IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Malware Detection using Data Mining Naïve Bayesian Classification Technique with Worm Dataset

Osama Mohammed Qasim[1], Karim Hashim Al-Saedi[2]

Postgraduate Candidate, Informatics Institute for Postgraduate Studies,

Commission for Computer & Informatics, Baghdad, Iraq[1]

Al-Mustansirya University Collage, Computer Science, Baghdad, Iraq[2]

Abstract: According to numerous increasing of worm malware in the networks nowadays, it became a serious danger that threatens our computers. Networks attackers did these attacks by designing the worms. A designed system model is needed to defy these threats, prevent it from multiplying and spreading through the network, and harm our computers. In this paper, we designed a detection system model for this issue. The designed system detects the worm malware that depends on the information of the dataset that is taken from Kaspersky company website, the system will receive the input package and then analyze it, the Naïve Bayesian classification technique will start to work and begin to classify the package, by using the data mining Naïve Bayesian classification technique, the system worked fast and gained great results in detecting the worm. By applying the Naïve Bayesian classification technique using its probability mathematical equations for both threat data and benign data, the technique will detect the malware and classify data whether it was threat or benign. The results of the experiments were 95% of worm detection accuracy and 98% of detection rate with 21% false positives, which makes it more accurate and effective to detect the worm malware by using the proposed dataset for this work.

Keywords: Network Security, Worm Detection, Malware, Naïve Bayesian, Data Mining.

I. INTRODUCTION

Network security is an important branch of computer science that protect the stored data in the computers, which it connected together by one network. Recently, the knowledge of the network became developed and common in our world, network attackers are increasing every day, and their threats are evolving as well [1]. Network security is a very important matter for foundations like universities, special projects and corporations. These foundations can supply many important functions for the countries safety [2]. Nowadays, the online services are very popular for the users. The users now can communicate with each other and share information, and knowledge among each other. Now these services are less expensive and more cooperative by using the Information Technology (IT) associations, and Internet Service Providers (ISPs) [3]. Malware may put network at danger. Malware is a program can install in the network and electronic devices like computers, smart phones and tablets that connected in the network. It damages these devices by accessing it illegitimately and destroy its personal data and information; for an example: Adware could do the malicious work [4]. Malware is the most dangerous threats to the networks. The malware could take many forms to do its attack, it always come as package and try to access to the network. Every day new types and forms of the malware are found. The malware programmers always make decisions about protecting their malware from anti malware programs like Kaspersky, McAfee, NOD, Norton and many anti viruses programs that we use in our PCs [5][6]. The malware threats are too many; the most serious one is worm. The worm can replicate itself and spread through the network very fast [7]. Nowadays, the world faces a major problem called worm especially the facilities and the network users. In spite of the detection techniques ability in detection it still have difficulties in detecting these worms [8]. Here the detection techniques role comes. Detection techniques are the most effective defense against the malware in the network. The malware defenders are the anti-viruses these days, it can detect the malware signature and prevent it from doing its malicious work [9][10]. In the Operating System (OS) keeping the safety, confidentiality and availability is very important and a hard task, because of the difficulties and impendences that the network faces in securing the data and information inside the network and keeping it safe from outside attacks, so it is very important these networks have a defense line against the outside attacks [11]. Electronic devices such as computers and smart phones; the malware can affect them by a huge number of malware that spread in these devices. Kaspersky Labs discovered many new types of mobile malware in 2015, it was 884.774 kinds, and this is three times more than in 2014 discovery which it were 295.539 types [12]. The most dangerous threat of malware is the worm; there are many types of it like Cross-Site Scripting (XXS) and java script that threatens the social networks like Twitter, Facebook and LinkedIn [13].

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

II. RELATED WORK

In 2013 Xiao, et al. proposed a system to detect new types of malware; they used API (Application Programing Interface) with OOA (Object-Oriented Associate) mining algorithm, which it used for association rules for detection matter. After many tests the proposed system showed great results and new different types of malware have been discovered [14].

Younghee Park, et al. presented a method to detect the malware that depend on the known behavior of graph that show the behavior of malware instances. This method uses clustering in detection; it clusters the set of separated behavior graphs and make a single behavior graph. The results included numerous detection rates with almost 0% of false alarms [15].

In 2015 Chun-I-Fan, et al. used methods of hooking to track dynamic signatures that the malware tries to hide by using data mining techniques. The technique detected different behaviors of malware and they compare it with the benign data. The detection rate was 95% of 80 attributes, which makes the technique they used increased detection rate with decreasing complexity [16].

In 2016 James B. Fraley, et al. discovered polymorphic types of malware threats by using data mining and feature extraction techniques. Their experimental results were 0.0030 of false low false positive rates, and the high true positive were 0.9978 for the unknown threats in a file with 4k size [17].

Tobias Wuchner, et al. in 2016 made a new study for detecting malware graph by applying frequency based graph-mining techniques to extract the malware features from malware graphs, which increased the detection efficiency of malware by more than 600% [18].

3. The Proposed System:

In this thesis, a detection system for worm malware has been proposed. The proposed system works as follow; it import the dataset and extract the features from it. Then it will classify the data whether it was threat or benign by using the Naïve Bayesian (NB) data-mining algorithm and make the detection according to this classification and then it will show the results as it shown in figure below:

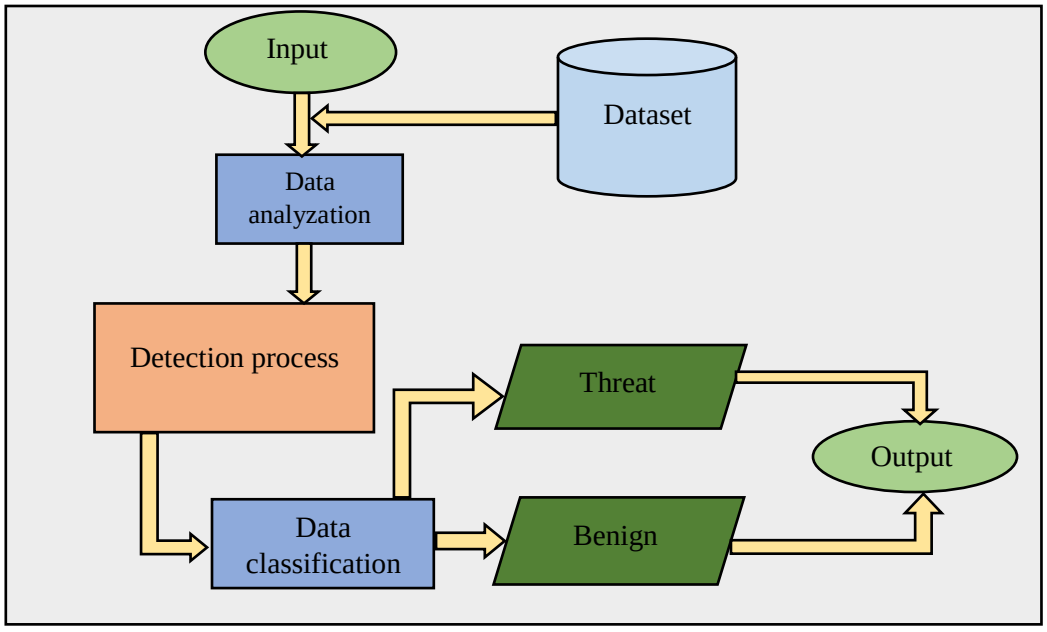

Fig 1 The Proposed System block diagram

The NB algorithm will analyze the data and check it whether it was benign or threat to make the detection and showing the results of this work. The NB algorithm classify the data by using probability equations:

$$p(C_k \mid x) = \frac{p(C_k) \cdot p(x \mid C_k)}{p(x)} \qquad \text{... Equation 1}$$

Where:

- $p(C_k \mid x)$ is posterior.
- $p(C_k)$ is prior.
- $p(x \mid C_k)$ is likelihood.
- $p(x)$ is evidence.

After having information from the dataset the data will be calculated by the probability equation to make the detection and classify the data. The calculations would be for the threat and benign data by forming the data into

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

values 1's and 0's the 1 is for the threat data and the 0 is for the benign ones. Seventeen types of threat data have been discovered during this work, they were all different types of malware; all of them were worm type. The algorithm has two counters (i and j) first one for threat and the other for benign data, when the system detects abnormal behavior the counter (i) will increment its value by 1; the same goes for j counter but for the benign data when normal behavior detected it will increment its value by 1. The equation will calculate the values of i and j and find how much i and j incremented and then it will classify the data and show the results of threat and benign. The accuracy of the NB algorithm increased for the detection by using the proposed dataset. So the system became more accurate and effective in detection.

The proposed algorithm pseudo code will be shown below and it explain all of written above:

Input: insert data to the algorithm to make classification
Output: classified data that has been given to the algorithm

1 $A_c = 0$, $A_{jc} = 0$ // initialize the item set
2 for i = 1 to A do
3 $c = b_i$
4 $A_c = A_c + 1$
5 for j = 1 to D do
6 if $x_{ij} = 1$ then // true = 1, false = 0
7 $A_{jc} = A_{jc} \mid 1$
8 else
9 $A_{jc} \mid 0$
10 end if
11 end for
12 end for
13 $P_c = A_c / A$, $P_{jc} = A_{jc} / A_c$

Algorithm 1 NB steps

III. RESULTS AND CONCLUSIONS DISCUSSION

The results of this work were gained from experiments on the dataset used for this work. We gained 95% of accuracy and 98% of detection rate with 21% of false alarms. The worm malware can use the computer ports as gateways to access the computer and invade the network. Our networks need protection from outside attackers to defy against their attacks, so strong systems are needed to detect and prevent the malware from breaking through the networks and do its malicious work.

REFERENCES

[1] A. . Fallis, “Neural Network Model,” J. Chem. Inf. Model., vol. 53, no. 9, pp. 1689–1699, 2013.
[2] M. S. Kacar and K. Oztoprak, “Network Security Scoring,” 2017 IEEE 11th Int. Conf. Semant. Comput., pp. 477–481, 2017.
[3] H. Li, P. W. C. Prasad, A. Alsadoon, L. Pham, and A. Elchouemi, “An improvement of Backbone Network security using DMVPN over an EZVPN structure,” pp. 14–16, 2016.
[4] A. Juan et al., “Native Malware Detection in Smartphones with Android OS Using Static Analysis , Feature Selection and Ensemble Classifiers,” pp. 67–74, 2016.
[5] E. Bocchi et al., “MAGMA network behavior classifier for malware traffic,” Comput. Networks, vol. 0, pp. 1–15, 2016.
[6] L. Jones, A. Sellers, and M. Carlisle, “CARDINAL : Similarity Analysis to Defeat Malware Compiler Variations,” 2016 11th Int. Conf. Malicious Unwanted Softw., 2015.
[7] L. Xue and Z. Hu, “Research of Worm Intrusion Detection Algorithm Based on Statistical Classification Technology,” 2015 8th Int. Symp. Comput. Intell. Des., pp. 413–416, 2015.
[8] M. Martens, H. Asghari, M. van Eeten, and P. Van Mieghem, “A time-dependent SIS-model for long-term computer worm evolution,” 2016 IEEE Conf. Commun. Netw. Secur., pp. 207–215, 2016.
[9] Y. Ye, T. Li, Q. Jiang, and Y. Wang, “CIMDS: Adapting postprocessing techniques of associative classification for malware detection,” IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., vol. 40, no. 3, pp. 298–307, 2010.
[10] C. J. Fung, D. Y. Lam, and R. Boutaba, “RevMatch : An Efficient and Robust Decision Model for Collaborative Malware Detection,” no. Cmd, 2014.
[11] W. Mao, Z. Cai, D. Towsley, Q. Feng, and X. Guan, “Security importance assessment for system objects and malware detection,” Comput. Secur., 2017.
[12] M. Ping, B. Alsulami, and S. Mancoridis, “On the Effectiveness of Application Characteristics in the Automatic Classification of Malware on Smartphones,” pp. 75–82, 2016.
[13] S. Gupta and B. B. Gupta, “Alleviating the proliferation of JavaScript worms from online social network in cloud platforms,” 2016 7th Int. Conf. Inf. Commun. Syst. ICICS 2016, pp. 246–251, 2016.
[14] D. Chen, “Detecting Hiding Malicious Website Using Network Traffic Mining Approach,” 2010 2nd Int. Conference Educ. Technol. Comput., 2010.
[15] Y. Park, D. S. Reeves, and M. Stamp, “Deriving common malware behavior through graph clustering,” Comput. Secur., vol. 39, pp. 419–430, 2013.
[16] C.-I. Fan, H.-W. Hsiao, C.-H. Chou, and Y.-F. Tseng, “Malware Detection Systems Based on API Log Data Mining,” 2015 IEEE 39th Annu. Comput. Softw. Appl. Conf., vol. 3, pp. 255–260, 2015.
[17] J. B. Fraley and M. Figueroa, “Polymorphic malware detection using topological feature extraction with data mining,” SoutheastCon 2016, pp. 1–7, 2016.
[18] O. Paper, “Leveraging Compression-based Graph Mining for Behaviorbased Malware Detection,” vol. XX, no. c, pp. 1–14, 2014.