DEVELOPMENT OF SECURITY STRATEGY IN MACHINE LEARNING

***N. Lingareddy, **Dr. Syed Umer**

**Research Scholar, **Research Supervisor,*
Dept. of Computer Science and Engineering,
Himalayan University, India.

ABSTRACT

It is no secret that Machine Learning as a Service (MLaaS) cloud platforms are becoming more popular as machine learning (ML) and deep learning (DL) methods improve. Outsourcing the training of Deep Learning (DL) models is increasingly being done via third-party cloud services, which need expensive computing resources (such as graphics processing units (GPUs), for example). Cloud-hosted ML/DL services are so widely used that attackers have a broad variety of attack surfaces from which to exploit the system. An assessment of cloud-hosted ML/DL models in terms of assaults and defenses is conducted in this paper. Security risks have been identified against a number of learning techniques in the past, including naive Bayes, logistic regression, decision trees and support vector machines (SVM). Due to the nature of the danger, we examine it from two different perspectives, namely the training and testing/inferring phases. Defensive machine learning methods are then divided into four categories: security assessment mechanisms, countermeasures in the training phase, countermeasures during testing or interpretation, data security, and privacy. Last but not least, we highlight five noteworthy themes in the research on machine learning security risks and protective methods, which need further study in the future.

Keywords: Cloud, Data, Security, Security, Machine, Learning

INTRODUCTION

Machine learning, in particular, is being used to attack and defend in the internet. Malware analysis is used by attackers to undermine defensive measures. Malware analysis is used in cyber security to build a strong defense against security risks, preventing or minimizing the effect or harm that occurs as a result of the attack. Intrusion detection, virus detection, cyber physical assaults, and data privacy protection are among the ML applications that have made use of unsupervised and supervised learning techniques extensively. On the one hand, unsupervised techniques look for structure and patterns in data without labeling them whereas supervised methods use examples that are labeled to teach themselves. To counter cyber-attacks - particularly those that are new or continuously changing - these techniques are ineffective. Also, detection and defensive responses are typically made after an assault, when traces of the attack may be collected and analyzed, and proactive security solutions are hampered as a result of this delay. After considerable damage to the cyber systems, 62 percent of the assaults have been detected.

A type of machine learning called reinforcement learning (RL) is the closest thing to human learning since it learns by exploring and exploiting an unfamiliar environment. In real-time and adversarial settings, RL is especially flexible and helpful. Because cyber-attacks are becoming more complex, fast, and pervasive, RL is well suited to cyber security applications of all kinds.

It is no secret that machine learning (ML) has been more popular over the last decade. Because these machine learning (ML) techniques—in particular, deep learning (DL)-based ML techniques—require a significant quantity of training data, they are resource demanding. High-performance graphics processing units (GPUs) and tensor processing units are often used to train Deep Learning models on big datasets. To compensate for the high cost of GPUs/Tensor Processing Units, machine learning as a service is usually outsourced to clouds (MLaaS).

A cloud computing service that offers machine learning services, such as predictive analytics, facial recognition, natural language services, and data modelling APIs, is known as Machine Learning as a Service. Users of MLaaS may submit their data and models to the cloud for training. Models may be put on the cloud for inference purposes as well as training; the usual MLaaS system architecture.

LITERATURE REVIEW

Ramani Sagar (2020) Many apps rely on machine learning (ML) to provide security and privacy. Problems as severe as real-time attack detection and data leaking vulnerability assessments are addressed using machine learning. For example, real-time decision-making, processing large amounts of data, reducing cycle times for learning, and cost-efficiency and error-free processing are all enhanced by machine learning. In this article, we examine the state-of-the-art methods for applying machine learning to security problems. We investigate various security applications from an ML model's point of view, and evaluate their accuracy outcomes on a variety of aspects. It offers a roadmap for an interdisciplinary study field by evaluating ML algorithms in security applications. By conducting adversarial assaults, attackers may escape the ML models, even with the most advanced technology and tools available today. Due to this, it is necessary to evaluate the susceptibility of ML models at the time of creation in order to be prepared for adversarial assaults. To this end, we also examine the various kinds of adversarial assaults on the ML models, as a complement to this. The threat model and defensive techniques against adversarial attack methods have been described so that security characteristics may be seen properly. En outre, we showed how attackers might launch adversarial assaults depending on what they knew about the computer model, as well as where they could target it. A number of distinct characteristics of adversarial assaults are also examined.

Thanh Cong Truong (2020) Facial recognition and image analysis are just a few of the areas where artificial intelligence methods have experienced significant growth in recent years. AI-based approaches in cybersecurity may enhance cyber defense systems and assist adversaries improve their assault tactics. Malicious actors, on the other hand, are aware of the new possibilities and will likely try to take advantage of them. On both the offensive and defensive side of cybersecurity, this article provides an outline of artificial intelligence's potential.

Luis Munoz-Gonz ˜ alez, (2018) Many contemporary apps rely on machine learning to gather useful information from a variety of sources. User quality of life is enhanced via customisation and resource optimization. Many procedures are automated as a result. As a result, attackers may try to take advantage of machine learning systems by exploiting their weaknesses. In various application areas, such attacks have previously been documented in the wild. By introducing malicious data or taking advantage of algorithms' vulnerabilities and blind spots, attackers may undermine machine learning systems described in this chapter. We also discuss the methods that may be used to minimize the effects of such assaults, as well as the difficulties associated with designing more secure machine learning systems in the future.

Nilaykumar Kiran Sangani (2017) Over time, the security threat environment has changed dramatically. Viruses, trojans, DoS, phishing, distributed DoS, etc. are all on the rise. Because of this, attackers have developed a new approach for their attack vector technique that is more focused on the weakest link in the security chain, i.e. the human being. The first thing an attacker thinks of when we discuss people is apps. A new generation of assaults and threats at the application layer have rendered traditional signature-based methods ineffective and obsolete. They are effective in defending organizations against perimeter and endpoint-driven assaults, but it is at the application layer where such defenses fail that attention and analysis should be directed. Identification of harmful user activity patterns that lead to a breach is one of the specific difficulties of protecting online applications. A dynamic and signature-independent approach for detecting such harmful use patterns inside apps is thus needed to address this issue Authors discuss technical elements of incorporating machine learning into apps for identifying harmful user behaviour patterns in this chapter.

Anthony D. Joseph (2017) In the intersection of machine learning and computer security, the study of learning in hostile settings is a new field. The great degree of complexity of the phenomena underpinning computer security and dependability has sparked interest in learning-based techniques for security and system design applications. To better comprehend the data gathered from these complex systems, learning techniques are increasingly being utilized since it becomes more difficult to achieve the required characteristics simply using statically built processes. However, opponents who alter their behavior in response to learning techniques may escape learning methods. Learning methods that are resistant to assaults and provide robustness guarantees have been the subject of little study. To explore methods, problems, and future research objectives for safe learning and learning-based security applications, experts from both the computer security and machine learning communities met at the Perspectives Workshop on "Machine Learning Methods for Computer Security". A number of priority study topics were highlighted as a consequence of the twenty-two presentations, workgroup sessions, and informal conversation that took place. There were a variety of open problems in the field, from traditional applications of machine learning in security (such as attack detection and analysis of malicious software) to methodological issues related to secure learning, particularly the development of new formal approaches with provable security guarantees. Security problems may also emerge in conjunction with data-driven techniques in other possible applications beyond the conventional area of computer security. Computer vision, spam detection in social media and sentiment analysis are some examples of such applications.

Survey Methodology for Security Applications

To perform the study, one must first identify the various kinds of applications in which machine learning classifiers are used. A taxonomy of various security applications where machine learning may be used to achieve the desired objective is shown in this section. Several survey articles based on machine learning have been presented, although the majority of them address security-related problems in machine learning applications. In light of the above, this study also examines privacy and security issues, as well as adversarial attacks on machine learning classifiers. Security-based apps are divided into various categories in order to create a survey that includes a thorough technique analysis for all elements of security. We also increase the types of classifiers that need to be analyzed, particularly for intrusion detection and prevention. Technology advancements in recent years have made it impossible to restrict the breadth of security applications. For the purposes of this study, we focused on security applications in which machine learning plays a crucial and important role Decision trees, logistic regression, and function approximations are among the statistical processes that are included in ML algorithms. A more influential kind of algorithm, this one may be employed in situations when categorization is vital.

ACM digital library, Web of Science, Scopus, and IEEE Xplore were utilized to obtain a thorough picture of the intersection between AI and cybersecurity. Additionally, Google Scholar was used. These databases have been indexed using a set of keywords relevant to the subject. They developed various keywords and keyword combinations for each search engine to maximize coverage.

DEEP REINFORCEMENT LEARNING PRELIMINARY

Instead of learning from examples like in the other prominent branch of ML, supervised techniques, RL defines agents by allowing them to create their own learning experiences by interacting with the environment. State, action, and reward are used to characterize RL (Fig. 1). Every action that the agent makes at each time step leads to two changes: the environment's present state has been altered, and he or she has earned or lost points. Given a state, the reward function tells the agent whether an action is good or harmful. In response to the incentives he receives, the agent learns to perform more positive acts and to filter out negative ones.

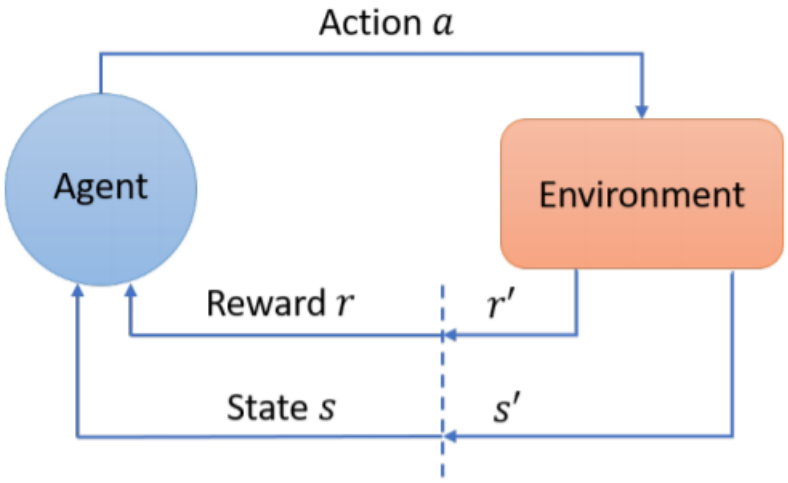

Fig. 1. Interactions between the agent and its environment in RL, characterized by state, action and reward

A lookup table or Q-table is required for Q-learning to record anticipated rewards (Q-values) of actions given a set of state conditions (Q-values). A significant amount of memory is required as the state and action spaces expand. Q-learning is inefficient for solving real-world issues since they typically include continuous state or action space. It is fortunate that deep learning has developed as a strong technique that can be used in conjunction with conventional machine learning approaches. Deep learning can develop a compact low-dimensional representation of raw high-dimensional data by using function approximation and representation learning techniques. As a result of Google DeepMind's pioneering research, deep learning and reinforcement learning were combined. Their proposal was to develop a deep Q network (DQN) that would utilize deep neural networks (DNN) to handle inputs with large dimensions.

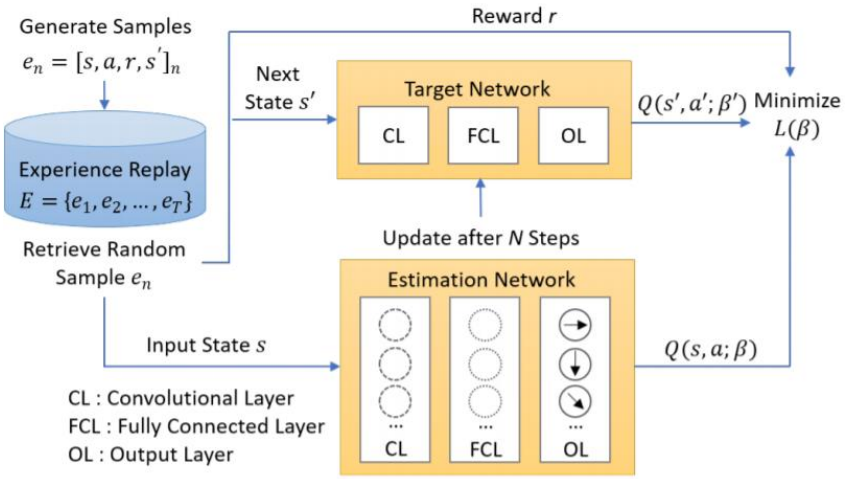

Fig. 2. DQN architecture with the loss function

Integration of machine learning with security mechanisms

An integration with current security instruments could provide significant qualitative improvements, given the effectiveness of learning-based methods for narrowly targeted security tasks. To be sure, achieving this level of integration is no easy job. As shown in Figure 3, learning-enhanced reactive security measures have an abstract design. Simple rule-based detection techniques may be used to manage the overwhelming majority of security-related data. Domain specialists can easily comprehend and manage rules, which are fast, inexpensive, and accurate. To handle carefully designed and unique input samples, however, rules are not strong enough. Despite making up a tiny portion of the overall input samples, these samples may cause significant harm if they are not prevented. Algorithms based on machine learning may play a key role in identifying previously unknown attack samples. It is possible to greatly increase the productivity of analysts by using the confidence intervals provided by learning techniques.

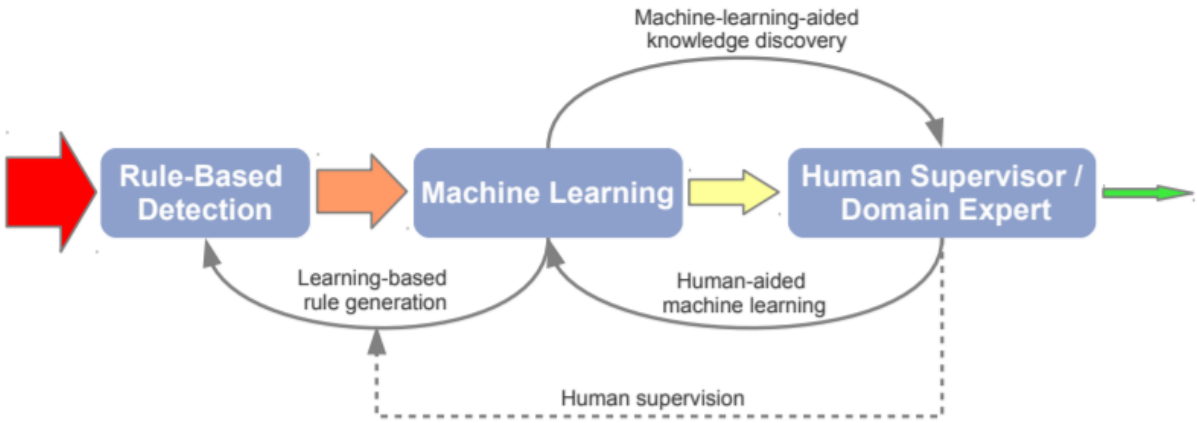

Figure 3: The abstract architecture of reactive security mechanisms utilizing machine learning

Machine Learning Vulnerability Analysis and Threat Model

System vulnerabilities are the primary source of security and privacy breaches. From both a security and privacy viewpoint, we outline various threats in our study. Techniques such as machine learning are widely used in security and privacy-related applications such as malware detection and pattern recognition as well as homomorphic encryption [150-151] as well as privacy preservation and statistical analysis of a database. To automate security or privacy breach detection, machine learning has emerged as a viable method. The security and privacy elements of a system cannot be addressed, as previously stated. ML methods are not without their flaws and limitations. In Figure 4, we have shown both an attack surface and a potential defensive scope for the ML life cycle. An adversarial scenario is used to demonstrate the features of potential assaults. The ML classifier's pre-processing, feature extraction, and model training phases are represented by the middle layer of Figure 4. The ML classifier's adversarial setup is countered in the top layer.

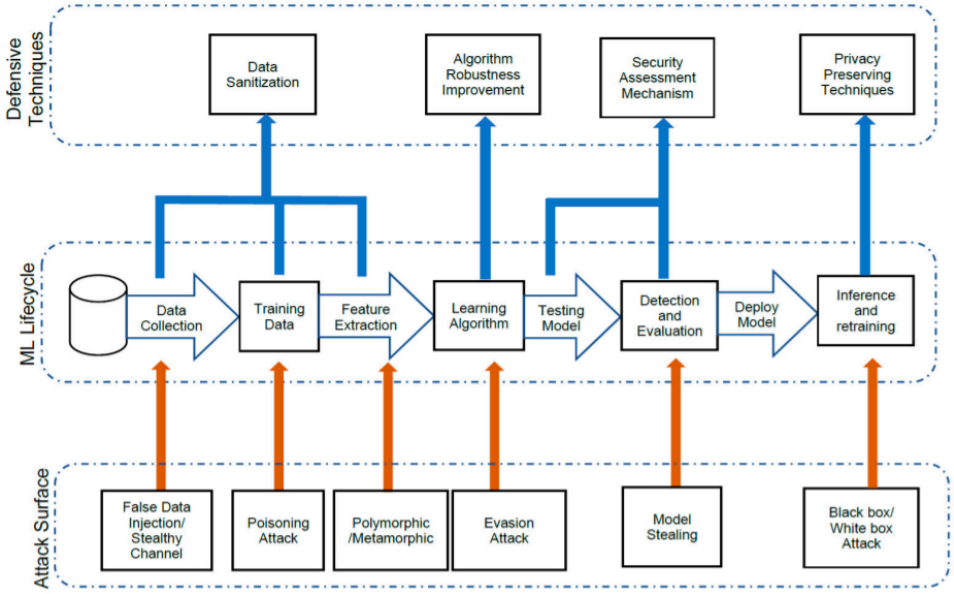

Figure 4. Threat model of the machine learning process

False data injection and covert channel assaults may be used to get access to the ML classifier, as shown in Figure 4. Classifiers that use machine learning must undergo training before they can make a particular classification of a dataset. By introducing hostile samples into the training datasets, the poisoning assault has undermined the integrity and availability of the ML models. A homomorphic strategy is one that can convert one feature vector object into another feature vector item in the actual world. They are designed to compromise the model's security by altering critical aspects of the algorithm and seizing control. Unauthorized adversarial models that may mislead the genuine model may be detected using a stolen model after the ML classifier has been deployed. In order to carry out this behaviour, the attackers mount application program interfaces by repeatedly submitting requests. As part of the inference step, the attacker model is divided into two groups based on the degree of understanding knowledge. Strong and sophisticated attackers may start a white-box attack by downloading and accessing the ML models and other data, whereas weak attackers can launch a black-box attack by utilizing APIs and filling in inputs to the system.

CONCLUSION

By using machine learning to improve the quality of security apps, new cyber-threats may harm vital data infrastructure. The performance of identifying an adversarial sample by collecting and forecasting hostile samples is a problem for machine learning-based security applications. From an attacker's and designer's viewpoint, the new models are becoming a study topic. Security in machine learning-based decision systems in hostile settings brings up a new study field with the fast rise in security incidents. As a result, the total error rate stays the same, and the assault goes undetected. Malicious users may take advantage of this in complex attacks by increasing false-negative rates and reducing false-positive rates proportionally. To effectively identify assaults on ML-based systems, this kind of problem must be addressed. Because machine learning algorithms operate on such a large

number of factors, current approaches of data privacy suffer from limited performance. Here, we conducted a comprehensive literature analysis on the security of cloud-hosted machine learning and deep learning (MLaaS) models in order to better understand the state of the art. Among the eight major publishers that contributed relevant articles to this project were ACM Digital Library, IEEE Xplore, ScienceDirect, international conference on machine learning, international conference on learning representations, journal of machine learning research, USENIX, neural information processing systems, and arXiv Articles were selected using a review protocol that included inclusion and exclusion formulas. We analysed the articles that met these criteria on two dimensions and provided thematic analyses of five attack and five defence themes, respectively, for each of the articles that were chosen. Our literature analysis also revealed several limits and hazards. Finally, we identified some unresolved problems that need additional research.

REFERENCES

1. Sarker, Iqbal & Kayes, A. S. M. & Badsha, Shahriar & Alqahtani, Hamed & Watters, Paul & Ng, Alex. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data. 7. 10.1186/s40537-020-00318-5.

2. Liu, Qiang & Li, Pan & Zhao, Wentao & Cai, Wei & Yu, Shui. (2018). A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View. IEEE Access. 6. 12103-12117. 10.1109/ACCESS.2018.2805680.

3. Julian Jang-Jaccard, A survey of emerging threats in cybersecurity, Journal of Computer and System Sciences 80 (2014) 973–993

4. Ramani Sagar, Applications in Security and Evasions in Machine Learning: A Survey, Electronics 2020, 9, 97

5. Thanh Cong Truong, Artificial Intelligence in the Cyber Domain: Offense and Defense, Symmetry 2020, 12, 410

6. Muñoz-González, Luis & Lupu, Emil. (2019). The Security of Machine Learning Systems. 10.1007/978-3-319-98842-9_3.

7. Sangani, Nilaykumar & Zarger, Haroot. (2017). Machine Learning in Application Security. 10.5772/intechopen.68796.

8. Qayyum A, Ijaz A, Usama M, Iqbal W, Qadir J, Elkhatib Y, Al-Fuqaha A (2020) Securing Machine Learning in the Cloud: A Systematic Review of Cloud Machine Learning Security. Front. Big Data 3:587139. doi: 10.3389/fdata.2020.587139

9. Thanh Thi Nguyen, Deep Reinforcement Learning for Cyber Security, arXiv:1906.05799v3 [cs.CR] 21 Jul 2020

10. Han, Y., Rubinstein, B. I., Abraham, T., Alpcan, T., De Vel, O., Erfani, S., ... and Montague, P. (2018, October). Reinforcement learning for autonomous defence in software-defined networking. In International Conference on Decision and Game Theory for Security (pp. 145-165)

11. Liu, X., Xu, Y., Jia, L., Wu, Q., and Anpalagan, A. (2018). Antijamming communications using spectrum waterfall: A deep reinforcement learning approach. IEEE Communications Letters, 22(5), 998- 1001

12. Zeng, K., Govindan, K., and Mohapatra, P. (2010). Non-cryptographic authentication and identification in wireless networks. IEEE Wireless Communications, 17(5), 56-62.

13. Shamshirband, S., Anuar, N. B., Kiah, M. L. M., and Patel, A. (2013). An appraisal and design of a multi-agent system based cooperative wireless intrusion detection computational intelligence technique. Engineering Applications of Artificial Intelligence, 26(9), 2105-2127.

14. Yau, D. K., Lui, J. C., Liang, F., and Yam, Y. (2005). Defending against distributed denial-of-service attacks with max-min fair server-centric router throttles. IEEE/ACM Transactions on Networking, 13(1), 29-42

15. Xu, X. (2010). Sequential anomaly detection based on temporal difference learning: Principles, models and case studies. Applied Soft Computing, 10(3), 859-867.