SVM Based Intrusion Detection Using Nonlinear Scaling Scheme

Xiaotang Tang*, Sheldon X.-D. Tan†, Hai-Bao Chen*
*Department of Micro and Nano Electronics, Shanghai Jiaotong University, Shanghai, China
†Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521

***Abstract*—Intrusion is one of major security issues of internet with the rapid growth in smart and Internet of Thing (IoT) devices, and it becomes important to detect attacks and set out alarm. In this paper, Support Vector Machine (SVM) with nonlinear input data scaling scheme is proposed to detect attacks, which is different than the existing linear scaling based machine learning methods. Experiments on the NSL-KDD dataset show that the performances of the proposed method are compared favorably with existing works. The detection rate from the new method is 82.2% for binary-classification, compared to 81.2% by existing Artificial Neural Networks (ANN) based works. For multi-classification, the proposed method shows similar performances of ANN. Further more, the detection rate of Denial of Service (DoS) is 86.5%, compared to 77.7% by existing ANN based works.**

I. Introduction

Intrusion is a major threat to security of Internet, which disrupts the normal information system activities. In order to identify attacks initially, using Intrusion Detection System (IDS) is essential. IDS plays an important role in system security and confidentiality, with detecting attacks to traffics among users and servers. Usually, IDS can be classified into host-based and network-based intrusion detection systems [1]. Host-based detection captures and analyzes data at the attacked system itself, while network-based detection captures and inspects online network data at the gateway or server before the attack's reaching.

Every attack can be seen as network anomaly in data flow, hence network data can be classified to normal or attack. The widely used and publicly available datasets for IDS are DARPA, KDD 99 and NSL-KDD datasets [2]. The KDD99 is originated and improved from DARPA 1998 in 1999, introduced by MIT Lincoln Laboratories. NSL-KDD is a dataset suggested to solve some inherent problems of KDD99 by eliminating redundant and duplicated records, though it may still not be a perfect representative of existing real networks, and it can be applied as an effective benchmark dataset to help to different intrusion detection method comparison. Attack types can be generally grouped into the following 4 categories [3]: Denial of Service (DoS) - A malicious attempt to block system by attacking the memory and computing resources; Probe - An attack to collect information about vulnerabilities of the target system; Remote to Local (R2L) - An unauthorized ability to get access to the target system and even dump data; User to Root (U2R) - An attack to gain administrative privilege for attackers as a normal user.

Several researches have been carried out on IDS and various anomaly detection techniques have been proposed. Ahmed et al. [4] reviewed intrusion detection works and datasets from 2009 to 2014, and compared performances of attack detection with classification, statistical and clustering methods. Many works used cross validation detect accuracy as evaluation, which is only for known attack. Yogita B. Bhavsar et al. [5] used Support Vector Machine (SVM) classifier on NSL-KDD and gained accuracy of 98.57% with 10-fold cross validation. Bajaj et al. [6] applied information gain model based on J48, SVM, NB Tree, Naïve Bayes and simpleCart methods and improved the time complexity by feature selection for binary-classification. Ibrahim et al. [7] applied Self-organizing Map (SOM) on both KDD99 and NSL-KDD datasets and showed accuracy of 75.5% on binary-classification of NSL-KDD test dataset. Ingre [8] applied Artificial Neural Networks (ANN) with Levenberg-Marquardt (LM) and BFGS quasi-Newton Back propagation (BFG) on NSL-KDD, which gained accuracy of 81.2% on binary-classification by feature reduction and accuracy of 79.9% on multi-classification using NSL-KDD test dataset.

In this paper, a SVM with new nonlinear scaling is proposed to detect intrusion in Section 2, which is different than the linear scaling as a stage of feature processing, and four types of kernel functions have been studied in SVM to obtain the best performance. Experimental results show that the detection rate of binary-classification with the proposed method reaches to 82.2%, compared to 81.2% by ANN based works [8]. And the detection rate of multi-classification is 79.6%, especially DoS detection rate is improved to 86.5%, compare to 77.7% by ANN based works.

II. The proposed SVM based detection with non-linear scaling scheme

Typically SVM was employed as classifier for intrusion detection. In this paper, we propose a new SVM based intrusion detection technique with the non-linear scaling for the training data as shown as Fig. 1. In this flow, instead of using traditional MinMax normalization, we use non-linear scaling scheme to process features of the data, which gets better detection results as shown later.

A. Motivation for nonlinear scaling

Typically, the MinMax normalization, which can be viewed as a linear scaling method, has been used widely in machine learning algorithms to process features of the data. The MinMax normalization is typically done via the following equation:

$$x' = \frac{x - min}{max - min} \tag{1}$$

The main problem with the MinMax scaling is that the scaling depends on the minimum and maximum values from the training and testing data. However, if the maximum and minimum values from training and testing data are not the

This work is supported in part by the Nature Science Foundation of China (NSFC) under No. 61604095, in part by a 985 research fund from Shanghai Jiao Tong University. *Correspondence author: Hai-Bao Chen. Email: haibaochen@sjtu.edu.cn.*

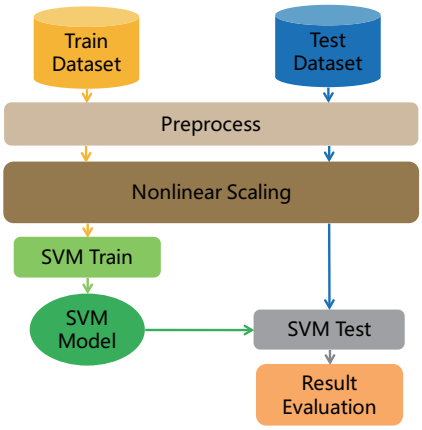


Fig. 1. Intrusion detection using SVM with nonlinear scaling

same, which will affect the final detection accuracy during the testing phase. Table I shows the difference of the minimum and maximum values from the training and testing data from NSL-KDD [2]. We can see that the minimum and maximum values are quite different in general. As a result, a better data-independent scaling should be used, which is the major motivation of this work.

TABLE I
FEATURE VALUE RANGE COMPARISON

Feature	Training Dataset		Testing Dataset	
	Min	Max	Min	Max
duration	0	2142658	0	2157715
srcbytes	0	381709090	0	62825648
dstbytes	0	5151385	0	1345927
urgent	0	1	0	3
hot	0	77	0	101
numcompromised	0	884	0	796
numroot	0	975	0	878
numfilecreations	0	40	0	100
numshells	0	1	0	5
numaccessfiles	0	8	0	4

To mitigate this problem, we propose two data-independent scaling methods, which are nonlinear in general.

Logistic scaling: A logistic scaling is a common "S" shape (sigmoid-like curve), which can be described by

$$x' = \frac{1}{1+e^{-x}} \tag{2}$$

where $x \in [0, +\infty)$, considering all the numerical feature values are positive in our problem, thus $x' \in [0.5, 1)$.

Arctan scaling: An arctan scaling has also sigmoid-like curve, which can be described as

$$x' = \arctan(x) \tag{3}$$

where $x \in [0, +\infty)$, thus $x' \in [0, \frac{\pi}{2})$ in our problem.

B. SVM for training

In machine learning, SVM [9] is a supervised learning model used for classification and regression analysis. The SVM model represents the patterns as points in space, mapping so that the patterns of the separate categories are divided by an exactly clear gap, which is as wide as possible. New patterns are then mapped into that same space, and predicted to belong to a certain category based on which side of the gap they fall. Moreover, kernel functions enable patterns to be operated in a high-dimension by mapping them to a space where the gaps can be created for classification. In our SVM implementation, four kernel functions [10], linear, polynomial, Radial Basis Function (RBF) and sigmoid, are employed for model building and training.

C. Detection accuracy analysis

The performance of SVM can be evaluated by using various measurements [11], standard parameters include True Positive (TP), False Positive (FP), True Negative(TN) and False Negative (FN), which can be presented by Confusion matrix and shown as Table II.

TABLE II
CONFUSION MATRIX

Class		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

Recall, also called detection rate, measures the system to detect positive instances correctly, formulated as

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

ACC, or called overall accuracy, is defined as a ratio of TP and TN to the total instances, formulated as

$$ACC = \frac{TP+TN}{FP+FN+TP+TN} \tag{5}$$

False Positive Rate (FPR), measures the system to misjudge negative instances as positive, formulated as

$$FPR = \frac{FP}{FP+TN} \tag{6}$$

Matthews Correlation Coefficient (MCC) [12], measures the quality of system binary classification, range -1 to +1, which +1 represents 100% prediction, -1 represents 0% prediction, and 0 represents no better than random prediction. MCC is formulated as

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{7}$$

III. EXPERIMENT AND RESULTS AND DISCUSSIONS

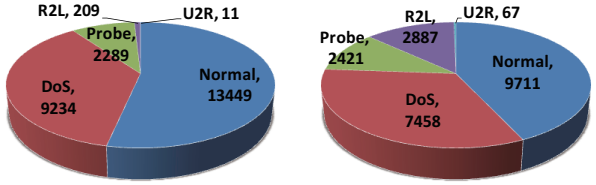
Experiments are performed on Intel(R) Core(TM) i5-4590 CPU @ 3.30GHz processor with 8GB RAM, and consist of several approaches, which differ in scaling methods in data processing and kernel functions used in SVM. Both binary-classification and multi-classification are experimented, with KDDTrain+_20Percent data for training and KDDTest+ data for testing. LIBSVM [13] is a library utilized for SVM training and testing.

A. Review of the NSL-KDD dataset

The raw NSL-KDD dataset needs to be preprocessed so that the data can suit the later SVM model well. NSL-KDD dataset is applied to our proposed method, which is in the form of 41 features and one class label. Hence, it precisely suits SVM, a supervised classifier. In NSL-KDD dataset, KDDTrain+_20Percent and KDDTest+ are respectively employed as training dataset and testing dataset. The training dataset contains 22 attack types where the testing dataset contains additional 17 attack types. Table III shows that these attack types can be grouped to DoS, probe, R2L and U2R. The distribution

TABLE III
ATTACK CATEGORY

Category	Attack Type
DoS	apache2,back,land,mailbomb,neptune,pod, processtable,smurf,teardrop,udpstorm
Probe	ipsweep,mscan,nmap,portsweep,saint,satan
R2L	ftp_write,guess_passwd,httptunnel,imap,multihop, named,phf,sendmail,snmpgetattack,snmpguess, spy,warezclient,worm,xlockp,xsnoo
U2R	buffer_overflow,loadmodule,perl,ps,rootkit, sqlattack,xterm;



(a) Train Dataset (b) Test Dataset

Fig. 2. Dataset distribution: (a) train dataset, (b) test dataset

of classes in training dataset and testing dataset is shown as Fig.2.

Except for the attack category classifying, the main task in data preprocessing is to transform symbolic labels and features to numerical values in NSL-KDD dataset. The labels need to be transformed by setting value -1 or 1 for binary-classification and value range from 1 to 5 for multi-classification. Table IV shows that the feature protocol type, service and flag need to be transformed. Protocol type is valued range from 1 to 3. To integrate all the services in both training dataset and testing dataset, since they are somewhat different to each other, service is valued range from 1 to 67, and flag is valued range from 1 to 11.

TABLE IV
FEATURE TRANSFORMATION

Feature	Numerical Value Transformation
Protocol Type	tcp=1,udp=2,icmp=3
Service	aol=1,auth=2,...,tftp_u=67
Flag	OTH=1,REJ=2,...,SH=11

B. Binary-classification study

Binary-classification classifies instances as normal or anomaly. Performances of 4 kernel functions with logistic scaling are shown as Table V, which tells that polynomial kernel performs well in train ACC 99.4% and test ACC 82.1% , and effectively spends 12s in training and 1s in testing. Performances of 4 kernel functions with arctan scaling are shown as Table VI, which tells that polynomial kernel performs well in train ACC 99.8% and test ACC 82.2% , and spends 131s in training and 1s in testing. Moreover, the normal and anomaly detection results of test dataset with polynomial kernel in three scaling methods are shown as Fig. 3, and it is obvious that nonlinear scaling (logistic/ arctan) performs better than linear scaling (MinMax), by declining normal detection FPR from 35% to 29% and increasing anomaly detection recall from 65.4% to 71%.

C. Multi-classification study

Multi-classification classifies instances as normal, DoS, Probe, R2L and U2R. Performances of 4 kernel functions with logistic and arctan scaling are shown respectively as Table VII and Table VIII. It is evident that polynomial kernel with logistic and arctan scaling perform well in accuracy of 79.2% and 79.6%, by spending 13s and 1657s in training and 1s and 2s in testing.

TABLE V
BINARY-CLASSIFICATION WITH LOGISTIC SCALING

Kernel	Train ACC (%)	Test ACC (%)	Time (s)	
			Train	Test
Linear	96.7	78.7	8.0	3.0
Polynomial	**99.4**	**82.1**	**12.0**	**1.0**
RBF	99.1	77.8	4.0	3.0
Sigmoid	93.5	78.2	21.0	18.0

TABLE VI
BINARY-CLASSIFICATION WITH ARCTAN SCALING

Kernel	Train ACC (%)	Test ACC (%)	Time (s)	
			Train	Test
Linear	97.5	77.4	8.0	3.0
Polynomial	**99.8**	**82.2**	**131.0**	**1.0**
RBF	98.6	76.7	6.0	4.0
Sigmoid	96.3	76.3	17.0	14.0

Each category detection result of test dataset with polynomial kernel in three scaling methods is shown as Fig. 4, and it's obvious that nonlinear scaling (logistic/ arctan) performs better than linear scaling (MinMax) in general, by increasing weighted average recall from 78.5% to 79.6%. Especially, DoS detection MCC is increased from 0.804 (MinMax) to 0.881 (arctan). Probe detection MCC is increased from 0.686 (MinMax) to 0.745 (arctan). U2R detection MCC is increased from 0.407 (MinMax) to 0.422 (arctan). However, R2L detection rate is somehow declined from 0.529 (MinMax) to 0.332 (arctan), since R2L detection has been an issue for machine learning based classification [4].

TABLE VII
MULTI-CLASSIFICATION WITH LOGISTIC SCALING

Kernel	Train ACC (%)	Test ACC (%)	Time (s)	
			Train	Test
Linear	97.5	77.5	6.0	3.0
Polynomial	**99.5**	**79.2**	**13.0**	**1.0**
RBF	98.7	77.7	6.0	5.0
Sigmoid	94.7	75.5	22.0	16.0

D. Comparison with existing methods

Since NSL-KDD is kind of new benchmark for IDS research, which lacks of sufficient results of related works to compare, considering of unified test dataset KDDTest+. Therefore, ANN based method proposed by Ingre [8] and SOM

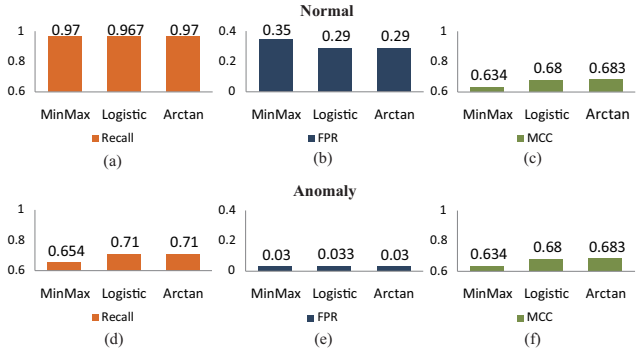


Fig. 3. Binary-classification results of test dataset with polynomial kernel

TABLE VIII
MULTI-CLASSIFICATION WITH ARCTAN SCALING

Kernel	Train ACC (%)	Test ACC (%)	Time (s)	
			Train	Test
Linear	98.3	77.8	4.0	2.0
Polynomial	**98.8**	**79.6**	**1657.0**	**2.0**
RBF	98.4	76.6	7.0	5.0
Sigmoid	96.4	77.1	19.0	14.0

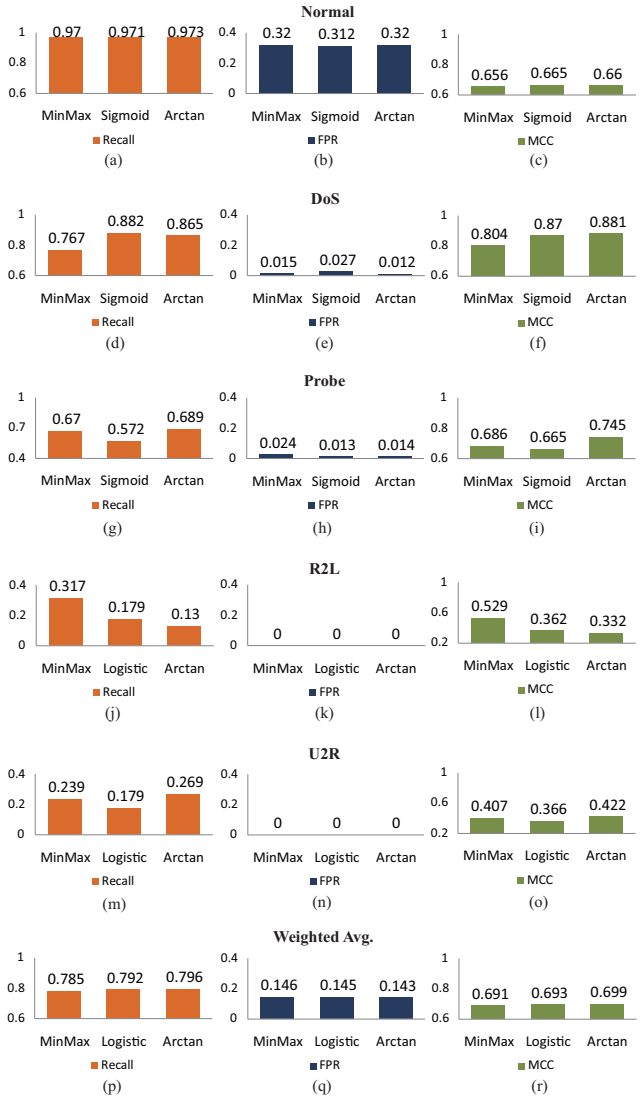


Fig. 4. Multi-classification results of test dataset with polynomial kernel

based method proposed by Ibrahim et al. [7] are comparable, which SOM based method only did binary-classification and ANN based method did both classifications. Table IX shows the proposed method performs best in binary-classification with recall 82.1% and FPR 14.3%. Especially, result comparison of DoS detection, the most demanding attack detecting, is shown as Table X, which tells that the proposed method by using arctan scaling outperforms ANN based method with recall 86.5% and FPR 1.2%.

IV. CONCLUSION

In this paper, we studied Support Vector Machine (SVM) based intrusion detection systems. We proposed to use the data-independent nonlinear scaling (logistic and arctan) for the processing of features of the input data instead of traditional data-dependent linear scaling method. Experimental results on the NSL-KDD dataset show that detection rates from the proposed method reach to 82.2% for binary-classification, compared to 81.2% by Artificial Neural Networks (ANN) based works. For multi-classification, the proposed method shows similar performances of ANN. For the detection of denial of services (DoS), which is the most common detecting attack type, the proposed method archives recall of 86.5% and keeps low false positive rate (FPR) to 1.2% by arctan scaling, compared to the recall of 77.7% and FPR of 1.3% by ANN based works.

TABLE IX
CLASSIFICATION COMPARISON

Method	Binary-classification		Multi-classification	
	Recall	FPR	Recall	FPR
Proposed Method	**0.821**	**0.143**	0.796	0.143
ANN [8]	0.812	0.150	0.799	NA
SOM [7]	0.755	0.058	NA	NA

TABLE X
DOS DETECTION COMPARISON

Method	Recall	FPR
Proposed Arctan	**0.865**	**0.012**
Proposed Logistic	0.882	0.027
ANN [8]	0.777	0.013

REFERENCES

[1] Phurivit Sangkatsanee, Naruemon Wattanapongsakorn, and Chalermpol Charnsripinyo. Practical real-time intrusion detection using machine learning approaches. *Computer Communications*, 34(18):2227–2235, 2011.

[2] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani. A detailed analysis of the kdd cup 99 data set. In *IEEE International Conference on Computational Intelligence for Security and Defense Applications*, pages 53–58, 2009.

[3] Kristopher Kendall. A database of computer attacks for the evaluation of intrusion detection systems. In *Darpa Off-line Intrusion Detection Evaluation, Darpa Information Survivability Conference & Exposition*, pages 12–26, 1999.

[4] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31, 2016.

[5] Yogita B. Bhavsar and Kalyani C. Waghmare. Intrusion detection system using data mining technique: Support vector machine. In *International Journal of Emerging Technology and Advanced Engineering*, volume 3, 2013.

[6] Karan Bajaj and Amit Arora. Improving the intrusion detection using discriminative machine learning approach and improve the time complexity by data mining feature selection methods. *International Journal of Computer Applications*, 76(1):5–11, 2013.

[7] D. T. Basheer L. M. Ibrahim and M. S. Mahamod. A comparison study for intrusion database (KDD99, NSL-KDD) based on self organization map (SOM) artificial neural network. *Journal of Engineering Science & Technology*, 8(1):107–119, 2013.

[8] B. Ingre and A. Yadav. Performance analysis of NSL-KDD dataset using ANN. In *2015 International Conference on Signal Processing and Communication Engineering Systems*, pages 92–96, Jan 2015.

[9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, September 1995.

[10] John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. China Machine Press,, 2005.

[11] David M W Powers. Evaluation: From precision, recall and F-Factor to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2:2229–3981, 2011.

[12] B. W. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta*, 405(2):442–451, 1975.

[13] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.