Feature Reduction for Intrusion Detection Using Linear Discriminant Analysis

Rupali Datti[1], Bhupendra verma[2]

[1] PG Research Scholar Department of Computer Science and Engineering, TIT, Bhopal (M.P.)

rupal3010@gmail.com

[2] Professor in Computer Science and Engineering, TIT, Bhopal (M.P.),

Abstract: Intrusion detection is one of core technologies of computer security. It is required to protect the security of computer network systems. Most of existing IDs use all features in the network packet to look for known intrusive patterns. Some of these features are irrelevant or redundant. A well-defined feature extraction algorithm makes the classification process more effective and efficient. The Feature extraction step aims at representing patterns in a feature space where the highest discrimination between legitimate and attack patterns is attained. The Classification step perform the intrusion detection task either by alerting if an observed pattern is described by an attack patterns model, usually called signature or misuse-based IDS, or by alerting if it is not described by a model of legitimate activity, usually called anomaly-based IDs.In this paper, Linear Discriminant Analysis algorithm is used to extraction of features for detecting intrusions and Back Propagation Algorithm is used for classification of attacks. Tests are done on NSL-KDD dataset which is improved version of KDD-99 data set. Results showed that the proposed model gives better and robust representation as it is able to transform features resulting in great data reduction, time reduction and error reduction in detecting new attacks.

Keywords: Linear Discriminant Analysis, NSL-KDD, Feature Extraction.

1. Introduction

With the tremendous growth of network-based services and sensitive information on networks, network security is getting more and more importance than ever. Intrusion poses a serious security risk in a network environment. The ever growing new intrusion type poses a serious problem for their detection. The human labeling of the available network audit data instances is usually tedious, time consuming and expensive. Anomaly detection and misuse detection [1] are two general approaches to computer intrusion detection system. Most of the existing IDs use all 41 features in the network to evaluate and look for intrusive pattern, some of these features are redundant and irrelevant. The drawback of this approach is time-consuming detection process and degrading the performance of ID system, thus we need to remove the worthless information from the original high dimensional database. To improve the generalization ability, we usually generate a small set of features from the original input variables by feature extraction. Feature extraction has basically two aims: First to shrink the original dimension of the feature vector to a reasonable size and second to eventually improve the classification accuracy by retaining the most discriminatory information and deleting the irrelevant and redundant information. Many current feature extraction techniques involve linear transformations of the original pattern vectors to new vectors of lower dimensionality. Linear Discriminant analysis feature reduction technique is novel approach used in the area of cyber attack detection. This not only reduces the number of the input features but also increases the classification accuracy and reduces the

training and testing time of the classifiers by selecting most discriminating features. We use Artificial Neural Network (ANN) classifier to compare the performance of the proposed technique.

In our experiment, we used NSL-KDD data set. It has solved some of the inherent problems of the KDD'99[2] which is considered as standard benchmark for intrusion detection evaluation [3]. The training dataset of NSL-KDD similar to KDD99 consist of approximately 4,900,000 single connection vectors each of which contains 41 features and is labeled as either normal or attack type ,with exactly one specific attack type . Empirical studies indicate that feature reduction technique is capable of reducing the size of dataset. The time and space complexities of most classifiers used are exponential function of their input vector size [4].

This paper organized as follows, in the second section we give an introduction to NSL-KDD dataset, section three gives the information about networking attacks, section four shows Importance of data reduction for intrusion detection systems**,** section fifth explains the proposed algorithm, section sixth gives a brief introduction to Linear Discriminant Analysis, section seventh shows the experimental set up and results finally in section eight conclusion is shown.

2. Introduction of NSL-KDD:

KDDCUP'99 is the mostly widely used data set for the anomaly detection. But researchers conducted a statistical Analysis on this data set and found two important issues which highly affects the performance of evaluated systems, and results in a very poor evaluation of anomaly detection approaches. To solve these issues, they have proposed a new data set, NSL-KDD, which consists of selected records of the complete KDD data set [5].

The following are the advantages of NSL-KDD over the original KDD data set:

- It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records.
- The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set. As a result, the classification rates of distinct machine learning methods vary in a wider range, which makes it more efficient to have an accurate evaluation of different learning techniques.
- The numbers of records in the train and test sets are reasonable, which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion. Consequently, evaluation results of different research works will be consistent and comparable.

NSL-KDD features can be classified into three groups [2]:

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) **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.

3) **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, etc.

b) "Same service" features: examine only the connections in the past 2 seconds that have the same service as the current connection.

The two aforementioned types of "traffic" features are called time-based. However, there are several slow probing attacks that scan the hosts (or ports) using a much larger time interval than 2 seconds, for example, one in every minute. As a result, these attacks do not produce intrusion patterns with a time window of 2 seconds. To solve this problem, the "same host" and "same service" features are re-calculated but based on the connection window of 100 connections rather than a time window of 2 seconds. These features are called connection-based traffic features.

Table 1 shows all the features found in a connection. For easier referencing, each feature is assigned a label (A to AO). Some of these features are derived features. These features are either nominal or numeric.

Table 1: Basic features of individual TCP connection

Label	Network Data Features
A	Duration
B	protocol_type
C	Service
D	flag
E	src_bytes
F	dst_bytes
G	land
H	wrong fragment
I	urgent

Table 2: Content features within a connection suggested by domain knowledge

Label	Network Data Features
J	Hot
K	num_failed_logins
L	logged_in
M	num_compromised
N	root_shell
O	su_attempted
P	num_root
Q	num_file_creations
R	num_shells
S	num_access_files
T	num_outbounds_cmds
U	is_hot_login
V	is_guest_login

Table 3: Traffic features computed using a two-second

Label	Network Data Features
W	Count
X	sev_count
Y	serror_rate
Z	sev_serror_rate
AA	rerror_rate
BB	srv_rerror_rate
AC	same_srv_rate
AD	diff_srv_rate
AE	srv_diff_host_rate
AF	Dst_host_count
AG	Dst_host_srv_count
AH	Dst_host_same_srv_rate
AI	Dst_host_diff_srv_rate
AJ	Dst_host_same_src_port_rate
AK	Dst_host_srv_diff_host_rate
AL	Dst_host_server_rate
AM	Dst_host_srv_serror_rate
AN	Dst_host_rerror_rate
AO	Dst_host_srv_rerror_rate

Table 1, Table 2 and Table 3 show all the features found in a connection. For easier referencing, each feature is assigned a label (A to AO).some of the features.

3. NETWORKING ATTACKS

The simulated attacks were classified, according to the actions and goals of the attacker. Each attack type falls into one of the following four main categories [5]:

i) Denials-of Service (DoS) attacks have the goal of limiting or denying services provided to the user,computer or network. A common tactic is to severely overload the targeted system. (e.g. apache, smurf, Neptune, Ping of death, back, mailbomb,udpstorm, SYNflood, etc.).

ii) Probing or Surveillance attacks have the goal of gaining knowledge of the existence or Configuration of a computer system or network.Port Scans or sweeping of a given IP address range typically fall in this category. (e.g. saint, portsweep,mscan, nmap, etc.).

iii)User-to-Root (U2R) attacks have the goal of gaining root or super-user access on a particular computer or system on which the attacker previously had user level access. These are attempts by a non-privileged user to gain administrative privileges (e.g. Perl, xterm, etc.).

iv)Remote-to-Local(R2L) attack is an attack in which a user sends packets to a machine over the internet, which the user does not have access to in order to expose the machine vulnerabilities and exploit privileges which a local user would have on the computer (e.g. xclock, dictionary, guest_password, phf, sendmail, xsnoop, etc.).

4. Importance of data reduction for intrusion detection systems:

IDS have become important and widely used tools for ensuring network security. Since the amount of audit data that an IDS needs to examine is very large even for a small network, classification by hand is impossible. Analysis is difficult even with computer assistance because extraneous features can make it harder to detect suspicious behavior patterns. Complex relationships exist between the features, which are practically impossible for humans to discover. IDS must therefore reduce the amount of data to be processed. This is extremely important if real-time detection is desired. Reduction can occur in one of several ways. Data that are not considered useful can be filtered, leaving only the potentially interesting data. Data can be grouped or clustered to reveal hidden patterns. By storing the characteristics of the clusters instead of the individual data, overhead can be significantly reduced. Finally, some data sources can be eliminated using feature reduction.

5. Proposed Algorithm:

LDA is a widely used feature dimension reduction method,it provides a linear transformation of n-dimensional feature vectors (or samples) into m-dimensional space ($m < n$), so that samples belonging to the same class are close together but samples from different classes are far apart from each other.

In our work we used LDA as a reduction tool and feed forward neural networks as a learning tool for the developed system, first to test the efficiency of the system after the removal of superfluous features and then to efficiently detect any intrusions.

Steps used in our algorithm

1. Data Preprocessing

Normalization is used for data preprocessing, where the attribute data are scaled so as to fall within a small specified range such as -1.0 to 1.0 or 0.0 to 1.0. If using neural network back propagation algorithm for classification, normalizing the input values for each attribute measured in the training samples will help speed up the learning phase.

2. Intermediate reduction Using Information Gain

Linear Discriminant Analysis (LDA) is a dimension reduction method which finds an optimal linear

transformation that maximizes the between-class scatter and minimizes the within class scatter. However, in under sampled problems where the number of samples is smaller than the dimension of data space, it is difficult to apply the LDA due to the singularity of scatter matrices caused by high dimensionality. In order to make the LDA applicable, several generalizations of the LDA have been proposed [6]. A common way to deal with the singularity problem is to apply an intermediate dimensionality reduction stage. Here we use INFORMATION GAIN [7] method for this purpose.

3. Dimensionality Reduction Using LDA

Linear discriminant analysis (LDA) is a classical statistical approach for supervised dimensionality reduction and classification. LDA computes an optimal transformation (projection) by minimizing the within-class distance and maximizing the between-class distance simultaneously, thus achieving maximum class discrimination. The optimal transformation in LDA can be readily computed by applying an eigen decomposition on the so-called scatter matrices. It has been used widely in many applications involving high-dimensional data [8].

4. Classification Using Back-Propagation Algorithm

Back-propagation algorithm is used for classification of attack classes as is capable of making multi-class classification.

6. Linear Discriminant Analysis

Goal of LDA is to

- Perform dimensionality reduction "while preserving as much of the class discriminatory information as possible".
- Seeks to find directions along which the classes are best separated.
- Takes into consideration the scatter within-classes but also the scatter between-classes.

Linear Discriminant Analysis (LDA) finds the vectors in the underlying space that best discriminate among classes [9]. For all samples of all classes the between-class scatter matrix S_B and the within-class scatter matrix S_W are defined by:

$$\boldsymbol{S_b} = \sum_{i=1}^{C} (\boldsymbol{\mu_i} - \boldsymbol{\mu})(\boldsymbol{\mu_i} - \boldsymbol{\mu})^T$$

$$\boldsymbol{S_w} = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (\boldsymbol{Y_j} - \boldsymbol{\mu_i})(\boldsymbol{Y_j} - \boldsymbol{\mu_i})^T$$

Where $\boldsymbol{M_i}$ is the number of training samples in class $\boldsymbol{i}$, $\boldsymbol{C}$ is the number of distinct classes, $\boldsymbol{\mu_i}$ is the mean vector of samples belonging to class $\boldsymbol{i}$ and $\boldsymbol{Y_j}$ represents the set of samples belonging to class $\boldsymbol{i}$ with $\boldsymbol{Y_j}$ being the $\boldsymbol{j}$ th data of that class. $\boldsymbol{S_w}$ represents the scatter of features around the mean of each class and $\boldsymbol{S_B}$ represents the scatter of features around the overall mean for all classes. The goal is to maximize $\boldsymbol{S_B}$ while minimizing $\boldsymbol{S_w}$, in other words, maximize the ratio

$$\mathbf{det} \mid \mathbf{S_B} \mid / \mathbf{det} \mid \mathbf{S_w} \mid$$

This ratio is maximized when the column vectors of the projection matrix are the eigenvectors of $\mathbf{S_w^{-1}}\,\mathbf{S_B}$. In order to prevent S_w to become singular, Information Gain is used as a preprocessing step.

7. Experimental setup and Results:

We ran our experiments on a system with a 2.20GHZ core 2 due processor and 3GB of RAM running windows XP. We used java programming for implementation. Feed forward back propagation neural network algorithm has been developed for training process.

The network has to discriminate the different kinds of anomaly –based intrusions. In this work 11850 training sample, 9652 sets of test samples with 41features are used. First test is applied on all 41 features with 25 and 15 hidden layers, 5 output neuron, and 0.5 learning rate are used for optimum results for classification of all data before feature reduction. By applying the reduction algorithm defined by Linear Discriminant analysis for 11850 samples 41 features are reduced to only 4 features this gives 97% reduction in input data and approximately 94% time reduction in training with almost same accuracy achieved in detecting new attacks. Feed forward back propagation neural network

architecture with 25 hidden layer and 5 output neuron was giving the optimum result for 4 reduced featured data.

In this experiment, five-class classification is done. The Normal data belongs to class 1, DOS belongs to class 2, probe belongs to class 3, user to super-user belongs to class 4, and remote to local belongs to class 5.

Following is the table which shows the results with Hidden layers 25 and 15.

Table 2. RESULT with Hidden Layers=25 and Step=100

	Before Features Reduction(With all 41 Features)	After Feature Reduction(With only 4 Features)	Target
Normal	7440	6956	6836
DOS	2547	2256	2625
Probe	1863	1332	1097
U2R	0	34	37
R2L	0	1272	1255
Training Time taken	540015	33500	

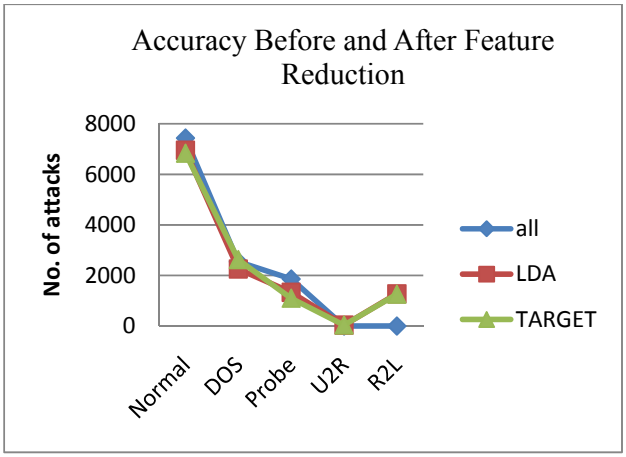

Figure 1. Accuracy Before and After Reduction with Hidden Layer=25 and Steps=100

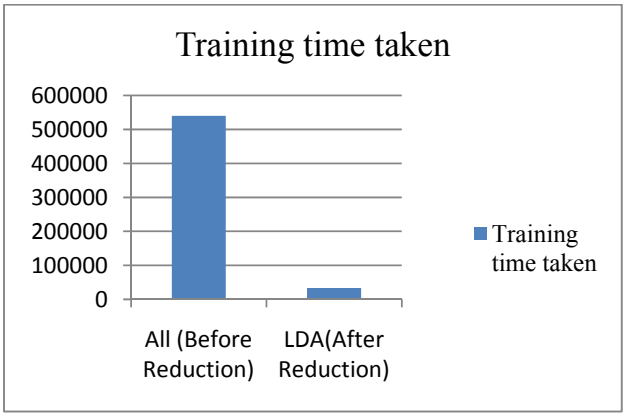

Figure 2. Training Time Before and After Reduction with Hidden Layer=25 and Steps=100

8. Conclusion:

Current intrusion detection systems (IDS) examine all data features to detect intrusion or misuse patterns. Some of the features may be redundant or contribute little (if anything) to the detection process. The purpose of this paper is to identify important input features in building IDS that is computationally efficient and effective.

Our experimental results show that the proposed model gives better and robust representation of data as it was able to reduce features resulting in a 97% data reduction and approximately 94% time reduction in training with almost same accuracy achieved in detecting new attacks. Meantime it significantly reduce a number of computer resources, both memory and CPU time, required to detect an attack .This shows that our Proposed algorithm is reliable in intrusion detection.

References

[1] F.sabahi, A.movaghar, “intrusion detection: A survey”, the third international conference on systems and network communications,2008.

[2] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali, A. Ghorbani, “ A Detailed Analysis of the KDD CUP 99 Data Set”, proceeding of IEEE symposium on computational Intelligence in security and defence application, 2009

[3] KDD Cup 1999. Available on http://kdd.ics.uci.edu/ Databases/kddcup 99/kddcup99.html, Ocotber 2007.

[4] R.O.Duda, P.E.Hart, and D.G.Stork, “Pattern Classification”, Vol. 1, Wiley, 2002.

[5] A.Sung & S.Mukkamala, "Identifying important features for intrusion detection using SVM and neural networks,”in symposium on application and the Internet, pp 209-216,2003.

[6] H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood,” Selecting Features for Intrusion Detection:A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets”, PST, 2005

[7] Jiawei han,Micheline kamber, Data Mining concepts and techniques, Elsevier Publication,II Edition,2006.
[8] Jieping Ye, Numerical Linear Algebra for Data Exploration Linear Discriminant Analysis,CSE 494 CSE/CBS 598 (Fall 2007).
[9] Kresimir Delac, Mislav Grgic, Sonja Grgic, Independent Comparative Study of PCA, ICA,and LDA on the FERET Data Set, University of Zagreb, FER, Unska 3/XII, Zagreb, Croatia, vol.15, 252-260, 2006.