International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 01, January 2023)

Cyber Security based on Machine and Deep Learning Techniques for NIDS

Anubhav

Assistant Professor, Department of BCA, Vidya Institute of Creative Teaching Meerut, UP, India

***Abstract*— The Internet of Things (IoT) has been rapidly evolving towards making a greater impact on everyday life to large industrial systems. Unfortunately, this has attracted the attention of cybercriminals who made IoT a target of malicious activities, opening the door to a possible attack on the end nodes. To this end, Numerous IoT intrusion detection Systems (IDS) have been proposed in the literature to tackle attacks on the IoT ecosystem, which can be broadly classified based on detection technique, validation strategy, and deployment strategy. This survey paper presents a comprehensive review of contemporary IoT IDS and an overview of techniques, deployment Strategy, validation strategy and datasets that are commonly applied for building IDS.**

***Keywords*— IoT, IDS, Cyber, Attack, Security, Internet.**

I. INTRODUCTION

Internet of Things (IoT) is interconnected systems of devices that facilitate seamless information exchange between physical devices. These devices could be medical and healthcare devices, driverless vehicles, industrial robots, smart TVs, wearables and smart city infrastructures; and they can be remotely monitored and regulated. IoT devices are expected to become more prevalent than mobile devices and will have access to the most sensitive information, such as personal information. This will result in increasing attack surface area and probabilities of attacks will increase. As security will be a vital supporting element of most IoT applications, IoT intrusion detection systems need also be developed to secure communications enabled by such IoT technologies.

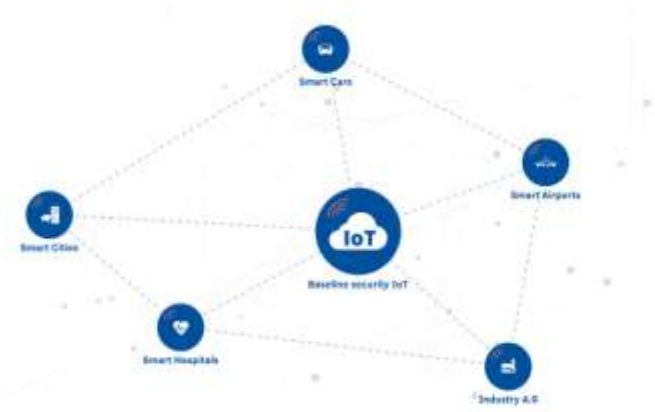

Figure 1: IOT smart infrastructure security

In the last few years, advancement in Artificial Intelligent (AI) such as machine learning and deep learning techniques has been used to improve IoT IDS (Intrusion Detection System). The current requirement is to do an up-to-date, thorough taxonomy and critical review of this recent work. Numerous related studies applied different machine learning and deep learning techniques through various datasets to validate the development of IoT IDS. But, it's still not clear that which dataset, machine learning or deep learning techniques are more effective for building an efficient IoT IDS.

II. LITERATURE SURVEY

H. Hou et al.,[1] To enhance the overall security of the Internet, an IDS based on hierarchical long short-term memory (HLSTM) networks is proposed.

With the introduction of HLSTM, the network can learn across multiple levels of temporal hierarchy over complex network traffic sequences. The system is evaluated on the well-known benchmark data set NSL-KDD for comparison with other existing methods. The experimental results demonstrate that compared with existing start-of-the-art methods, our system has better detection performance for different types of cyberattacks. In addition, the low-frequency network attack types have higher classification accuracy and a lower false detection rate.

P. Feng et al.,[2] construct approximate call graph from function invocation relationships within an Android application to represent this application, and further extract intra-function attributes, including required permission, security level and statistical instructions information, to form the node attributes within graph structures. Then, we use graph neural network (GNN) to generate a vector representation of the application, and then malware classification is performed on this representation. We conduct experiments on real-world application samples. The experimental results demonstrate that our approach implements high effective malware detection and outperforms state-of-the-art detection approaches.

S. Liu et al.,[3] addresses the automatic binary-level software vulnerability detection problem by proposing a deep learning-based approach. The proposed approach consists of two phases: binary function extraction, and model building. First, we extract binary functions from the cleaned binary instructions obtained by using IDA Pro. Then, we employ the attention mechanism on top of a bidirectional long short-term memory for building the predictive model. To show the effectiveness of the proposed approach, we have collected datasets from several different sources. We have compared our proposed approach with a series of baselines including source code-based techniques and binary code-based techniques. We have also applied the proposed approach to real-world IoT related software such as VLC media player and LibTIFF project that used on Autonomous Vehicles. Experimental results show that our proposed approach betters the baselines and is able to detect more vulnerabilities.

Y. Jin et al.,[4] focus on these peculiarities and propose a method for detecting malware infected computers by monitoring unintended DNS traffic on wireless networks by collaboration with DHCP (Dynamic Host Configuration Protocol) server.

By deploying the proposed system on campus wireless networks, the computers within DHCP configured environment can be detected when they are infected by some types of malware and it attempts to communicate with the corresponding C&C servers using DNS (Domain Name System) protocol. In this paper, we describe the detailed design of the proposed method and the future work includes prototype implementation as well as evaluations.

B. Peng et al.,[5] proposed a K-NN classification algorithm, which is used to match the characteristic vectors of network data packets in the new energy plant and station system, which realizes the anomaly detection of network attack scenarios, malformed messages and irregular business instructions in the new energy plant and station system. Finally, a simulation experiment environment of the new energy plant and station system is built to verify the method proposed in this paper. The experimental results show that the algorithm has high ability of anomaly detection and low false alarm rate. It is of great significance to improve the level of network security protection of new energy plants and stations, and to ensure the safe and stable operation of new energy plants and stations.

W. Bi et al.,[6] FCPAs, dynamic characteristics of Area Control Error (ACE) are utilized to detect the compromised data. Compared with FCPAs, VCPAs are more deceptive. A relation-based (RB) feature extraction method is introduced to distinguish the signals compromised by VCPAs from the normal ones. A detection model that does not require compromised samples is developed with the aid of support vector domain description. In the end, a comprehensive detection scheme is designed to detect both FCPAs and VCPAs on the LFC system.

K. Liu et al.,[7] reviews the problem of instrusion detection for Smart Home and different approach to detect instrusion. A hybrid instrusion detection method based on Convolutional Neural Networks(CNN)and K-means is proposed in this paper. At smart home device node, K-means is used to generate the rule base by clustering, then Principal Component Analysis(PCA)is used to extract the dimensionality reduced features. During the test process, PCA is also used to extract the dimensionality reduced features, the feature matching is performed with the rule base to determine the intrusion data. At the smart home server side, a CNN model is proposed to detect the specific type of intrusion.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 01, January 2023)

Y. Jin et al.,[8] propose a client based anomaly traffic detection and blocking mechanism by monitoring DNS name resolution per application program. In the proposed mechanism, by the collaboration of DNS proxy and packet filter, DNS traffic is monitored on the client and the traffic destined to the IP addresses obtained without DNS name resolution or the traffic from unrecognized programs will be detected and blocked. In addition, in order to mitigate false positive detection, an alert-window will be shown to let the users decide whether to allow the traffic or not. We implemented a prototype system on a Windows 7 client and confirmed that the proposed mechanism worked as expected.

R. Velea et al.,[9] discuss a hybrid approach that leverages CPU and GPU compute capabilities in order to accelerate pattern matching for malware signatures. The solution presented focuses on improving performance and reducing power consumption of string matching algorithms on devices such as ultrabooks and laptops.

S. Merat et al.,[10] The main focus of this work is the improvement of machine learning where a number of different types of computer processes can be mapped in multitasking environment. A software mapping and modelling paradigm named SHOWAN is developed to learn and characterize the cyber awareness behaviour of a computer process against multiple concurrent threads. The examined process start to outperform, and tended to manage numerous tasks poorly, but it gradually learned to acquire and control tasks, in the context of anomaly detection. Finally, SHOWAN plots the abnormal activities of manually projected task and compare with loading trends of other tasks within the group.

S. Han et al.,[11] Cyber-physical systems (CPSs) integrate the computation with physical processes. Embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa. CPS was identified as one of the eight research priority areas in the August 2007 report of the President's Council of Advisors on Science and Technology, as CPS will be the core component of many critical infrastructures and industrial control systems in the near future. However, a variety of random failures and cyber attacks exist in CPS, which greatly restrict their growth. As a result, the effort will be made to discuss how to appropriately apply the intrusion detection mechanism to CPS in this work.

M. Bousaaid et al.,[12] The advent of information and communication technologies (ICT) in the domain of education represents a real opportunity for spreading knowledge.

Many results have already been obtained, which aimed mainly to facilitate the arrangement of pedagogical contents by large and massive deployment of digital environments of work. The development of technologies of multimedia, linked to that of Internet and democratization of high output has made henceforth E-learning possible for learners being in virtual classes and geographically distributed. The quality and quantity of asynchronous and synchronous communications are the key elements for E-learning success. It is important to have a propitious supervision to reduce the feeling of isolation in E-learning. This feeling of isolation is among the main causes of loss and high rates of stalling in E-learning.

III. IOT INTRUSION DETECTION SYSTEMS TECHNIQUES

IoT Intrusion is defined as an unauthorised action or activity that harms the IoT ecosystem. In other words, an attack that results in any kind of damage to the confidentiality, integrity or availability of information is considered an intrusion. For example, an attack that will make the computer services unavailable to its legitimate users is considered an intrusion. An IDS is defined as a software or hardware system that maintains the security of the system by identifying malicious activities on the computer systems. The main aim of IDS is to identify unauthorised computer usage and malicious network traffic which is not possible while using a traditional firewall. This results in making the computer systems highly protective against the malicious actions that compromise the availability, integrity, or confidentiality of computer systems.

A. Signature-based intrusion detection systems (SIDS)

Signature intrusion detection systems (SIDS) utilize pattern matching techniques to find a known attack; these are also known as Knowledge-based Detection. In SIDS, matching methods are used to find a previous intrusion. In other words, when an intrusion signature matches the signature of a previous intrusion that already exists in the signature database, an alarm signal is triggered. For SIDS, the host's logs are inspected to find sequences of commands or actions which have previously been identified as malware. SIDS has also been labelled in the literature as Knowledge-Based Detection or Misuse Detection. Traditional methods of SIDS have difficulty in identifying attacks that span multiple packets as they examine network packets and perform matching against a database of signatures.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 01, January 2023)

With the increased sophistication of modern malware, extracting signature information from multiple packets may be required. With this, IDS needs to bring the contents of earlier packets as well. For creating a signature for SIDS, generally, there have been several methods where signatures are created as state machines, formal language string patterns or semantic conditions.

B. Anomaly-based intrusion detection system (AIDS)

AIDS has attracted a lot of scholars because of its feature to overcome the limitation of SIDS. In AIDS, a normal model of the behavior of a computer system is created using machine learning, statistical-based or knowledge-based methods. Any significant deviation between the observed behavior and the model is regarded as an anomaly, which can be interpreted as an intrusion. This kind of technique works on the fact that malicious behaviour is different from typical user behaviour. The behaviour of abnormal users that differentiates from the standard behaviour is defined as an intrusion. There are two phases in the development of AIDS: the training phase and the testing phase. In the training phase, the normal traffic profile is used to learn a model of normal behaviour. In the testing phase, a new data set is used to develop the system's capacity to generalise to previously unseen intrusions. AIDS can be sub-categorized based on the method used for training, for instance, statistical-based, knowledge-based and machine learning-based.

The main advantage of AIDS is the ability to identify zero-day attacks because recognizing the abnormal user activity does not rely on a signature database. AIDS triggers a danger signal when the examined behavior deviates from normal behavior. Furthermore, AIDS has a number of benefits. First, they can discover internal malicious activities. If an intruder starts making transactions in a stolen account that are unidentified in the typical user activity, it creates an alarm. Second, it is challenging for a cybercriminal to recognize what is a normal user behavior without producing an alert as the system is constructed from customized profiles.

C. Machine Learning based Technique

Machine learning is the process of extracting knowledge from large quantities of data. Machine learning models comprise of a set of rules, methods, or complex "transfer functions" that can be applied to find interesting data patterns or to recognise or predict behaviour.

Machine learning techniques have been applied extensively in the area of AIDS. To extract the knowledge from intrusion datasets, different algorithms and techniques such as clustering, neural networks, association rules, decision trees, genetic algorithms, and nearest neighbour methods are utilized.

Some prior research has examined the use of different techniques to build AIDSs. Examined the performance of two feature selection algorithms involving Bayesian networks (BN) and Classification Regression Trees (CRC) and combined these methods for higher accuracy.

Techniques of feature selection using a combination of feature selection algorithms such as Information Gain (IG) and Correlation Attribute evaluation. They tested the performance of the selected features by applying different classification algorithms such as C4.5, naïve Bayes, NB-Tree and Multi-Layer Perceptron. A genetic-fuzzy rule mining method has been used to evaluate the importance of IDS features. NIDS by using the Random Tree model to improve accuracy and reduce the false alarm rate.

Various AIDSs have been created based on machine learning techniques as shown in Fig. 4. The main aim of using machine learning methods is to create IDS that requires less human knowledge and improve accuracy. The quantity of AIDS which makes use of machine learning techniques has been increasing in the last few years. The main objective of IDS based on machine learning research is to detect patterns and build an intrusion detection system based on the dataset. Generally, there are two categories of machine learning methods, supervised and unsupervised.

IV. Conclusion

This paper presented a critical review of IoT intrusion detection system methodologies, deployment strategy, validation strategy, Dataset and technologies with their advantages and limitations. Several intrusion detection systems have been proposed to detect IoT attacks are reviewed. However, such approaches may have the problem of detecting all IoT attacks due to IoT architecture. To develop reliable IoT IDS based on heterogeneous device categories, novel IDS must be developed. We recognize some elements that have a vital feature in the building of reliable IDS for the IoT. First, be low on false alarms due to the large volume of data. Second, be highly adaptive to extreme IoT communication systems due to unexpected behavior in IoT sensors that once appeared usual may start considering attacks.

REFERENCES

[1] H. Hou et al., "Hierarchical Long Short-Term Memory Network for Cyberattack Detection," in IEEE Access, vol. 8, pp. 90907-90913, 2020, doi: 10.1109/ACCESS.2020.2983953.

[2] P. Feng, J. Ma, T. Li, X. Ma, N. Xi and D. Lu, "Android Malware Detection Based on Call Graph via Graph Neural Network," 2020 International Conference on Networking and Network Applications (NaNA), 2020, pp. 368-374, doi: 10.1109/NaNA51271.2020.00069.

[3] S. Liu, M. Dibaei, Y. Tai, C. Chen, J. Zhang and Y. Xiang, "Cyber Vulnerability Intelligence for Internet of Things Binary," in IEEE Transactions on Industrial Informatics, vol. 16, no. 3, pp. 2154-2163, March 2020, doi: 10.1109/TII.2019.2942800.

[4] Y. Jin, M. Tomoishi and N. Yamai, "Anomaly Detection by Monitoring Unintended DNS Traffic on Wireless Network," 2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), 2019, pp. 1-6, doi: 10.1109/PACRIM47961.2019.8985052.

[5] B. Peng, Q. Wang, X. Li, J. Cai, J. Fei and W. Chen, "Research on Abnormal Detection Technology of Real-Time Interaction Process in New Energy Network," 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2019, pp. 433-440, doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00092.

[6] W. Bi, K. Zhang, Y. Li, K. Yuan and Y. Wang, "Detection Scheme Against Cyber-Physical Attacks on Load Frequency Control Based on Dynamic Characteristics Analysis," in IEEE Systems Journal, vol. 13, no. 3, pp. 2859-2868, Sept. 2019, doi: 10.1109/JSYST.2019.2911869.

[7] K. Liu, Z. Fan, M. Liu and S. Zhang, "Hybrid Intrusion Detection Method Based on K-Means and CNN for Smart Home," 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 2018, pp. 312-317, doi: 10.1109/CYBER.2018.8688271.

[8] Y. Jin, K. Kakoi, N. Yamai, N. Kitagawa and M. Tomoishi, "A Client Based Anomaly Traffic Detection and Blocking Mechanism by Monitoring DNS Name Resolution with User Alerting Feature," 2018 International Conference on Cyberworlds (CW), 2018, pp. 351-356, doi: 10.1109/CW.2018.00070.

[9] R. Velea and Ş. Drăgan, "CPU/GPU Hybrid Detection for Malware Signatures," 2017 International Conference on Computer and Applications (ICCA), 2017, pp. 85-89, doi: 10.1109/COMAPP.2017.8079736.

[10] S. Merat and W. Almuhtadi, "Artificial intelligence application for improving cyber-security acquirement," 2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), 2015, pp. 1445-1450, doi: 10.1109/CCECE.2015.7129493.

[11] S. Han, M. Xie, H. Chen and Y. Ling, "Intrusion Detection in Cyber-Physical Systems: Techniques and Challenges," in IEEE Systems Journal, vol. 8, no. 4, pp. 1052-1062, Dec. 2014, doi: 10.1109/JSYST.2013.2257594.

[12] M. Bousaaid, T. Ayaou, K. Afdel and P. Estraillier, "Hand gesture detection and recognition in cyber presence interactive system for E-learning," 2014 International Conference on Multimedia Computing and Systems (ICMCS), 2014, pp. 444-447, doi: 10.1109/ICMCS.2014.6911197.