ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 10, October 2017

Preventing Cyber Crimes by Using Data Mining Techniques

Prof. Rukaiya Shaikh[1], Aman Memon[2], Manoj Kumar[3], Ismaeil Pathan[4]

Department of Computer Engineering, Al-Ameen College of Engineering, Koregaon Bhima, Pune, India[1,2,3,4]

ABSTRACT :-Globally the internet has been accessed by enormous people within their restricted domains. When the client and server exchange messages among each other, there is an activity that can be observed in log files. Log files give a detailed description of the activities that occur in a network that shows the IP address, login and logout durations, the users behaviour etc. There are several types of attacks occurring from the internet. Our focus of research in this paper is Denial of Service (DoS) attacks with the help of pattern recognition techniques in data mining. Through which the Denial of Service attack is identified. Denial of service is a very dangerous attack that jeopardizes the IT resources of an organization by overloading with imitation messages or multiple requests from unauthorized users.
The survey of the various technologies used for this purpose is done in this paper. The technology will help in careful investigation of the crime and group them in the form of a cluster.

KEYWORDS:- Crime, cluster, Data Mining, Data Collection, Denial of Service, Log File, Cyber Crimes, Data mining, outliers, Association rules.

I. INTRODUCTION

Cyber Security is that branch of Computer Technology that deals with security in cyberspace. Cyberspace refers to the description of policies regarding the networks and computer systems. The policies laid out in the Cyber security are for the reason of avoiding the malicious activity or unauthorized access to secured information. Since the emergence of high structured networks , there arises a concern about how intelligently these networks are secured. These issues are major concerns in the internet era. Cyber security is concerned with protecting IT resources like server, network etc. from performing illegal activities or fraudulent acts. Data mining is also applicable to problem solving or network intrusions. Therefore in this paper we focus the applications of data mining for cyber security applications.

Indexing and crawling are two important aspects. If the content does not include indexing and crawling, then update of data will not occur properly within time, and the chance of duplicates values will be increased.

II. OVERVIEW OF CYBER CRIME DATA MINING

Data Mining: Data mining deals with the discovery of unexpected patterns and new rules that are “hidden” in large databases. The use of data mining in this paper is to give the structured data from unstructured data of judge. In this paper the Data Mining techniques of crime in two directions they are as follows:-

1. Classification of Cyber Crime
2. Clustering Technique of Cyber Crime

1. Classification of Cyber Crime

Cyber Crime: Crime is defined as “an act or the commission of an act that is forbidden, or the omission of a duty that is commanded by a public law and that makes the offender liable to punishment by that law”. Crime is referred to as a comprehensive concept that is defined in both legal and non-legal sense.

Classification of Cyber Crime IPC ACT 1. Offences by Public Servant 2. Destruction of Electronic Record 3. Cheating 4. Forgery 5. Data Theft 6. Criminal Breach of fraud+Creadit, Debit Card 7. Counterfeiting (Currency, Stamps, property) 8. False Electronic Evidence IPC ACT 9. Copyright Act 1957 10. Trade Marks Act 1999

DOI: 10.15680/IJIRCCE.2017. 0510053

2. **Clustering Technique of Cyber Crime Clustering**: Data clustering is a process of putting similar data into groups. A clustering algorithm partitions a data set into several groups such that the similarity within a group is larger than among groups. Clustering can also be considered the most important unsupervised learning technique; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. There are so many techniques used in clustering, in this paper only K-means algorithm is used. K-Means Clustering Algorithm: K-means algorithm mainly used to partition the clusters based on their means. Initially number of objects are grouped and specified as K clusters. The algorithm clusters observations into K groups, where K is provided as an input parameter. It then assigns each observation to clusters based upon the observation proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. In this paper the use of K-means algorithm is the process of getting a structured data from a unstructured data. The working of algorithm is explained as follows:
   k : pre-determined number of clusters Algorithm
   (Step 0: determine value of k
   Step 1: Randomly generate k random points as initial cluster centers
   Step 2: Assign each point to the nearest cluster center
   Step 3: Re-compute the new cluster centers
   Repetition step: Repeat steps 2 and 3 until some convergence criterion is met (usually that the assignment of points to clusters becomes stable).)

III. METHODOLOGY

In our research paper, we have shown the concept of data mining techniques to identify cyber-attacks. Our focus of attention would be on "finding patterns" in a log file (records that occur in the system) which shows the sequence of events. From this log file we identify patterns. To start with, we use the clustering technique to discover the type of cyber-crime, Denial of service (DoS) attacks. As we know that clustering is grouping of data that has similar features. So this grouping helps to discover similar patterns of data that occur constantly in the log file.

- Step 1: Evaluate the log file.
- Step 2: Mine the date with time
- Step 3: Scan the data
- Step 4: Add the found data in the main file.

When the above procedure is carried out, we will record that data which contains normal patterns and also abnormal patterns (malicious). By using the clustering technique we identify the data that occur repeatedly . System Configuration: In order to run our obtained data, we use the Windows Server to maintain the database. Initially we run the data that contains zero attacks and then add them to the master file or log file. The ICMP (Internet Control Message Protocol) will make the system inactive by sending voluminous amount of "ping" command. Now the data that contains the normal activities and the data that contains attacks are passed through the technique that we have proposed. If the observations of the log file show normal behaviour then they will be ignored. If the observations show multiple requests of the same transaction, then this data will be directed through our algorithm "Apriori" and will be shown in the attack logs. This algorithm will detect if similar patterns of requests exist in the normal records prior to consider it as attack. If the algorithm finds out the pattern and or finds the number of request for the same transaction more than the threshold value it is considered as an attack and it sends signal or message to the administrator about the suspected attack.

- **System Architecture:-**

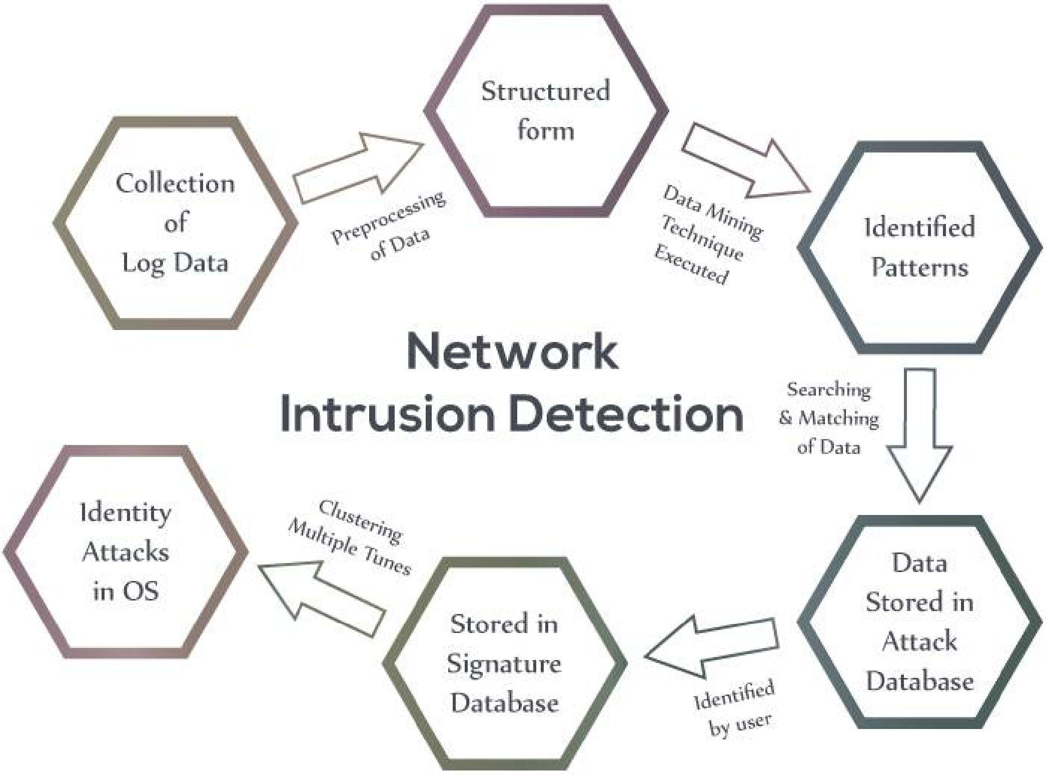
Structured form
Collection of Log Data
Preprocessing of Data
Data Mining Technique Executed
Identified Patterns
Network Intrusion Detection
Searching & Matching of Data
Identity Attacks in OS
Clustering Multiple Tunes
Stored in Signature Database
Identified by user
Data Stored in Attack Database

Fig.System Architecture

- **Log Analysis:-**

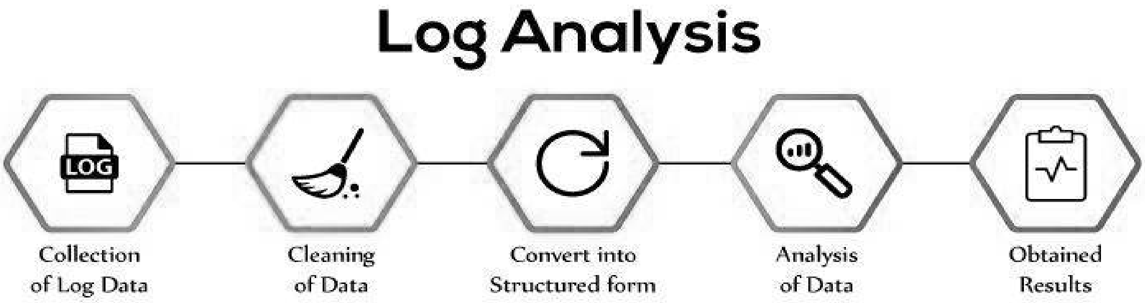

Fig.Log Analysis

IV. ALGORITHM

- Step 1: Start
- Step 2: Let the Count=0, set the threshold value. The threshold value can be set based on the working environment.
- Step 3: Check if the counts of matched rules have crossed the threshold value. • If true, intimate the administrator assuming as an attack. • If false, continue.
- Step 4: Check whether new event is recorded in log file. • If no new event found, wait • If event_found, go to step 2.

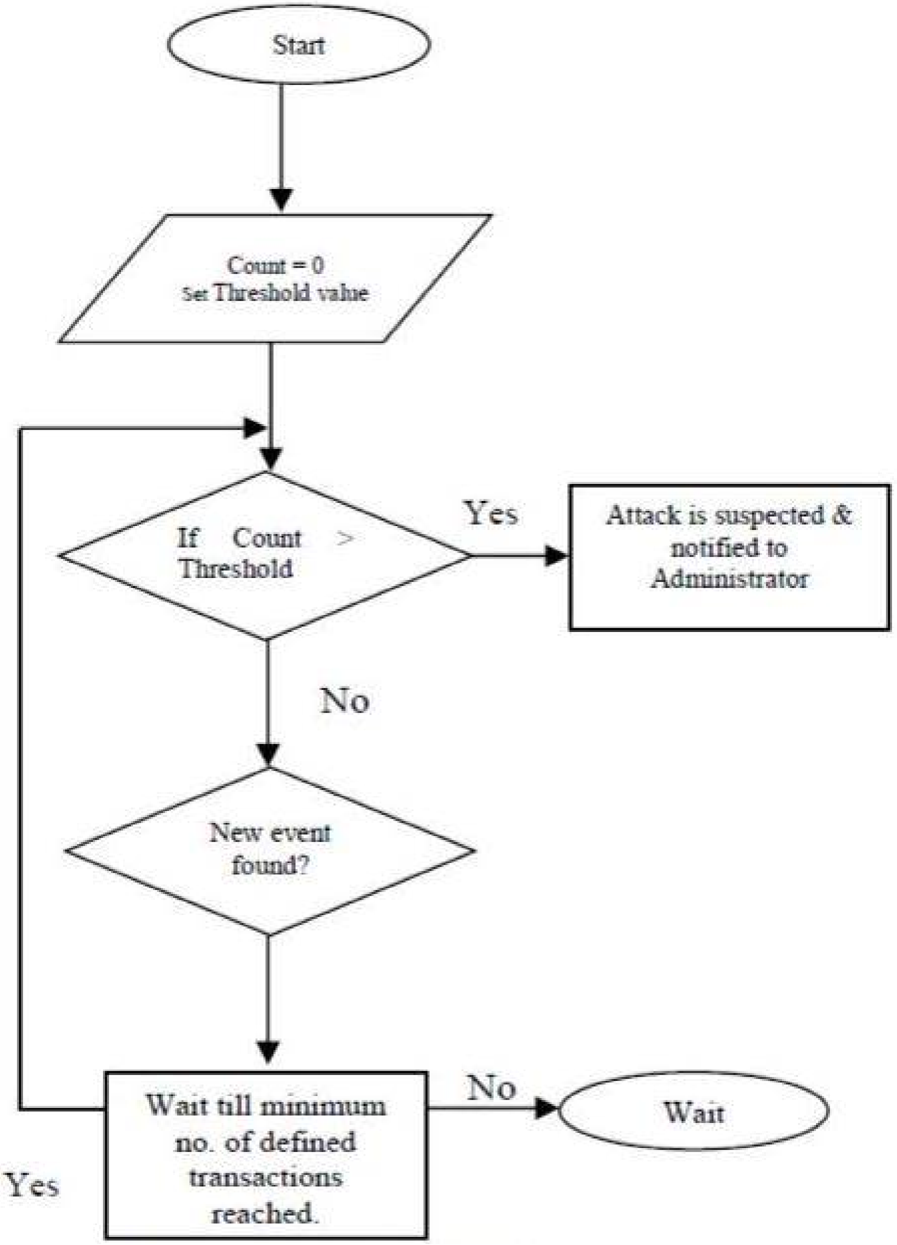
Start
Count = 0
Set Threshold value
If Count > Threshold
Yes
Attack is suspected & notified to Administrator
No
New event found?
Wait till minimum no. of defined transactions reached.
No
Wait
Yes

Fig. . Flowchart for detecting DoS attacks.

DOI: 10.15680/IJIRCCE.2017. 0510053

V. Conclusion and Future Work

In this paper we have applied the data mining techniques for identifying the Denial of Service attack. This type of attack is very dangerous as it jeopardizes the IT resources. It makes the server busy by imitation messages and repeated queries. The server is congested by traffic packets, in order to mitigate the server performance. In this research paper, we have discussed about Cyber security, cyber-crimes their types, clustering, outliers and pattern recognition. We have applied the famous data mining technique called as pattern recognition on the log file. We set a threshold value. If the number of similar requests are received at the server, which is greater than the threshold value, we assume this as an attack and the administrator is been informed. By this approach we can identify the denial of service attack easily as in DoS attack, the attacker or the hacker sends same multiple requests in order to mitigate the server performance.

REFERENCES

1. Know Your Enemy: Learning About Security Threats, 2nd Edition.ISBN: 0321166469. The Honeypot Project 2004.
2. M.Khan , S.K.Pradhan, M.A.Khaleel, “Outlier Detection for Business Intelligence using data mining techniques”, International journal of Computer Applications (0975 -8887), Volume 106- No. 2, November 2014.
3. Masud, M.M, Gao,J.Khan, ”Peer to Peer Botnet Detection for Cyber Security: A Data Mining Approach”. In proceedings: Cyber-security and information Intelligence research workshop. Oakridge national Laboratory, Oakridge May 2008.
4. Internet Security Threat Report, Volume 21, April 2016, Symantec Crime Report.
5. Ibrahim Salim, T.A.Razzack,”A study on IDS for Preventing denial of service attack using outliers’ techniques”, 2nd IEEE international conference on Engineering and technology, March 2016.
6. S.S Rao, SANS Institute Infosec Reading Room.” Denial of service Attack and mitigation techniques: Real time implementation with detailed analysis”, 2011.
7. Data Mining: Concepts and Techniques, Third Edition, Jiawei Han and Micheline Kamber, ISBN-13, 9780123814791.
8. Mining of Massive Data Sets, Anand Rajaraman, Jure Leskovec, Jeffrey D. Ullman,2014
9. A. Klein, F. Ishikawa, and S. Honiden. Efficient heuristic approach with improved time complexity for qos-aware service composition. In ICWS, pages 436–443. IEEE, 2011.
10. Tripathy, M.Khan, M.R.Patra, H.Fatima, P.Swain, “Dynamic web service composition with QoS clustering” IEEE, International Conference on Web services, 2014.
11. D. E. Brown, "The regional crime analysis program (RECAP 1998) : A Frame work for mining data to catch criminals," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol.3, pp.2848-2853
12. IEEE proceedings Rushinek, A, Rushinek, SF (1993). “Using Experts for Detecting and Litigating Computer Crime”. Managerial Auditing Journal. 8.7:19-22. Security Focus. Florida.
13. Anshu Sharma, Shilpa Sharma, An Intelligent Analysis of web Crime Data Using Data Mining, International Journal of Engineering and Innovative Technology, 2012.
14. S. Yamuna, N. Sudha Bhuvaneswari, Data Mining Technique to Analyse and Predict Crime, The International Journal of Engineering And Science, 2012.
15. Malathi. A, Dr. S. Santhosh Baboo, Anbarasi. A, An Intelligent Analysis of a City Crime Data Using Data Mining, Internation Conference on Information and Electronics Engineering, 2011
16. Devesh Bajpai, Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview, Journal of Environmental Science, Computer Science and Engineering & Technology, 2012
17. R.G Uthra Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview, Journal of Data Mining Technique to Analyze Crime Data, International Journal for Technological Research in Engineering , 2013