AN EFFICIENT INTRUSION DETECTION USING RELEVANCE VECTOR MACHINE

V. Jaiganesh\(^1\), Dr. P. Sumathi\(^2\)

\(^1\)Doctoral Research Scholar, Department of Computer Science, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India.

\(^2\)Doctoral Research Supervisor, Assistant Professor, PG & Research Department of Computer Science, Government Arts College, Coimbatore, Tamilnadu, India.

ABSTRACT

Internet becomes a globally used public network. Internet causes tremendous growth in the business to reach the end users. On the other hand, the usage of networks has paved the way for intruders to attack the communication path and to steal the valuable asset (data) of any organization. Hence in order to protect the organization data, Intrusion Detection System (IDS) offers protection from external users and internal attackers. Intrusion detection is the process of examining the events which happens in a computer system or network and evaluates them for signs of possible events, which are imminent threats of violation of computer security policies, standard security practices and acceptable use policies. In the proposed method, an effective intrusion system can be applied using unity-based normalization to standardize data and Relevance Vector Machine (RVM) for classification. The experiment is carried out with the help of WEKA by using KDD Cup 1999 dataset and the results indicate that the proposed technique can achieve higher detection rate and very low false alarm rate than the regular SVM algorithms.

Keywords: Cascade forwards back propagation, Intrusion Detection System (IDS), Relevance Vector Machine (RVM),

I. INTRODUCTION

Internet created several ways to negotiate the stability and security of the systems connected to it. Even though static defense mechanisms such as firewalls and software updates can afford a reasonable level of security, new dynamic mechanisms should also be employed. Examples of such dynamic mechanisms are intrusion detection systems and
network analyzers. Intrusion detection aims to achieve the specific goal of detecting attacks whereas network analysis determines the changing trends in computer networks and connected systems. [1] Hence network analysis is a generic tool that helps system administrators to discover what happens on their networks.

Intrusion Detection System is software or hardware systems that automate the process of monitoring and inspecting the events that takes place in a computer network to reveal malicious activity. To provide a security infrastructure for most organizations due to the drastic increase in the severity of attacks occurring in the network, intrusion detection plays an additional necessary role. Intrusion detection permits organization to guard their systems from the threats that come with increasing network connectivity and trust on information systems. [2] Intrusion detection attacks are segmented into two groups,

- Host-based attacks [3-5] and
- Network-based attacks [6, 7].

In case of host-based attacks, the intruders aim at a particular machine and attempt to get access to privileged services or resources on that specific machine. Recognition of these kind of attacks typically uses routines that acquire system call data from an audit-process which monitors all system calls made with the support of each user. It is extremely complicated for legitimate users to use various network services by purposely occupying or disrupting network resources and services in case of network-based attacks. Intruders attack these system by transmitting huge amounts of network traffic, consuming familiar faults in overloading network hosts and networking services, etc. Recognition of these kind of attacks uses network traffic data (i.e., tcpdump) to look at traffic addressed to the machines being monitored.

Several intrusion detection systems are available and they do not meet the challenges of a susceptible internet atmosphere [8, 9]. In the current scenario, an IDS is much essential for a modern computer system. IDS can be categorized into two major groups:

- Misuse detection and
- Anomaly detection.

A misuse detection system traces intrusion activities that follow recognized patterns. These patterns explain a suspect collection of sequences of activities or operations that can possibly be dangerous. The major drawback of this detection is that it doesn’t have the capability to trace or detect new kind of intrusions (certain events that have never occurred in the past). Abnormality detection system examines event data and identifies pattern of activities that appear to be ordinary. An event which lies outside of the patterns is regarded as a possible intrusion [10].

The Relevance Vector Machine (RVM) is a Bayesian learning model for regression and classification of identical functional form to the Support Vector Machine (SVM). RVM can be generalized well and provide inferences at low computational cost. The proposed method employs RVM classification.

The paper can be arranged as follows : Section II provides the related works involved in intrusion systems and the techniques used in it. Section III reveals the proposed methodology and section IV gives the experimental results of the proposed work.
II. RELATED WORKS

Security is considered as a major issue in networks since the network has been dramatically extended. Internet attacks are increasing nowadays. Intrusion detection systems have been used along with the data mining techniques to detect intrusions. Ektefa et al., [11] aimed to use data mining techniques including classification tree and support vector machines for intrusion detection. The result of this approach indicates that the C4.5 algorithm is better than SVM in detecting network intrusions and false alarm rate in KDD CUP 99 dataset.

The victory of any Intrusion Detection System (IDS) is a major problem due to its nonlinearity and the quantitative or qualitative network traffic data stream with irrelevant and redundant features. Selecting the effective and key features to IDS is a major topic in information security. SVM has been employed to provide potential solutions for the IDS problem. Though, the practicability of SVM is affected due to the difficulty in selecting appropriate SVM parameters. Particle swarm optimization (PSO) is an optimization method which has strong global search capability and it is easy to implement. Wang et al., [12] proposed PSO–SVM model which is applied to an intrusion detection problem using the KDD Cup 99 data set. The typical PSO is used to find free parameters of SVM and the binary PSO is to obtain the optimum feature subset at building intrusion detection system. The observation results reveals that the PSO–SVM method can achieve higher detection rate than regular SVM algorithms in the same time.

III. METHODOLOGY

The proposed methodology used for employing Intrusion detection system is explained in this section. The figure 1 shows the steps involved in the proposed methodology.

![Diagram](image)

**Fig 1.** Steps involved in the proposed method
1. Data collection and Preprocessing

The proposed IDS is experimented using the Waikato Environment for Knowledge Analysis (WEKA) and the dataset used is KDD Cup99 dataset. WEKA, a complete set of Java class libraries that execute several state-of-the-art machine learning and data mining approaches [13]. KDD Cup99 dataset comes from DARPA 98 Intrusion Detection Evaluation handled by Lincoln laboratory at MIT [14].

Both training and testing data are divided into following three protocol types such as TCP, UDP or ICMP in order to train and test the data separately. The number of remaining data which are repeating has been deleted. The number of training data for TCP and UDP will be still large. Therefore some number of data has to be deleted randomly. The data to be deleted were chosen mostly from “normal” labeled data from the dataset. [15] Still there were some attacks remaining in testing data set that were not in the training data set. These can be tested using RVM classification.

2. Normalization

Normalizing data means to make the data value within unity (1), hence all the data values will range from 0 to 1. But some models confuses at the value of zero. It is because an arbitrary range of 0.1 to 0.9 is chosen instead of zero. To overcome this limitation, a unity-based normalization technique is employed in the proposed method. [16] The following equation is used to implement a unity-based normalization:

\[ X_{i-1 to 1} = \frac{X_i - \frac{X_{\text{max}} + X_{\text{min}}}{2}}{\left(\frac{X_{\text{max}} + X_{\text{min}}}{2}\right)} \]  

(1)

Where \( X_i \) indicates each data point I, \( X_{\text{min}} \) represents the minima among all the data points, \( X_{\text{max}} \) represents the maxima among all the data points, \( X_{i-1 to 1} \) represents the data point i normalized between 0 and 1.

3. Relevance Vector Machine

The Relevance Vector Machine (RVM) was introduced by [17] as a Bayesian counterpart to the SVM has made tremendous growth in the Machine Learning community due to its simplicity and applicability. The Relevance Vector Machine (RVM) presents an empirical Bayes treatment of function approximation by kernel basis expansion. RVM attains a sparse representation of the approximating function by structuring a Gaussian prior distribution in a way that implicitly creates a sparsity pressure on the coefficients appearing in the expansion. The use of independent Gamma hyperpriors yields a product of independent marginal prior for the coefficients and hence it achieves the desired sparsity.

In order to reduce the dimensionality of the hyperparameter space, specify a prior structure which reflects the possibility of correlation between the hyperparameters of the coefficients distribution and hence it is possible to segregate a unique solution.

RVM has been used for classification in the proposed method. Relevance vector machine (RVM) is a special case of a sparse linear model in which the basis functions are formed by a kernel function \( \varphi \) centred at the different training points:
This model is similar in form to the support vector machines (SVM), the kernel function in the above equation does not satisfy the Mercer’s condition and it requires $\phi$ to be a continuous symmetric kernel of a positive integral operator.[18]

Multi-kernel RVM is an extension of the RVM model. It consists of different types of kernels $\phi_m$ and it is expressed as:

$$y(x) = \sum_{i=1}^{m} \sum_{i=1}^{N} w_i \phi_m (x - x_i)$$  (3)

The sparseness property enables choosing proper kernel automatically at each location by pruning all irrelevant kernels, hence it is possible that two different kernels remain on the same location.

Assume a two-class problem with training points $X = \{ X_1, \ldots, X_N \}$ and corresponding class labels $t = \{ t_1, \ldots, t_N \}$ with $t_i \in \{0, 1\}$. Applying the Bernoulli distribution, the likelihood (the target conditional distribution) can be expressed as:

$$p( t|w) = \prod_{i=1}^{N} \sigma \{ (y(x_i)) \}^{t_i} [1 - \sigma \{ (y(x_i)) \}]^{1 - t_i}$$  (4)

Where $\sigma(y)$ - logistic sigmoid function

$$\sigma(y(x)) = \frac{1}{1 + \exp (-y(x))}$$  (5)

Consider $\alpha_i^*$ denotes the maximum a posteriori (MAP) estimate of the hyperparameter $\alpha_i$. The MAP approximate for the weights is denoted by wMAP and it can be obtained by maximizing the posterior distribution of the class labels given the input vectors. It is equivalent to maximizing the objective of the function given by:

$$J(w_1, w_2, \ldots, w_N) = \sum_{i=1}^{N} \log p (t_i|w_i) + \sum_{i=1}^{N} \log p (w_i|\alpha_i^*)$$  (6)

where the first term indicates the likelihood of the class labels and the second term indicates prior on the parameters $w_i$. Those samples associated with nonzero coefficients $w_i$ which is called relevance vectors will contribute to the decision function.

The gradient of the actual function J with respect to w is given by:

$$\nabla J = -A^*w - \varphi^T (f - t)$$  (7)

Where $f = [ \sigma(y(x_1)), \ldots, \sigma(y(x_N)) ]^T$, matrix $\varphi$ has elements $\varphi_{i,j} = K(x_i, x_j)$. The Hessian of J is
\[ H = \nabla^2 (J) = -(\varphi^T B \varphi + A^*) \]  
\[ \text{(8)} \]

Where \( B = \text{diag}(\beta_1, \ldots, \beta_N) \) is a diagonal matrix with \( \beta_i = \sigma(y(x_i))[1 - \sigma(y(x_i))] \).

The posterior is approximated around \( w_{\text{MAP}} \) by a Gaussian approximation with covariance

\[ \sum = -(H|w_{\text{MAP}})^{-1} \]  
\[ \text{(9)} \]

and mean is given by,

\[ \mu = \sum \varphi^T B t \]  
\[ \text{(10)} \]

RVM has several advantages which includes the number of relevance vectors can be much smaller than that of support vectors, RVM does not need the tuning of a regularization parameter (\( C \)) as in SVM during the training phase. Thus the proposed dataset can be classified using RVM classifier.

iv. EXPERIMENTAL RESULTS

KDD Cup99 is an audited set of standard dataset which includes training and testing set. Data has the following four major groups of attacks

i. Denial-of-Service (DoS) like apache2, smurf, pod, etc.
ii. Remote-to-Local (R2L) like worm, phf, imap, etc.
iii. User to Root (U2R) like rootkit, perl and so on.
iv. Probing like portsweep, nmap, etc.

Attack detection can be calculated by using the following metrics:

i. False Positive (FP): Matches the number of detected attacks but it is actually normal.
ii. False Negative (FN): Corresponds to the number of detected normal instances but it is really an attack. These attacks are the major target of intrusion detection systems.
iii. True Positive (TP): Corresponds to the number of detected attacks and it is in fact attack.
iv. True Negative (TN): Matches to the number of detected normal instances and it is actually normal.

1. Performance Measures

The performance measure evaluated are used in the proposed KSVM with LM against SVM is

- Detection rate and
- False-alarm rate

The intrusion detection system accuracy is computed based on the detection rate and false alarm rate.
2. Detection Rate Comparison

Detection rate indicates the percentage of detected attack among all the attack data, and it is given as:

\[
\text{Detection Rate} = \frac{TP}{TP + TN} \times 100
\]  \hspace{1cm} (11)

Fig 2. Comparison of Detection Rate on Four Attacks

The results of detection rate for different types of attacks is shown in fig 2. From the results it is observed that in case of DoS attacks, detection rate for RVM obtains better results in all other attacks.

3. False Alarm Rate Comparison

False alarm rate indicates the percentage of normal data which is wrongly considered as attack, and it is defined as follows:

\[
\text{False Alarm Rate} = \frac{FP}{FP + TN} \times 100
\]  \hspace{1cm} (12)

Fig 3. Comparison of False Alarm Rate on Four Attacks
The results of false alarm rate for different types of attacks is shown in fig 3. From the figure it is observed that for DoS attacks, false alarm rate for RVM is lesser in all other attacks. Thus the experimental results proved that the proposed RVM obtains better results.

v. CONCLUSION

At present, security inside the network communication is of a important thing. Being the information that the datas are considered as one of the valuable asset for an organization, providing security in opposition to the intruders is very essential. Intrusion detection system tries to identify security attacks of intruders by investigating several data records observed in processes on the network. In this paper, unity-based normalization is proposed to standardize data and Relevance Vector Machine (RVM) is proposed for efficient classification. The experiment is exposed in WEKA by using KDD Cup 1999 dataset and the results indicate that the proposed system can provide better detection rate and low false alarm rate than the KSVM with LM. As a future work, various training algorithms are employed to improve its performance.

REFERENCES

International Journal of Computer Engineering and Technology (IJCET), ISSN 0976-6367(Print), ISSN 0976 – 6375(Online) Volume 4, Issue 1, January- February (2013), © IAEME

2/10


[18] Dimitris G. Tzikas, Liyang Wei, Aristidis Likas, Yongyi Yang and Nikolas P. Galatsanos, “ATutorial on Relevance Vector Machines For Regression and Classification with Applications”.

[19] B.Venkateswara Reddy, Dr.P.Satish Kumar, Dr.P.Bhaskar Reddy and B.Naresh Kumar Reddy, “Identifying Brain Tumour From MRI Image Using Modified FCM and Support VECTOR MACHINE” International journal of Computer Engineering & Technology (IJCET), Volume 4, Issue 1, 2013, pp. 244 - 262, Published by IAEME.


AUTHOR BIOGRAPHIES

V. JAIGANESH is working as an Assistant Professor in the Department of Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore, Tamilnadu, India. and Doing Ph.D., in Manonmaniam Sundaranar University, Tirunelveli. Tamilnadu, India. He has done his M.Phil in the area of Data Mining in Periyar University. He has done his post graduate degrees MCA and MBA in Periyar University, Salem. He has presented and published a number of papers in reputed conferences and journals. He has about twelve years of teaching and research experience and his research interests include Data Mining and Networking.

Dr. P. SUMATHI is working as an Assistant Professor, PG & Research Department of Computer Science, Government Arts College, Coimbatore, Tamilnadu, India. She received her Ph.D., in the area of Grid Computing in Bharathiar University. She has done her M.Phil in the area of Software Engineering in Mother Teresa Women’s University and received MCA degree at Kongu Engineering College, Perundurai. She has published a number of papers in reputed journals and conferences. She has about Sixteen years of teaching and research experience. Her research interests include Data Mining, Grid Computing and Software Engineering.