ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Effective Analysis on Remote to User (R2L) Attacks Using Random Forest Algorithm

S. Revathi*1, Dr. A. Malathi2

*1 Ph.D. Research Scholar, PG and Research, Department of Computer Science, Government Arts College, Coimbatore-18, India

2 Assistant Professor, PG and Research, Department of Computer Science, Government Arts College, Coimbatore-18, India

revathisujendran86@gmail.com

Abstract

This paper focus on analysis of Remote to User attack using random forest algorithm. The R2L attack occurs when the attacker tries to send packets to a machine over a network who have no account on it. The R2L attack leads to vulnerability issues to access secured information from the machine. In intrusion detection system the R2L attacks plays a vital role in accessing unauthorized information and it affects security issues more effectively. This paper proposed a new concept of analyzing individual attacks of R2L using efficient machine learning random forest algorithm which shows high efficiency.

Keyword: Intrusion detection, Remote-to-User attack, Random Forest..

Introduction

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are violations of threats, computer security policies, acceptable use policies, or standard security practices [1]. Intrusion prevention is the process of performing intrusion detection and attempting to stop detected possible incidents. Intrusion detection primarily focused on identifying possible incidents, logging information, attempting to stop, and reporting to security administrators. In addition, organizations use IDPSs for other purposes, such as identifying problems with security policies, documenting existing threats, and deterring individuals from violating security policies [2]. IDPSs have become a necessary addition to the security infrastructure of nearly every organization. The four major attack categories are denial of service, probe, User to Root and Remote to Local attacks [3]. This paper focused on Remote to User attacks where the attackers tries to access machine over the network who are the unauthorized user.

The rest of the paper is organized as in section II is explains the Remote to User attack in NSL-KDD dataset. Section III explains about Random Forest algorithm. Section IV shows experimental analysis and section V draws some conclusion and future works.

Dataset Description

Many Researcher used DARPA 98 [4] and KDDcup99 [5] dataset to examine intrusion detection using various methodologies. The major drawback 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], such drawback leads to new version of KDD dataset as NSL-KDD dataset [7] which has been used as effective benchmark dataset to analysis R2L attacks using machine learning random forest algorithm. 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 Remote-to-User in both training and testing dataset.

Table1: Attacks in Remote-to-User

Attack Names	Training Attacks	Testing Attacks
guess_passwd	53	1231
ftp_write	8	3
imap	11	1
Phf	4	2
multihop	7	18
warezmaster	20	944
warezclient	890	0
Spy	2	0
named	0	17
Sendmail	0	14
snmpgetattack	0	178
snmpguess	0	331
udpstorm	0	2
worm	0	2
xlock	0	9
Xsnoop	0	4
Total	995	2756

The Dictionary, Ftp-Write, Guest and Xsnoop attacks tries to exploit weak or misconfigured system security policies. The Xlock attack involves social engineering attacks to be successful, the attacker must successfully spoof a human operator into supplying their password to a screensaver that is actually a trojan horse.

Random Forest Algorithm

The random forests are an ensemble of unpruned classification or regression trees, generates many decision tree. Each tree is constructed by a different bootstrap sample from the original data using a tree classification algorithm [8]. After the forest is created, a replacement object that must be classified is place down every single tree within the forest for classification. Every tree provides a vote concerning the category of the object. The forest chooses the class with the most votes. By injecting randomness at each node of the grown tree, it has improved accuracy. RF algorithm is given below:

Step 1. Choose T number of trees to grow

Step 2. Choose m wide range of variables used to split each node, m<<M, where M is the number of input variables.

Step 3. Grow trees, while growing each tree do the following:

(a) Construct a sample of size N from M training cases with replacement and grow a tree because of this new sample.

(b) When growing a tree at each and every node select m variables at random from M and bust them out to have the best split.

(c) Grow the tree to a new maximal extent. There isn't really pruning.

Step 4. To classify point X collect votes from every tree on the inside forest and then suddenly use majority voting available to you the class label.

Random Forest has been applied in various fields such as modelling, prediction and intrusion detection system. Zhang and Zulkernine [9] implemented RF in their hybrid IDS to detect known intrusion. They used outlier detection to detect unknown intrusion. RF handles large and unbalanced dataset and does not over fit.

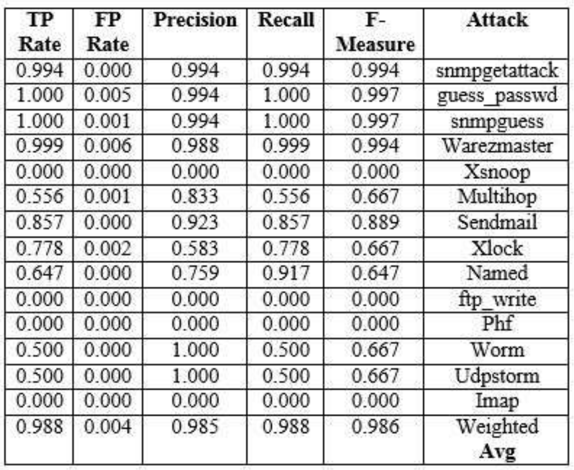
Experimental Result Analysis

The effective analysis of the proposed work has been performed on Weka tool [10] using NSL-KDD dataset. The number or training and testing dataset for R2L is 995 and 2756 which reduce the performance of intrusion detection on experimentation. It consist of 41 attribute which is completely used for experiment. The analyses result has been evaluated based on correctly and incorrectly classified instance with mean square error is shown in table2. Random forest of 10 trees, each constructed using 6 random features. The Out of bag error be 0.0221 and the time taken to build the model is 0.33 sec. The statistical calculation for the cross validation is given below

Table 2: Cross Validation result

Correctly Classified Instances	98.8389 %
Incorrectly Classified Instances	1.1611%
Mean absolute error	0.003

Table 3 Performance of R2L Attacks



TP Rate	FP Rate	Precision	Recall	F-Measure	Attack
0.994	0.000	0.994	0.994	0.994	snmpgetattack
1.000	0.005	0.994	1.000	0.997	guess_passwd
1.000	0.001	0.994	1.000	0.997	snmpguess
0.999	0.006	0.988	0.999	0.994	Warezmaster
0.000	0.000	0.000	0.000	0.000	Xsnoop
0.556	0.001	0.833	0.556	0.667	Multihop
0.857	0.000	0.923	0.857	0.889	Sendmail
0.778	0.002	0.583	0.778	0.667	Xlock
0.647	0.000	0.759	0.917	0.647	Named
0.000	0.000	0.000	0.000	0.000	ftp_write
0.000	0.000	0.000	0.000	0.000	Phf
0.500	0.000	1.000	0.500	0.667	Worm
0.500	0.000	1.000	0.500	0.667	Udpstorm
0.000	0.000	0.000	0.000	0.000	Imap
0.988	0.004	0.985	0.988	0.986	Weighted Avg

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

The precision, recall and f-value based on true positive and false positive for individual attacks are listed in table 3.

Conclusion

The proposed work shows that R2l attacks works more efficiently using random forest algorithm. The paper also explains briefly about various attacks types that present in R2l, various machine learning technique has their own merits to improve classification accuracy and to build pattern classification. From the above result it's clear that Random Forest 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 it 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.

Reference

[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] Breiman.L, 2001,"Random forests," Mach. Learn., vol. 45, pp. 5–32.

[9] Zhang.J and Zulkernine.M, 2005,"Network intrusion detection using random forests," in Proc. 3rd Annu. Conf. Privacy, Secur. Trust (PST), St.Andrews, NB, Canada, pp. 53–61.

[10] Weka – Data Mining Machine Learning Softwarehttp/www.cs.waikato.ac.nz/ml/weka /