ISSN:0975-9646

A System for Detection of Distributed Denial of Service (DDoS) Attacks using KDD Cup Data Set

Janhavi Kaskar, Ruchit Bhatt, Rohit Shirsath

Department of Information Technology,Dr. D. Y. Patil College of Engineering, Pune, India

***Abstract*: Distributed denial-of-service (DDoS) attacks are a major security threat, the prevention of which is very hard, like when it comes to highly distributed daemon-based attacks. The early discovery of these attacks, although difficult, is necessary to protect network resources as well as the end users. In this paper, we address the problem of DDoS attacks and present the foundation and algorithms of our IDS. The base of our system is composed of intrusion detection systems (IDSs) which use the KDD Cup dataset to detect intrusion. The IDS scans all the files being transmitted from the routers for malicious content and known virus signatures. The evaluation of our system, using the KDD testing dataset, shows a better ratio of detecting attacks and a low false positives ratio. It also supports easy modifiability, scalability and usability.**

Keywords: Detection, distributed denial- of-service (DDoS), network security, Intrusion Detection System (IDS), KDD Cup 99 dataset.

I.INTRODUCTION

A. IDS

The process of monitoring the events occurring in a computer system or network and analyzing them for sign of intrusions is known as intrusion detection. Information is the best asset of today’s corporate world. And as such, efforts have to be taken to maintain the integrity, confidentiality and availability of the information. IDS are the first steps to maintaining the CIA triad (i.e. confidentiality, integrity and availability) of any information pertaining to any company. The concept of intrusion detection was first described by Anderson to aid conventional computer security approaches. Anderson defined an intrusion or a threat to be a deliberate unauthorized attempt to:

1. access unauthorized information,
2. manipulate information and harm its integrity, or
3. render a system unreliable or unavailable to its authorized users.

Most intrusion detection systems are generally software based, designed to detect anomaly in behavior of the user. People use these systems to keep scanning the events occurring in a computer system or network, and for the analysis of the system events, detection of suspected intrusion, and then flagging an intrusion.

A typical intrusion detection system consists of three functional components: an information source, an analysis engine and a decision maker. The function of the information source is to generate a stream of event records. This component is also known as an event generator. It monitors different data sources and converts it into data that are well formatted and suitable for analysis. The data sources can be mainly divided into three categories: first, data sources dealing with operating systems, like system calls and system logs; second, network packet traffic monitors which generate raw network packets; and lastly, data collectors of various applications.

The analysis engine finds the onset of attack. A decision maker applies some rules on the result of the analysis engine, and decides what actions should be taken based on the reactions of the analysis engine. The major use of the decision maker is to improve the usability of an intrusion detection system.

*B. **IDS Terminology***

Some important Intrusion Detection concepts are:

1.) Attack/Intrusion:

An act carried out by one adversary, the intruder, against another adversary, the victim. The intruder carries out an attack with a specific objective in mind. The attack can be a passive attack or an active attack. A passive attacker only works to gather information about the victim whereas the active attacker wishes to harm the victim and his machine. From this point of view, an attack is a set of one or more actions that may have one or more security breaches. From the perspective of an intruder, an attack is a mechanism to fulfill an objective.

2.) Intruder:

Intruder is a person who is responsible for carrying out an attack. Attacker is a common synonym for intruder. The words attacker and intruder apply only after an attack has been successfully launched. A would-be intruder may be referred to as an adversary. Since the role of intruder is decided by the victim of the intrusion and is therefore based on the victim's definition of security breach, there can be no common classification of the events as being threatening or not.

3.) Vulnerability:

A feature or a combination of features of a system that allow an adversary to manipulate the system and place it in a state that is undesirable to the wishes of the stakeholders responsible for the system and increases the chances or magnitude of unwanted behavior in or of the system.

4.) Exploit:

It is the process of using a system vulnerability to breach a security policy. A tool or defined technique that could be used to violate a security policy is often referred to as an exploit.

5.) False negative:

An event in which IDS fails to identify an intrusion when one has in fact occurred.

6.) False positive:

It is an event, wrongly identified by the IDS as being an intrusion when none has occurred.

II. Literature Survey

Intrusion detection systems must have the ability to distinguish between normal and abnormal user activities, and to discover intrusions in time. The activity of the user in IDS may fall into any one of the behaviors specified in figure 2.1.

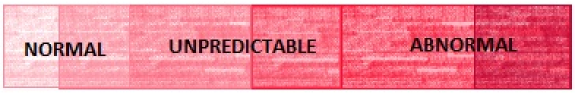

Figure 1: Behavioral diagram

In order to classify actions, intrusion detection systems use analysis approach. An analysis approach is a technique used by IDS to determine whether or not an attack has been mounted on the system. There are two major types of analysis approaches:

A. Anomaly-based Detection Approach

An automatically developed profile is generated by the IDS that collects and analyses the different characteristics of system behavior over a period of time and forms a statistically correct profile of such behavior. An anomaly might include

1. unexplained reboots or changes to system variables
2. users logging in at unexpected and out of the way hours
3. users logging in from unfamiliar location on the network
4. multiple failed login attempts

Anomaly detectors generate profiles that represent normal usage and then use logged behavior data to detect a possible mismatch between profiles and recognize possible intrusion attempts.

Advantages:

1.ability to detect abuse of user privileges

B. Misuse-based detection Approach

Misuse-based detection approach identifies signatures or patterns related to known malicious content or malwares. It also includes signature analysis which is an interpretation of a series of packets that are defined, in advance, as a representative of a known attack. Signature-based IDS carries out a simple pattern matching process. It is also known as pattern-based IDS.

Advantages:

1.relatively low rate of false alarms.

III. Introduction To Kddcup Data Set

Since 1999, KDDCUP'99 has been the most commonly used data set for the testing and training of anomaly detection methods. This data set is constructed based on the data captured in the DARPA'98 IDS evaluation program. DARPA'98 is about 4 gigabytes of compressed raw (binary) tcpdump data of 7 weeks of network traffic, which can be processed into about 5 million connection records. KDD training dataset consists of approximately 4,900,000 single connection vectors each of which is made up of 41 features and is labeled as either normal or an attack, with only one specific attack type.

The intrusions fall in one of the following four categories:

1.)Denial of Service Attack (DoS): is an attack in which the attacker tries to make some service or memory resource too busy or too full to handle legitimate requests, or denies authorized users access to a machine, i.e. it hampers availability of services.

2) User to Root Attack (U2R): is a class of attacks in which the intruder starts out with access to any particular user account on the system (perhaps gained by guessing passwords, a dictionary attack, or brute force attack) and is able to manipulate some vulnerability to gain unauthorized access to the system, which may be over the privileges of the user.

3) Remote to Local Attack (R2L): occurs when an intruder who has the power to transmit packets to a system over a network but who does not have an authorized account on that machine exploits some vulnerability to gain local access as an authorized user of that machine.

4) Probing Attack: is an attempt to gather information about a network of computers for the apparent purpose of eluding its security protocols.

TABLE I:ATTACK TYPES AND SIZE IN 10%KDD DATA SET

Category	Attack Type
Normal	Normal(97277)
DOS	Back(2203), Land(21), Neptune(107201), Pod(264), Smurf(280790), Teardrop(979)
U2R	Buffer_overflow(30), loadmodule(9), perl(3), Rootkit(10),
R2L	ftp_write(8), Guess_passwd(53), Imap(12), Multihop(7), Phf(4), Spy(2), Warezclient(1020), Warezmaster(20)
Probe	Ipsweep(1247), Nmap(231), Portsweep(1040), Satan(1589)

KDD'99 features can be classified into three main groups:

1) Basic features: this category binds all the characteristics that can be extracted from a TCP/IP connection.

2) Traffic features: this category includes the features that are calculated with respect to a window of interval and is divided into two groups:

2.1) Same host features: these examine only the connections in the last 2 seconds that have the same destination host as the current connection, and calculate statistics dealing with protocol behavior, service, etc.

2.2) Same service features: examine only the connections in the last 2 seconds that have the same service as the current connection.

3) Content features: unlike most of the DoS and Probing attacks, the R2L and U2R attacks do not display any intrusion frequent sequential patterns. This is so because the DoS and Probing attacks involve many connections to some particular host(s) in a very minute period of time; but the R2L and U2R intrusions are embedded in the data field of the packets, and normally involve only a single connection. To detect these kinds of attacks, we need some features to be able to depict the suspicious behavior in the data field, for e.g., number of failed login attempts. These features are called content features.

TABLE II: KDD'99 FEATURE DESCRIPTION

Feature No.	Feature Name	Description
1.	Count	No. of connections to the same host as the current connection in the last two seconds
2.	destination bytes	Bytes sent from destination to source
3.	diff srv rate	percentage of connections to different services
4.	dst host count	count of connections having the same destination hosts
5.	dst host diff srv rate	percentage of different services on the current host
6.	dst host rerror rate	percentage of connections to the current host that have an RST error
7.	dst host same src port rate	percentage of connections to the current host having the same src port
8.	dst host same srv rate	percentage of connections having the same destination host and using the same service
9.	dst host serror rate	percentage of connections to the current host that have an S0 error
10.	dst host srv count	count of connections having the same destination host and using the same service
11.	dst host srv diff host rate	percentage of connections to the same service coming from different hosts
12.	dst host srv rerror rate	percentage of connections to the current host and specified service that have an RST error
13.	dst host srv serror rate	percentage of connections to the current host and specified service that have an S0 error
14.	Duration	Duration of the active connection.
15.	Flag	Status flag of the connection
16.	Hot	No. of "hot" indicators
17.	is guest login	1 if the login is a "guest" login; Otherwise 0
18.	is host login	1 if the login belongs to the "host"; otherwise 0
19.	Land	1 if connection is from/to the samehost/port; Otherwise 0
20.	logged in	1 if successfully logged in; otherwise 0
21.	num access files	No. of operations on access control files
22.	num compromised	No. of compromised conditions
23.	num failed logins	No. of failed logins
24.	num file creations	No. of file creation operations
25.	num outbound cmds	No. of outbound commands in an ftp session
26.	num root	No. of "root" accesses
27.	num shells	No. of shell prompts
28.	protocol type	Connection protocol (e.g. tcp, udp).
29.	rerror rate	percentage of connections that have "REJ" Errors
30.	root shell	1 if root shell is obtained; otherwise 0
31.	same srv rate	percentage of connections to the same service
32.	serror rate	percentage of connections that have "SYN" Errors
33.	Service	Destination service (e.g. telnet, ftp)
34.	src bytes	Bytes sent from source to destination
35.	srv count	No. of connections to the same service as the current connection in the last two seconds
36.	srv diff host rate	percentage of connections to different hosts
37.	srv rerror rate	percentage of connections that have "REJ" errors
38.	srv serror rate	percentage of connections that have "SYN" Errors
39.	su attempted	1 if "su root" command attempted; otherwise 0
40.	Urgent	No. of urgent packets
41.	Wrong fragment	No. of wrong fragments

IV. PROPOSED SYSTEM

Our system is basically a secure platform for sharing of files. Any user can create his/her profile on the system and then upload new data or download data that is already present on the system. The user can also request the admin for some specific data which the admin can later upload. KDDCUP rule mapping is applied at the server side to check for any abnormalities in the data being received or transmitted. Our system basically concentrates on the KDDCUP entries which deal with denial of service attacks.

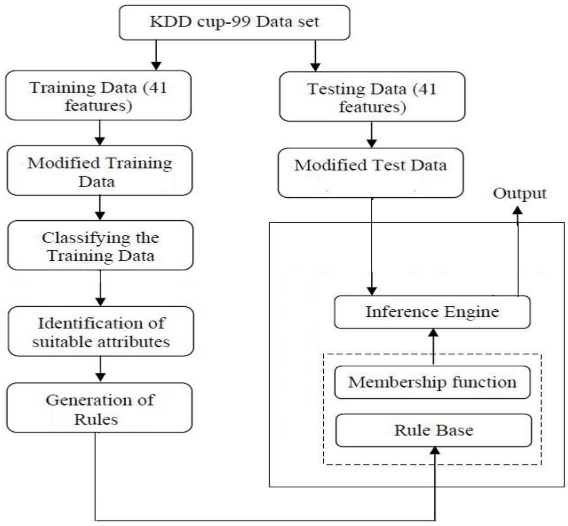

Figure 2: System Flow

Modules of our system are:

1. Authentication:

In this module, functionality is provided for admin and user login. A new user can register on the system. The system registers the user's user id, password, email id, etc. In this way, the user's profile is maintained.

2. Application User Interface:

In this module, the user interface is designed and implemented. The user interface consists of a website which helps the user to navigate through the platform and upload and download files. The user can also post comments and queries.

The admin can view the user requests, upload requested files, and unlock blocked users, handle the virus signature database, view the transmission of files through the router.

3. Server side Verification:

Whenever a particular file is received at the server, its details like size, source IP address, destination IP address, protocol of file transmission, etc, are extracted and stored in the database. These details are verified whenever the file leaves the router, to check that integrity of the data is maintained.

Additionally, image files and PDF files are watermarked with the source IP address and destination IP address.

Also, the KDDCUP rule verification occurs at the server side and router side to monitor the network for signs of intrusion.

In case any user uploads a file which is found to be malicious, then, that file is blocked along with that particular user. The user cannot upload any further files to the system. The user can however, request the admin to unblock him and then the admin can decide whether to do so or not.

4. Bandwidth Calculator:

There is a bandwidth calculator which is used to monitor the active network communication links, and measure the number of bytes received and sent at the router ends.

V. Algorithmic Study

Step I: User registers on the system.

Step II: User uploads file to the system.

Step III: The file is scanned for any malicious content. If the file is clean, then the file is uploaded to the system, else, the file and the user are both blocked. If user is blocked, go to step X.

Step IV: If the file is a pdf file or an image file, then it is watermarked with the source and destination IP address and its details like file size etc, are entered into the database.

Step V: For any other file, the details are inserted into the database and then the file is available for download.

Step VI: Simultaneously, the files being uploaded and downloaded are scanned at the router side to ensure that their integrity is maintained.

Step VII: KDDCUP rule mapping also occurs at the server and router side to monitor the system for any data leakage or smurf attack.

Step VIII: In case an attack is detected, the user is informed of the attack and the attacker is blocked off at the router.

Step IX: The admin can login and view all these router details as well as user requests for different files. The admin can also manage the virus signature database.

Step X: The admin checks if any file or user was blocked due to a false positive and if so, then unblocks the user and the file.

VI. Conclusion

In conclusion, this paper has presented a secure platform for sharing of files for individual users. It is also fairly protected from external intrusions. It is highly modifiable and scalable and has high usability.

VII. Future Study

As of now, whenever data is transmitted from the client side to the server side or vice-a-versa, it is done so in a plain format. In the future, this system could be improved by encrypting the data that is being transmitted on the network. As a result of this, in case of data leakage, the intruder would not be able to gain any important information.

Acknowledgement

We would like to express a deep sense of gratitude to our respected project guide Prof. Girish Kala, whose help and guidance was a beacon of inspiration for us. At the outset, we would like to thank the Principal of our college, Dr. Shirbahadurkar, for providing us with the necessary resources to build our project. We also wish to express our gratitude towards H.O.D. of Information Technology, Prof. Ravi Patki and our respected project co-ordinator Prof. Shubhangi Sonone for their wholehearted help and guidance towards completion of this project. And last but not the least, we would also like to express our sincere thanks to the teaching staff of IT Dept., for their co-operation and support at every stage of this project.

REFERENCES

[1] Shanmugavadivu R., Nagarajan N., "Network Intrusion Detection System using fuzzy logic", in Indian Journal of Computer Science and Engineering (IJCSE), ISSN : 0976-5166, Vol. 2 No. 1,2011.

[2] P. Srinivasu, P.S. Avadhani, V. Korimilli, P. Ravipati, "Approaches and data processing techniques for Intrusion Detection Systems", Vol. 9, No. 12, 2009.

[3] Yao, J. T., S.L. Zho, and L.V. Saxton, "A study on fuzzy intrusion detection", in Proceedings of the Data Mining, Intrusion Detection, Information Assurance, And Data Networks Security, SPIE, Vol. 5812, 2005.

[4] Dr.**S.**Siva Sathya, Dr. R.Geetha Ramani, K**.**Sivaselvi, "Discriminant Analysis based Feature Selection in KDD Intrusion Dataset" in International Journal of Computer Applications (0975 – 8887) Volume 31– No.11, October 2011.

[5] Shaik Akbar, Dr.J.A.Chandulal, Dr.K.Nageswara Rao, G.Sudheer Kumar, "Troubleshooting Techniques for Intrusion Detection System using Genetic Algorithm", in International Journal of Wisdom Based Computing, Vol. 1 (3), December 2011.

[6] Kamlesh Lahre, Tarun Dhar Diwan, Suresh Kumar Kashyap, Pooja Agrawal, "Analyze Different approaches for IDS using KDD 99 Data Set", in International Journal on Recent and Innovation Trends in Computing and Communication, ISSN 2321 – 8169, Volume: 1 Issue: 8, 645 – 651.