ISSN 2395-1621

Potentiate the Detection-Rate of Network Intrusion Detection using Adaboost Algorithm

[#1]Ankita Chowdhury, [#2]Rupali Bhanuse, [#3]Shradha Birajdar, [#4]Devendra Ghorsad

[1]ankita.mousumi@gmail.com,
[2]rupsbhanuse@gmail.com,
[3]shradha.birajdar@gmail.com,
[4]devendra.ghorsad4@gmail.com

[#1234]AISSMS IOIT, Kennedy Road, Near RTO, Pune-411001

ABSTRACT

With the enormous growth of network-based services and users of the Internet, it is important to keep the data and transactions in the Internet more secure. In this article, an Adaboost algorithm for network intrusion detection system with combination of multiple weak classifiers is designed. In this paper, first a conventional online Adaboost process is used where decision stumps are used as weak classifier. In the second algorithm, online Adaboost process is used and online Gaussian mixture models (GMMs) are used as weak classifier. In addition to the algorithm proposed particle swarm optimization (PSO) and support vector machine (SVM) is used. A distributed intrusion detection framework is proposed, in which a local parameterized detection model is constructed in individual node using the online Adaboost algorithm. The global detection model is constructed in each node by combining the local parametric models using a minimum number of samples in the node, which is used to detect intrusions.

ARTICLE INFO

Article History

Received :9th April 2016

Received in revised form :

11th April 2016

Accepted : 13th April 2016

Published online :

15th April 2016

Keywords: Adaboost, Decision Stumps, Online GMM, KDD'99, Network Intrusion.

I. INTRODUCTION

Internet plays a vital role in communication between people. To ensure a secure communication between two parties, we need a security system to detect the attacks very efficiently. Network intrusion detection serves as a major system to work with other security system to provide protection to the computer networks. The main focus of network intrusion detection techniques is to catch, look into the various header parts and data portion of the packets and classify the attack packets from the normal packets. There are mainly two types of intrusion detection systems namely misuse based detection and anomaly based detection. The anomaly based detection system first learns normal user activities and then alerts all user behaviors that differentiate from the already learned activities. The misuse based detection mechanism uses the definite standard patterns of attacks to detect intrusions by representation of the same type of attacks.

In this, we are using KDD'99 as our training data set, and attacks from them as testing data. We use decision stumps as weak classifiers in the first instance. GMM classifier is constructed for each classifier. With this the local models are constructed at each node. After this, PSO and SVM are used for reducing and combining the results. This results in the creation of the global models.

II. RELATED WORK

1. J. B. D. Caberera, B. Ravichandran, and R. K. Mehra Examines the application of statistical traffic modeling for detecting novel attacks against computer networks. In this paper it is discuss the application of network activity models and application models using the 1998 DARPA Intrusion Detection Evaluation data set. Network activity models monitor the volume of traffic in the network, while application models describe the operation of application protocols.

2. W. Lee, S. J. Stolfo, and K. Mork-

This paper describes a data mining framework for adaptively building Intrusion Detection (ID) models. The central idea is to utilize auditing programs to extract an extensive set of features that describe each network connection or host session, and apply data mining programs to learn rules that accurately capture the behavior of intrusions and normal activities. These rules can then be used for misuse detection and anomaly detection.

3. H. G. Kayacik, A. N. Zincir-heywood, and M. T. Heywood-

An approach to network intrusion detection is investigated, based purely on a hierarchy of Self-Organizing Feature Maps. Our principle interest is to establish just how far such an approach can be taken in practice. To do so, the KDD benchmark dataset from the International Knowledge Discovery and Data Mining Tools Competition is employed

CHALLENGES IN EXISTING SYSTEMS

1) Network environments and the attacks training data changes rapidly over time, as new types of attack appear. In addition, the size of the training data expands over time and can become very large. Most existing algorithms for training intrusion detection are offline. The intrusion detector must be reinforced periodically in batch mode in order to keep up with the changes in the network. This reinforcing is time consuming.

2) There are various types of attributes for network connection data, including both categorical and continuous ones, and the value ranges for different attributes differ greatly—from {0, 1} to describe the normal or error status of a connection, to specify the number of data bytes sent from source to destination. The combination of data with different attributes without loss of information is major to maintain the accuracy of intrusion detectors.

3) In traditional centralized intrusion detection, in which all the network data are sent to a central site for processing, the raw data communications occupy considerable network bandwidth. There is a estimation burden in the central site and the privacy of the data obtained from the local nodes cannot be protected.

Dataset Analysis:

KDDCup99 training dataset is about four giga bytes of compressed binary TCP dump data from seven weeks of network traffic, processed into about five million connections record each with about 100 bytes. The two weeks of test data have around two million connection records.

Each KDDCup'99 training connection record contains 41 features and is labeled as either normal

or an attack, with exactly one specific attack type. The attack types are grouped into attack categories in order to combine similar attack types into a single category which could improve the detection rate. Table 1 shows the number of records for each attack category in the training and testing datasets respectively.

Table 1: Number of samples in the KDDCup'99 data set

Dataset	Normal	Attacks				Total
		Probe	Dos	R2L	U2R	
Training Set	97278	4107	391458	1126 52		494021
Testing Set	60593	4166	229853	16189 228		311029

KDD'99 features can be classified into three groups:

1) Basic features: this category encapsulates all the attributes that can be extracted from a TCP/IP connection. Most of these features leading to an implicit delay in detection.

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

a) "Same host" features: examine only the connections in the past 2 seconds that have the same destination host as the current connection, and calculate statistics related to protocol behavior, service,

b) "Same service" features: examine only the connections in the past 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 don't have any intrusion frequent sequential patterns. This is because the DoS and Probing attacks involve many connections to some host(s) in a very short period of time; however the R2L and U2R attacks are embedded in the data portions of the packets, and normally involves only a single connection. To detect these kinds of attacks, we need some features to be able to look for suspicious behavior in the data portion, e.g., number of failed login attempts. These features are called content features.

III.DISTRIBUTED INTRUSION DETECTION FRAMEWORK

There have been many survey of the field Dynamic DIDS. Intrusion detection which contain two models: Local Model and Global Model. Fig. gives an overview of framework that consists of the local models, and global models.

1. Local Models: Local model is constructed into each node by using weak classifiers and Adaboost-based training. So that each node contains a parametric model that consists of the parameters of the weak classifiers and the ensemble weights.

2. Global Models: It is constructed by combining all local parametric models by using PSO and SVM based algorithms. Global models are used to update local models and then updated models are shared by other nodes.

IV. ONLINE ADABOOST-BASED LOCAL INTRUSION DETECTION MODELS

The classical Adaboost algorithm carries out the training task in batch mode. By using training set a number of weak classifiers are generated. The final strong classifier is an ensemble of weak classifiers.

Weak Classifiers: Weak classifier consist two types.

3.1. Decision stumps and normal behaviors for classifying attacks.

The limitation of weak classifier is that the decision stumps do not consider the different types of attacks. This cause the influence in the performance of the Ad boosts method.

3.2. Online GMMs that model a distribution of values of each factor component for each attack type. Online GMM: For each type of attack or normal samples, we use a GMM. Let s Ɛ {+1, -1,-2,....., -N} be a sample label where +1 represents normal samples and 1,-2,....., -N represent different types of attacks where N is number of different type of attacks, s represent the jth element of sample.

Where = number of GMM components indexed by i, w=weight, μ= mean, and σ= standard deviation. Where the computational complexity of the online GMM for one sample is O(k), which is higher than the decision stumps. Design of the weak classifiers and the strong classifier, as shown in Fig.

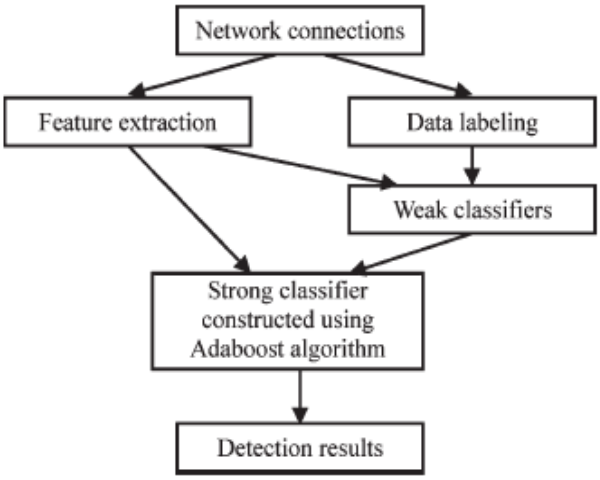

Fig.Framework of Algorithm

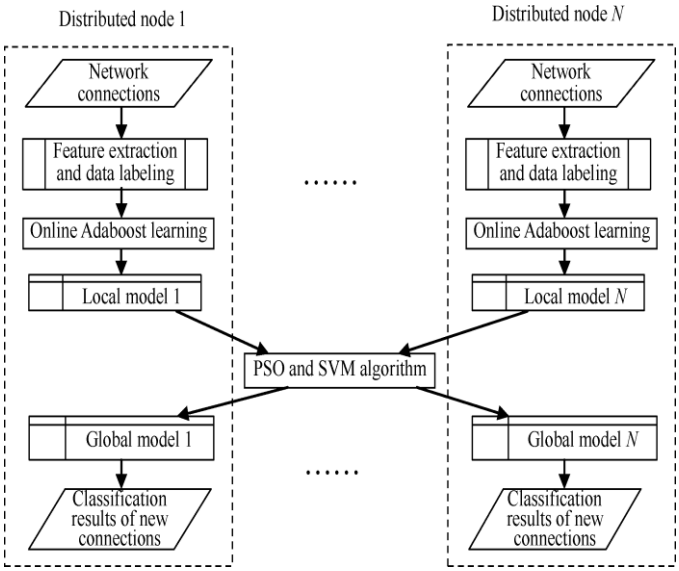

Fig. Overview of the Intrusion Detection framework

V. COMPARATIVE STUDY OF VARIOUS ALGORITHMS TESTED ON THE KDD CUP 1999 DATA SET

Algorithm	Features	Training data		Test data	
		Detection rate (%)	False Alarm rate	Detection rate (%)	False Alarm rate
Decision stumps +Traditional online Adaboost	Only continuous	98.68	8.35	90.05	13.76
	Continuous +Categorical	98.93	2.37	91.27	8.38
Online GMM + Our online Adaboost	Only continuous	98.79	7.83	91.33	11.34
	Continuous +Categorical	99.02	2.22	92.66	2.67

VI. CONCLUSION

This paper has presented a survey and comparative study of the Adaboost-based algorithms that have been proposed towards the improvement of the Dynamic DIDS. We have shown the how online Adaboost algorithm has been helpful for new type of attacks and the way of how it works. The main objective of system is to improve the detection rate and reduce false alarm rate.

Finally we propose the comparative study of algorithms tested on the KDD CUP 1999 dataset.

REFERENCES

[1] Weiming Hu, Jun Gao, Yanguo Wang, Ou Wu, and Stephen Maybank," Online Adaboost-Based Parameterized Methods for Dynamic Distributed Network Intrusion Detection," IEEE

TRANSACTIONS ON CYBERNETICS, VOL. 44, NO. 1, JANUARY 2014

[2] D. Smallwood and A. Vance, "Intrusion analysis with deep packet inspection: Increasing efficiency of packet based investigations," in Proc. Int. Conf. Cloud Service Computing, Dec. 2011, pp. 342–347.

[3] D. Denning, "An intrusion detection model," IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222–232, Feb. 1987.

[4] W. Lee, S. J. Stolfo, and K. Mork, "A data mining framework for building intrusion detection models," in Proc. IEEE Symp. Security Privacy, May 1999, pp. 120–132.

[5] M. Qin and K. Hwang, "Frequent episode rules for internet anomaly detection," in Proc. IEEE Int. Symp. Netw. Computing Appl., 2004, pp. 161–168.

[6] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," in Proc. Int. Joint Conf. Neural Netw., vol. 2. 2002, pp. 1702–1707.

[7] S. Parthasarathy, A. Ghoting, and M. E. Otey, "A survey of distributed mining of data streams," in Data Streams: Models and Algorithms. C. C. Aggarwal (Ed.) New York: Springer, Nov. 2006.

[8] M. E. Otey, A. Ghoting, and S. Parthasarathy, "Fast distributed outlier detection in mixed-attribute data sets," Data Ming Knowl. Discovery, vol. 12, no. 2–3, pp. 203–228, May 2006.

[9] B. Pfahringer, "Winning the KDD99 classification cup: Bagged boosting," SIGKDD Explorations, vol. 1, no. 2, pp. 65–66, 2000.

[10] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, Aug. 1997.

