A Survey on SVM Classifiers for Intrusion Detection

R.Ravinder Reddy
Asst.professor
CSED-CBIT
Hyderabad, India

B.Kavya
Student, M Tech
CSED-CBIT
Hyderabad, India

Y Ramadevi, Ph.D
Professor & Head
CSE-CBIT
Hyderabad, India

ABSTRACT
Intrusion detection is an emerging area of research in the computer security and networks with the growing usage of internet in everyday life. An Intrusion Detection is an important in assuring security of network and its different resourses. Intrusion detection attempts to detect computer attacks by examining various data records observed in processes on the network. Recently data mining methods have gained importance in addressing network security issues, including network intrusion detection. Intrusion detection systems aim to identify attacks with a high detection rate and a low false positive. Here, we are going to propose Intrusion Detection System using data mining technique: Support Vector Machine (SVM). Support vector machine-based intrusion detection methods are increasingly being researched because it can detect novel attacks. But solving a support vector machine problem is a typical quadratic optimization problem, which is influenced by the feature dimensions and number of training samples. In this paper how the support vector machines are used for intrusion detection are described and finally proposed a solution to the inrusion detection system.

General Terms
Support vector machines, Intrusion Detection, Classification.

Keywords
Kernel functions, soft margins, Multi class SVM

1. INTRODUCTION
With the development of the internet and its wide applications in all domains of everybody's life, intrusion detection is becoming a critical process in computer network security. Intrusion detection systems (IDS) attempts to recognize and notify the user's activity as either normal or anomaly (or intrusion) by comparing the network connection records to the known intrusion patterns obtained from the human experts. Support vector machine is a machine learning method that is widely used for data analyzing and pattern recognizing. The algorithm was invented by Vladimir Vapnik (1995) [7] and the current standard incarnation were proposed by Corinna Cortes and Vladimir Vapnik. Support vector machine is a new kind of machine learning algorithm proposed recently which is based on structural risk minimization of statistical learning theory. Many researchers verified that SVM performed well in intrusion detection classification [9]. SVM has been widely used for intrusion detection as a classical pattern recognition tool. Network Intrusion Detection using SVM is better than artificial neural network [8][12].
Support vector machine (SVM) is used for classification in IDS due to its good generalization ability and non linear classification using different kernel functions and performs well as compared to other classifiers. However, when applying SVM on high dimension and large-scale dataset, such as network connection dataset, it often suffers memory storage and time consuming problem because a SVM solver will solve a dual quadratic optimization problem [11].
There are three phases in the construction of the SVM intrusion detection systems. The first phase is the pre processing phase, which processes the randomly selected raw TCP/IP dump data using automated parsers and converts it into machine readable form. The second phase is the training phase in which the SVMs are trained on different types of attacks and normal data. The data has a total of 41 input features and can be classified into two categories normal (+1) and attack (-1). The SVM will be trained with both the type of data: normal as well as intrusive data. The final phase is the testing phase. This training phase involves measuring the performance of the data being tested.

2. PRELIMANARIES
2.1 Intrusion Detection System
The popularity of using Internet contains some risks of network attacks. Intrusion detection is one major research problem in network security, whose aim is to identify unusual access or attacks to secure internal networks. The network intrusion detection techniques are important to prevent our systems and networks from malicious behaviours [1]. However, the traditional network intrusion prevention systems such as firewalls, user authentication and data encryption have failed to completely protect networks and systems from the increasing and sophisticated attacks and malwares. An intrusion can be defined [2] as "the act of causing obtrusion or an inappropriate situation". Intrusions are caused by: Attackers accessing the systems, Authorized users of the systems who attempt to gain additional privileges for which they are not authorized, Authorized users who misuse the privileges given to them. Intrusion detection is "the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions. It is also defined as attempts to compromise the confidentiality, integrity, availability, or to bypass the security mechanisms of a computer or network".

Intrusion detection approaches are commonly divided into two categories: signature or misuse detection and anomaly detection [3][4]. Misuse detection is the ability to identify intrusions based on a known pattern for the malicious activity .These known patterns are referred to as signatures. The idea of misuse detection is to establish a pattern or a signature form so that the same attack can be detected. Thus, the main drawback of misuse detection is it cannot detect new types of attacks. The IDS has a pattern database that includes signatures of possible attacks. If the system matches the data with the attack pattern, the IDS regards it as an attack. Consequently, misuse detection provides a low false positive rate (the rate of misclassified normal behaviour).

Anomaly detection is the attempt to identify malicious traffic based on deviations from established normal network traffic patterns. Anomaly detection requires storage of normal usage behaviour and operates upon audit data generated by the operating system. With the anomaly detection approach, one represents patterns of normal behaviour, with the assumption that an intrusion can be identified based on some deviation from this normal behaviour. When such a deviation is observed, an intrusion alarm is produced. Anomaly detection is capable of catching new attacks, yet it suffers a higher false positive rate.

Intrusion detection is categorized into two, host based and network based approaches. Network intrusion detection system (NIDS) [6] detects intrusions by continuously monitoring network traffic by connecting to network hub or switch which is configured for port mirroring, or network tap. NIDS uses sensors to capture all network traffic and to monitor individual packets to identify whether it is normal or attack. Network based IDS is installed on network elements like routers to monitor the network traffic. In network-based (NIDS), the packets are collected from the network. An example of a NIDS is Snort.

A host-based intrusion detection system (HIDS) [5] is an intrusion detection system that monitors and analyzes the internals of a computing system as well as the network packets on its network interfaces. Host based IDS is installed on each system for monitoring of malicious activities locally .Host-based intrusion detection system (HIDS) uses agent as a sensor on a host that identifies intrusions by analyzing system calls, application logs, file-system modifications (binaries, password files, etc.) and other host activities and state. OSSEC is an example for Host based intrusion detection system.

3. PROCESS FLOW

A support vector machine classifier for intrusion detection can be constructed on the following steps:

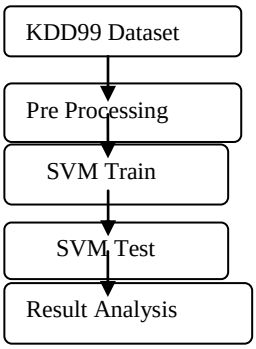

Fig 1: The flow of SVM

First, data pre-processing is utilized to data arrangement. After that training data set is sent to the SVM Train to build the model file. Finally, the system uses the SVM Test to predict the accuracy and False alarm rate.

3.1 Dataset Analysis

Under the sponsorship of Defence Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL), MIT Lincoln laboratory has collected and distributed the datasets for the evaluation of researches in computer network intrusion detection systems. The KDDCup99 dataset is subsets of the DARPA benchmark dataset. Each KDDCup'99 training connection record contains 41 features and is labelled as either normal or an attack, with exactly one specific attack type. The various attack types in the datasets are grouped into attack categories in order to combine similar attack types into a single category which could improve the detection rate. The dataset contains 24 attack types that could be classified into four major categories, namely, Denial of Service (Dos), Probing (Probe), Remote to Local (R2L) and User to Root (U2R).

4. SUPPORT VECTOR MACHINE

In this paper the different approaches for the SVM are discussed for intrusion detection. Based on the structural risk minimization principle, SVM is a new machine learning method presented by Vapnik (1995) [13]. Generalization ability of SVM is obviously superior to other traditional learning methods. This basic SVM deals with two-class problems, known as Binary classification problems in which the data are separated by a hyper plane defined by a number of support vectors. Support vectors are a subset of training data used to define the boundary between the two classes. Each instance in the training set contains one "target value" (class labels: Normal or Attack) and 41 features. The goal of SVM is to produce a model which predicts target value of data instance in the testing set which consists of only features. To achieve this goal, we have used Radial Basis Function (RBF) kernel functions [29, 31] available with SVM.

In situations where SVM cannot separate two classes, it solves this problem by mapping input data into high-dimensional feature spaces using a kernel function [14, 33]. In high-dimensional space it is possible to create a hyper plane that allows linear separation (which corresponds to a curved surface in the lower- dimensional input space). Accordingly, the kernel function plays an important role in SVM. The kernel functions can be used at the time of training of the classifiers which selects support vectors along the surface of this function. SVM classify data by using these support vectors that outline the hyper plane in the feature space. In practice, various kernel functions can be used, such as linear, polynomial or Gaussian.

The SVM is already known as the best learning algorithm for binary classification [11][15][16]. However, it is not the reason that we have chosen SVM. The most significant reason we chose the SVM is because it can be used for either supervised or unsupervised learning. The SVM, originally a type of pattern classifier based on a statistical learning technique for classification and regression with a variety of kernel functions [7, 19], has been successfully applied to a number of pattern recognition applications [15]. Recently, it has also been applied to inform security for intrusion detection [17][8]. Another positive aspect of SVM is that it is useful for finding a global minimum of the actual risk using structural risk minimization, since it can generalize well with kernel tricks even in high-dimensional spaces under little training sample conditions. The SVM can select appropriate setup parameters because it does not depend on traditional empirical risk such as neural networks. In the case of supervised SVM learning, it has relatively fast processing and high detection performance when compared to existing artificial neural networks and the unsupervised SVM , as shown in[24][25]. However, one of the main disadvantages of the supervised method is that it requires labelled information for efficient learning. Moreover, it cannot deal with the

relationship between consecutive variations of learning inputs without additional pre-processing. Therefore, Taeshik Shon and Jongsub Moon have proposed the real time intrusion detection system using Enhanced SVM, which combines soft margin SVM using supervised learning and one-class SVM approach using the unsupervised learning. The enhanced SVM approach inherits the advantages of both SVM approaches, namely high performance and unlabelled capability.

The SVM is generally used as a supervised learning method. Vapnik proposed the initial idea of SVM for the separable case (hard margin SVM) in which the positive and negative samples can be definitely separated by a unique optimal hyper plane with the largest margin. However, this algorithm will find no feasible solution when applied to the non-separable case. Cortes and Vapnik extended this idea to the non-separable case (soft margin SVM or the so called standard SVM) by introducing positive slack variables $\{\xi\}$ i=1,…,l .In order to decrease misclassified data, a supervised SVM approach with a slack variable is called soft margin SVM. Additionally, single class learning for classifying outliers can be used as an unsupervised SVM. After considering both SVM learning schemes, an Enhanced SVM approach is proposed.

4.1 Soft Margin SVM: Supervised SVM

We begin by discussing a soft margin SVM learning algorithm written by Cortes [7], which is sometimes called c-SVM. This SVM classifier has a slack variable and penalty function for solving non-separable problems. First, given a set of points $x_i \in R^d$, i=1,……,l and each point x_i belongs to either of two classes with the label $y_i \in \{+1,-1\}$.These two classes can be applied to anomaly attack detection with the positive class representing normal and negative class representing abnormal. Suppose there exists a hyper-plane $w^T x_i + b = 0$ that separates the positive examples from the negative examples. That is, all the training examples satisfy:

$w^T x_i + b \geq +1$ for all $x_i \in P$

$$w^T x_i + b \leq -1 \quad \text{for all} \quad x_i \in N \tag{1}$$

w^T is an adjustable weight vector, x_i is the input vector and b is the bias term.

Equivalently:

$$y_i.(w^T.x_i - b) \geq 1 \ \ \forall i, = 1 \ldots . N \tag{2}$$

In this case, we say the set is linearly separable.

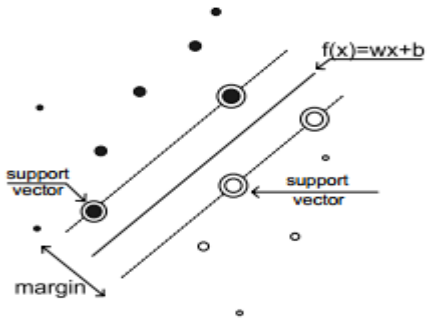

Fig 2: Separable hyper-plane between two datasets

In Fig. 2, the distance between the hyper-plane and f(x) is $\frac{1}{\|w\|}$ .The margin of the separating hyper-plane is defined to be $\frac{2}{\|w\|}$. Hence, the learning problem is reformulated as minimizing $\|w\|^2 = w^T w$ is subject to the constraints of linear separation shown in Eq.(3). This is equivalent to maximizing the hyper-plane distance between the two classes, of which the maximum distance is called the support vector. The optimization is now a convex quadratic programming problem.

$$Minimize_{w,b} \quad \phi(w) = \frac{1}{2}\|w\|^2 \tag{3}$$

Subject to $y_i.(w^T.x_i - b) \geq 1 \ \ \forall i, = 1 \ldots . N$ is convex in w and the constraints are linear in w and b, we can be sure that this problem has a global optimum solution. This has the advantage that parameters in quadratic programming (QP) affect only the training time, and not the quality of the solution. This problem is tractable, but anomalies in internet traffic show characteristics of non-linearity, and as a result they are more difficult to classify. In order to proceed to such non-separable and non-linear cases, it is useful to consider the dual problem as outlined below.

The Lagrange for this problem is

$$L(w, b, \lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{p} \lambda_i (y_i (w^T.x_i + b) - 1) \tag{4}$$

Where $\lambda = (\lambda_{1,\ldots\ldots,}\lambda_l)^T$ are the Lagrange multipliers, one for each data point. The solution to this quadratic programming problem is given by maximizing L with respect to $\lambda \geq 0$ and minimizing with $\frac{1}{2}\|w\|^2$ respect to w and b. Note that the Lagrange multipliers are only non-zero when $y_i(w^T.x_i + b) = 1$, and the vectors for this case are called support vectors, since they lie closest to the separating hyper-plane. However, in the non-sep-arable case, forcing zero training error leads to poor generalization. We introduce the soft margin SVM using a vector of slack variables $\xi = (\xi_{1\ldots\ldots}\xi_I)^T$ that measure the amount of violation of the constraints (5), taking into account the fact that some data points may be misclassified.

The equation is now

$$Minimize_{w,b,\xi} \ \ \phi(w, b, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i^k \tag{5}$$

subject to $y_i(w^T.\phi(x)_i + b) \geq 1 - \xi_i \geq 0,$

where C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the training error. The effects of C are crucial. If C is too small, insufficient stress is placed on fitting the training data. If C is too large, the algorithm will overfit the dataset.

In practice, a typical SVM approach, such as the soft margin SVM, showed excellent performance more often than other machine learning methods[25,19].For intrusion detection applications, supervised machine learning approaches based on SVM methods proved superior to intrusion detection approaches using artificial neural networks[8,25]. Therefore, the high classification capability and processing performance of the soft margin SVM approach is useful for anomaly detection. However, it is not appropriate to use the soft-margin SVM method for detecting novel attacks in Internet traffic since it requires pre acquired learning information for supervised learning procedure. Such pre-acquired learning information is divided into normal and attack traffic with labels separately.

4.2 One-class SVM: Unsupervised SVM

SVM algorithms can also be adapted into an unsupervised learning algorithm called one-class SVM, which identifies outliers among positive examples and uses them as negative examples [17].In anomaly detection, if we consider anomalies as outliers, then the one-class SVM approach can be applied to classify anomalous packets as outliers. This SVM method does not require pre-existing knowledge for classification.

The main idea is that the algorithm maps the data into a feature space H using an appropriate kernel function, and then attempts to find the hyper-plane that separates the mapped vectors from the origin with maximum margin. Given a training dataset $(x_1, y_1),\ldots,(x_l, y_l) \in R^N \times \{+1,-1\}$,let $\emptyset: R^N \rightarrow H$ be a kernel map that trans-forms the training examples into the feature space H. Then, to separate the dataset from the origin, we need to solve the following quadratic programming problem (6).

$$Minimize_{w,b,\xi}\ \phi(w,b,\xi) = \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i^k - \rho \quad (6)$$
$$\text{Subject to } y_i.(w^T.\varphi(x_i)) \geq \rho - \xi_i\ , \xi_i \geq 0\ \forall i, = 1 \ldots . N$$

Where v is a parameter that controls the trade-off between maximizing the distance from the origin and containing most of the data in the region related to the hyper-plane, and corresponding to the ratio of outliers in the training set. Then the decision function f(x) = sgn ((w.ϕ(x)+b)− ρ) will be positive for most examples xi contained in the training set.

In practice, even though the one-class SVM has the capability of outlier detection, the approach is more sensitive to a given dataset than other machine learning schemes [17,24].This means that it is very important to decide on an appropriate hyper-plane for classifying outliers.

4.3 Enhanced SVM: supervised and unsupervised SVM

In the problem domain, most of the Internet traffic exhibits normal features, and the amount of anomalous traffic is relatively small. To put it simple, we can see that the number of outliers we hope to detect is relatively small in comparison with normal traffic. Therefore, from this point of view, the one-class SVM need not always make the single classifier to maximize the distance between the origin and outliers. An enhanced SVM method with the capability of one-class SVM through modifying soft margin SVM. The enhanced SVM approach can be run on the soft margin SVM, and enhanced SVM approach will not only have higher detection rate and faster processing performance than the soft-margin SVM but also possess the unsupervised learning feature of one-class SVM.

In one-class SVM learning (6), the important parameters for deciding the hyper-plane are$\|w\|, \rho, \frac{1}{vl}\sum_{i=1}^{l}\xi_i^k$. Each parameter has the following meaning: $\|w\|$ has to be decerased, ρ has to be increased in order to obtain maximum margin between the origin and hyper-plane. In the case of $\frac{1}{vl}\sum_{i=1}^{l}\xi_i^k$,this parameter is related to the violation of outliers and has to be decreased. Moreover, in soft margin SVM learning (5), the bias term is generally related to the distance between the origin and hyper-plane. As the bias term is decreases, the hyper-plane approaches the origin similar to the behaviour of the one-class SVM classifier. The bias term b of (5) is deleted, and then we derived Eq(7) as follows:

$$Minimize_{w,b,\xi}\ \ \phi(w,b,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k \quad (7)$$
$$\text{Subject to } y_i.(w^T.\varphi(x_i)) \geq 1 - \xi_i\ , \xi_i \geq 0\ \forall i, = 1 \ldots . l$$

Therefore, by considering these parameters of one-class SVM and soft margin SVM, we derive an Enhanced SVM that provides the unlabeled classification capability of one-class SVM as well as the high detection performance of supervised soft margin SVM. If we compare (7) with (6),

Soft margin SVM without a bias ≃one-class SVM

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k \cong \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i^k - \rho \quad (8)$$

$$\text{Subject to } y_i.(w^T.\varphi(x_i)) \geq 1 - \xi_i \cong y_i.(w^T.\varphi(x_i)) \geq \rho - \xi_i, 0 < V < 1\ ,1< l, 0 \leq \rho \quad (9)$$

In the minimization condition (8), $C\sum_{i=1}^{l}\xi_i^k$ of soft margin SVM is the trade-off value to adjust the training error to obtain maximum margin between the origin and the hyper-plane. $\frac{1}{vl}\sum_{i=1}^{l}\xi_i^k$ of one-class SVM is also related to the extent of violation of outliers. Thus, we can approximate each term as a similar value by controlling C of soft margin SVM. In the subject condition (9) , the ρ of the one-class SVM is a parameter for maximizing the margin between the origin and the hyper-plane. However, we do not need to consider maximizing the ρ value, because anomalous data of the domain can be classified by the hyper-plane near the origin. In the end, we can regard the value of ρ as a very small number like '1'. We finally derive an Enhanced SVM, by combining all these parameter approximations in Eq. (10).

$$\text{Minimize}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k - \rho, \quad C \cong \frac{1}{VL}, \quad (10)$$
$$\text{subject to } y_i.(w^T.\varphi(x_i)) \geq \rho - \xi_i,\ 0 \leq \rho \cong 1,$$

The minimization condition and subject condition of (10) have to satisfy the approximated conditions of $C \cong \frac{1}{VL}$ and $\rho \cong 1$, respectively. Therefore, in the proposed SVM approach (10), we can expect the unlabeled learning feature of one-class SVM, the relatively low false alarm and high detection rate of soft margin SVM.

Optimization classification function is:

$$f(x) = sgn\{w.x + b\} = sgn\left\{\sum_{i=1}^{l}\lambda_i y_i x_i . x + b\right\} \quad (11)$$

Nonlinear problems can be converted into linear problems in higher dimensions space through nonlinear transformation and optimal hyper plane can be got in higher dimensions space. However, this transformation may be a little complicated, and it is usually not easy to attain. According to some theory in functional analysis, as long as one kernel function k(x_i , x_j) meets the conditions of Mercer, it will correspond to inner product in a certain transformation space. Therefore, after nonlinear transformation, the linear classification can be realized, and operation and the complexity of algorithm do not increase.

4.4 Kernel Functions

In Support vector machines kernel functions are crucial in deciding the performance of the classifier [32, 33]. There are several kernel functions are available, from those the popular are consider here.

Consider the kernel function $K(x_i, x_j) = \left(\phi(x_i), \phi(x_j)\right)$, and then classification function is

$$f(x) = sgn\{\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b\} \quad (12)$$

The following kernel is radial basic function (RBF), which will be used in the simulation experiment.

$$\text{RBF: } K(x,x_i) = \exp(-\gamma|x - x_i|^2), \gamma > 0$$

Here, γ is kernel parameter.

Table1. Some Commonly Used Kernel Functions

Gaussian Radial Basis Function(RBF)	$K(x,y)=e^{(-(x-y)^2/2\sigma^2)}$
Exponential Radial Basis Function	$K(x,y)=e^{(-(x-y)/2\sigma^2)}$
Hyperbolic Tangent (sigmoid)	K(x,y) = tanh(b(x,y)+c)
Polynomial	$K(x,y) = (1 + x^T.x)^p$
Fourier Series	$K(x,y) = \frac{\sin(\delta+\frac{1}{2})(x-y)}{\sin(\frac{1}{2}(x-y))}$
Two-layer perception	$\text{Tanh}(s_0 x^T.x_i + s_1)$

4.5 Multiclass SVMs

Support Vector Machines are binary classifiers in their basic form. Their theoretical advantages and their practical success motivated researchers to investigate extensions to multiclass problems. Tarun Ambwani reports the multi-class SVM [20, 21, 22, 31] to intrusion Detection. With the term multiclass, we refer to problems in which any instance is assigned exactly one class label. Such problems are called mutually exclusive multiclass problems. They can be divided into two categories, those that consider the whole dataset with all the classes at once and solve the multiclass problem directly and those that decompose the problem into constructing several binary classifiers and combining their output. Examples of direct approaches are those presented by and Crammer and Singer (2000), such methods however present numerical difficulties due to the large number of variables that need to be optimized and are rather difficult to implement. For these reasons, so far, the approaches that decompose the problem into binary classification have been more popular and widely used. The standard method is the one-against-all method. For each class a binary SVM classifier is constructed, discriminating the data points of that class against the rest. Each classifier yields a decision value for the test data point and the classifier with the highest positive decision value assigns its label to the data point.

Multi-SVM classifiers are applied to intrusion detection because of multi-types existing in network. 'One-against-one', 'One-against-all' and 'Binary tree' are the popular methods in SVM multi-class classification [26,27]. 'Binary tree' SVM classification algorithm needs only $k - 1$ two-class SVM classifiers for a case of k classes, while 'One-against-all' SVM classification algorithm needs k two-class SVM classifiers where each one is trained with all the samples and 'One-against-one' SVM classification algorithm needs $k(k - 1)/2$ two class SVM classifiers where each one is trained on data from two classes . Obviously less two-class classifiers help to expedite the rate of training and recognition. Thus, 'Binary tree' SVM classification algorithm is adopted to construct detection model for intrusion detection.

4.6 Incremental SVM

An Incremental SVM Support Vector Machines (SVMs) can be used to learn with large amounts of high dimensional data. Syed et al first proposed the incremental learning algorithm of SVM.However, computing a SVM is very costly in terms of time and memory consumption. It is a good idea to learn incrementally from previous SVM results. Compared to the number of training examples, in most cases the number of Support Vectors is very small. SVMs can compress the data of the previous batches to their Support Vectors in incremental learning. This incremental learning approach with SVMs has been investigated. Compare to non incrementally trained SVMs, incrementally trained SVMs behave well. The data is provided in several batches. For each new batch of data a SVM is trained on the new data and the Support Vectors from the previous learning step. After each training step, a preliminary result will be produced by the algorithm. So time and memory consumption are controlled. Incremental techniques have been found widespread used in SVM. Incremental SVM learning is particularly attractive in an online system [28], and for active learning. In an online system, the data is often collected continuously in time. Significant effort has been spent in the recent years on development of online SVM learning algorithms .The elegant solution to online SVM learning is the incremental SVM which provides a framework for exact online learning.

Research on the incremental SVM learning algorithm in the intrusion detection system is of significance. Firstly, it is difficult for traditional security and defense strategies to meet the ever changing needs of the network security and shot of static protection technologies. Secondly current intrusion detection methods are mostly non-incremental learning algorithm. As the accumulation of the new incremental samples, the training time expenses will continue to increase.Thirldly incremental learning can rapidly learn from the new incremental samples to modify the existing model. Time consumption is relatively small. Finally compared with non-incremental learning algorithms, incremental learning algorithms are in a relatively small number of studies, especially the incremental SVM algorithm. In order to reduce the number of samples and shorten the training time, the algorithm introduces the sample selection process, and applies it to network intrusion detection.

It divided the training set into N subsets, after training on a subset; the algorithm only retained the support vectors and discarded the other samples, and then added them into next subset to form a new training subset and trained on the new training subset.

5. SUPPORT VECTOR MACHINE WITH ROUGH SET

Rough set theory is used for the future reduction**.** The 41 features of kddcup99 dataset are reduced to 15 features by using the RST. So that we get the good accuracy and less false alarm rate. Using the reduct computed the new dataset and tested the dataset on the SVM and rough set based SVM. Here it shows the good result for intrusion detection.

Table 2.Comparison of Accuracy and False alarm rate values for SVM and Rough set with SVM.

	Accuracy	False alarm rate
SVM	97.734	2.318
Rough set with SVM	97.740	2.311

6. CONCLUSION

The major reasons for using SVMs in intrusion detection system is speed: as real-time performance is of primary importance to IDSs, any classifier that can potentially run "fast" is worth considering and the another reason is scalability, SVMs are relatively insensitive to the number of data points and the classification complexity does not depend on the dimensionality of the feature space, so they can potentially learn a larger set of patterns and thus be able to scale better than neural networks. Also SVM provides a standard mechanism to fit the surface of the hyper plane to the data by utilizing the kernel function. Finally, SVMs give highly accurate classification of the patterns. Feature selection or attribution reduction can help reduce the SVM classification time and saving memory space effectively.In future genetic algorithm and rough set theory combinely apply to SVM for enhancing the performance and accuracy of real time intrusion detection.

7. REFERENCES

[1] Intrusion Detection – Wikipedia, the free encyclopedia. Available at http://en.wikipedia.org/wiki/Intrusion detection

[2] Axelsson, S.: Research in intrusion detection systems: a survey. Technical Report TR 98-17 (revised in 1999). Chalmers University of Technology, Goteborg, Sweden(1999)

[3] Lee W and Stolfo S., "Data Mining techniques for intrusion detection", In: Proc. of the 7th USENIX security symposium, San Antonio, TX, 1998

[4] Dokas P, Ertoz L, Kumar V, Lazarevie A, Srivastava J, and Tan P., "Data Mining for intrusion detection", In: Proc. of NSF workshop on next generation data mining, 2002

[5] de Boer P., Pels M. "Host-Based Intrusion Detection Systems".Availablehttp://staff.science.uva.nl/~delaat/snb-2004-2005/p19/report.pdf

[6] Scarfone K., Mell P. "Guide to Intrusion Detection and Prevention Systems". Available at http://csrc.nist.gov/publications/nistpubs/80094/SP80094.pdf, 2007.

[7] C. Cortes and V. Vapnik, "Support-vector network," Machine Learning, vol. 20, pp. 273–297, 1995

[8] S. Mukkamala, G.1. Janoski, A.H. Sung. Intrusion Detection Using Neural Networks and Support Vector Machines. In Proceedings of IEEE International Joint Conference on Neural Networks, Vol 2, Honolulu, 2002.5, pp. 1702-1707.

[9] Dong Seong Kim, Ha-Nam Nguyen,Jong Sou Park Genetic algorithm to improve SVM based network intrusion detection system. In 19th International Conference on Advanced Information Networking and Applications, Vol.2, Taiwan, 2005.3, pp.155–158.

[10] Hansung Lee, Jiyoung Song, Daihee Park. Intrusion Detection System Based on Multi-class SVM. Lecture Notes in Computer Science, vol.3642, Springer Berlin, 2005.9, pp.511-519.

[11] V. N. Vapnik. The nature of statistical learning theory. Springer Verlag, New York. NY, 1995

[12] CannadyJ.,"Artificial Neural Networks for Misuse Detection. National Information Systems Security Conference", (1998).

[13] C.J.C. Burges, A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, vol 2(2), Springer US, 1998, pp.121-167.

[14] K.-P. Lin and M.-S. Chen, "Efficient kernel approximation for large-scale support vector machine classification," in Proceedings of the Eleventh SIAM International Conference on Data Mining, 2011, pp. 211–222

[15] H. Byun, S.W. Lee, A survey on pattern recognition applications of support vector machines, International Journal of Pattern Recognition and Artificial Intelligence 17 (2003) 459–486

[16] Amit Konar, Uday K. Chakraborty, Paul P. Wang, Supervised learning on a fuzzy Petri net, Information Sciences 172 (2005) 397–416

[17] B. Scho ¨lkopf, Estimating the support of a high-dimensional distribution, Neural Computation 13 (2001) 1443–147

[18] K.A. Heller, K.M. Svore, A. Keromytis, S.J. Stolfo, One class support vector machines for detecting anomalous windows registry accesses, in: Proc. The workshop on Data Mining for Computer Security, Melbourne, FL, 2003, pp. 281–289

[19] T. Joachims, Estimating the Generalization Performance of an SVM efficiently, in: Proc. the Seventeenth International Conference on Machine Learning, San Francisco, CA, 2000, pp. 431–438

[20] Hsu, C., Lin, C., "A comparison on methods for multi-class support vector machines", Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, (2001)

[21] Weston, J. and Watkins, C. Support vector machines for multi-class pattern recognition. Proceedings 7th European Symposium on Artificial Neural Networks, 1999.

[22] Xu, P. and Chan, A. An efficient algorithm on multi-class support vector machine model selection. Proceedings of the International Joint Conference on Neural Networks, 4:3229–3232, 2003

[23] KDDCUP'99dataset,availableat http://kdd.ics.uci.edu/dataset/kddcup99/kddcup99.htm

[24] B.V. Nguyen, An Application of Support Vector Machines to Anomaly Detection, CS681 (Research in Computer Science – Support Vector Machine) report, 2002

[25] S. Dumais, H. Chen, Hierarchical classification of Web content, in: Proc. The 23rd annual international ACM SIGIR conference on Research and development in information retrieval, Athens, Greece, 2000, pp. 256–263

[26] D. Srivastava, L. Bhambhu, Data classification using support vector machine, J.Theoret. Appl. Inf. Technol. 12 (1) (2010) 1–7

[27] C.W. Hsu, C.C. Chang, C.J. Lin, A Practical Guide to Support Vector Classification[EB/OL], 2010 http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

[28] S.-J. Horng, P. Fan, Y.-P. Chou, Y.-C. Chang, Y. Pan, A feasible intrusion detector for recognizing IIS attacks based on neural networks. Computers & Security, 2008. 27(3-4): 84-10

[29] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," Neural Computation, vol. 15,no. 7, pp. 1667–1689, 2003.

[30] E. M. Gertz and J. D. Griffin, "Support vector machine classifiers for large data sets," Argonne National Laboratory, Tech. Rep.ANL/MCS-TM-289, 2005.

[31] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research, 2:265–292, 2001.

[32] Joachims, T., Making Large-Scale SVM Learning Practical.Advances in Kernel Methods – Support Vector Learning, 1999.

[33] N. Cristiani and J. Shawe-Taylor. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge, Cambridge University Press, 2000.

[34] S. Abe. Support Vector Machines for pattern classification. London, Springer, 2005.