JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

IOT DEEP LEARNING BASED DETECTION OF CYBER SECURITY THREATS

[1]V.Navya Sree, [2]K. Hemanthi, [3]K. Swarupa Rani

[1] Associate Professor, Potti Sriramulu Chalavadi MallikarjunaRao College of Engineering and Technology, Vijayawada. navya.sree@gmail.com
[2]Assistant Professor, Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram hemanthikallam@gmail.com
[3]Prasad V Potluri Siddhartha Institute of Technology, Vijayawada swarupapvpsit@gmail.com

Received: 16 March 2020 Revised and Accepted: 16 June 2020

ABSTRACT

IoT is a revolutionary technology that brings together the world's living and non-living things. IoT deployment is growing rapidly but cybersecurity remains a loophole, so that it is likely to lead to numerous cyber-attacks and it is very important for the achievement of each system that the system is totally secure something else the user might not use the technology. DDoS assault has recently targeted a large number of IoT networks and contributed to massive losses. In this article we have proposed a consolidated methodology for the identification of pilfered records from programming and malware all through the IoT organize. It is proposed to characterize pilfered programming utilizing source code literary theft utilizing the TensorFlow profound neural system. To channel boisterous information and to additionally improve the significance of every token concerning the counterfeiting of the source code, the tokenization and gauging techniques. This method is likewise used to distinguish literary theft in source code. Google Code Jam (GCJ) accumulates the dataset to explore the robbery of utilizations. Furthermore, the profound neural system is utilized to recognize vindictive contaminations by color image representation in the IoT network. The samples of malware are collected from the experimental Maling dataset. The findings show that the classification efficiency of the approach being proposed for evaluating cyber security risks in IoT is higher than state-of-the-art methods.
KEYWORDS: IoT, cyber security, Google Code Jam (GCJ)

1. INTRODUCTION

The Internet of Things (IoT) proliferation has increased considerably in societies worldwide in recent years. As the amount of interconnected loT equipment reached $27 billion in 2017, these IoT equipment will increase exponentially with consumer demand, with a capacity expected to reach $125 billion by 2030[1]. Specific intelligent city applications are linked to massive, real-world devices, which in reality have very significant urban benefits[2]. The large numbers of loT devices in various service types, architectures, applications and protocols (e.g. Wifi, wired, mobile, cellular, Bluetooth) contribute to the challenge of potential IoT network management [3], [4]. There are therefore serious cyber security threats and vulnerabilities to these Internet integration protocols in order to attack information about everyday citizen activities. Such cyber threats may not be allowed on the LOT system unless the authorized client or executive (for example Miria botnet) knows about them[5].

In brilliant city ventures, there are two major security issues. The principal challenge is the means by which to distinguish zero-day assaults from an assortment of loT conventions in the keen cloud server farm if significant dangers are avoided IoT frameworks. The second is the best approach to identify digital assaults [5] shrewdly (for example IoT malware assaults, and so on.) in the IoT organize previously destroying an intelligent community. Today, most loT sensors collect all information through the huge amount of data collected on cloud servers. At present, the IOT network systems have minimal resources and less features (e.g. smart watches, intelligent lights, intelligent locks, etc.) which are not used by IOT network applications.

The Thing Internet is the new innovative technology that links the entire planet with the Web. IoT technology ensures that our personal , professional and social lives are enhanced and helped[1]. IoT consists of a worldwide system of intelligent objects without any human intervention, which is awesome because like every other network it is vulnerable to cyber attacks. An important technique for the detection of cyber attacks on any

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

network is the intruder detection system (IDS). Many of the new IDS bases their preparation and cyber attack detection algorithm on a machine-learning network. Fog computing improves central distributed computing development in which the scattered hubs in mist are associated with curios in the IoT framework and addresses the scalabling bottlenecks, lower QoS (Quality of Service) high transmission capacity use and distributed computing 's high idleness requirement. Mist to-hub is reasonable to consolidate and achieve the IoT arranges practically. The figure. 1 shows the mist to-hub model design with appropriated equal figuring giving data to circulated mists by supplying IDS measurements, controls, and stocking closer to IoT network artifacts.

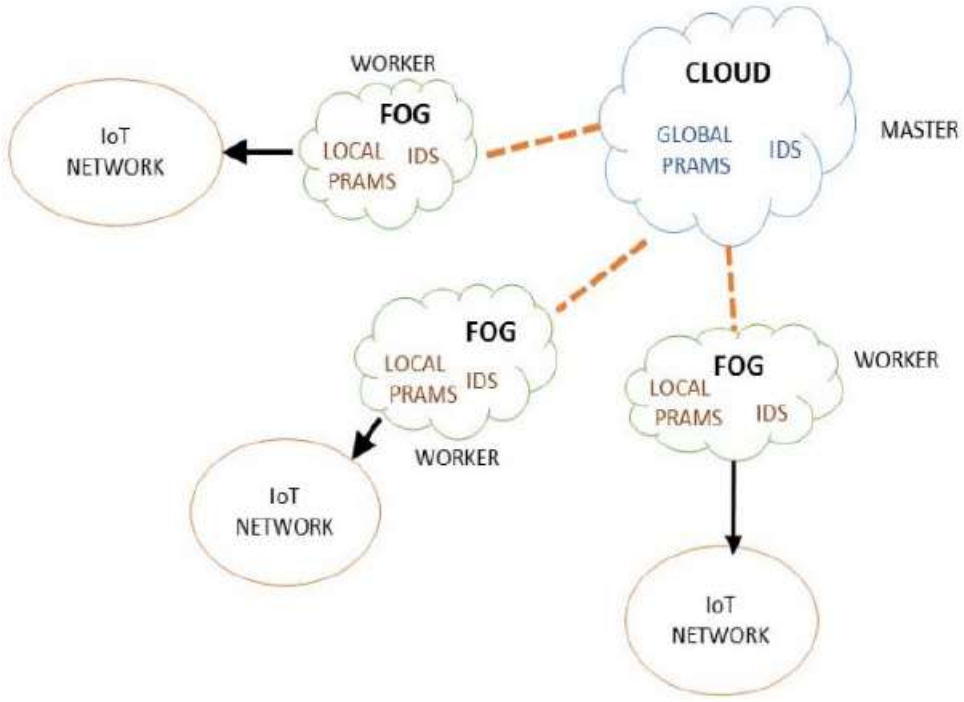

Fig. 1. Fog-to-Node architecture for IoT network

Due to always accessible on the network, IoT devices can be used for a free attack. The IoT-cloud can easily be targeted for malicious use and protection through contamination of malware and pirated apps [4, 5]. The piracy of software is software development through the illegal reuse of source codes and disguising them as the first form of others' work. The saltine can utilize figuring out strategies to duplicate the rationale of the first programming, and afterward structure a similar rationale in another sort of source code [6]. It represents a significant danger to Internet assurance, empowering access to unhindered pilfered programming downloads, open source code and pilfered variants advancement and advertisement. This is rising exponentially every year and causes the tech industry tremendous economic loss[7]. The 2016 Report of the Business Software Alliance (BSA) indicates the open theft proportion is around 39 percent, with yearly business harms of up to $52.2 billion. Numerous sorts of examination indicated that every product contains appropriated rationale source codes extending from 5 % to 20% [8, 9]. Keen unoriginality methods are required if the copied source code is to be caught in pilfered programming. A few procedures of written falsification are recommended, including clone identification , acknowledgment of similitudes with source code,analysis of software glitches and investigation of software marker. Structure and text analysis are mainly such techniques. The structure-based approach investigates the fundamental structure of source, syntax, graph and sub-routine function call graphs. It is therefore restricted to a particular one Word structure for programming. Therefore, if a cracker reuse the logic of the original software to another form of programming language, specific structural activity makes it difficult to capture it. IoT cloud industrial services can be used in the design of intelligent software plagiarism and malware detection technologies for securing and protecting intelligent devices.

2. LITERATURE REVIEW

In the fields of image processing, speech recognition, healthcare and so on, deep learning methods were used, which offered better performance compared with other methods. An in-depth learning approach is proposed in order to detect distributed fog-to-thing attacks. This paper illustrates the inconvenience of cloud IoT networking as it is centralized and not suitable for the big IoT network because it requires the edge of the network to be

processed for cyber security. The field of big data areas has shown profound awareness, so the massive IoT network which creates massive data can be fog-to-node methods for the IoT network. This work is done by a stacked car encoder along with softmax as a classifier on NSLKDD data sets and is compared with a low level model of learning based on performance metrics including precision, false warning levels and detection rate. As an increase in accuracy and efficacité of attack detection, the author has also shown the efficacy of the distributed parallel computing used in fog to node model.

In combination with SVM (Support Vector Machine), the author proposes a self-teaching deep learning authentication encoder for network intrusion. Deep learning is unattended to reduce the training and testing time and improve efficiency by increasing the precision of the SVM classification as a functional selection technique. In addition to contrasting it with other shallow machine learning algorithms like J48, the author has contrasted the proposed method of binary and multi-class classification. In terms of performance, as with accuracy compared to other proposed methods , the proposed method provided better results. A comparison was suggested between the shallow and the deep neural network. In addition he compared the results achieved by KDDCup-'99 'datasets based on performance measures, such as accuracy, precisions and recall, to compare and train the proposed study methods at a learning rate of 0.1. In their findings, deep learning is exciting technology for the field of cyber security and the deep-neural network model carried out with 3 layers has been best performed compared to other models. A deep learning model is proposed for the detection of a Bidirectional Long Term Neural Network (BLSTMRNN) and comparable to LSTM, an RNN model. For this work , the author generated data for four attack vectors used by mirai botnet. Our approach on four attack vector including mirai, udp, dns and ack has been checked and validated. The approach proposed shows good results for vectors with accuracies of mirai, udp and dns, respectively 99%, 98%, and 98%, but does not perform well for vectors of ack attack comparatively, for which more details can be suggested.

An illustrated literature survey and a brief lesson on cyber security machine learning and profound learning methods is provided. They addressed several problems in IDS databases and the complexities of using machine learning and cyber security. They addressed several problems. The Author raised the issue of training both of the techniques for updating network data very quickly, and this led to the retraining of the models so that lifelong training was suggested as future work.

3. PROPOSED METHODOLOGY FOR CYBER SECURITY THREATS AND PROTECTION

MALWARE THREAT DETECTION

1) DATA PREPROCESSING

The images in color are generated in raw binary files to turn a question of image classification into malware detection. The proposed research distinguished between state-of-the-art methods in which malware binary files transform into a 256 color grayscale image. It depends not on any devices such as disassembler / decompiler for reverse engineering. In contrast to gray images with only 256 colours, colored images will access better functionality. Within the grouping of malware families, the stronger characteristics of malware images may be superior. Most methods of malware detection based on machine learning algorithms have previously produced better results with gray imagery. The color images are turned to visualizing the grayscale and extraction techniques have then been used to classify the type of malware. The efficiency of classification is improved by methods of function reduction to decrease functionality collection. The findings showed that machine learning algorithms are not a better choice when using color pictures to detect malware since they produce exponential values. Huge malware datasets exceed the deep learning algorithms, as these approaches can use filters to automatically decrease noise. The color pictures use deep learning methods to produce better results.

Malware binary file conversion into a color image requires four steps. The first is to generate hexadecimal strings (0-15) from raw binary files. The second is to divide a hexadecimal stream into an 8-bit slice, which calculates each eight-bit segment by an unsigned integer (0-255). Thirdly, the 8-bit vector is then turned into a double-dimensional space matrix. Third, every eight-bit integer is composed of red , green and blue shaded colors and creates two-dimensional spaces. The entire data pre-processing steps are shown in Fig.1.

2) DEEP CONVOLUTIONAL NEURAL NETWORK

It is suggested to conduct a detailed malware data analysis by the Deep Convolutionary Neural Network (DCNN). As shown in Figure 1, the DCNN contains five modules. For the built neural network model, the Input layer is used to obtain training images. The first step is to lower the noise and boost signal characteristics using a convolution layer. Secondly, the pooling layer is used to reduce the overhead data, which preserves valuable information. Second, to translate the two-dynamic array into one-dimensional and input it to the particular

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

classification, the fully integrated layer is used. Third, the malware families are marked by means of the respective images classification.

3) CONVOLUTION LAYER

Through the the image of parameters by using the convolution layer, the appropriate features are extracted. Convolution sheet, i.e., invariance of definition, invariance of rotation and invariance of size. This reduces the over-fit and gives the central design the overall concept. As seen in the following equation, the input of the convolutionary layer is multiplied.

$$x_j^l = f\left(\sum_{i \in M_j} x_j^{l-1} * k_{ij}^l + b_j^l\right)$$

POOLING LAYER

In general, the term sub-sampling layer is used for bundling layer, which offers two approaches , i.e. maximum and average bundling. The reverse propagation does not interrupt it and reduces the effect of image distortion as much as possible. It also reduces the functionality factor and increases the proposed functioning of DCNN as shown in the following equation

$$x_j^l = f\left(down\left(x_j^{l-1}\right) + b_j^l\right)$$

SOFTWARE PIRACY THREAT DETECTION

The primary goal of the proposed profound learning approach is to procure pirated software from different sources. A profound learning technique has been developed to identify plagiarism in different outlets. The pirated copy of the plagiarized version is used by the cracker, as shown in Figure 2. The logic of the software used in this version. To reduce the size of the data and extract useful features for further steps, the source codes are tokenized in preprocessing steps. The Keras API Deep Learning Algorithm TensorFlow Platform is used to detect plagiarism between different kinds of source codes using extracted meaningful properties. The data collection consists of 400 separate sources from 100 programmers, and is compiled by GCJ. The dataset is stored in the GCJ database.

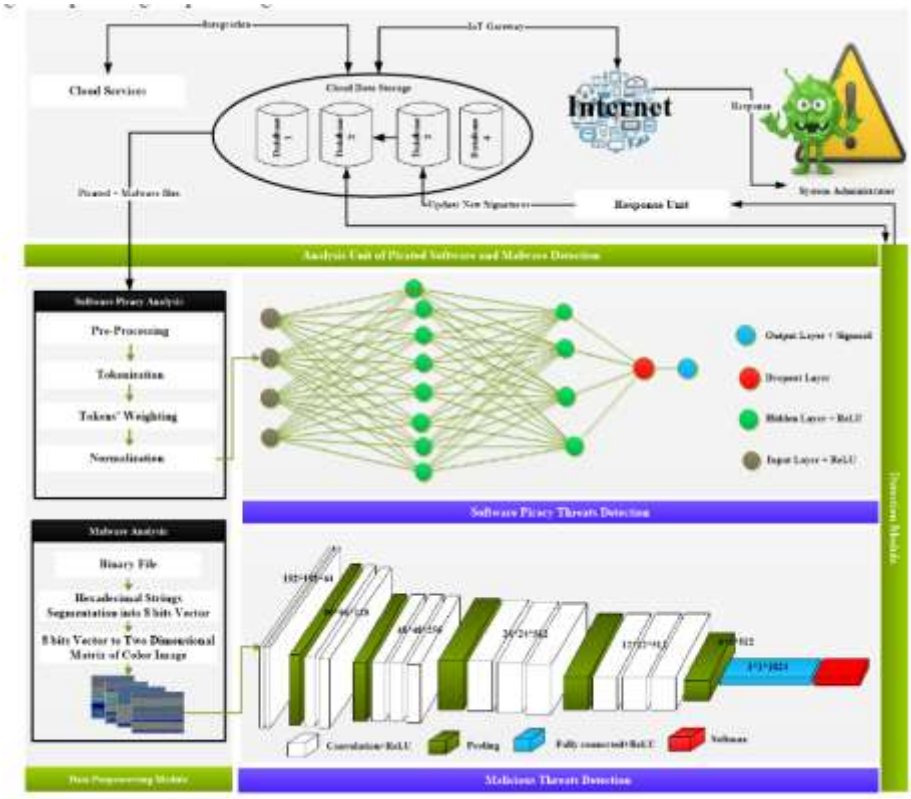

FIGURE 2: Architecture model for Cyber Security Threats prediction in IoT

ISSN- 2394-5125 VOL 7, ISSUE 18, 2020

DEEP LEARNING WITH TENSORFLOW FRAMEWOKR

The TensorFlow is a learning machine that is used in a compact environment for high-level computation. Through calling TensorFlow 's special application programming interface (API), we can implement various types of machines and deep learning algorithms. This has various layer types that can be configured for complex calculations, data creation and monitoring of the state of each system. The profound learning method is built to classify related source codes using TensorFlow frames in different types of programming languages. Different codes are then collected for pirated device detection. As an entry into the deep learning model, weighting values are used. The dense layer is also known as a completely connected input and input layer. Three dense layers of 100, 50 and 30, respectively, are built. The first layer is for processing data in the form of an input variable. Every neuron receives information from previously connected layers. For the output variable to target the plagiarized code, the 4th thick layer is used.

In the sense of activation and failure, optimization and learning error, the deep learning approach is improved by drop out layer. The problem of overlay is also solved with dropout. Input variables are enabled by the rectification method (ReLU) to obtain obtained data patterns. In the following equation, it is expressed mathematically as the positive part of its argument:

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

4. CONCLUSION

In this paper, the dream of a clever vulnerability detection system to improve conventional IDS in cleverly built loT applications is studied. In this article we suggested four different models of deep education and compared them with algorithms for master learning. We find that the CNN+LSTM hybrid model performs with 97.16 percent accuracy better than other deep-learning models and algorithms. The MLP deep learning model is also less profound in the data collection. With the exception of MLP, we find that the accuracy achieved using these three deep-learning approaches is more than 95.00 percent.

5. REFERENCES

[l] J. Howell. Number of connected iot devices will surge to 125 billion by 2030, ihs markit says - ihs technology. [Online]. Available: https://technology.ihs.com/596542/, last accessed: 11/07/2018.

[2] E. Borgia, "The internet of things vision: Key features, applications and open issues," *Computer Communications*, vol. 54, pp. 1-31, 2014.

[3] F. Restuccia, S. D'Oro, 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, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis *et al.*, "Understanding the mirai botnet;' in *USENIX Security Symposium*, 2017, pp. 1092-1110.

[6] B. B. Zarpelao, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in internet of things;' *Journal of Network and Computer Applications*, vol. 84, pp. 25-37, 2017.

[7] J. Santos, P. Leroux, T. Wauters, B. Volckaert, 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.

[8] A. Yousefpour, G. Ishigaki, and J. P. Jue, "Fog computing: Towards minimizing delay in the internet of things," in *Edge Computing (EDGE), 2017 IEEE International Conference on.* IEEE, 2017, pp. 17-24.

[9] 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.

[10] A.Vanamala Kumar, B. Narasimha Swamy, G. Lalitha Kumari, Y. Surekha *"ANALYSIS OF MACHINE LEARNING BASED CYBER SECURITY" in* Test Engineering and Management ,Vol-83, ISSN-0193-4120 Apr-2020

[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.