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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

Data Leakage Prevention by KDD-99 Intrusion Detection System

Dev Desai[1], Rohan Vadsola[2], Mihir Brahmbhatt[3], Nayan Mali[4]

Student, Information Technology, Sigma Institute of Technology, Vadodara, India[1,2,3]

Assistant Professor, Information Technology, Sigma Institute of Technology, Vadodara, India[4]

Abstract: It is not possible to detect all the security attacks together. The detection that is available is on various platforms. This in turn exploits security problems which are way too difficult and expensive to resolve. The integrity, availability and confidentiality of a system is compromised by an intrusion. The Intrusion Detection System (IDS) is used to detect such intrusions in a network. For this the Data Mining is integrated with the IDS to detect such intrusions. Algorithms such as EDADT algorithm, varying HOPERAA algorithm, Hybrid IDS model, and semi supervised approach are used for detection of such attacks. All the algorithms are quite accurate and false alarm rate is reduced.

Keywords:Anomaly based algorithm, classification algorithm, data communication, Denial of Service attack, Intrusion Detection, EDADT algorithm, HOPERAA algorithm, hybrid IDS model, semi-supervised approach.

I. INTRODUCTION

The people using computers and other devices that interact with the internet are increasing day by day. So, to keep a track of all those devices is quite difficult and time consuming. Nowadays to intrude a system is quite simple and doesn't take much time. A system can be intruded in a matter of seconds, and after the intrusion the footsteps can be easily cleaned. Powerful IDS can easily differentiate between intrusive and non-intrusive records. IDS was first introduced by James Anderson in the year 1980 [1].

The existing systems are quite vulnerable and can be easily breached, and to solve this isn't quite possible. IDS is based on misuse detection and anomalies detection. So to solve this anomalies and create a better and secure system is in process. It is a challenging task for the network administrator and security experts. The priority of this task is also quite high.

So the security experts and network administrator comes up with their own intrusion detection systems. But recently the focus on KDD has increased. KDD has 41 features for every record with 37 different types of attack. So it takes IDS long time to detect all the attacks and the methods.

So, here the proposed method selects only the important feature sets and reduces the IDS for overhead. And finally the IDS and data mining techniques are combined in order to detect the dynamic intrusion and keep the system secure.

II. RELATED WORK

In the previous version of KDD there were 41 features for every record with 37 different types of attack [2]. So it takes IDS long time to detect all the attacks and the methods. This makes the KDD process quite slow and less effective. The present system doesn't include redundant data records in the set and so duplicate records are eliminated. Here other algorithms such as Anomaly based algorithm, classification algorithm, EDADT algorithm, HOPERAA Algorithm etc are used for detection of such attacks.

The KDD is based on dictionary based attacks [3]. Which means KDD uses a good knowledge of data mining if we want it to put to proper use.

III. CLASSIFICATION

The following algorithms has been used for KDD and IDS. EDADT algorithm (Efficient Data Adapted Decision Tree): First the local and global values of n number of iterations are found out to find the optimal solution. The optimal solution can be found out by finding the average values from the dataset. Then unique values are found out and they are searched if they belong to the same class or not.

Then, if the values (n) belong to the same class, then they are spit into intervals (m), and n must be greater than m.

But if the values belong to different class then the probability is found out. Then it is checked if the probabilitiesbelongs to the same class or not. If it belongs to the same class, the class value of each class is replaced with the class value of the number with the class value of highest probability. Then the value must be spitted into s intervals and this process is repeated for each value in the data set. Then the normalized information gain is found out for each attribute and then the highest normalized information gain is used to form the best attribute of the class. Then the best attributes are used to form the sublets and from those nodes the child nodes are generated.

HOPERAA algorithm:

To reduce and remove the effects of the distributed denial of service attack varying HOPERAA algorithm is used. It includes the following aspects:

DOI 10.17148/IJARCCE.2015.44103

- Contact initiation part.
- Data transmission part.

Contact initiation part:
An activity is done at the client side. A request to the server is initiated for connection from the client side. And further another request is made for further communication. Then the server divides the range of port number into n number of intervals and ports are spilt for every intervals. Every client has a unique port number that is known only to that client and the server.

Data transmission part:
Here the client sends messages to the server. Then the client receives a reply from the server during contact initiation part. The port sequence varies based on the state of message. As soon as the open port reaches its time limit a new port is allotted to that client. The timer which is placed at the client is set to zero as soon as it receives a reply from the server.

EDADT Algorithm:
In data mining technology we can detect old as well as new attacks which are yet unknown. This is quite helpful for dynamic IDS.

The following table represents the rule structure of KDD-99.

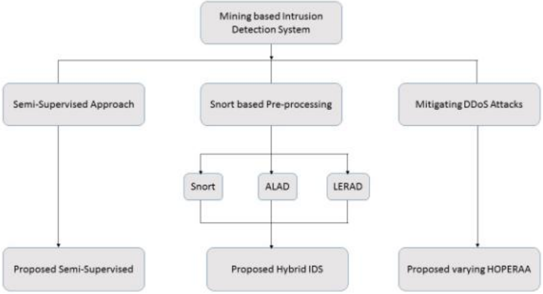

Fig. 1 Representation of EDADT algorithm.

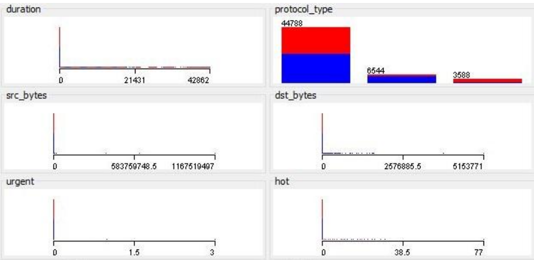

Fig. 2 Various analysis done for KDD-99 in Weka.

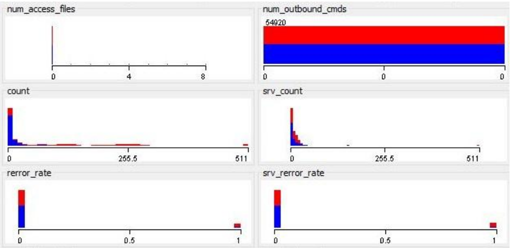

Fig. 3 Various analysis done for KDD-99 in Weka.

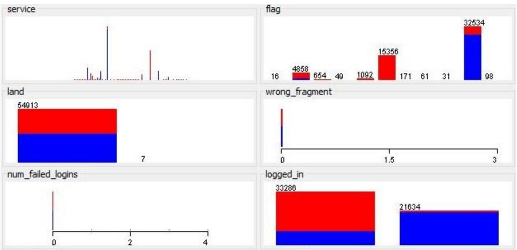

Fig. 4 Various analysis done for KDD-99 in Weka.

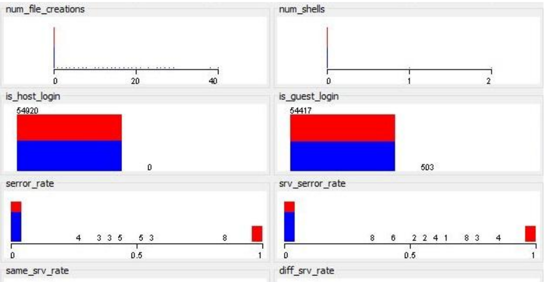

Fig. 5 Various analysis done for KDD-99 in Weka.

IV. ANALYSIS

To use the KDD-99 along with the IDS the following techniques and algorithms should be learnt.

- Data Mining.
- Anomaly based algorithm.
- classification algorithm.
- Denial of service attack
- EDADT algorithm
- HOPERAA Algorithm
- Hybrid IDS model
- Semi-Supervised approach.

The previous version of KDD were quite lengthy and it consumed a lot of time. In that time the dynamic intruder could easily compromise the system before it can be detected [4]. So this technique is introduced to reduce the reaction time and the overhead of the system. For this the data mining technique was used, along with the EDADT algorithm and varying HOPERAA Algorithm. This all is combined to form the hybrid IDS model.

V. CONCLUSION

The proposed method explains why the data mining technique is necessary to be integrated with the IDS. This paper shows how can we introduce the above methods and the integrate them to form a better intrusion detection system which is quite effective, faster and low costing. With the help of this the system can be more secure from intruders. After normalization, Discretion and feature selection this method can be used to improve the security of a system in large number of ways [2]. It is also noticed that the KDD-99 process is quite faster and more secure than any of the previous methods.

ACKNOWLEDGMENT

We would like to extend our gratitude to our respected teacher – **Mr. Nayan Mali**, for his constant support and for imparting us with the knowledge and helping us throughout the paper.

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

REFERENCES

[1]. Anderson JP. Computer security threat monitoring and surveillance. In: Technical report, Fort Washington, Pennsylvania: James P Anderson co; 1980.

[2]. G.V. Madiammai, M. Hemalatha, “Effective approach toward Intrusion Detection System using data mining techniques”, Egyptian Informatics Journal (2014) 15, 37-50.

[3]. Anish Das, Rashmi Bikash Nayak, “A Divide and Conquer Deature Reduction and Feature Selection Algorithm in KDD Intrusion Detection Dataset”, Chennai and Vivekanandha College of Technology for women, Third International Conference on Sustainable Energy and Intelligent System (seiscon 2012),VCTW, Tiruchengode, Tamilnadu, India on 27-29 December, 2012.

[4]. Aditya Harbola, Jyoti Harbola, Kunwar Singh Vaisla, “ Improved Intrusion Detection in DDoS Applying feature selection Using Rank & Score of Attributes in KDD-99 data set”, 2014 IEEE 978-1-4799-6929-6/14.

DOI 10.17148/IJARCCE.2015.44103