ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.5

Detecting Worms Based on Data Mining Classification Technique

Jitendra Jain[1], Dr. Parashu Ram Pal[2]
Research Scholar[1], Professor[2]
Department of Faculty of Computer Science[1]
Pacific Academy of Higher Education & Research University, Udaipur, Rajasthan, India[1]
Lakshmi Narain College of Technology, Bhopal, M.P, India[2]

Abstract:
A computer worm is a program that self- propagates across a network exploiting security or policy flaws in widely-used services .One of the essential requirements of cyber security is to provide information security. Cyber security involves protecting information by preventing, detecting, and responding to attacks. Cyber security is also known to as information technology security. The main focus of cyber security is protection of computers, networks, programs and data from unauthorized access, change or destruction. Internet worms pose a serious threat to computer security. Data mining tools and techniques are used by many researchers in the recent years to detect unknown worms. A number of classifiers have been built for very high accuracy rates. By using data mining, malwares can be analyzed and detected. In this paper Naïve Bayesian classifies a unknown worm to a particular type is used. The three classes to classify the unknown worm will be used. Some parameters which will be based on experimental analysis have been used.

Keywords: Cyber Security, its threads and security, viruses and its types, testing parameters.

I. INTRODUCTION

Cyber security (also known as information security) is the practice of protecting information from unauthorized user, disclosure, disruption, modification or destruction. Computer and communication systems repeatedly suffer security and privacy attacks. Information security technology is an essential component for protecting public and private computing infrastructures. Advancement in technology is making people more oriented towards frequent use of information technology resulting in more usage of online resources which in turn is giving rise to a large number of security threats to these resources.

II. SECURITY THREADS AND ATTACKS

Common ingredients of cyber crime are the malicious code such as viruses, worms, and Trojan horses. Threat that attempts to modify a system, its resources, its data or its operations are known as active threads .Threat that attempts to learn or make use of information from a system but does not attempt to alter the system passive thread. Some Common threads are shown with the help of figure 1. They are follows:

a. Viruses: - Computer virus is a self replicating code (including possibly evolved copies of it) that infects other executable programs. Viruses usually need human intervention for replication and execution. Some of them are Boot Sector Virus, File Virus Macro Virus.

b. Worms - Worms propagate without user intervention and start by exploiting software vulnerability. Similar to viruses, worms can spread through email, web sites, or network-based software. The key characteristic of worm is that it propagates automatically.

c. Trojan horses - A Trojan horse program is software that does not let the user know its actual consequences. For example, a program which claims that it will speed up your computer may actually be sending confidential information to a remote intruder.

d. Hacker, Attacker, Intruder, or Denial of Service - These terms are applied to the people who seek to exploit weaknesses in software and computer systems for their own gain. Although it is difficult to comment on one's intention for doing this because they may or may not cause direct harm to the end user but denial of service definitely deprives the end user to be properly served.

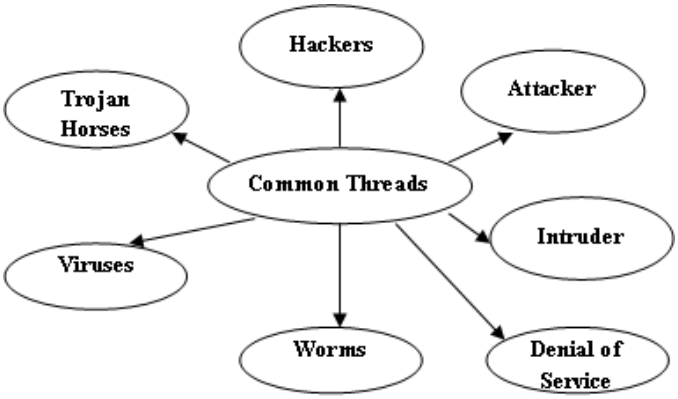

Figure 1 Common Thread

III. DATA MINING FOR CYBER SECURITY

Data mining has many applications in information security. Data mining techniques are being used to identify suspicious individuals and groups, and to discover which individuals and groups are capable of carrying out harmful activities. Cyber

security is concerned with protecting computer and network systems from malicious software including Trojan horses and viruses. Data mining is also being applied to provide solutions such as intrusion detection and auditing. Cyber security is the area that deals with protecting from cyber terrorism. Cyber attacks include access control violations. [1] Various data mining techniques have been applied for detection and prevention of security attacks on the system. With the advancements in the area of information security, the applications of data mining has also increased immensely to various other areas of information security and are not restricted to just intrusion detection and prevention systems. [2] Network intrusion detection is another area which requires immediate attentions, as the number of intrusion attacks are increasing. It is a unique form of computer-generated threat analysis to identify nasty actions that could compromise the integrity, confidentiality, and availability of information resources. Intrusion detection mechanisms based on data mining are extremely useful in discovering security breaches. [3] A number of data mining algorithms have been proposed to deal with the information security and privacy problems, by using approaches like classification, frequent pattern mining, and clustering methods to do intrusion detection, anomaly detection, and privacy preserving. Application of these data mining methods have resulted in stimulating results that has concerned many researchers in both data mining and information security areas.

IV. LITERATURE SURVEY

Several method and algorithm have been developed in the past year we have study some of them in our survey. In 2013 Usukhbayar Baldangombo et al. proposed "A Static Malware Detection System Using Data Mining Methods". They proposed a static malware detection system using data mining techniques SVM and Naïve Bayes classifiers. They used static analysis to extract valuable features of Windows PE file. They extract raw features of Windows executables which are PE header information, DLLs, and API functions inside each DLL of Windows PE file. They used the concepts of machine learning and data-mining, and construct a static malware detection system. [4] In 2014 Milan Jain et al. proposed "Malicious Code Detection through Data Mining Techniques". They proposed three algorithms named as RIPPER, Naives Bayes approach, and Multi-Naïve Bayes using data mining techniques and the comparison of these algorithms. They used the comparison between these three methods. Separate features of RIPPER, Naive Bayes, and a Multi-Classifier system where used. These Data mining techniques perform better than traditional techniques approach.[5][6] In 2015 Mahesh N Gunjal et al. "Data Mining for Malicious Code Detection and Security System Application". They successfully implemented data mining techniques such as K-nearest neighbor, Multi-navie byes and Ripper algorithm for malicious code detection such as worm detection, managing firewall policies. This algorithm detects and removes all threats include non real-time threats and real-time threats. They implemented these techniques for credit card fraud detection and biometrics related applications. [7] [8] In 2015 Abhay Pratap Singh proposed "Improving the Malware Detection Ratio Using Data Mining Techniques". They proposed a data mining approach which improves the malware detection ratio with high precision, whereas current anti-virus detection technologies which is based on signature, code emulation, and anomaly based, these are the technologies are failed to detect malware. [9] In 2016 Monire Norouzi et al. "A Data Mining Classification Approach for Behavioral Malware Detection". They proposed different classification methods in order to detect malware based on the feature and behavior of each malware. A dynamic analysis method has been presented for identifying the malware features. They used malware behavior executive history in XML file with WEKA tool. They showed the performance efficiency as well as training data and test. They also applied this approach to real case study data set using WEKA tool. Also our proposed data mining approach is more efficient for detecting malware and behavioral classification of malware can be useful to detect malware in a behavioral antivirus.[10]

V. CLASSIFYING WORM USING NAIVE BAYESIAN CLASSIFIERS

In this paper data mining classification techniques to classify a known worm have been used. Some parameters in proposed approach have also been used. Here three classes to unknown worm has been used. The data has been collected by using NP AV net protector. Figure 2 shows the types of scheduler used. Figure 3 shows the types of protection. Figure 4 shows the types of scan frequency.

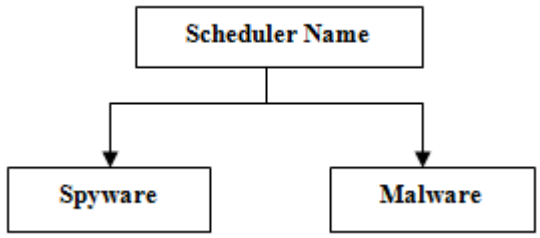

Figure.2. Types of scheduler used

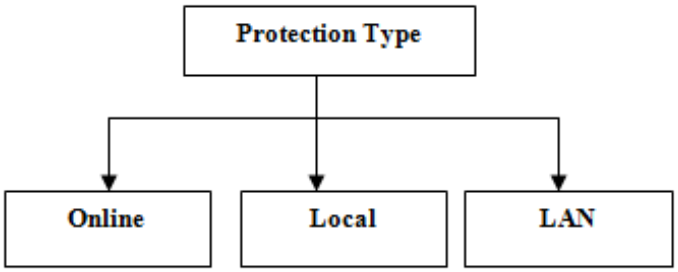

Figure.3. Protection type

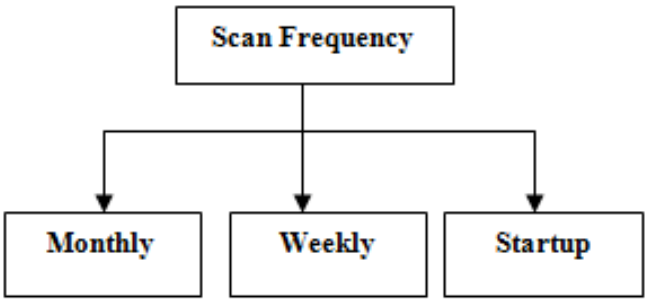

Figure.4. Scan Frequency type

TABLE.1. TESTING PARAMETER WITH VIRUS

S No.	Scheduler	Protection	Scan Frequency	Scan Drive	Worm Categories
1	Spyware	Online	Monthly	Removal Drive	EXE
2	Malware	Online	Weekly	Memory	Script
3	Spyware	LAN	Monthly	Removal Drive	EXE
4	Malware	Online	Monthly	Removal Drive	EXE
5	Spyware	LAN	Monthly	Memory	Script
6	Spyware	Online	Monthly	Removal Drive	EXE
7	Malware	Online	Weekly	Memory	Script
8	Spyware	LAN	Monthly	Removal Drive	EXE
9	Malware	Online	Monthly	Removal Drive	Script
10	Spyware	LAN	Monthly	Removal Drive	Script

Find the probability of heart attach yes and no
P (Worm Categories =exe) = 5/10 = 0.5
P (Worm Categories =Script) = 5/10 = 0.5
P (Scheduler = Spyware | Worm Categories =exe) = 4/5 = 0.8
P (Scheduler = Spyware | Worm Categories =Script) = 2/5 = 0.4
P (Scheduler = Malware | Worm Categories =exe) = 1/5 = 0.2
P (Scheduler = Malware | Worm Categories =Script) = 3/5 = 0.6

P (Protection = Online | Worm Categories =exe) = 3/5 = 0.6
P (Protection = Online | Worm Categories =Script) = 3/5 = 0.6
P (Protection = LAN| Worm Categories =exe) = 2/5 = 0.4
P (Protection = LAN | Worm Categories =Script) = 2/5 = 0.4

P (Scan Frequency = Monthly | Worm Categories =exe) = 5/5 = 1
P (Scan Frequency = Monthly | Worm Categories =Script) = 3/5 = 0.6
P (Scan Frequency = Weekly | Worm Categories =exe) = 0/5 = 0
P (Scan Frequency = Weekly | Worm Categories =Script) = 2/5 = 0.4

P (Scan Drive = Memory | Worm Categories =exe) = 0/5 = 0
P (Scan Drive = Memory | Worm Categories =Script) = 3/5 = 0.6
P (Scan Drive = Removal Drive | Worm Categories =exe) = 5/5 = 1
P (Scan Drive = Removal Drive | Worm Categories =Script) = 2/5 = 0.4

Scheduler = Malware, Scan Frequency = Monthly, Protection =Online then Worm Categories =exe
P (Scheduler = Malware | Worm Categories =exe) = 1/5 = 0.2
P (Scan Frequency = Monthly | Worm Categories =exe) = 5/5 = 1
P (Protection = Online | Worm Categories =exe) = 3/5 = 0.6
0.2*0.6*1=0.12
P (Scheduler = Malware | Worm Categories =Script) = 3/5 = 0.6
P (Scan Frequency = Monthly | Worm Categories =Script) = 3/5 = 0.6
P (Protection = Online | Worm Categories =Script) = 3/5 = 0.6
0.6*0.6*0.6=0.216VI. EXPERIMENTAL ANALYSIS

NP AV antivirus and perform scanning for infected drives with different files have been used for this paper. Some parameters during scan for virus infected drive have also been used for experimental analysis. Virus scan using different time interval has been performed. Figure 5 and Figure 6 show time interval. Two types of virus categories exe and script categories have been used and also shown with the help of figure 7 and figure 8.

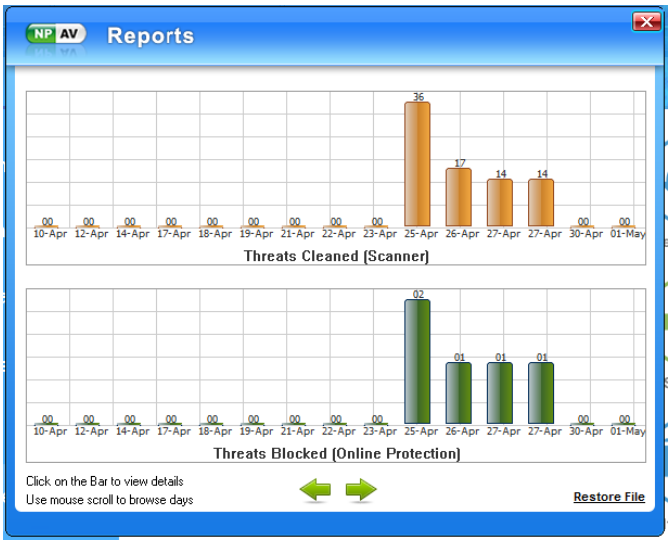

Figure. 5. Virus scan weekly

Figure. 6. Virus scan monthly

VII. CONCLUSION

After performing several tests for virus using different parameters, a data set has been created. The naive Bayesian classifier on this data set has been applied. Classifiers which are capable to classify a unknown tuple to a particular of virus has been built. Simple calculation is used. In future this approach can further extend for more parameter and number classes.

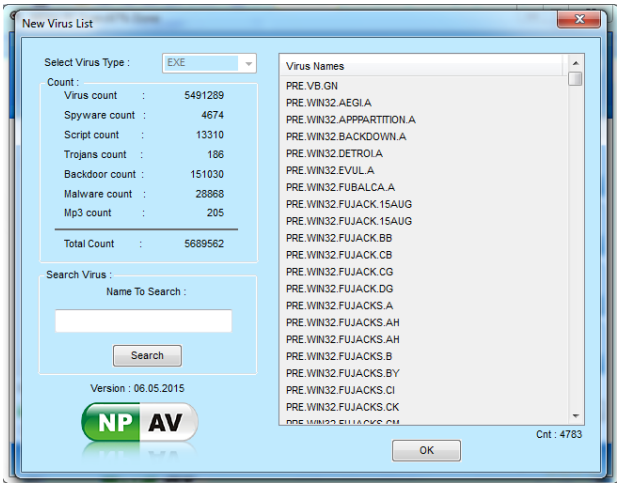

Figure.7. exe virus types

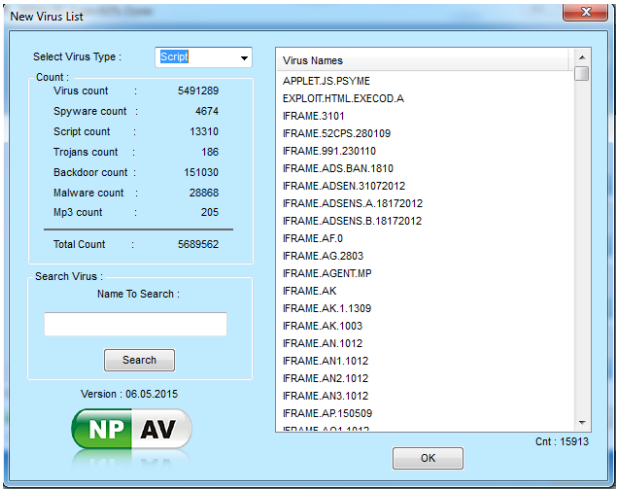

Figure.8. Script Virus Types

VIII. REFERENCES

[1]. Muazzam Siddiqui, Morgan C. Wang, and Joohan Lee "Detecting Trojans Using Data Mining Techniques" D.M.A. Hussain et al. (Eds.): IMTIC 2008, CCIS 20, pp. 400–411, 2008. c Springer-Verlag Berlin Heidelberg Available: "http:// stars. library. ucf.edu/cgi/ view content. cgi?article= 4709& context =etd"

[2]. K.Rajitha "Comprehensive Study And Analysis Of Malicious Website Detection Techniques" in International Journal of Computer Application (2250-1797) Volume 6– No.5, September- October 2016 Available: http://rspublication. com/ijca/2016/oct16/2.pdf

[3]. Matthew G. Schultz and Eleazar Eskin "Data Mining Methods for Detection of New Malicious Executables" in Department of Computer Science Columbia University fmgs,eesking@cs.columbia.edu Available: http://citeseerx.ist. psu.edu/viewdoc/download?doi=10.1.1.13.4514&rep=rep1&type=pdf

[4]. Usukhbayar Baldangombo1, Nyamjav Jambaljav1, and Shi-Jinn Horng "A STATIC MALWARE DETECTION SYSTEM USING DATA MINING METHODS" in International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 4, July 2013. Available: https://arxiv.org/abs/1308.2831

[5]. Milan Jain "Malicious Detection Using Multiple Classification Algorithms & Their Comparison Using Different Clustering Techniques" in Volume 4, Issue 8, August 2014 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering. Available: https://www.ijarcsse.com/docs/papers/Volume_4/8_August2014/V4I8-0120.pdf

[6]. Milan Jain "Malicious Code Detection through Data Mining Techniques" in International Journal of Computer Science & Engineering Technology (IJCSET) ISSN: 2229-3345 Vol. 5 No. 05 May 2014 Available: http://www.ijcset. com/docs/IJCSET14-05-05-092.pdf

[7]. Mahesh N Gunjal "Data Mining For Malicious Code Detection System" Multidisciplinary Journal of Research in Engineering and Technology, Volume 2, Issue 3 2-2-3-7-2015 Available: http://www.mjret.in/V2I3/M2-2-3-7-2015.pdf

[8]. Mahesh N Gunjal "Data Mining For Malicious Code Detection And Security System Application" in International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015 ISSN 2091-2730 Available: http://pnrsolution.org/Datacenter/Vol3/Issue3/113.pdf

[9]. Abhay Pratap Singh "Improving The Malware Detection Ratio Using Data Mining Techniques" in Second International Conference on Science, Technology and Management, September 2015. Available: http://data. conferenceworld.in/ICSTM2/P852-857.pdf

[10]. Monire Norouzi et al " A Data Mining Classification Approach for Behavioral Malware Detection" in Journal of Computer Networks and Communications Volume 2016, March 2016, Artcile No. 1, March 2016. Available: http://dl.acm .org /citation.cfm?id=2984887