 ISSN: 2454-4248
94 – 95

Marionettwork Security

Pooja Marsonia
Final year , CSE
ITM Universe
Vadodara, India
pooja.marsonia@gmail.com

Nupur Mishra
Final year , CSE
ITM Universe
Vadodara, India
mishranupur2601@gmail.com

Khyati Pandya
Final year , CSE
ITM Universe
Vadodara, India
khyatipandya39@gmail.com

Shailvee Vaghela
Final year , CSE
ITM Universe
Vadodara, India
Shailyvaghela.sv@gmail.com

Pradeep Laxkar
Head , CSE Department
ITM Universe
Vadodara, India
pradeep.laxkar@gmail.com

Abstract— Malware is one of the most terrible and major security threats facing the Internet today. It has been surveyed that around 71% of attacks are done using Trojan horses and DOS attacks. So, Trojan horses and DOS attacks are the main focus of the system. Trojan horse are the most dangerous malwares, it never shows their presence on the pc or laptop it works at the back end and provides all the credentials and sensitive information to the attacker, it also opens door for other malwares like, viruses, worms, spyware, adware etc. Report informs about currently working software in market and comparison between new system developed and current systems and describes how newly developed system is better than current systems. This report gives best option for real time network security analyzing.

Keywords— *clustering, feature extraction, k-means, machine learning, spark, Trojan horse.*

*****

I. INTRODUCTION

As some surveys and researches shows that threats of cyber-attacks for any industries or organizations are increasing at alarming rates. 42% of organizations claims that they are very likely to experience cyber-attack in 2017. Cyber security facts were proved very important for the problem definition selection. Project domain was decided after going through many surveys and researches done in last 2 to 3 years.Purpose of the project is to develop a system that will run on the server and it can analyze real time data packets from the network for any malicious activity.Previous systems are able to scan real time network data packets but with poor detection rate and but not as fast as our system, some previously working systems are not even able to identify new instances of malwares[6-8]. Using high speed Apache Spark, an open-source cluster-computing framework makes the system 100 times faster than systems working on Hadoop and 500 times faster than other systems. Machine learning will provide enhancement in the detection rates as well as it will update the model as it learns about new Trojans and use that knowledge in further scanning. After combining these two important modules we get best features of our system those are high efficiency, high detection rates, low false positive rate, and high speed.

As the phrase says that "Prevention Is Better than Cure", in this project we intent to make a system which will analyze real time network data packets and identify the malicious packets with high speed and competence and prevent any cyber threat on the network.

II. LITERRATURE REVIEW

The paper "Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Feature" [1] has proposed malware detection system using deep neural network. System contains three main components. First is feature extraction which extracts 4 different types of features, second is deep neural network which is composed of 1 input layer 2 hidden layer and 1 output layer and third is score calibration which calibrates score using Bayesian rule that can be realistically interpreted as approximating the probability that the file is actually malicious. It has usable detection rate with extremely low false positive rate that too on real world training data. System has 95% detection rate at 0.1% false positive rate and it can detect newly created or previously undetected malwares.

The paper "Unknown Malware Detection Using Network Traffic Classification" [2] have presented an end-to-end supervised based system for detecting malware by analyzing network traffic. Evaluation of the system shows that many unknown malware incidents could have been detected at least a month before their static rules were introduced to either the Snort [9] or Suricata [10] systems. Features were collected from four different layers those are transaction, session, and flow and conversation window. Feature extractor was implemented on top of WIRESHARK and output is feature vector in form of CSV format file that are sent to classification algorithm. Three different classification algorithms are used Naïve Bayes, decision tree (J48) and Random Forest. Random forest algorithm was stable for most of experiments. CFS (Correlation Feature Selection) algorithm has been used for feature selection using weka selection library. Real time data has also been used for the test but results were very poor.

"Malware Detection with Deep Neural Network Using Process Behavior" [3] present malware process detection method based on process behavior in possible infected terminals using deep neural network. Recurrent neural network (RNN) was trained from log files for feature extraction from process behavior and Concurrent neural

 ISSN: 2454-4248
94 – 95

network (CNN) was trained for classifying feature images which were generated by the extracted features from RNN. Best results for AUC was 0.96.

Paper "An Improved Cluster Analysis Algorithm Using for Network Traffic Flow" [4] has researched An improved cluster analysis algorithm of combining SVM with supervised subset density clustering is proposed in this paper, and minimize the training set of SVM by means of clustering is researched. Results show that the algorithm reduces the iteration time of the whole training process without compromising the accuracy and generalization capacity of the algorithm obviously. K-means is widely applied in many fields. But its performance is sensitive to the initial center positions. Most improved K-means algorithms rely on sample density to select initial centers

"Detection of Trojan Horses by the Analysis of System Behavior and Data Packets" [5] paper propose process explorer which is used to identify the malicious executables and to determine whether they are Trojans or not. As Trojan needs to call different API functions and process explores gives alert to the user whenever API call is made by other executables. Process Explorer tool is very similar to task manager, it includes information such as: (i) if the process is packed or not, (ii) if the process resides in auto start location, (iii) the process-timeline, (iv)if the process is digitally verified and (v) whether the process contains any (dll) injected into core system processes. This software can only work on client's system and not on server.

III. SYSTEM ARCHITECTURE

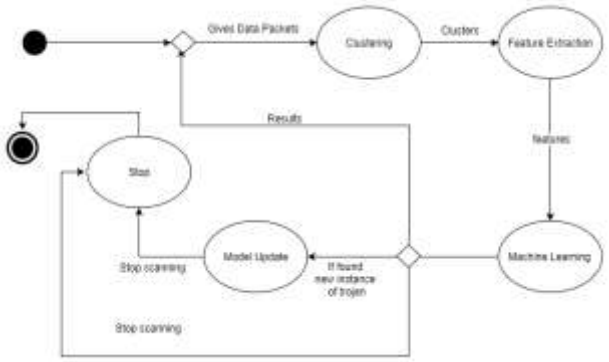

Fig.1: State transiction diagram

IV. PROPOSED SYSTEM

Proposed system will collect real time data from network and scan it for any threats like Trojan Horse or DOS attacks. It will eliminate or clean the infected node. System contains main 4 modules: (i) Clustering (ii) Feature Extraction (iii) Machine Learning

(iv) Model Update.

Clustering: This module will operate on real time data packets which are collected and apply clustering on those data packets. Apache spark will be used for clustering because of the lightning fast speed and supportiveness of multiple languages. Apache spark is 50 times faster than Hadoop and it can work with data streams effectively and easily. K-means algorithm will be used for clustering [11].

Feature Extraction: This module will analyze clusters and will extract features from those clusters for further examinations. Feature extraction involves reducing the amount of resources required to describe a large set of data.

Machine Learning: In this module features will be given to the machine learning model for identifying If the cluster is malicious or not. By machine learning only system will be able to detect new samples of malware here Trojan and use them in future.

Model Update: whenever newly designed malware is detected, its signature or features are added to the model and model is updated so that it can be used in future for identifying those malware here Trojans again. Results will be stored using mongo dB and NoSQL. Database will hold records about scans, Trojans detected and Nodes scanned.

V. CONCLUSIONS

We have concluded after some market surveys and similar project analysis that this system will provide more security then the systems currently working for organizations and institutes, this system is able to provide more security and that too with lightning fast speed, the system itself is as secure as a skull so that no one can crack it's security. As the system uses machine learning with spark it will be able to identify new Trojan instances and store them in database for further reference.

REFERENCES

[1] J. Saxe and K. Berline "Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Features" published in Malicious and Unwanted Software (MALWARE), 2015 10th International Conference on 20-22 Oct. 2015.

[2] D. Bekerman, B. Sharpira, L. Rokach and A. Bar "Unknown malware detection using network traffic classification" published in Communications and Network Security (CNS), 2015 IEEE Conference on 28-30 September 2015.

[3] S. Tobiyama, Y. Yamaguchi, H. Shimada, T. Ikuse and T. Yagi"Malware Detection with Deep Neural Network Using Process Behavior" in 2016 IEEE 40th Annual Computer Software and Applications Conference.

[4] S. Yong, S. Zhen-Chao, Z. Ran, Z. Geng and L.Shi-Dong "An Improved Cluster Analysis Algorithm Using for Network Traffic Flow" published in Computer Science & Education (ICCSE), 2015 10th International Conference on 22-24 july 2015.

[5] V. K. Gudipati, A. Vetwal, V. Kmar, A. Adeniyi and A. Abuzneid "Detection of Trojan Horses by the Analysis of System Behavior and Data Packets" published on Systems, Applications and Technology Conference (LISAT), 2015 IEEE Long Island on 1 may 2015.

[6] Z. Shumei and J. Yanru, "The Model of Trojan Horse Detection System Based on Behavior Analysis," in Multimedia Technology (ICMT), 2010 International Conference on, 2010, pp. 1-4.

[7] Y. Liu, L. Zhang, J. Liang, S. Qu and Z. Ni "Detecting Trojan Horses based on System Behavior Using Machine Learning Method" published in Machine Learning and Cybernetics (ICMLC), 2010 International Conference on 11-14 July 2010

[8] Q. Jie, Y. Huijuan, S. Qun, and Y. Fuliang, "A Trojan Horse Detection Technology Based on Behavior Analysis," in Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th International Conference on, 2010, pp. 1-4.

[9] "k-means clustering": https://en.wikipedia.org/wiki/K-means_clustering

[10] "Snort": https://www.snort.org/.

[11] "Suricata": http://suricata-ids.org/.