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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 4, April 2014

Detecting User-To-Root (U2R) Attacks Based on Various Machine Learning Techniques

S. Revathi [1] Dr. A. Malathi[2]

Ph.D. Research Scholar, PG and Research, Department of Computer Science, Government Arts College, Coimbatore, India[1]

Assistant Professor, PG and Research, Department of Computer Science, Government Arts College, Coimbatore, India[2]

Abstract: Intrusion detection mainly focused on four major attack category such as denial of service, probe, user-to-root, and remote-to-local. This paper focused on user-to-root attack, which the attacker tries to access normal user account and gains root access information of the system. The U2R attacks leads to several vulnerability such as sniffing password, a dictionary attack and social engineering attacks. This paper makes a comparative study analyses for U2R attacks based on several popular machine learning techniques such as navie bayes, random forest, J48, random tree, JRIP and Multilayer perceptron to achieve better accuracy and to reduce mean square error for individual attacks that belongs to user to root category.

Keyword: Intrusion detection, User-to-Root attack, Random Forest, Multilayer perception, J48

I. INTRODUCTION

Intrusion detection system was mainly used for prevention of security. As every association is employing network communication, for transferring data using data packet within the network which is prone to intrusion or interference of the unauthorized user that violates the security of the network link established between two systems [1]. So for preventing this intrusion its detection is the highest priority. This is the extensively studied topic in computer research in recent years. The nature of the data packets or content of the data packets is studied to so as to classify the different type of packets, specially normal or non-intruded packet and intruded packet [2]. The Deployment of effective IDS systems is extremely challenging. For specific environment, there will be generation of thousands of alerts, with most of these alerts being incorrect and thus are false alerts. However, it is not obvious whether the alert is positive or negative until after they have been investigated thereby creating a large burden on the IT department. The four major attack categories are denial of service, probe, User to Root and Remote to Local attacks [3]. This paper focused on User to Root attacks where the attackers tries to access limited privilege of the machine.

The rest of the paper is organized as in section II is explains the user to root attack in NSL-KDD dataset. Section III explains several machine learning techniques. Section IV shows experimental analysis and section V draws some conclusion and future works.

II. DATASET DESCRIPTION

Commonly many Researcher used DARPA 98 [4] and KDDcup99 [5] dataset to examine intrusion detection using various methodologies. The major statistical degradation in these dataset are its huge dataset size which leads to dimensionality problem and the reputation of data which result in poor evaluation of detection was proposed by Tavalleein [6] . These problems leads to new version of KDD dataset as NSL-KDD dataset [7] which has been used as effective benchmark dataset to compare various machine learning techniques. The main advantage of NSL KDD dataset are [7]

- No redundant records in the train set
- No duplicate record in the test set
- The selected records is inversely proportional to the percentage of original records in KDD data set.

The training dataset is consist of 21 different attacks out of the 37 present in the test dataset. Most novel attacks are present in test dataset which are not present in training data. The 4 major attack categories: DoS, Probe, U2R and R2L. Table 1 shows the major attacks in user-to-root in both training and testing dataset.

Table1: Attacks in User-to-Root

Attack Names	Training Attacks	Testing Attacks
Buffer_overflow	30	20
Loadmodule	9	2
Perl	3	2
Rootkit	10	13
httptunnel	0	133
Ps	0	15
Sqlattack	0	2
Xterm	0	13
Total	52	200

The major attack in U2R is buffer overflow which copies too many data into static buffer without checking whether the data will exactly fit into program [8]. The loadmodule attack makes system server into dynamically two loadable kernel that currently running the program to create special device in the directory to use those module. The Ps attacks leads to an exploitable race condition in the actions of a single program, or two or more programs running simultaneously. The attacker execute arbitrary code to access root privilege. The Xterm attack exploits a buffer

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 4, April 2014

overflow in the Xaw library distributed in Redhat Linux and allows an attacker to execute arbitrary instructions with root privilege.

III. MACHINE LEARNING TECHNIQUES

Data Mining is a non-trivial extraction of implicit, previously unknown, and possible useful information from data. It used to finds hidden pattern in large volumes of data [9, 10]. It is an interdisciplinary filed which involves association, classification, clustering and visualization to design pattern. Now- a- days data mining is mainly used to solve problem of network intrusion based security attack [11]. As data point of view, intrusion detection is a data analysis process. It maps data item into one of several pre-defined categories. The algorithms normally output "classifiers", in the form of either decision trees or as rule generation. An ideal application in intrusion detection will be to gather sufficient "normal" and "abnormal" audit data from user program, and then apply a classification algorithm to learn about desire classifier that will determine the audit data as belonging to the normal class or the abnormal class [12]. Nevertheless of good anomaly detection methods are used, the main problem as high false alarm rates is difficult in finding features, and high performance requests still exist. Therefore, various machine learning schemes are used to investigates detection process for user to root type attacks. Some of the classification algorithm that most commonly used to classify the dataset are Multi-layer perceptron, J48, Random forest, JRIP and Navie Bayes [13, 14].

IV. EXPERIMENTAL RESULT ANALYSIS

The comparative analysis of the proposed work has been performed on Weka tool [15] using NSL-KDD dataset. The number or training and testing dataset for U2R is minimum as 52 and 200 which reduce the overall performance of intrusion detection on experimentation. It consist of 41 attribute which are completely used for analysis. The experimental result has been evaluated based on three major parameter as accuracy, mean square error and time which are shown in table2.

Table2: Performance of various learning techniques

Learning techniques	Accuracy (%)	Mean square error	Time (sec)
Navie bayes	73.07	0.1314	0.01
Random forest	86.53	0.1457	0.04
Random tree	76.92	0.1154	0.01
J48	84.61	0.106	0.09
MLP	88.46	0.0846	0.78
JRIP	73.07	0.159	0.05

From the above table it's clear that multi-layer perception performs better in all three aspect to detect U2R attacks, since the time taken to detect attack may increase which improves detection accuracy and drastically reduce mean square error. The paper also examined individual attacks and the below table shown the performance for individual attack using MLP.

Table3: Performance for individual attacks in U2R

Attacks	Precision	Recall	Fvalue
Buffer_overflow	0.853	0.967	0.906
Rootkit	0.889	0.800	0.842
Loadmodule	1.000	0.667	0.800
Perl	1.000	1.000	1.000
SQLattack	0.885	0.136	0.894
Xterm	1.000	0.987	0.965
Ps	0.865	0.768	1.000

The precision, Recall and Fvall are used to calculate the accuracy of the learning techniques and from the above table it shows that multi-layer perceptron shows high accuracy in detecting intrusive activity.

V. CONCLUSION

This paper presents a comparative analysis between various machine learning techniques such as navie bayes, J48, Random Forest, Multi-layer perceptron, Random tree and to detect User-to-Root attack. The paper also explains briefly about various attacks types that present in U2R. Each machine learning technique has their own merits to improve classification accuracy and to build pattern classification. From the above result it's clear that multilayer perceptron performs better than other existing machine learning techniques. Individual attack classification has also been analyzed in this paper, since the dataset has only limited number of records which may reduce overall detection performance, the main aim of this paper is to examine individual attack completely. Future work includes testing other attacks and how it works on other real time environment.

REFERENCES

[1] T. Crothers, Implementing Intrusion Detection Systems: A Hands-On Guide for Securing the Network, Wiley, 2003.
[2] R. Bace and P. Mell, NIST Special Publication on Intrusion Detection Systems, National Institute of Standards and Technology, 2001.
[3] C. E. Landwehr, A. R. Bull, J. P. McDermott, and W. S. Choi, “A taxonomy of computer program security flaws,” ACM Comput. Surv., vol. 26, no. 3, pp. 211–254, 1994.
[4] J. McHugh, “Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory,” ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262–294, 2000.
[5] KDD Cup 1999. Available on http://kdd.ics.uci.edu/Databases/kddcup 99/kddcup99.html, Ocotber 2007.
[6] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set”, In the Proc. Of the IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009), pp. 1-6, 2009.
[7] “Nsl-kdd data set for network-based intrusion detection systems.” Available on: http://nsl.cs.unb.ca/KDD/NSLKDD. html, March 2009.
[8] http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/docs/attackDB.html
[9] Jiawei Han and Micheline Kamber “Data mining concepts and techniques” Morgan Kaufmann publishers .an imprint of Elsevier .ISBN 978-1-55860-901-3. Indian reprint ISBN 978-81-312- 0535-8.
[10] T. Lappas and K. P. “Data Mining Techniques for (Network) Intrusion Detection System," January 2007.
[11] Term “INTRODUCTION OF DATA MINING”,”Data Mining: What is Data Mining “, sourcefromhttp://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

[12] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, Dan Steinberg, " Top Ten Data Mining Algorithms", Knowledge and Information Systems Journal, Springer-Verlag London, vol. 14, Issue 1, pp. 1-37, 2007.
[13] Kohavi, R.: Scaling up the accuracy of naïve-bayes classifier: A decision-tree hybrid. In: Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining, pp.202–207. AAAI Press, Menlo Park (1996).
[14] Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993).
[15] Weka – Data Mining Machine Learning Software http/www.cs.waikato.ac.nz/ml/weka/