted 28 Jan 2022, Published online: 04 Mar 2022

- https://doi.org/10.1080/08839514.2022.2037254

ABSTRACT

Cyberattacks are becoming more sophisticated and ubiquitous. Cybercriminals are inevitably adopting Artificial Intelligence (AI) techniques to evade the cyberspace and cause greater damages without being noticed. Researchers in cybersecurity domain have not researched the concept behind AI-powered cyberattacks enough to understand the level of sophistication this type of attack possesses. This paper aims to investigate the emerging threat of AI-powered

cyberattacks and provide insights into malicious used of AI in cyberattacks. The study was performed through a three-step process by selecting only articles based on quality, exclusion, and inclusion criteria that focus on AI-driven cyberattacks. Searches in ACM, arXiv Blackhat, Scopus, Springer, MDPI, IEEE Xplore and other sources were executed to retrieve relevant articles. Out of the 936 papers that met our search criteria, a total of 46 articles were finally selected for this study. The result shows that 56% of the AI-Driven cyberattack technique identified was demonstrated in the access and penetration phase, 12% was demonstrated in exploitation, and command and control phase, respectively; 11% was demonstrated in the reconnaissance phase; 9% was demonstrated in the delivery phase of the cybersecurity kill chain. The findings in this study shows that existing cyber defence infrastructures will become inadequate to address the increasing speed, and complex decision logic of AI-driven attacks. Hence, organizations need to invest in AI cybersecurity infrastructures to combat these emerging threats.

Introduction

Cyberattacks are pervasive and are often regarded as one of the most tactically significant risks confronting the world today (Dixon and Eagan 2019). Cybercrimes can engender disastrous financial losses and affect individuals and organizations as well. It is estimated that a data breach costs the United States around 8.19 million Dollars and 3.9 Million Dollars on average, and the annual effect on the global economy from cyberattack is approximately 400 Billion Dollars (Fischer 2016; Kirat, Jang, and Stoecklin 2018). A Cyberattack is the intentional exploitation of computer systems, networks, and businesses. With increasingly sophisticated cybersecurity attacks, cybersecurity specialists are becoming incapable of addressing what has become the most significant threat climate ever before (Chakkaravarthy et al., 2018).

The sophistication of cyberattack techniques poses an existential danger to enterprises, essential services, and organization infrastructures, with the power to interrupt corporate operations, wipe away critical data, and create reputational damage. Today's current wave of attacks outwits and outpaces humans and even includes Artificial Intelligence (AI). Cybercriminals will be able to direct targeted attacks at unprecedented speed and scale while avoiding traditional, rule-based detection measures thanks to what's known as "offensive AI" (DarkTrace, 2021). A new generation of cybercriminals has emerged, one that is both subtle and secretive, which will influence the future of cybersecurity. The new generation of cyber threats will be smarter and capable of acting independently with the help of AI. Future cyberattack methods will be able to be aware of their surroundings and make informed decisions based on the target environment. The potential of AI to learn and adapt will usher in a new era of scalable, custom-made, and human-like assaults (Thanh and Zelinka 2019).

ent utilization of AI by embedding some hypothetical concepts within digital, physical and political security domains. Researchers have established a few concepts that showed the potential of an automatic exploit generation in state-of-the-art applications. Malicious actors are utilizing fuzzy models to develop a next-generation malware capable of learning from its environment, continuously updating itself with new variants, and infecting vulnerable and sensitive computer infrastructures without being noticed (Kaloudi and Li 2020). Malicious actors can utilize these concepts to deploy a new type of sophisticated and stealthy cyber weaponries.

References

1. DarkTrace. 2021. The Next Paradigm Shift AI-Driven Cyber-Attacks. DarkTrace Research White Paper. https://www.oixio.ee/sites/default/files/the_next_paradigm_shift_-_ai_driven_cyber_attacks.pdf (accessed June 9, 2021). [Google Scholar]
2. Anderson, H. S., J. Woodbridge, and B. Filar. 2016. Deepdga: Adversarially-tuned domain generation and detection. In Proceedings of the ACM Workshop on Artificial Intelligence and Security, Vienna, Austria, 13–21. [Crossref], [Google Scholar]
3. Babuta, A., M. Oswald, and A. Janjeva. 2020. Artificial Intelligence and UK National Security Policy Considerations. Royal United Services Institute Occasional Paper. [Google Scholar]
4. Bahnsen, A. C., I. Torroledo, L. Camacho, and S. Villegas. 2018. DeepPhish: Simulating malicious AI. In APWG Symposium on Electronic Crime Research, London, United Kingdom, 1–8. [Google Scholar]
5. Bilal, M., A. Gani, M. Lali, M. Marjani, and N. Malik. 2019. Social profiling: A review, taxonomy, and challenges. *Cyberpsychology, Behavior and Social Networking* 22 (7):433–50. doi:https://doi.org/10.1089/cyber.2018.0670. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]
6. Bocetta, S. 2020. Has an AI cyberattack happened yet? https://www.infoq.com/articles/ai-cyberattacks/ (accessed December 9,2020). [Google Scholar]
7. Brundage, M., S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T . Zeitzoff, B. Filar, et al. 2018. *The malicious use of artificial intelligence: forecasting, prevention, and mitigation*. Oxford: Future of Humanity Institute. [Google Scholar]
8. Bursztein, E., J. Aigrain, A. Moscicki, and J. C. Mitchell. 2014. The end is nigh: generic solving of text-based CAPTCHAs. 8th Usenix workshop on Offensive Technologies WOOT '14, San Diego, CA, USA. [Google Scholar]
9. Cabaj, K., Z. Kotulski, B. Księżopolski, and W. Mazurczyk. 2018. Cybersecurity: trends, issues, and challenges. *EURASIP Journal On Information Security*. doi:https://doi.org/10.1186/s13635-018-0080-0. [Crossref], [Web of Science ®], [Google Scholar]
10. Cani, A., M. Gaudesi, E. Sanchez, G. Squillero, and A. Tonda (2014). Towards automated malware creation. Proceedings of The 29Th Annual ACM Symposium On Applied Computing, Gyeongju Republic of Korea, 157–60. doi: https://doi.org/10.1145/2554850.2555157. [Crossref], [Google Scholar]
11. Chakkaravarthy, S. S, D. Sangeetha, V. M. Rathnam, K. Srinithi, and V. Vaidehi. 2018. Futuristic cyber-attacks. *International Journal of Knowledge-Based and Intelligent Engineering Systems* 22 (3):195–204. doi: https://doi.org/10.3233/kes-180384. [Crossref], [Web of Science ®], [Google Scholar]
12. Chen, J., X. Luo, J. Hu, D. Ye, and D. Gong. 2018. An Attack on Hollow CAPTCHA Using Accurate Filling and Nonredundant Merging. *IETE Technical Review*, 35(sup1):106–118. doi:https://doi.org/10.1080/02564602.2018.1520152. [Taylor & Francis Online], [Web of Science ®], [Google Scholar]
13. Chung, K., Z. T. Kalbarczyk, and R. K. Iyer. 2019. Availability attacks on computing systems through alteration of environmental control: Smart malware approach. Proceedings of the 10th

ACM/IEEE International Conference on Cyber-Physical Systems, Montreal Quebec, Canada, 1–12. [Crossref], [Google Scholar]
14. Dheap, V. 2017. AI in cybersecurity: A balancing force or a disruptor? https://www.rsaconference.com/industry-topics/presentation/ai-in-cybersecurity-a-balancing-force-or-a-disruptor (accessed February 13, 2020). [Google Scholar]
15. Dixon, W., and N. Eagan. 2019. AI will power a new set of tools and threats for the cybercriminals of the future. https://www.weforum.org/agenda/2019/06/ai-is-powering-a-new-generation-of-cyberattack-its-also-our-best-defence/ (accessed December 3, 2020). [Google Scholar]
16. Fischer, E. 2016. Cybersecurity issues and challenges: In brief. . https://fas.org/sgp/crs/misc/R43831.pdf (accessed December 1, 2020). [Google Scholar]
17. Gao, H., M. Tang, Y. Liu, P. Zhang, and X. Liu. 2017. Research on the security of microsoft's two-layer captcha. *IEEE Transactions On Information Forensics And Security* 12 (7):1671–85. doi:https://doi.org/10.1109/tifs.2017.2682704. [Crossref], [Web of Science ®], [Google Scholar]
18. Hamadah, S., and D. Aqel. 2020. Cybersecurity becomes smart using artificial intelligent and machine learning approaches: An overview. *ICIC Express Letters, Part B: Applications* 11 (12):1115–1123. doi:https://doi.org/10.24507/icicelb.11.12.1115. [Google Scholar]
19. Hitaj, B., P. Gasti, G. Ateniese, and F. Perez-Cruz. 2019. PassGAN: A deep learning approach for password guessing. *Applied Cryptography and Network Security* 11464:217–37. doi:https://doi.org/10.1007/978-3-030-21568-2_11. [Crossref], [Google Scholar]
20. Hu, W., and Y. Tan. 2021. Generating adversarial malware examples for black-box attacks based on GAN.https://arxiv.org/abs/1702.05983 (accessed August 12, 2021). [Google Scholar]
21. John, S., and T. Philip. 2018. Generative models for spear phishing posts on social media. NIPS Workshop On Machine Deception, California, USA. arXiv:1802.05196 [Google Scholar]
22. Kaloudi, N., and J. Li. 2020. The AI-based cyber threat landscape. *ACM Computing Surveys* 53 (1):1–34. doi:https://doi.org/10.1145/3372823. [Crossref], [Web of Science ®], [Google Scholar]
23. Kirat, D., J. Jang, and M. Stoecklin. 2018. DeepLocker concealing targeted attacks with AI locksmithing. https://www.blackhat.com/us-18/briefings/schedule/index.html#deeplocker—concealing-targeted-attacks-with-ailocksmithing-11549 (accessed December 4, 2020). [Google Scholar]
24. Lee, K., and K. Yim. 2020. Cybersecurity threats based on machine learning-based offensive technique for password authentication. *Applied Sciences* 10 (4):1286. doi:https://doi.org/10.3390/app10041286. [Crossref], [Google Scholar]
25. Li, C., X. Chen, H. Wang, P. Wang, Y. Zhang, and W. Wang. 2021. End-to-end attack on text-based CAPTCHAs based on cycle-consistent generative adversarial network. *Neurocomputing* 433:223–36. doi:https://doi.org/10.1016/j.neucom.2020.11.057. [Crossref], [Web of Science ®], [Google Scholar]
26. Meng, G., Y. Xue, C. Mahinthan, A. Narayanan, Y. Liu, J. Zhang, and T. Chen. 2016. Mystique. Proceedings of the 11Th ACM On Asia Conference On Computer and Communications Security, Xi'an, China, 365–76. doi:https://doi.org/10.1145/2897845.2897856. [Crossref], [Google Scholar]

27. Moher, D., A.Liberati, J.Tetzlaff, and D. G.Altman. 2010. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. International Journal of Surgery 8(5): 336–341. doi:https://doi.org/10.1016/j.ijsu.2010.02.007. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

28. Ney, P., K. Koscher, L. Organick, L. Ceze, and T. Kohno. 2017. Computer security, privacy, and dna sequencing: compromising computers with synthesized DNA, privacy leaks, and more. USENIX Security Symposium, Vancouver, BC, Canada, 765–79. [Google Scholar]

29. Noury, Z., and M. Rezaei. 2020. Deep-CAPTCHA: A deep learning based CAPTCHA solver for vulnerability assessment. ArXiv, abs/2006.08296. [Crossref], [Google Scholar]

30. Petro, D., and B. Morris. 2017. Weaponizing machine learning: Humanity was overrated anyway. DEF CON. [Google Scholar]

31. Rigaki, M., and S. Garcia. 2018. Bringing a GAN to a knife-fight: adapting malware communication to avoid detection. IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA. doi:https://doi.org/10.1109/spw.2018.00019. [Crossref], [Google Scholar]

32. Seymour, J., and P. Tully. 2016. Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter. https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-EngineeringAutomated-E2E-Spear-Phishing-On-Twitter-wp.pdf (accessed December 21, 2020). [Google Scholar]

33. Sood, A., S. Zeadally, and R. Bansal. 2017. Cybercrime at a Scale: A Practical Study of Deployments of HTTP-Based Botnet Command and Control Panels. *IEEE Communications Magazine* 55 (7):22–28. doi:https://doi.org/10.1109/mcom.2017.1600969. [Crossref], [Web of Science ®], [Google Scholar]

34. Tang, M., H. Gao, J. Yan, F. Cao, ZhangZ., L. Lei, M. Zhang, P. Zhou, X. Wang, X. Li, and L . X. Jiawei. 2016. A simple generic attack on text captchas. Proceedings 2016 Network And Distributed System Security Symposium, San Diego, California. doi:https://doi.org/10.14722/ndss.2016.23154. [Google Scholar]

35. Thanh, C., and I. Zelinka. 2019. A survey on artificial intelligence in malware as next-generation threats. *MENDEL* 25 (2):27–34. doi:https://doi.org/10.13164/mendel.2019.2.027. [Crossref], [Google Scholar]

36. Trieu, K., and Y. Yang. 2018. Artificial intelligence-based password brute force attacks. Proceedings of Midwest Association for Information Systems Conference, St. Louis, Missouri, USA, 13(39). [Google Scholar]

37. Truong, T., I. Zelinka, J. Plucar, M. Čandík, and V. Šulc. 2020. Artificial intelligence and cybersecurity: past, presence, and future. *Advances In Intelligent Systems And Computing* 351–63. doi:https://doi.org/10.1007/978-981-15-0199-9_30. [Crossref], [Google Scholar]

38. Usman, M., M. Jan, X. He, and J. Chen. 2020. A survey on representation learning efforts in cybersecurity domain. *ACM Computing Surveys* 52 (6):1–28. doi:https://doi.org/10.1145/3331174. [Crossref], [Web of Science ®], [Google Scholar]

39. Xu, W., D. Evans, and Y. Qi. 2018. Feature squeezing: Detecting adversarial examples in deep neural networks. Proceedings 2018 Network and Distributed System Security Symposium, San Diego, California, USA. doi:https://doi.org/10.14722/ndss.2018.23198. [Crossref], [Google Scholar]

40. Yao, Y., B. Viswanath, J. Cryan, H. Zheng, and B. Zhao. 2017. Automated crowdturfing attacks and defenses in online review systems. Proceedings Of The 2017 ACM SIGSAC Conference On Computer And Communications Security, Dallas Texas, USA. doi:https://doi.org/10.1145/3133956.3133990. [Crossref], [Google Scholar]
41. Ye, G., Z. Tang, D. Fang, Z. Zhu, Y. Feng, P. Xu, X. Chen, and Z. Wang. 2018. Yet another text captcha solver. Proceedings of The 2018 ACM SIGSAC Conference On Computer And Communications Security, Toronto, Canada. doi:https://doi.org/10.1145/3243734.3243754. [Crossref], [Google Scholar]
42. Yu, N., and K. Darling. 2019. A low-cost approach to crack python CAPTCHAs using AI-based chosen-plaintext attack. *Applied Sciences* 9 (10):2010. doi:https://doi.org/10.3390/app9102010. [Crossref], [Google Scholar]
43. Zhou, X., M. Xu, Y. Wu, and N. Zheng. 2021. Deep model poisoning attack on federated learning. *Future Internet* 13 (3):73. doi:https://doi.org/10.3390/fi13030073. [Crossref], [Web of Science ®], [Google Scholar]
44. Zouave, E., M. Bruce, K. Colde, M. Jaitnee, I. Rodhe, and T. Gustafsson. 2020. Artificially intelligent cyberattacks.https://www.statsvet.uu.se/digitalAssets/769/c_769530-l_3-k_rapport-foi-vt20.pdf (accessed December, 21 2020). [Google Scholar]

Contexts and Context-awareness Revisited from an Intelligent Environments Perspective

Juan Carlos Augusto
Applied Artificial Intelligence
Published online: 2 Mar 2022