METHODS AND MEANS OF PROTECTING INFORMATION SYSTEMS FROM HACKING BASED ON MACHINE LEARNING

Yusupov B.K.

Head of the Department "Information Technologies" of the Military Institute of Information and Communication Technologies and Communication, Uzbekistan

Ziyoratov Sh.O.

Master's degree, Faculty of Cyber-Security,Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan

https://doi.org/10.5281/zenodo.7445044

Abstract. *The cyber physical systems integrate the sensing, computation, control and networking processes into physical objects and infrastructure, which are connected through the Internet to execute a common task. Cyber physical systems can be applied in various applications, like healthcare, transportation, industrial production, environment and sustainability, and security and surveillance. However, the tight coupling of cyber systems with physical systems introduce challenges in addressing stability, security, efficiency and reliability. The machine learning (ML) security is the inclusion of cyber security mechanism to provide protection to the machine learning models against various cyber attacks. The ML models work through the traditional training and testing approaches. However, execution of such kind of approaches may not function effectively in case if a system is connected to the Internet. As online hackers can exploit deployed security mechanisms and poison the data. This data is then taken as the input by the ML models. In this article, we provide the details of various machine learning security attacks in cyber physical systems. We then discuss some defense mechanisms to protect against these attacks. We also present a threat model of ML security mechanisms deployed in cyber systems. Furthermore, we discuss various issues and challenges of ML security mechanisms deployed in cyber systems. Finally, we provide a detailed comparative study on performance of the ML models under the influence of various ML attacks in cyber physical systems.*

Keywords: *Machine learning, Cubersecurity, attacks, CPS.*

МЕТОДЫ И СРЕДСТВА ЗАЩИТЫ ИНФОРМАЦИОННЫХ СИСТЕМ ОТ ВЗЛОМА НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Аннотация. *Киберфизические системы интегрируют сенсорные, вычислительные, управляющие и сетевые процессы в физические объекты и инфраструктуру, которые связаны через Интернет для выполнения общей задачи. Киберфизические системы могут применяться в различных приложениях, таких как здравоохранение, транспорт, промышленное производство, окружающая среда и устойчивость, а также безопасность и наблюдение. Однако тесная связь киберсистем с физическими системами создает проблемы в обеспечении стабильности, безопасности, эффективности и надежности. Безопасность машинного обучения (ML) — это включение механизма кибербезопасности для обеспечения защиты моделей машинного обучения от различных кибератак. Модели машинного обучения работают с использованием традиционных подходов к обучению и тестированию. Однако выполнение такого рода подходов может работать неэффективно, если система подключена к Интернету. Поскольку онлайн-хакеры могут использовать развернутые механизмы безопасности и отравить данные. Затем эти данные используются в качестве входных* *данных моделями ML. В этой статье мы подробно расскажем о различных атаках на безопасность машинного обучения в киберфизических системах. Затем мы обсудим некоторые защитные механизмы для защиты от этих атак. Мы также представляем модель угроз механизмов безопасности ML, развернутых в киберсистемах. Кроме того, мы обсуждаем различные проблемы и проблемы механизмов безопасности ML, развернутых в киберсистемах. Наконец, мы предоставляем подробное сравнительное исследование производительности моделей машинного обучения под влиянием различных атак машинного обучения в киберфизических системах.*

Ключевые слова: *Машинное обучение, Cubersecurity, атаки, CPS.*

Introduction. Cyber physical systems (CPS) are systems that collaborate computational entities (i.e., sensors and actuators) in connection with the physical world and the associated processes. This further facilitates the data-accessing and data-processing services available on the Internet. CPS can be used in various types of applications (i.e., smart home, smart healthcare and transportation systems, etc.). Though CPS can be deployed in various applications at the same time, they have many Challenges related to security and privacy because various attacks (i.e., malware injection, impersonation, man-in-the-middle (MiTM), leakage of secret data, unauthorized data updates, etc.) are possible. Sometimes we use the machine learning (ML) models in the CPS-based applications to draw useful outcomes from the collected data of the sensors. Therefore, the role of ML models is very important here, and their predictions and outcomes should be accurate. However, the existence of various attacks may affect the performance of the ML models, and thus, they may produce wrong outcomes. ML is a field of computing that utilizes computational algorithms to train machines to be able to perform various tasks that further automate the workload without explicitly intervening at each step and limiting human interaction with the process. ML has its applications in various domains (for example, medicine, agriculture, and natural disaster prediction and management). Moreover, it can be integrated and utilized in various domains, like the Internet of Things (IoT), cyber physical systems, cyber security, computer vision, image processing, robotics, natural language processing, and many more.

Since the presented work is related to ML security, therefore it is essential to explain the phases of ML tasks. Fig. 1 depicts various phases of ML tasks in the cyber physical systems. A basic ML task can be divided into two phases: (1) training phase and (2) deployment phase (testing phase) [11]. Their details are given below.

- Training phase: The task starts with accumulation of data from reputable and authorized sources (i.e., sensors). The often humongous data is analyzed and prepared for training. It utilizes different techniques, like cleaning, augmenting, and segmenting. It basically involves converting irregularities and missing values in the dataset into consistent data that can be processed further. Next, the dataset and problem statement is analyzed. The class labels and features of the dataset are understood, and correlation (degree of relationship) between different features is visualized. After that, the data is split into two parts for training and testing purposes of the model keeping in mind the correlation between features and the target prediction required the selection of suitable ML algorithm. The algorithm is the intuition behind the model, and it provides the prediction output of the input data based on the value of the feature through a mathematical formula. The training data is given to the ML model to train it with the feature values, and then the possible pattern is made. This pattern is calculated, and the parameters for prediction are calculated. Now, the testing phase starts, and the calculated parameters are used on the test data in order to carry out the new predictions. The accuracy score of the ML algorithm is calculated using the test data to find the prediction capability of the algorithm. Hypertuning is performed on the best algorithm by tweaking its formula to get the best iteration of the ML algorithm. Once this hyper tuning is performed and the model gives a satisfactory accurate prediction, the trained model finally deployed.

• Deployment phase: The deployed model after hyper tuning is supplied with the real-time data. The trained model will provide prediction output on input new data. The model may use the Application Programming Interface (API) to interact with the users where we can feed the data through it and obtain the predictions based on training done under the training phase. The results of predictions and findings are then summarized and presented in the illustrated manner for future analysis and decision-making purpose.

Materials. Information security is the methodology of protecting information and sensitive data from security risks (i.e., unauthorized access and usage, modification, inspection, and deletion of the information). Information security in the cyber physical systems is provided on the basis of CIA Triad, which comprises techniques like confidentiality, integrity, and availability. Confidentiality (or privacy) involves restricting access to the information. Its usage is much needed in order to protect information from being accessed or modified by malicious entities. This can be achieved by utilizing encryption techniques, including public-key cryptography and security tokens. Moreover, integrity (or data integrity) involves maintaining the trustworthiness and dependability of the information. It is practiced to retain the usability of data and prevail it to be usable for other tasks. It could be achieved by using the techniques like version and access controlling, hashing and compliance checks, and keeping data checksums. Furthermore, availability is the practice of accessibility of information for retrieval and usage by authorized entities. It is required to maintain the information consistently through the maintaining systems which hold them. It could be achieved through server monitoring, redundancy, resolving software issues, and maintaining contingency protocols to deal with Denial-of-Service (DoS) or distributed DoS (DDoS) attacks. These are the basic characteristics that we cover under information security. However, we need to take care of other important properties like authentication, access control, authorization, forward secrecy, backward secrecy, data freshness, etc.

Methods. In the following, we provide the details of machine learning (ML) security in cyber physical systems. ML security is the inclusion of cyber security techniques to safeguard the integrity and privacy of an ML model from cyber attacks. It utilizes the various defense mechanisms to prevent the subjection of the model from attacks that further prevent sensitive information from getting breached. It also stops any disaster-related to the prediction of wrong outcomes. With the vast extent of ML being used and fueling software that affects lives of billions, these days protecting our vital data, and for smooth deployment of services that directly or indirectly utilize ML security is becoming an important field of study. Protection against malicious activities is a vital aspect of the ML task. Securing our ML process is very important. For instance, when we work with sensitive data, the correct training of data is essentially required. With the substantial growth of technologies and development in Big data, securing all such types of data and protecting ML tasks are the ever occurring tasks, which need to be resolved with utmost priority.

Results. ML security operates with the help of various cyber security mechanisms, which are deployed there to safeguard the integrity and privacy of an ML model against the various threats and attacks. It uses different safeguarding schemes to prevent the subjection of the model from attacks. This prevents sensitive information from getting breached and also prevents the system from producing bad outcomes (predictions). The primary motivation behind the survey article is to summarize the research work and case studies done in the field of ML security in the cyber physical systems. Such a communication environment is being used in various domains (i.e., healthcare, security and surveillance, retailing, industrial automation, control and support, intelligent transportation system, etc.,). The correct prediction and privacy of users' data are essentially required. In machine learning, we use a model, which is called as ML model, and is used for the purpose of prediction of some phenomena. During the literature survey, it has been identified that the ML models are vulnerable to various types of attacks (i.e., dataset poisoning attack, model poisoning attack, privacy breach, runtime disruption attack, and membership inference attacks). Due to the enormous use of ML, it becomes essential to protect its models against the various possible attacks. Therefore, we focus on various attacks that are associated with the ML models. The different mechanisms of these attacks have also been discussed, along with some possible solutions to prevent them. This research work will be helpful for the researchers to make machine learning more secure and robust.

The research contributions of this work are summarized below.

• We present a threat model of ML security in the cyber physical systems, in which we provide the details of all threats associated with the ML models.

• We then discuss various issues and challenges of ML security in the cyber physical systems.

• Next, we discuss the mechanisms of various attacks related to the ML security.

• We also discuss some possible solutions that can be used to protect the ML security.

• Furthermore, we provide a comparative study on performance of the ML models under the influence of various attacks that can be also deployed for the cyber physical systems.

Conclusion. We provided the details of various Machine learning security attacks (i.e., dataset poisoning attack, model poisoning attack, privacy breach, membership inference attack, runtime disruption attack), which are possible on the machine learning models deployed in the cyber physical systems. We also discuss some of the defense mechanisms which can be deployed to protect against these attacks. We then presented the threat model of ML security, in which we provided the details of all threats

REFERENCES

1. Jang-JaccardJulian et al. A survey of emerging threats in cybersecurity
2. J. Comput. System Sci. (2014)
2. AfuwapeAfeez Ajani et al. Performance evaluation of secured network traffic classification using a machine learning approach Comput. Stand. Interfaces(2021)
3. KamilarisAndreas et al. Deep learning in agriculture: A survey Comput. Electron. Agric.(2018)

4. RahmanZiaur et al. Blockchain-based security framework for a critical industry 4.0 cyber-physical system IEEE Commun. Mag.(2021)
5. RaoAakarsh et al. Probabilistic threat detection for risk management in cyber-physical medical systems IEEE Softw. (2018)
6. KordestaniMojtaba et al. Observer-based attack detection and mitigation for cyberphysical systems: A review IEEE Syst. Man Cybern. Mag. (2021)
7. GiraldoJairo et al. Security and privacy in cyber-physical systems: A survey of surveys IEEE Des. Test (2017)
8. HumayedAbdulmalik et al. Cyber-physical systems security-A survey IEEE Internet Things J. (2017)