IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-2, Issue-8 E-ISSN: 2347-2693

A Comprehensive Survey on Methods Implemented For Intruder Detection System

B. Kiranmai[1*] and A. Damodaram[2]

[1*] *Dept. of CSE, Nishitha College of Engg. & Tech., Greater Hyderabad, India*
[2] *Dept. of CSE and Director AAC, JNTUH, Hyderabad, India*

www.ijcseonline.org

Received: 14 July 2014 Revised: 22 July 2014 Accepted: 16 August 2014 Published: 31 August 2014

***Abstract*— Intrusion recognition is the act of discovering undesirable visitors on a system or a system. An IDS can be a piece of set up software or a physical equipment that watches system visitors in order to identify undesirable action and activities such as unlawful and harmful visitors, visitors that goes against security plan, and visitors that goes against appropriate use policies. Intruder detection system can be implemented using various data mining approaches. This paper summarizes intrusion motives and some of the methods used and implemented for intrusion detection system. This paper also reviewed about processing environment and type of data required for evaluation of Intruder detection system.**

Keywords*—*Intruder Detection System; Data Mining; Kddcup99;

I. INTRODUCTION

Traditional methods of program designs identification are based on the saved designs of known designs. They recognize style by analyzing the program connection features to the style design that are offered by human professionals. The main drawback of conventional methods is that they cannot recognize unknown style. Even if a new style of the designs were discovered, this new style would have to be individually customized into program. It is also capable of identifying new designs to some degree of likeness to the discovered ones, the neurological systems are generally considered as an effective way to adaptively classify designs, but their high computations durability and the long training times considerably limit their programs, especially for the style identification problem, where the amount of related data is very important.

The program relies on users' actions in order to draw out features and then guidelines allow the program to categorize any infrequent activity as an attack act.

II. INTRUDERS MOTIVES [1]

Network intruders are gaining access through unauthorized access to networking devices through physical, system and remote attempts. The intruder uses some outdated exploits that are ineffective against up-to-date patched hosts. The intruders can be an insider or an outsider. The insider can have some access to certain areas or network.

The intruder forms are 1) masquerader 2) misfeasor 3) clandestine.

1) To perform network scanning to find out vulnerable hosts in the network
2) To install an FTP server for distributing illegal content on network

Corresponding Author: *kiranmai, kiranmaimtech@gmail.com*

A) Tools used by Hackers and Intruders:

1) Trojan horse: It acts as a back door to get access.
2) Virus: They are self-replicating spreads and crashes the system
3) Worm: Self-replicating and does not attach itself to other code.
4) Vulnerability scanners: To check computers on a network for known weaknesses
5) Sniffer: This is an application that captures password and other data in transit either the computer or over the network.
6) Exploit: This is an application to take advantages of known weaknesses.
7) Root kit: This tool is for hiding the fact that a computer security has been compromised.

III. PROCESSING ENVIRONMENT FOR IDS'S:

Intrusion Detection system can be processed either in offline or online (on fly)[1][2].

A) Offline Processing :

The use of data mining techniques in IDSs usually implies analysis of the collected data in an offline environment. There are important advantages in performing intrusion detection in an offline environment, in addition to the real-time detection tasks typically employed. In off-line analysis, it is assumed that all connections have already finished and, thus, we can compute all the features and check the detection rules one by one. The estimation and detection process is generally very demanding and, therefore, the problem cannot be

addressed in an online environment because of the various the real time constraints.

B) On-the fly processing:

With on the fly processing, IDS performs online verification of system events. Generally, a stream of network packets is constantly monitored constantly. With this type of processing, intrusion detection uses the knowledge of current activities over the network to sense possible attack attempts (it does not look for successful attacks in the past).

Given the computation complexity, the algorithms that are used here are limited to quick and efficient procedures that are often algorithmically simple. This is due to a compromise between the main requisite – attack detection capability and the complexity of data processing mechanisms used in the detection itself.

At the same time, construction of an on-the-fly processing IDS tool requires a large amount of RAM (buffers) since no data storage is used. Therefore, such an IDS may sometime miss packets, because realistic processing of too many packets is not available.

IV. TYPE OF STATIC DATA USED IN INTRUSION DETECTION [13]

A) KDD CUP 99

KDD CUP 99 dataset is taken from DARPA (Defense Advanced Research Projects Agency) which has been used widely to asses anomaly detection methods since 1999. KDD CUP 99 training set and test set contain respectively 4898431 and 311027 intrusion and normal records. Intrusion types in this dataset are divided into four groups: Denial of Service Attack (DoS): is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine. User to Root Attack (U2R): is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system.

Remote to Local Attack (R2L): occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine. Probing Attack: is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls.

B) DARPA98 data set [14]

The DARPA/MIT Lincoln lab evaluation dataset has been used to test a large number of intrusion detection systems. The DARPA98 can be used to test host- based systems, network-based systems, and signature and anomaly detection systems.

V. A SURVEY ON DATA MINING TECHNIQUES USED FOR INTRUSION DETECTION METHOD

Various data mining techniques are used in finding intrusions detection system.

Some of them are

A) Classification
B) Clustering
C) Association

A) Classification Related Work:

Classification involves finding rules that partition the data into disjoint groups. The input for classification is the training dataset, whose class labels are already known. A set of classification rules generated by such a classification process, which can be used to classify future data and develops a better understanding of each class in the data base.

There are several classification discovery models. They are [3]

1) Classification by Decision Tree Induction
2) Bayesian Classification
3) Rule Based Classification
4) Classification by Back Propagation
5) Support Vector Machines
6) Genetic Algorithms

Some of the techniques implemented under Classification model are discussed.

1) Implementing Rule based Genetic Algorithm as a Solution for Intrusion Detection System [4].

In this [4] authors implemented intrusion detection system in two phases. In the first phase, the Learning stage, rule set is generated for detecting intruders using network audit data. The second phase, the best rule set with highest fitness value is used for detecting intruders in the Internet

world. This paper presents the Genetic Algorithm for the Intrusion detection system for detecting DoS, R2L, U2R, Probe from KDD99CUP data set. This provides a high rate of the rule set for detecting different types of attacks. The results of the experiments are good with and an 83.65% of average success rate and got satisfied.

2) Network Intrusion detection using Navie Bayes [5][6].

In Bayesian classification, we have a hypothesis that the given data belongs to a particular class. We then calculate the probability for the hypothesis to be true. This is among the most practical approaches for certain types of problems. The approach requires only one scan of the whole data. Also, if at some stage there are additional training data, then each training example can incrementally increase/decrease the probability that a hypothesis is correct. Thus, a Bayesian network is used to model a domain containing uncertainty.

3) Intrusion Detection Using Neural Networks and Support Vector Machines [7].

In this the [7] authors used Support Vector Machine for Intrusion Detection System. The construction of an SVM intrusion detection system consists of three phases are Preprocessing, Training, and Testing. Preprocessing: using automated parsers to process the randomly selected raw TCP/IP dump data in to machine-readable form. Training: in this process SVM is trained on different types of attacks and normal data. Testing is measure the performance on testing data. With SVM‘s could make only binary Classification which is a drawback of the system.

B) Clustering Related Work

Clustering is a technique for finding similarity groups in data called as clusters. it groups data instances that are similar to (near) each other in one cluster and data instances that are very different (far away) from each other into different clusters. Clustering is also called as unsupervised learning.[3]

Categorization of clustering methods: [3]

1) Partitioning methods
2) Hierarchical methods
3) Density based methods
4) Grid based methods
5) Model based methods

Some of the methods implemented under clustering are

1) Optimized Sampling with Clustering Approach for Large Intrusion Detection Data [8]

In this paper authors [8] have been used K-Mean evolution clustering method for intrusion detection. This method decreases the overhead of performing the detection over whole datasets. As it work on partition of data sets and this results in increase in the processing speed of the clustering method. From the experimental result it is shown that using this method 620 clusters are formed from 5000 data points with standard deviation of .03.But sometimes it gives wrong result because it performs detection on the sample datasets.

2) A clustering method for Intrusion Detection [9][10]

In this paper[9][10] authors used another modification to K-Mean clustering algorithm has been proposed. This modified K Mean clustering algorithm is called as Y-Mean clustering. It overcomes the shortcoming of K-mean clustering mainly number of cluster dependency and degeneracy Y-Mean clustering.

2)An Intrusion Detection System Based on the Clustering [10] [11]

In this paper authors [11] implemented Ensemble In a new clustering algorithm called ensemble algorithm has been introduced which combines different clustering algorithms. The author has tested already exiting algorithms using different parameters. EAIDS pre-process the datasets and then create the classifier based on EA algorithm.

3)A clustering-based method for unsupervised intrusion detections [14].

In this paper authors [14] considered the outlier factor of clusters for measuring deviation degree of a cluster. Data classification is performed by an improved nearset neighbour method.The time complexity of CBUID is linear with the size of dataser and the number of attributes.

20 Association Related work [3]

An association rule is a pattern that states when *X* occurs, *Y* occurs with certain probability. Proposed by Agrawal et al in 1993. It is an important data mining model studied extensively by the database and data mining community. Initially used for market basket analysis to correlate items purchased by customers.

An association rule is an implication of the form:
$X \rightarrow Y$, where X, $Y \subset I$, *and* $X \cap Y = \varnothing$
Various Frequent mining methods are:

1) Apriori algorithm
2) Mining by partitioning data
3) Mining frequent item sets without candidate generation
4) Vertical data format

1) A Data Mining Method for Adaptive Intrusion Detection Method [12].
In this paper authors implemented the association rules algorithm following the ideas of Apriori (Agrawal and Srikant, 1994). Briefly, an item set *X* is frequent if *support X min support*. The Apriori algorithm starts by finding all length 1 frequent item sets, and then iteratively generates length *k* _ frequent item sets from the length *k* frequent item sets. At each iteration, the algorithm first uses a *join* step to generate length *k*+1 candidate item sets, each from 2 length *k* frequent item sets; then a *prune* step to filter out a candidate item set if one of its length *k* subsets is not frequent; finally the *support* values for the remaining candidate item sets are counted to form the set of frequent *k*.

VI. CONCLUSION

This paper reviewed various data mining techniques that are implemented in Intrusion detection system. However improvements has to developed for processing on-fly data.

VII. REFERENCES

[1] Asmaa shaker,Ashroor 2011 International conference on Future Information Technology IPCSIT vol.13 (**2011**) © (2011) IACSIT Press, Singapore.
[2] Singh, S. and S. Kandula, "Argus - a distributed network-intrusion detection system," Undergraduate Thesis, Indian Institute of Technology, **May 2001.**
[3] Jiawei Han and Micheline Kamber Data Mining Concepts and Techniques Second Edition Morgan Kauffman Publishers **,2006**
[4] Shaik Akbar, Dr.K. Nageswara Rao, Dr.J.A. Chandulal IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.8, August **2011**pp **138-144**
[5] Mrutyunjaya Panda, Manas Ranjan Patra IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December **2007** pp **258- 263**
[6] P.Jenson, "Bayesian networks and decision graphs", Springer, New-york, USA, **2001**.
[7] Srinivas Mukkamala, Guadalupe Janoski, Andrew Sung 0-7803-7278-6/02 ©**2002** IEEE
[8] Nani Yasmin1, Anto Satriyo Nugroho2, Harya Widiputra3," Optimized Sampling with Clustering Approach for Large Intrusion Detection Data", International Conference on Rural Information and Communication Technology **2009** Pp.**56-60**
[9] Yu Guan and Ali A. Ghorbani, Nabil Belacel,"Y-Mean: A Clustering method For Intrusion Detection", 1CCECE 2003, pp.1-4
[10] Fangfei Weng, Qingshan Jiang, Liang Shi, and Nannan Wu,"An Intrusion Detection System Based on the Clustering Ensemble", IEEE International workshop on 16-18 April **2007**,pp.**121 – 124**
[11] Kusum kumara Bharati,Sanyam Shukla, Swetha Jain Special Issue of IJCCT Vol.1 Issue 2, 3, 4; 2010 for International Conference [ACCTA-2010], 3-5 August **2010**
[12] Wenkee Lee, Salvatore J. Stolfo, Kui W. Mok c **2000** Kluwer Academic Publishers. Printed in Netherlands.
[13] Rahimeh Rouhi , Farshid Keynia, Mehran Amiri Journal of Computer Sciences and Applications, **2013**, Vol. 1, No. 3**, 33-38**
[14] Shengi YiJiang,Xiaoyu Song, Hui Wang, Jian-Jun Han,Qing-Hua Li Science direct © **2005** Elsevier pp **802-810**

Authors profile

B. kiranmai persuing Ph.D in C.S.E at JawaharlalNehruTechnologicalUniversity. Working as a Assoc.prof in Nishitha college of Engg.& Tech.

Dr. A. Damodaram working as a Professor of CSE Dept at JNTU,Hyderabad. Currently he is Director of Academic Audit Cell. He is aso the chairman, Board of Studies for CSE and IT departments.