IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

A Solution to Detecting Botnets using Convolutional Neural Networks and Support Vector Machine Algorithms

Vipul Jha[1], Omkar Katule[2], Tanvi Bajad[3], Shreyas Agadi[4], Priyanka Bendale[5]
Students, Department of Computer Engineering[1,2,3,4]
Faculty, Department of Computer Engineering[5]
Sinhgad College of Engineering, Pune, Maharashtra, India

Abstract: *A botnet is an Internet-connected network of devices and nodes that spread malware software, such as Trojan horses, viruses, and worms. Recently, numerous approaches for detecting and combating mobile malware have been developed. Our model, on the other hand, is distinct from previous models. We're using a dataset that we found on the Kaggle website. The findings we obtained were obtained using machine learning techniques such as CNN and SVM. We have a range of attack or non-attack scenarios, as well as any subtypes that may occur. The proposed system is a web-based tool that predicts App/URL botnets with high accuracy.*

Keywords: Convolutional Neural Network, Support Vector Machine, Botnet, Attacks, Web application

I. INTRODUCTION

The term 'botnet' is derived from the expressions "roBOT NETwork", alluding to a victim being controlled by an attacker. Botnet utilization has risen impressively lately. Botnets are an assortment of PCs associated with the web with a lot of transfer speed and computational limit. The assailant, frequently known as the botmaster, can order enormous organizations of botnets from different areas to start assaults. Circulated refusal of administration (DDoS) assaults, email spam, key logging, and secret key breaking are signs of a botnet. Botnets are presently one of the most genuine dangers to the Internet. Botnets have an assortment of components that make them practically special with regards to design, capacities, and specialized execution. A bot herder (or a botmaster), or more than one order and control servers, and different, maybe in millions, controlled hubs are generally present to control the bot. A botnet is a kind of web worm that comprises of contaminated hubs that execute orders while endeavoring to stay away from discovery by anti-malware programming. The botmaster is the botnet's preeminent authority. The assistance is right now under danger. As a rule, a client can deal with a piece of the botnet hubs. The client's guidance set is commonly a subset of the full guidance set. In truth, the genuine attacker is whoever controls the botnet (the botmaster) at a specific time.

II. LITERATURE SURVEY

Sarnsuwan et al. [1] proposed a strategy to recognize the malware by using data mining, where it incorporates the use of data assessment technique for tracking down dark data by significant associations and models in gigantic educational lists. These mechanical assemblies can consolidate genuine models, mathematical estimations and AI techniques. So that, data mining contains more than social affair and administering data. Sarnsuwan et al. [22] used three data mining computations that are C4.5 Decision Tree, Random Forest, and Bayesian association. NBD [2-4] essentially understands the traffic network in the request and control articulation of each botnet, considering the way that the social characteristics are extraordinary in connection between two articulations. NBD focuses for the most part investigate two sorts of authoritative direct: stream features and the speed of frustration affiliation. The computations that depend upon used stream incorporates that consolidate the amount of uplink and downlink of data packages, the ordinary length of uplink and downlink of data divides, amount of uplink and downlink, transmission bytes, the term time of data stream, the best length of downlink and uplink of data packages, the full scale length of stacked data bundles in a stream, the speed of the length of data packages in uplink and downlink, and the typical length of downlink and uplink of data bundles. As of now, investigators are adding

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

mind association and AI to NBD to perceive dark botnet traffic organization. Furthermore, this methodology is a hot investigation point in the examination of botnet traffic and area [5].

Lashkari et al. [6] ran the pernicious malwares and harmless applications on authentic telephones, making an effort not to adjust the real runtime lead of forefront malware tests that can assist with recognizing the environment of the emulator. To get an exhaustive layout of our malware tests, Lashkari et al. [28] had made a specific circumstance for each malware order. The structure approach similarly described three states of data variety to beat the secretiveness of bleeding edge malware. The structure approach contained three stages.

In [7], the results showed that the AFSA shown amazing execution in work improvement, and the ability of applying the AFSA in smoothing out issue were uncovered too. Furthermore, in [8], the researchers proposed a sort of component determination and backspread network for botnet ID; in any case, using an AFSA got together with a SVM classifier could yield common execution. In this survey, an organized model was suggested joining an AFSA computation and a SVM. The proposed procedure was used to recognize the essential features choosing the case of a botnet.

III. BACKGROUND

3.1 Convolutional Neural Network

Convolutional neural network (CNN) is a class of deep learning neural networks. It can be thought of as an algorithm of machine learning that takes an input image, assigns biases or learnable weights to the multiple aspects in the image, and is then able to differentiate them from each other. It is crucial to understand that Artificial Neural Networks, is not capable of extracting features from the object or image. Due to this incapability, a combination of convolution and pooling layers is introduced to enhance it. In the same manner, the pooling layers and the convolution aren't enough to perform classification; hence a fully connected Neural Network comes into picture.

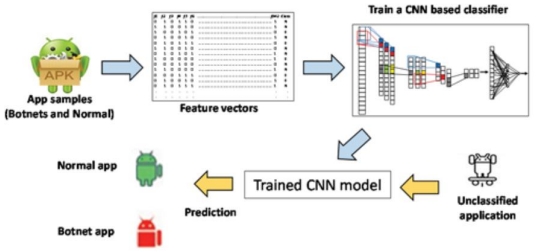

Figure: Training and prediction with the CNN-based model for botnet detection phase of the system.

3.2 Support Vector Machine

A support vector machine (SVM) is a supervised machine learning algorithm that is generally used for both regression and classification problems. But, SVM are majorly used in classification problems. SVM is developed on the idea of locating a hyperplane that can best divide a dataset into two classes or partitions. These Support vectors are the data points that are nearest to the hyperplane, the points of a dataset that, when deleted, would simultaneously alter the position of the class-dividing hyperplane. Due to this, they are considered the critical elements of a dataset. The distance between the nearest data point and the dividing hyperplane from either of the sets is known as the margin. The main aim of SVM is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, while resulting in a higher chance of new input data being classified correctly.

3.3 Problem Statement of System

With the rising notoriety of cell phones and Internet surfing, especially those in view of Android and different Websites, there has been a huge expansion in the downloading and sharing of outsider projects and client created content, delivering handsets and frameworks helpless against various kinds of malwares. As far as security and finances, a ton of exploration is required in this subject. Malware makers or programmers, have tracked down simpler and productive strategies to go after working frameworks that are generally utilized, are open-source, and the ones that don't forestall or deny the establishment of programming from any hurtful outsider source. We are fostering an electronic application to beat some of such similar issues and identify the bots.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

IV. Implementation Details Of Module

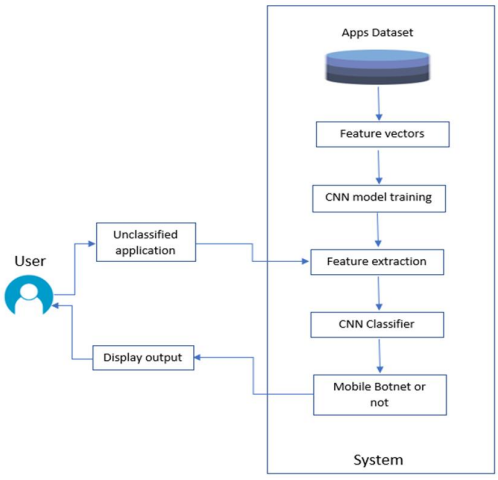

Figure: System Architecture for Botnet Detection

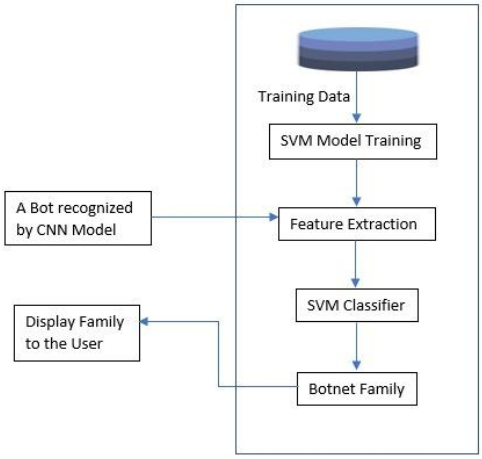

Figure: System Architecture for Botnet Family Classification

The proposed system contains following:

4.1 Pre-processing

The system will load the data, check for cleanliness, and then trim and clean given dataset for analysis and explorations. There might be instances when the data which has been collected might contain missing values, leading to inaccurate and inconsistent results. Hence, to prevent such inaccuracy, data needs to be pre-processed so as to improve the performance and efficiency of the algorithm along with gaining better results. This includes conversion of variables whenever required, and removing any outliers if present.

4.2 Building the Classification Model

Predicting the sentimental analysis by using supervised machine learning techniques, like SVM algorithm, makes the prediction model highly effective, accurate and efficient with the major reason being that SVM algorithm provides better results in classification problems.

4.3 Working

The web frontend contains HTML structure where user enters the information/URL of apk/web-app. The web frontend transfers this information given by the user, which is then shipped off to the backend framework of the system. The script

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

running in the backend retrieves the information passed by frontend. The backend processes that information with the assistance of a few strategies and techniques prompting to different elements being anticipated, for example, whether botnet is identified or not, and if it is, then what family does it belong to. The backend sends this result information to the frontend after the grouping step and contingent upon the outcome, the system returns the answer and the result is displayed to the user. Creating a pop-up window to alert the user in case botnet is detected, and sending a mail to a separate developer to inform of such detected botnets are some new features that have been added to this system.

4.4 Dataset

To build and train our proposed model, we are using two datasets; one for botnet detection(data.csv), and other for botnet family classification(family-db.csv). The number of rows and columns in data.csv are around 2,00,000 and 2 respectively. Similarly, the number of rows and columns in family-db.csv are around 75,000 and 4 respectively.

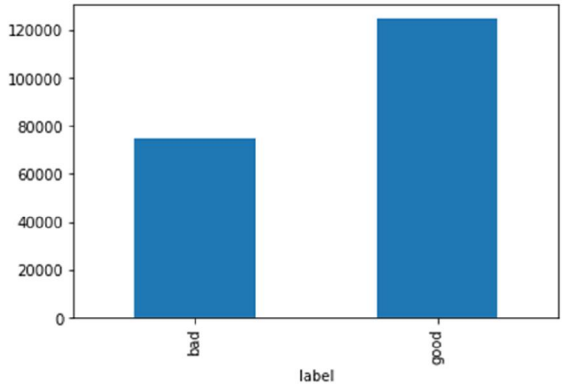

Figure: Frequency count for good and bad links in data.csv

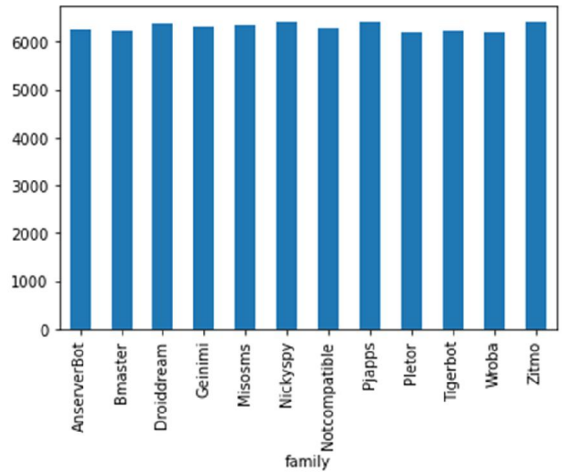

Figure: Frequency count for botnet families of bot affected links in family-db.csv

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

V. RESULTS

Figure: Home Page

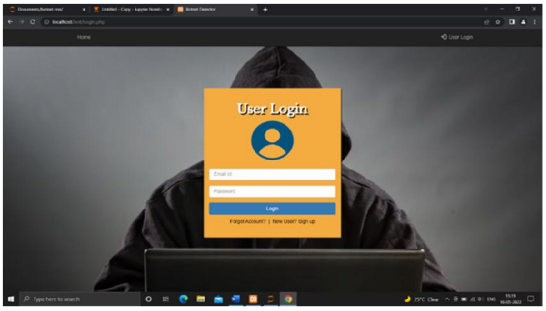

Figure: Login Page

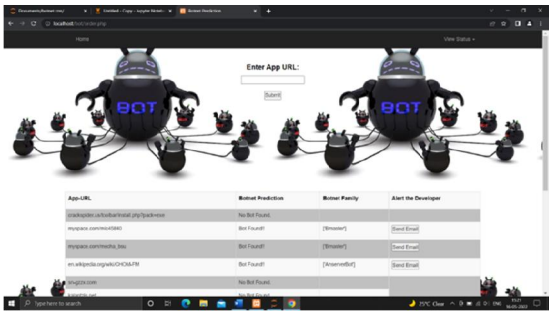

Figure: Check Botnet

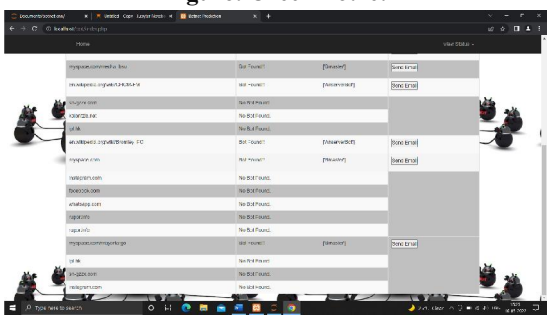

Figure: Prediction Page

VI. CONCLUSION

Even though many different approaches and techniques have been attempted and implemented till date, but most of them have their own share of disadvantages such as: Signature based systems have inability to effectively detect previously unseen botnets, or features extracted from apps only include app permissions, or the computational overhead for converting features

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

vectors into images prior to model training, etc. Our proposed model takes a step ahead to deal with such problems. Our system uses the Convolutional Neural Networks to predict if a URL is botnet affected, and Support Vector Machines to predict its family. Using these two specific machine learning techniques not only helped us build an efficient system, but also proved that our model outperforms several popular machine learning classifiers evaluated on the same dataset. The results indicate that our proposed CNN-SVM based model can be used to detect new, previously unseen botnets more accurately than the other models, along with throwing some light on the families the bots belong to. For future work, we will aim to improve the model training process by adding some key influencing parameters that will jointly result in a more optimal performing CNN-SVM model.

REFERENCES

[1]. Sarnsuwan, N. Charnsripinyo, C., Wattanapongsakorn, N., "A new approach for internet worm detection and classification", In 6th International Conference on Networked Computing, 2012.

[2]. Shanthi, K., Seenivasan, D., "Detection of botnet by analyzing network traffic flow characteristics using open-source tools". In Proceedings of the 9th IEEE International Conference on Intelligent Systems and Control (ISCO '15), India, 2015

[3]. Kirubavathi, G., Anitha, R., "Botnet detection via mining of traffic flow characteristics", Computers and Electrical Engineering, 2016

[4]. Zhang, J., Perdisci, R., Lee, W., Luo, X., Sarfraz, U., "Building a scalable system for stealthy P2Pbotnet detection", IEEE Transactions on Information Forensics and Security, 2015

[5]. Chen, R., Niu, W., Zhang, X., Zhuo, Z., Lv, F., "An Effective conversation-based botnet detection method. Mathematical Problems in Engineering", 2017

[6]. Lashkari, A., Draper-Gil, G., Mamun, M., Ghorbani, "Characterization of traffic using time-based features". In the proceeding of the 3rd International Conference on Information System Security and Privacy, 2017

[7]. H. Chen, S. Wang, J. Li, and Y. Li, "A hybrid of artificial fish swarm algorithm and particle swarm optimization for feedforward neural network training", in Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering, 2007.

[8]. J. L. Liao and K. C. Lin, "A Study of Feature Selection Integrated with Back- Propagation Network for Botnet Detection", National Chung Hsing University, Taichung, Taiwan, 2013.

[9]. Ahmad Karim, Rosli Salleh and Syed Adeel Ali Shah "DeDroid: A Mobile Botnet Detection Approach Based on Static Analysis", IEEE

[10]. Vikramajeet Khatri, "Mobile Guard Demo", IEEE Zubaile Abdullah and Madihah Mohd Saudi "Mobile Botnet Detection: Proof of Concept", IEEE

[11]. AV-Comparatives Security Survey, 2019, Security Survey2019en.pdf

[12]. Amro, B.: "Personal Mobile Malware Guard PMMG: a mobile malware detection technique based on user's preferences". IJCSNS International Journal of Computer Science and Network Security, Vol. 18, No. 1, pp. 18–24 (2018)

[13]. Idrees, F., Rajarajan, M., Conti, M., Chen, T., Rahulamathavan, Y.: "Pindroid: a novel android malware detection system using ensemble learning methods". Computers Security, Vol. 68, pp. 36–46 (2017)

[14]. Chaba, S., Kumar, R., Pant, R., Dave, M.: "Malware Detection Approach for Android systems Using System Call Logs", arXiv preprint arXiv:1709.0880 (2017)

[15]. McLaughlin, N., Martinez del Rincon, J., Kang, B, et al.: "Deep android malware detection". In Proc. of the Seventh ACM on Conference on Data and Application Security and Privacy, pp. 301–308 (2017)