International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue6- June 2013

Analyzing Intrusion Detection Using Machine Learning Adaboost Algorithm: An Observations Study

Ms.S.S.Kazi
Second Year (M.E.)
Computer Engineering
Sipna's COET,Amravati

Dr.P.R.Deshmukh
Professor
Dept. Of Computer Sci & Engg
Sipna's COET, Amravati

Abstract

The current approaches for intrusion detection have some problems that adversely affect the effectiveness of the Intrusion Detection System. Current approaches often suffer from relatively high false-alarm rates. As most network behaviors are normal, resources are wasted on checking a large number of alarms that turn out to be false. Secondly their computational complexities are oppressively high. The adaboost algorithm gives better results for intrusion detection in this respect. The paper here mainly has its focus on these results.

Keywords: Detection Rate, False Alarm Rate, False Positive Rate, False Negative Rate

Introduction

Intrusion detection on the internet is all the while a heated field in computer science since its initiation from 1987, in which a grand amount of research has been done. An intrusion is defined as any set of actions that attempt to compromise the integrity, confidentiality, or availability of a resource. It is defined to be the potential possibility of a deliberate unauthorized attempt to access information, manipulate information, or render a system unreliable or unusable. Adaboost algorithm initially mainly used in Face Recognition but it also gives much satisfactory results when applied as a Intrusion Detection Techniques. This basis framework to use Adabooost in Intrusion Detection System is suggested by Wei Hu is used here [1].

Intrusion dataset: The KDD CUP 99 data set

The KDD CUP'99 data set was built using TCP packets collected by Defense Advanced Research Projects Agency of U.S. (DARPA) during 1998 intrusion detection evaluation program. Data packets that form a complete session are gathered in a single feature vector or connection record. Data mining techniques and frequently occurring patterns were used to identify features to detect various attack categories. KDD records have forty one features. There are broadly classified four kinds of features present in KDD data set, namely, basic, content, time-based, and host-based features.

The following adaboost algorithm is applied on KDD CUP 99 intrusion data set in order to get the set of observations. The obtained observations are then compared the detections results that are obtained by using Support Vector Machine model [2] in intrusion detection and other detection methods [1].

Adaboost Algorithm

Given: (x_1,y_1)……..(x_n,y_n) where $y_i \epsilon \{+1,-1\}$

Initialize weights: $D_1(i)=1/n$ $(i=1,….,n)$

For t=1,…..,T:

1. Choose a weak classifier h_t which minimizes the weighted error:

h_t= arg min $\varepsilon_j = \sum^{n}_{i=1} D_t(i)I[y_i \neq h_j(x_i)]$

$h_j \epsilon H$

2. If $\varepsilon_t = \min_j \varepsilon_j > ½$, set T=t-1 and stop loop.

3. Choose an α_t

4. Update the weights.

$D_{t+1}(i) = (D_t(i) \exp(-\alpha_t y_i h_t(x_i)))/ Z_t$

Where Z_t is a normalization factor assuring D_{t+1} is a distribution.

The strong classifier is:

$H(x) = \text{sign} (\sum^{T}_{t=1} \alpha_t h_t(x))$

Observations

This paper uses the framework proposed in [1] and applies adaboost algorithm on KDD CUP 99 data set. The code implementation is in java using Netbeans IDE and uses various training and testing data sets. It came across following sets of observations.

Number of samples of various types in Training Data

Normal	Attack				Total
	DOS	U2R	R2L	PROBE	
	242027	39	5993	2377	
60593	250436				311029

Number of samples of various types in Testing Data

Normal	Attack				Total
	DOS	U2R	R2L	PROBE	
	117608	21	2901	1150	
28865					150545

Results of Adaboost at various iterations for Training Data

Iteration	Detection Rate	False Alarm Rate	False Positive Rate	False Negative Rate
3	95.92%	4.08%	0.88%	3.21%
7	96.32%	3.68%	0.79%	2.89%
14	96.73%	3.27%	0.70%	2.57%
18	97.14%	2.86%	0.61%	2.25%
21	97.55%	2.45%	0.53%	1.93%
27	97.96%	2.04%	0.44%	1.61%
34	98.36%	1.64%	0.35%	1.28%
36	98.77%	1.23%	0.26%	0.96%
38	98.77%	1.23%	0.26%	0.96%
40	**98.77%**	**1.23%**	**0.26%**	**0.96%**

Results with improved initialized weights

r	Training Set		Test Set	
	FPR (%)	DR (%)	FPR (%)	DR (%)
0	99.99	100	100	100
0.1	7.35	99.32	8.25	98.99
0.2	4.89	99.29	5.43	98.92
0.3	3.71	99.14	3.83	98.01
0.4	2.34	99.07	2.68	97.99
0.5	**0.26**	**98.77**	**0.45**	**97.89**
0.6	0.19	98.23	0.29	97.01
0.7	0.16	98.29	0.25	96.92
0.8	0.12	98.1	0.19	96.21
0.9	0.09	97.94	0.12	95.25
1	0	0.89	0.08	0.91

Conclusion.

The intrusion detection system designed by using adaboost algorithm has better performance as compared to other detection methods. It also has low computational complexity, high detection rate and low false alarm rate and hence can have wider domain of applicability and hence can be treated as a considerably good intrusion detection system.

Result Comparison		
Methods	**FPR (%)**	**DR (%)**
Genetic Clustering	0.3	79
Bagged C5	2.19-3.99	91.81
RSS-DSS	0.27-3.5	89.2-94.4
SOM	2.19-3.99	90.94-93.46
SVM	6.0-10	91-98
Adaboost	**0.45**	**97.89%**

References

[1] "Network based Intrusion Detection Using Adaboosl Algorithm" Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05) ,Wei Hu and Weiming Hu

[2] "An Enhanced Support Vector Machine Model for Intrusion Detection" JingTao Yao, Songlun Zhao, and Lisa Fan

[3] Z. Zhang and H. Shen: Online training of svms for real time intrusion detection based on improved text categorization model. Computer Communications. February 2005

[4] C. Elkan. Results of the kdd99 classifier learning contest. SIGKDD Explorations, 1(2):63-64, 2000.

[5] D. Denning. An intrusion-detection model. IEEE Transactions on Software Engineering. 13(2):222-232, February 1987.

[6] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Computer and System Sciences, August 1997

[7] P. Hong, Zhang, and T. Wu "Intrusion detection method based on rough set and svm" algorithm. In Proceedings of International Conference on Communications, Circuits and Systems, volume 2, pages 1127-1 130, June 2004

[8] H. G. Kayacik, A. Zincir-Heywood, and M. Heywood. On the capability of an som based intrusion detection system. In Proceedings of the International Joint Conference on Neural Networks, volume 3, pages 1808-1813, July 2003.

[9] W.Lee, S.J.Stolfo. A framework for constructing features and models for intrusion detection systems. ACM Transactions on Information and System Security, 3(4):227- 261, November 2000.