International Journal of Engineering, Management, Humanities and Social Sciences Paradigms (IJEMHS)
(Volume 30, Issue: Special), April 2018
An Indexed and Referred Journal with Impact Factor: 2.75
ISSN (Online): 2347-601X
www.ijemhs.com

Intelligent Intrusion Detection Systems with Machine Learning Models for Detecting Cyber Threats in IoT Networks

Subham Kumar Sahoo, Assistant Professor, Department of Computer Science & Engineering, Krupajal Engineering College, Bhubaneswar
Shantiswarup Mahapatra, UG Student, Department of Computer Science & Engineering, Krupajal Engineering College, Bhubaneswar
Sushree Priyadarshini Dash, UG Student, Department of Computer Science & Engineering, Krupajal Engineering College, Bhubaneswar
Anupurba Das, UG Student, Department of Computer Science & Engineering, Krupajal Engineering College, Bhubaneswar

***Abstract*—As the internet has flourished, the connectivity between devices has increased. This has propelled the Internet of Things technology to reach new heights. But sadly, with these advancements, cyber-attacks have also increased. Consequently, these attacks have posed significant threats to every potential IoT technology user and IoT device. Such malpractices can result in substantial losses of capital and intellectual property alike. The need for developing a robust system for detecting Cyber Security threats has become a crucial operation to prevent such losses. The proposed paper establishes Deep-Learning techniques to help detect malicious attacks on an IoT architecture and prevent unwanted intrusion we've described a single strategy for identifying stolen data from software and malware throughout the Internet of Things (IoT) network in this post. The TensorFlow platform is a prime candidate to develop Deep Learning algorithms to classify stolen programming with source code literary theft. Google Code Jam (GCJ) is an international programming competition administered by google. GCJ collects data annually to examine the true nature of theft of utilizations. The use of the Mailing Dataset is prevalent to gather the malware samples. The use of Deep Learning techniques to detect Cyber Security threats presents a novel yet efficient approach to solving practical issues and shows promise for the future.**

Keywords—Intelligent Intrusion Detection System; Network;IoT;Deep Learning;

I. INTRODUCTION

In recent years, the expansion of the Internet of Things (IoT) has advanced substantially in civilizations all over the world. Consumer demand is expected to fuel the exponential increase in the number of connected IoT devices, with a capacity of $125 billion by 2030[1]. There are many real-world gadgets that are linked to smart city apps, and these devices have a lot of power in the actual world. The advantages of living in a city Large numbers of IoT devices in different service kinds, topologies, and applications [2]. And IoT networks provide a variety of challenges due to the type of protocols that may be used. In charge [3,4]. These Internet-connected devices thus face significant cyber security risks and vulnerabilities. Procedures of integration for the assault of information on the daily actions of citizens. These online dangers may not materialize. Only approved clients and executives (for example, those working on the Miria botnet) have access to the LOT system [5]. There are many real-world gadgets that are linked to smart city apps, and these devices have a lot of power in the actual world. The main issue is figuring out how to To determine whether or not an attack is zero-day, look for a variety of IoT standards in the cloud server farm. IoT frameworks help protect users from potential hazards. The second method is the most effective for detecting cyberattacks. To put it another way, Examples of this kind of attack include IoT malware attacks and others. The Internet of Things (IoT) has already orchestrated the demise of a sophisticated community. The vast majority of Internet of Things (IoT) sensors currently obtain all of their data from cloud storage servers. IoT network systems now have limited resources and capabilities (e.g., smartwatches, drones, etc.). Other IoT network devices (smart bulbs, intelligent locks, etc.) don't utilize these.

New revolutionary technology connects the whole worldto the Internet called The Thing Internet. We may enhance and help our personal, professional, and social lives as a result of the Internet of Things (IoT). As with any network, the Internet of Things (IoT) is susceptible to cyber assaults since it comprises a global system of intelligent devices that does not need human interaction. An intruder detection system is a critical tool for detecting cyber assaults on any network (IDS). Many modern IDS uses a machine-learning network to help them prepare for and identify cyber-attacks. Fog computing addresses IoT framework and scalability limits by enhancing central distributed computing development by connecting mist hubs to IoT curiosity and reducing QoS (Quality of Service) high transmission capacity utilization. Consolidating and achieving IoT arrangements realistically is possible with Mist to-hub. The mist-to-hub model architecture is depicted in Fig. 1 with a corresponding figure that gives data to circulating mists by

bringing IDS measurements, controls, and stocking closer tothe mists' IoT network artefacts.

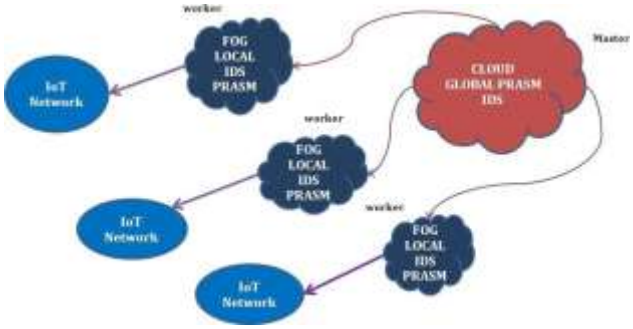

Figure 1: IoT network fog-to-node architecture

Internet of Things (IoT) devices are always connected to thenetwork., a free assault is possible. Contamination with malware and pirated applications may readily target the IoT cloud for criminal usage and protection [6,7]. Software piracy is the unlawful re-use of source codes and the subsequent disguise of those programs as the original work of others. You can use saltine strategies to reproduce the logic of your first programming, and then use saltine strategies to construct the logic of another programming in a different type of source code[8]. Because it permits free access to stolen programming downloads, open-source code,and the development and commercialization of stolen versions, it poses a severe danger to Internet security. This grows exponentially every year, causing enormous economic losses for the IT industry[9]. According to the Business Software Alliance's (BSA) 2016 Annual Report, open theft affects 39% of businesses, resulting in losses of up to $52.2 billion annually. According to various types of analysis, from 5% to 20% of a product's reasoning source codes were found to be appropriated. The duplicated source code must be detected using unoriginality techniques that are as keen as possible. Some recommended ways for written falsification include clone detection, acknowledgment of resemblance with source code, investigation of software defects, and research of the software marker. The most prevalent techniques are structure and text analysis. To understand more about their underlying structure, the source, syntax, graphs, and sub-routine function call graphs are all studied using a structure-based technique. As a result, programming can only be done using a single Word structure. This makes catching a cracker who has reused the logic of the original software in a new programming language more challenging. IoT cloud industrial services may be used to create intelligent software plagiarism and malware detection solutions for securing and protecting intelligent devices.

The Internet of Things (IoT) is a complicated system that has many interconnected components. As a result, it's challenging to keep the IoT system's security requirements current with such a large attack surface. To meet the security requirement, solutions must take a comprehensive approach. However, most Internet of Things (IoT) devices operate without human intervention. As a result, an unauthorized

person may get physical access to these gadgets. A wireless network may allow an intruder to listen in on a communication channel and obtain confidential information. Due to their limited processing and power resources, IoT devices cannot implement sophisticated security mechanisms.[10,11] Due to the IoT system's inclusion in a cyber-physical system and its reliance on constantly adapting and surviving exactly and predictably with safety as a critical priority, complex security structures have emerged in the IoT due to limited computation, communication, and power resources as well as the trustworthiness of its interaction with the physical domain, particularly the behavior of the physical environment in unpredictable and unanticipated situations, in particular, complex security structures have emerged in the IoT due to the limited computation, communication, and The Internet of Things (IoT) ecosystem also introduces new attack vectors. The interconnected and linked ecosystems of the Internet of Things present such attack surfaces.. IoT systems are more vulnerable to hacking than other computer systems as a result, and the conventional approach may not work for them [7].

II. LITERATURE REVIEW

Deep learning techniques were utilized in many areas, such as image processing, voice recognition, healthcare, and so on since they performed better than previous approaches. To identify distributed fog-to-thing assaults, an in-depth learning method is suggested. Cyber security is not appropriate for an extensive IoT network, as this study shows. Cloud IoT networking is inconvenient since it is centralized. Fog-to-node techniques for the IoT network have demonstrated deep awareness in the big data fields, so the vast IoT network that generates vast data may be used. Classifiers such as stack encoder and softmax are used to classify NSLKDD data sets. . On the basis of performance criteria such as accuracy, false alarm levels, and detection rate, the findings are compared to low-level learning models.The fog to node model's distributed parallel processing has also been found to increase attack detection accuracy and efficacy, according to the author.

Using SVM, the author describes how to create a self-teaching deep learning authentication encoder for network intrusion. (Support Vector Machine). As a practical selection method, deep learning reduces training and testing time while improving accuracy in the SVM classification. The suggested technique of binary and multi-class type has been compared with existing shallow machine learning algorithms like J48. The proposed technique outperformed previous approaches in terms of both performance and accuracy. The shallow and deep neural networks have been compared. He also analyzed the results of KDDCup-'99 datasets based on performance criteria including accuracy, precision, and recall, using a learning rate of 0.1 to compare and train the recommended research methodologies. Researchers have discovered that deep learning is a promising new technique for cyber security and that a

three-layer deep-neural

The network model outperformed all others. BLSTMRNN (Bidirectional Long-Term Neural Network) detection is easier using this deep learning model, which compares well to another RNN model called LSTM. The author produced data for four Mirai botnet attack vectors for use in this paper. Using four attack vectors such as Mirai, UDP, DNS, and ack has proven effective. The suggested technique performs well for vectors with Mirai, UDP, and DNS accuracy of 99 percent each; however, further information may be supplied for vectors of ack attack, since it does not perform well comparable. A brief presentation on cyber security machine learning and deep learning approaches, as well as an illustrated literature survey, is offered. The problems of applying machine learning and cyber security in IDS databases were among the topics discussed. They discussed a variety of issues. The author retrained the models and offered lifetime training as a future option after noting the challenge of training both approaches for frequently updated network data. This was agreed upon.

PROPOSED METHODOLOGY

A. Data Pre-processing

To make virus identification a matter of image classification, the colour graphics are constructed from raw binary data. Using state-of-the-art approaches like those provided in this paper, a 256-color grayscale image is created from malware binary files.. Reverse engineering does not need the use of special tools like a disassembler or decompiler. Colored pictures, as opposed to grey ones with just 256 colours, have greater usefulness. Malware pictures with more incredible features may be preferred within a particular family grouping. Most machine learning-based virus detection techniques historically had superior outcomes when using grayscale images. The color pictures have been converted to grayscale and visualized using extraction methods to identify the malware. Methods of function reduction enhance classification efficiency by reducing the collection of functionalities. Researchers discovered that when detecting malware using color images, machine learning algorithms don't perform any better since they generate exponential numbers. Deep learning algorithms can't handle large malware datasets since they can't utilize filters to reduce noise automatically. Deep learning techniques are used to improve the quality of the color images.

A malicious binary file is converted into a colour picture in four phases. Converting raw binary data into hexadecimal strings is the initial step (0-15). Second, an 8-bit slice is created by splitting a hexadecimal stream into eight segments and multiplying each segment by an unsigned integer (0-255). The 8-bit vector is then turned into a two-dimensional space matrix in the third phase. Adding a third element, every eight-bit integer is made up of two-dimensional red, green, and blue areas.

B. Deep Convolutional Neural Network

The Deep Convolutional Neural Network should conduct a thorough investigation of malware data (DCNN). The DCNN has five modules. The input layer is where the neural network model's training photos are collected. To minimize noise and increase signal quality, the first step is to apply a convolution layer. Second, the pooling layer is used to reduce the quantity of overhead data while still maintaining critical data. Second, the fully integrated layer reduces the two-dimensional dynamic array to one dimension and feeds it into the classification. Finally, malware families are discovered using an image classification method.

A. Convolution Layer

The relevant characteristics are retrieved by employing a convolution layer on top of a picture with parameter values. It's a convolution sheet if the definition is invariant, the rotation is constant, and the uniform size. As a result, there is less over-fitting, and the core design now has a clearer vision. The following equation shows that the coevolutionary layer's input is multiplied.

$$x^l_J = f(\sum_{i=M} x_j^{j-i} * k_{ij}^l + b_j^l)$$

j

B. Pooling Layer

The bundling layer provides two methods, i.e., maximum and average bundling, and is often referred to as a sub-sampling layer. As far as feasible, the picture distortion is reduced by using reverse propagation, which does not disrupt the original transmission. In addition, it decreases the functionality factor while increasing the suggested DCNN functionality, as indicated in the equation below.

$$x^l_J = f(\text{down}\,(x_j^{l-1}) + b_j^l$$

C. Software Piracy Threat Detection

Obtaining pirated software is the main objective of the suggested deep learning strategy. A sophisticated learning method has been devised to detect plagiarism in various media. Figure 2 shows how the he cracker takes advantage of a pirated copy of the plagiarised version—the software's logic in this iteration. Tokenizing source codes in preprocessing stages reduces the amount of data and extracts valuable information for subsequent steps. The Keras API Deep Learning Algorithm on the TensorFlow platform is used to detect plagiarism using a deleted meaningful characteristic. The data, which originates from 400 distinct sources and 100 programmers, is compiled by GCJ. The information will be saved in the GCJ database for future reference. The capacity to track Internet of Things (IoT) devices might aid in the prevention of zero-day attacks. Machine learning and deep learning classify how IoT components and devices interact with one another in the IoT

ecosystem as normal or abnormal behavior. (ML/DL). This data may be gathered and analyzed to determine how the system usually interacts so that malicious behavior can be detected early on. As a result, machine learning and deep learning approaches may be useful in forecasting new attacks, which are frequently mutations of earlier assaults, because they may intelligently foresee future unknown attacks by learning from current instances of former episodes. That's why DL/ML- enabled security-based intelligence is needed in IoT systems rather than simply enabling secure connectivity between devices.

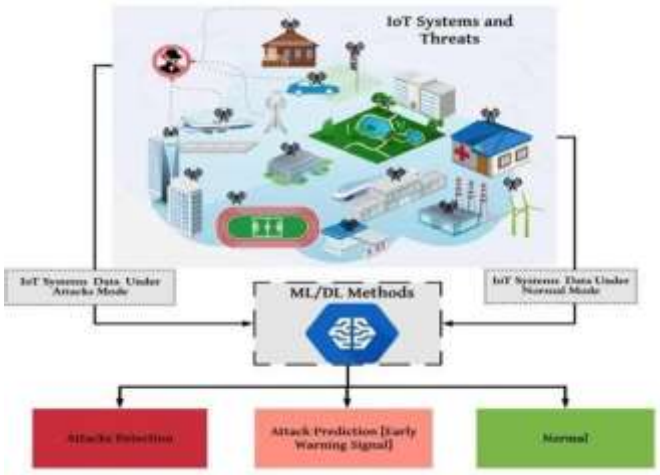

Figure 2: IoT cyber security threat model architecture

I DEEP LEARNING WITH TENSORFLOW FRAMEWORK

It's possible to build a wide range of machine types and deep learning algorithms using the TensorFlow API. Complex computations, data generation, and system monitoring are possible with the many layer types this offers. TensorFlow frames are used in many programming languages to identify related source codes using the advanced learning technique. For the identification of pirated devices, several principles are then gathered. Weighting values are utilized as an entrée into the deep learning model. The dense layer is sometimes called input and put as a fully linked layer. It is necessary to construct three dense layers, each 100 cells thick. A data input variable is processed at the lowest level of the stack.

Previously linked layers provide and receive information to each neuron. The 4th thick layer targets the copied code in the output variable.

Drop out layer improves deep learning inactivation and failure, optimization, and learning error. Dropout solves the overlay issue as well. The rectification technique (ReLU) makes it possible to use input variables to get produced data patterns. If it's the affirmative half of an argument, we can write it mathematically as:

$$f(x) = x^{+} = \max(0, x)$$

III. CONCLUSION

Researchers in this article look at the possibility of an intelligent vulnerability detection system, which would be an improvement above traditional intrusion detection systems (IDS). This article compares and analyses four different deep education models employing master learning algorithms. The CNN+LSTM hybrid model surpasses existing deep-learning models and methodologies by 97.16 percent, according to our findings. In terms of data gathering, the MLP deep learning model falls short. MLP's accuracy is less than 95.00%, while the other two deep-learning methods' accuracy is far above that mark. Here, the idea of an intelligent vulnerability detection system to enhance traditional IDS in intelligently designed IoT applications is explored. This article presents four alternative deep education models and contrasts using master learning algorithms. Compared to existing deep-learning models and methods, ore accuracy is achieved using the CNN+LSTM hybrid model (97.16 percent). In terms of data gathering, the MLP deep learning model falls short. MLP's accuracy is less than 95.00%, while the other two deep-learning methods' accuracy is far above that mark.

REFERENCES

[1] J. Howell. The number of connected IoT devices will surge to 125 billion by 2030, last accessed: 11/07/2018,IHS Markit says - his technology. Available online: https://technology.ihs.com/596542 / ,
[2] E. Borgia, "The internet of things vision: Key features, applications, and open issues," vol. 54, pp. 1-31, 2014.,Computer Communications,
[3] F. Restuccia, and T. Melodia, "Securing the internet of things: New perspectives and research challenges," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 1-14, 2018.
[4] J. A. Stankovic, "Research Directions for the internet of things," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 3-9, 2014.
[5] M. Antonakakis, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric,
[6] J. A. Halderman, M. Kallitsis et al., "Understanding the Mirai botnet;' in USENIX Security Symposium, 2017, pp. 1092-1110.
[7] B. B. Zarpelao, R. S. Miani, C. T. Kawakami, and S. C. de Alvarenga, "A survey of intrusion detection in the internet of things;' Journal of Network and Computer Applications, vol. 84, pp. 25-37, 2017.
[8] J. Santos, P. Leroux, T. Wauters, and F. D. Turck, "Anomaly detection for smart city applications over 5g low power wide area networks," in NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, 2018, pp. 1-9.
[9] A. Yousefpour, G. Ishigaki, "Fog computing: Towards minimizing delay in the internet of things," in Edge Computing (EDGE), 2017 IEEE International Conference on. IEEE, 2017, pp. 17-24.
[10] A. Abeshu and N. Chilarnkurti, "Deep learning: the frontier for distributed attack detection in fog-to-things computing," IEEE Communications Magazine, vol. 56, no. 2, pp. 169-175, 2018.
[11] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," Computer Networks, vol. 57, no. 10, pp. 2266-2279, 2013.
[12] A.Vanamala Kumar, G. Lalitha Kumari, Y. Surekha “ANALYSIS OF MACHINE LEARNING BASED CYBER SECURITY” in Test Engineering and Management , Vol-83, ISSN-0193-4120 Apr-2020.