Analysis of Botnet using SVM

Arundhathi Das[1]

[1]*DDMCA Student, Department of MCA, Sree Narayana Guru Institute of Science and Technology, Kerala, India*

---***---

Abstract - *Botnets are common nowadays with the extension of the web and commonly occur in many cyber-attacks turns to serious threats to network services and users' properties. With the rapid development of the net of Things applications, the botnet can easily make use of IoT devices for larger-scale attacks. Domain name system & DNS & is extensively used by the botnet to organize the connection between bots and their corresponding command-and-control. In order to avoid the track of the C&C through the DNS information, some sophisticated schemes are used by the botnet is a typical one. In this paper, the activities of botnet domain names which just use the fast-flux as the connection method between bots and C&C, are deeply analyzed from multiple aspects. The work of this paper aims to provide guidance for future botnet detection supported real statics and experiments.*

KeyWords: Botnet Detection, machine learning

1.INTRODUCTION

Botnets are group of devices (e.g., computers) which are infected with particular malicious software and then these devices can be remotely run to launch large-scale malicious attacks for example. With the rapid development and wide deployment of the Internet of Things, Various insecure devices will be connected into the public Internet and they can be easily controlled by the botnets. Then IoT based botnet becomes one of the significant security issues in the future Internet[2].

Domain Name System service provider was criticised by the Distributed Denial of service. The attacking traffic extent 1.2Tbps and brought down many notable websites including Twitter, Netflix, CNN and many others in Europe and the US. This event was caused by a botnet which was construct out from a ragtag collection of IoT related devices. The botnet was consist of all manner of Internet-connected devices from home routers to digital video recorders.

On one hand, the IoT devices are always not be installed with sophisticated software to protect the botnet, and on the other hand, the IoT devices are always not been updated in time and then the vulnerabilities or bugs can be simply used by the botnet. Using dynamic DNS, a botmaster which control the whole botnet can keep running the Command-and Control server even when its current name cannot be accessed. Typically, the botmaster cause a series of domain names based on an algorithm and it can dynamically updates the location of C&C server with moving the active DNS entry. Then the bots can always request the active domain name to access the C&C server to listen to the command even when the C&C server vary its location. Excepting the enlarge detection difficulty under dynamic DNS, bots will request the new IP address of the C&C server with the fresh domain name after the bots uncouple from the old C&C server, and this will result in an increase of DNS queries. The conventional schemes to expose a botnet using Intrusion Detection Systems or scanners are not effective enough.

2. RELATED WORKS

In 2007, the first seminar on botnets was hold. Since then, the community keeps trying to introduce different solutions in order to detect the botnet exactly and efficiently and some typical systems also have been established (e.g. Bot Hunter by Gu et al. [1]. The Internet Corporation for Assigned Names and Numbers (ICANN) Security and Stability Advisory Committee (SSAC) [6] has also liberated the report which presented the similar DNS technique used by botnet. In the recent years, some researchers concentrated on the analysis of botnet traffic in order to reveal the botnets timely and dispose them more efficiently. The detection schemes can be divided into two types: passive detection schemes and active detection schemes. The passive detection schemes need the process of accumulation, detection and reaction of evidence and they may handle the following typical approaches: 1) Packet inspection approaches: based on the pre-defined model of the botnet, these approaches will equivalent the characteristics of packets such as port number and payload. 2) DNS-based approaches: based on the fact that bots may drive DNS queries to detect and access the C&C server, DNS monitoring is used as an productive manner to detect botnet [4]–[5]. However, these past works mainly target on monitor-based traffic features related with botnet. The DNS behaviors should be evaluate from the specific authoritative server because the traffic information is assigned in multiple resolvers. In addition, the features from different aspects, such as name structure and registered Resource Records (RRs), should be examined simultaneously. And that is just the inspiration of our work. 3) Honeypot: it is represented as an environment where vulnerabilities can be deliberately affect in order to recognize attacks and intrusions [7]. Active detection schemes are mainly used to adapt or cutoff botnet behavior and there are the following two types of active detection approaches: 1)Sinkholing: it is an approach to cutoff the controller and bots for example with the hijack of the botnet domain names. 2) Infiltration: based on the extraction of the botnet protocol messages, the protocol reverse-engineering can be performed accordingly and then the fence against botnet can be executed actively. In our work, we concentrate on the deep analysis of the botnet domain name characteristics and then attempt to detect the botnet domain names precisely and accurately. In this way, the botnet domain names can be refined out and there solution can be immediately suspended. And the work in this paper is an expansion based on our previous achievement with extended feature analysis and expanded dataset in order to validate the effectiveness of these features and detect the botnet domain name more effectively[10].

3. METHODOLOGY

Here we are using SVM classification. The aim of the support vector machine algorithm is to find a hyper plane in an N-dimensional space (N — the number of features) that distinctly categorize the data points. To disparate the two classes of data points, there are many possible hyper planes that could be chosen. Our purpose is to find a plane that has the maximum margin, i.e. the ultimate distance between data points of both classes. Enlarge the margin distance provides some reinforcement so that future data points can be classified with more confidence. Hyperplanes are decision boundaries that used to classify the data points. Data points dropping on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane build upon the number of features. If the number of input features is 2,then the hyperplane is due a line. If the number of input features is 3, then the hyperplane shift a two-dimensional plane. Support vectors are data points that are closer to the hyperplane and guide the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will vary the position of the hyperplane. These are the points that help us construct our SVM. The botnet detection steps are shown in Fig-1.

3.1 MODULES

Preprocessing

The dataset consist of various botnet data's. The preprocessing steps helps to convert the data into system understandable format.

Architecture

The algorithm used is SVM. SVM works by calculating data to a high-dimensional feature space so that data points can be classified, even when the data are not otherwise linearly separable. A separator between the categories is launch, then the data are transformed in such a way that the separator could be drawn as a hyperplane.

Feature Extraction

Every classes exhibit their own features. In this features each family features can be extracted. Every classes differ in their datas.

Training

The training of the system can be carried out on this phase. The extracted features and pattern (algorithm) can help the system to learn information from the available dataset. After training process, a file can be created which is called as model file. This model file include content which are learned by the system.

Testing

In testing phase, the model file can be loaded first. Then user can input link. The given data can be preprocessed by the system and predict the botnet based on the model file.

Login and Registration

The user can register to the system with basic details. The user can login to the system using username and password.

3.2 DATASET

- Here dataset is Text. Text consist of

1. Attack
2. Normal

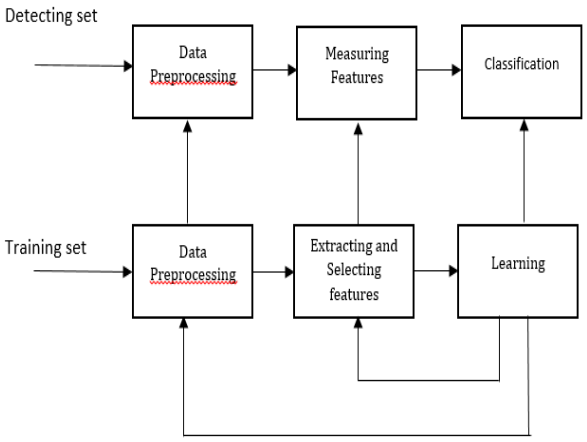

Fig 1 – Botnet Detection Model

4. CONCLUSION

In this paper, a feature selection method for detecting botnet viruses is proposed, which is the SVM method. Here, we deeply examined the characteristics of botnet domain name resolution activities from multiple aspects. The term 'bot' is used when we have to represent some automated tasks that are carry out without user intervention. But as this term is used regarding hacking, to specify a new breed of malicious threats, we will learn about it with every detail. The botnet, from the cybercrime point of view, is an automated cyber army with few computers connected to the Internet without their owner's knowledge. The results displayed that only four DNS querying types were used. we will deploy this model under larger DNS data size and use more sophisticated learning algorithms to detect the malicious domain names for protecting critical infrastructure [8], [9].in order to reduce the time required to carry out the classification, a parallel SVM algorithm may be used. This system helps to find the attacking links and helps to understand the links better.

REFERENCES

[1] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee, "BotHunter: Detecting malware infection through IDS-driven dialog correlation," in Proc. 16th USENIX Secur. Symp., 2007, pp. 1–16.

[2] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other Botnets," Computer, vol. 50, no. 7, pp. 80–84, 2017.

[3]N.Hoque,D.K.Bhattacharyya,andJ.K.Kalita,"Botnetin DDoSattacks: Trends and challenges," IEEE Commun. Surveys Tuts., vol. 17, no. 4, pp. 2242–2270, 4th Quart., 2015.

[4] J. Nazario and T. Holz, "As the net churns: Fast-flux botnet observations," in Proc. 3rd Int. Conf. Malicious Unwanted Softw., Oct. 2008, pp. 24–31.

[5] S. Lysenko, O. Pomorova, O. Savenko,A. Kryshchuk, and K. B. Lysenko, "DNS-basedanti-evasiontechniqueforbotnetsdetection,"inProc.8thInt. Conf.Intell.DataAcquisitionAdv.Comput.Syst.Technol.Appl.(IDAACS), Sep. 2015, pp. 453–458.

[6] Security Stability Advisory Committee (SSAC), SSAC Advisory Fast-Flux Hosting DNS, ICANN, Los Angeles, CA, USA, 2008.

[7] C. Costarella, S. Chung, and B. Endicott-Popovsky, "Hardening a honeynet against honeypot-aware botnet attacks: Toward secure cloud," in Proc. 3rd Int. Conf. Cloud Secur. Manage., 2015, p. 135.

[8] L. Maglaras, K. H. Kim, H. Janicke, M. A. Ferrag, S. Rallis, P. Fragkou, A. Maglaras, and T. J. Cruz, "Cyber security of critical infrastructures," ICT Exp., vol. 4, no. 1, pp. 42–45, 2018.

[9] A. Shenfield, D. Day, and A. Ayesh, "Intelligent intrusion detection systems using artificial neural networks," ICT Express, vol. 4, no. 2, pp. 95–99, Jun. 2018.

[10] Bijalwan, Anchit & Chand, Nanak & Pilli, Emmanuel & Challa, Rama. (2016). Botnet Analysis Using Ensemble Classifier. Perspectives in Science. 8. 10.1016/j.pisc.2016.05.008.