IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Prediction of Varied Network Attacks using Data Mining Techniques – A Review

R. Thilakavathi[1], P. Kalaivani[2]

Assistant Professor, Department of Applied commerce, Park's College, Tirupur, India [1, 2]

Abstract: In computer networks, an attack is an attempt to seize, immobilize, destroy, modify, or gain unauthorized access to or make unauthorized use of an asset. Network attacks can cause network services slow, temporarily unavailable, or down for a long period of time. The main aim of this work is to find different types of network attacks and detection techniques to resolve malicious attacks in the network flow. It is necessary for users and network administrator to detect these attacks before they cause damage to the system. Data mining techniques will help the network administrators to identify new network attacks in current scenario.

Keywords: Networks, malicious attacks, Data Mining

I. INTRODUCTION

Securing the network is the major challenge in this information era from the various types of network threats and attacks [1]. The threats are classified based on their behaviour such as leakage: unauthorized access of information available in the network [2]. Tampering: modifying the information without permission of the author. Vandalism: making malfunction over a normal execution of a system. The various types of attacks such as eavesdropping: collecting the replica information without obtaining permission to the arbiter. Masquerading: making conversation using through others identity without permission of others. Message tampering: modifying and altering the information while travel on the communication media. Man-in-the-middle attack: is a one type of message interfering in which an attacker interrupt the very first message in an exchange of encrypted keys to establish a secure channel [3][4]. The attacker substitutes compromised keys that enable them to decrypt subsequent messages before reconfiguring them in the correct keys and passing them on. Replying: This is one type of attack that stores intercept messages then sends these messages later. This attack may be effective even with authenticated and encrypted messages [3]. Denial of service: makes the transmission channels and systems as busy as possible by sending garbage data for denying the service [5].

The knowledge about these attacks is acquired from the huge volume of network data with data mining tools. This knowledge facilitates the security system to identify the attackers or hackers based on their behaviour in a network. The behaviour of the attackers and hackers are studied and identified by two types of learning strategies namely supervised and unsupervised learning.

In data mining based network security approach, the network sniffing or scanning software collects the data about the activities of the attacker. The collected data are learnt by the supervised learning algorithm and the predictive model is built. This model predicts and detects the attackers and hackers.

II. PREDICTING NETWORK THREATS

To protect network infrastructure, users need to predict the types of network threats which is vulnerable. Security experts use a model called STRIDE to classify network threats.

A. Spoofing

These are attacks that are aimed at obtaining user account information. Spoofing identity attacks affect data confidentiality.

B. Tampering

These are attacks that are aimed at modifying information. Data tampering ends up affecting the integrity of data. A man – in – middle attack is a form of data tampering.

C. Repudiation

It takes place when the user performs some form of malicious action on a resource and then later denies carrying out that particular activity. Network administrators have no evidence to back up their suspicious.

D. Information disclosure

In this type of attacks private and confidential information are made available to individual who should not have access to particular information. It impacts data confidentiality and network resource confidentiality.

E. Denial of service

These attacks affect the availability of data and network resources and services. It aimed at preventing legitimate uses from accessing network resources and data.

F. Elevation of privilege

Elevation of privilege occurs when an attacker escalates privileges to obtain a high level of access like administrative privileges in an attempt to gain control of network system.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

III. DETECTION TECHNIQUES

In the network attack a detection technique varies depend upon the nature and deployment of network infrastructure. The following are the various detection mechanisms which detect the different types of network attacks.

A. **Host Based Detection**

The host based detection systems monitors and analyzes the internals of a computing system rather than its external interfaces [6]. Such systems might detect internal activity such as which program accesses what resources and attempts illegitimate access. An example is a word processor that suddenly and inexplicably starts modifying the system password database.

B. **Network Based Detection**

A network is connected to the rest of the world through the Internet. The Network based detection system reads all incoming packets or flows, trying to find suspicious patterns. For example, if a large number of TCP connection requests to a very large number of different ports are observed within a short time, we could assume that someone is committing a 'port scan' at some of the computer(s) in the network[6].

C. Mining Based Detection

Data mining look for hidden patterns and trends in data warehouse that is not immediately apparent from summarizing the data, and there is no query involved but use the concept interestingness criteria i.e specification of data such as Frequency, Rarity, Correlation, Length of occurrence, Consistency, Repeating/ periodicity, abnormal behaviour, and other patters of interestingness. The algorithms which are used for intrusion detection based on data mining techniques are listed as follows

Association rule

Association rules mining identifies association among database attributes and their values. It is a pattern-discovery technique which does not serve to solve classification problems nor predict problems. Association rule mining requires two thresholds i.e Minimum support and Minimum Confidence. Example: Apriori for mining Association rules Algorithm.

Classification

Classification is the process of learning a function that maps data objects to a subset of a given class set. There are two goals of classification, First finding a good general mapping that can predict the class of so far unknown data objects with high accuracy. Second to find a compact and understandable class model for each other classes.

Clustering techniques

Clustering group's data elements into different groups based on the similarity between within a single group Cluster partitions the data set into clusters or equivalence classes. Cluster methods divided into two categories based on the cluster structure namely Non Hierarchical and Hierarchical –connection oriented.

Decision Tree

Decision tree initially builds a tree with classification. Each node represents a binary predicate on one attribute, one branch represents the positive instances of the predicate and the other branch represents the negative instances. Construction of Decision Tree does not require any domain knowledge and can handle high dimensional data.

Genetic Algorithms

Method: learning examples are stored in relational database that are represented as relational tuples. It solves the problems with multiple solutions and easily transferred to existing models

K Nearest Neighbour

An object classification process is achieved by the majority vote of its neighbours. The object is being assigned to the class most common amongst its k nearest neighbours. If k=1, then the object is simply assigned to the class of its nearby neighbour. Its Implementation tasks are simple and Easy for parallel implementations.

Support vector Machine

Method: A support vector machine is a classification and regression technique it constructs a hyper plane or set of hyper planes in a high or infinite dimensional space. It is able to model complex and nonlinear decision boundaries. [4]

IV. DATA MINING TOOLS

In this section, various open source data mining tools for analyzing and building the predictive model for the unlabeled data classification and prediction are discussed.

A. WEKA (Waikato Environment for Knowledge Analysis)

The university of Waikato New Zealand developed this open source tool in Java technology. This consists of a collection of machine learning algorithms such as Clustering, Feature selection/Attribute subset selection, Classification, Association Rule mining etc. The Weka provides four interfaces namely Explorer, Experimenter, Knowledge Flow, Simple CLI (command-line interface) to work with machine learning algorithm and datasets. The Explorer provides a platform for data exploration. The Experimenter provides a platform to perform Experiments for conducting statistical tests among the learning schemes. The Knowledge Flow provides a Graphical User Interface to implement the functionalities available in explorer. The Simple CLI provides a simple command-line interface to execute the Weka commands [7].

B. Orange

The open source Orange tool contains a variety of machine learning and data mining algorithms with routines to data exploration. It works with Python and C++ it works for the functionalities such as decision trees, attribute subset

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

selection, boosting and bagging .It gives a platform for visual programming to use the visual component widgets, this explores the data for analysis. The widgets modularity connects the communication media to exchange the data packets automatically for data analysis. This orange is used for many data analyzing applications [8].

C. R Tool:

Initially, the Ihaka and Gentleman from University of Auckland, New Zealand developed R tool in 1996. R provides a platform to compute the statistics for data analysis. , R works with the Unix, Windows, Mac platform. The formal work flow for a data mining task is carried out through the following steps [9].

· Load a Dataset and select features
· Explore the data in understandable format
· Distribution of Test
· Transform the data to suit the modeling
· Build the Models
· Evaluate the models with dataset
· Review the Log in data mining task.

D. Keen Tool

The Keen (Knowledge Extraction based on Evolutionary Learning) carry outs many data mining tasks includes regression, supervised, unsupervised and evolutionary learning algorithms. It consist of my library functions for pre and post processing method for data manipulation and soft computing mythology for helps to the scientific research in the area of machine learning. It also supports the fuzzy and genetic algorithm to do research in the area of data mining and various applications related to data analysis [10].

V .CONCLUSION

This paper explored and analyzed the various challenges of threats and attacks in networks in this recent era. Various data mining tools for learning and building the predictive models and various supervised learning classification algorithms for learning the network data are used to identify the behaviour of the attackers and hacker. There is much research scope for data mining based detection to find a solution for detecting emerging network attacks.

REFRENCES

[1] M. Schäfer, V. Lenders, and I. Martinovic, "Experimental analysis of attacks on next generation air traffic communication," in Applied Cryptography and Network Security, 2013, pp. 253–271.
[2] N. Meng, J. Wang, E. Kodama, and T. Takata, "Reducing data leakage possibility resulted from eavesdropping in wireless sensor network," International Journal of Space-Based and Situated Computing, vol. 3, no. 1, pp. 55–65, 2013.
[3] T. Denning, T. Kohno, and H. M. Levy, "Computer security and the modern home," Communications of the ACM, vol. 56, no. 1, pp. 94–103, 2013.
[4] T.-H. Lin, C.-Y. Lin, and T. Hwang, "Manin- the- Middle Attack on 'Quantum Dialogue with Authentication Based on Bell States'," International Journal of Theoretical Physics, pp. 1–5, 2013.
[5] Z. Tan, P. Nanda, R. P. Liu, A. Jamdagni, and X. He, "A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," IEEE Transactions on Parallel and Distributed Systems, vol. 99, no.1, p. 1, 2013.
[6] Monowar H. Bhuyan, D. K. Bhattacharyya and J. K. Kalita, "Network Anomaly Detection: Methods, Systems and Tools", IEEE Communications Surveys & Tutorials, 2013.
[7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10–18, 2009.
[8] J. Dem\vsar, B. Zupan, G. Leban, and T. Curk, Orange: From experimental machine learning to interactive data mining. Springer, 2004.
[9] G. J. Williams, "Rattle: A data mining GUI for R," The R Journal, vol. 1, no. 2, pp. 45– 55, 2009.
[10] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesús, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, and V. M. Rivas, "KEEL: A software tool to assess evolutionary algorithms for data mining problems," Soft Computing, vol. 13, no. 3, pp. 307–318, 2009.
[11] D. A. A. G. Singh and E. J. Leavline, "IATARPA: Implementation of anonymity threat avoidance routing protocol architecture for MANET," in Advanced Computing (ICoAC), 2011 Third International Conference on, 2011, pp. 321–326.
[12] Jeyanthi N, Vinithra J,Sneha, Thandeeswaran R and N.Ch. Sriman NarayanaIyengar, "A Recurrence Quantification Analytical approach to Detect DDoS Attacks", International Conference on Computational Intelligence and Communication Systems, 2011.
[13] B.S. Kiruthika Devi, G. Preetha, S. Mercy Shalinie, "DDoS Detection using Host-Network based Metrics and Mitigation in Experimental Testbed", ICRTIT, 2012.
[14] Y. Xie, S. Tang, X. Huang, C. Tang, X. Liu, "Detecting latent attack behavior from aggregated Web traffic", Computer Communications 36, Pg. 895–907, 2013.
[15] P. Jongsuebsuk, N. Wattanapongsakorn, C. Charnsripinyo, "Network Intrusion Detection with Fuzzy Genetic Algorithm for Unknown Attacks", IEEE International Conference on Information Networking, 2013.
[16] Ahmad Sanmorino, Setiadi Yazid, "DDoS Attack Detection Method and Mitigation Using Pattern of the Flow", IEEE International Conference of Information and Communication Technology, 2013.
[17] Sumaiya Thaseen, Ch. Aswani Kumar, "An Analysis of Supervised Tree Based Classifiers for Intrusion Detection System", Proceedings ofIEEE International Conference on Pattern Recognition, Informatics and Mobile Engineering, 2013.