An Competent Intrusion Detection System using Relevance Vector Machine

V. Jaiganesh
Doctoral Research Scholar,
Department of Computer Science,
Manonmaniam Sundaranar University,
Tirunelveli, Tamilnadu, India.

P. Sumathi,Ph.D
Doctoral Research Supervisor,
Assistant Professor,
PG & Research Department of Computer Science,
Government Arts College,
Coimbatore, Tamilnadu, India

ABSTRACT
Nowadays everyone needs Internet for their useful. So internet becomes a public network in the world wide. Intrusion Detection (ID) is the art of detecting inappropriate, incorrect, or anomalous activity. It is a security service that monitors and analyzes system events for the idea of finding, and providing near real-time or real-time notice of, attempts in order to access system resources in an unauthorized manner In proposed system the `relevance vector machine' (RVM), a model of same functional form to the popular and state-of-the-art `support vector machine' (SVM). It demonstrate that by exploiting a probabilistic Bayesian learning framework and it can derive accurate prediction models which typically utilize dramatically fewer basis functions than a comparable SVM. These include the benefits of probabilistic predictions, automatic estimation of `nuisance' parameters and the facility to utilize arbitrary basis functions. The experiment is carried out with the help of MATLAB and 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
Intrusion Detection System (IDS), Relevance Vector Machine (RVM)

1. INTRODUCTION
Network instruction detection is the 'burglar alarms' or 'intrusion alarms' of the computer and network security field system. The chief idea of this to defend a system by using a combination of an alarm that sounds whenever the site's security has been compromised, and an entity – most often a site security officer that can respond to the alarm and take the appropriate action. An IDS tries to identify attempts to hack or break into a computer system or try to misuse it. IDSs may check packets passing over the network, monitor system files, observe log files, or set up deception systems which attempt to trap hackers. [1]

Intrusion detection is needed in today's computing environment because it is impossible to keep pace with the current and potential threats and vulnerabilities in the computing system techniques. This type of surroundings is constantly evolving and changing fueled by new technology and the internet system. To make matters inferior, intimidation and vulnerabilities in this environment are also constantly developing. Intrusion detection products are tools to assist in managing threats and vulnerabilities in this changing environment.

Intrusion detection attacks are segmented into two groups they are Host-based attacks[2-4] and Network-based attacks[5-7]

Categories of IDS

There are many ways to categorize IDS:

Misuse detection vs. anomaly detection: In misuse detection, these analyze the information it gathers and compares it to large databases of attack signatures. Basically, it looks for a specific attack that has already been documented. Similar to a virus detection system, this type of detection software is only as good as the database of attack signatures that it uses to compare packets against. In anomaly detection in this that the system administrator defines the baseline their normal and the state of the network's breakdown, traffic load, typical packet size and protocols. This type of detector monitors network segments to compare their state to the normal baseline and look for anomalies.

Network-based vs. host-based systems: In a network-based system, it analyzes the individual packet flowing through the network. It can detect malicious packets that are designed to be overlooked by a firewall's simplistic filtering rules. In a host-based system the activity on each individual computer or host examined by the IDS.

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 scheme employs RVM classification.

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

2. RELEATED WORK
Instruction detection is the process of monitoring and analyzing events occurred in a computer or network and presenting the results to the administrator [8, 9]. The related research on intrusion detection started in the early 1980's stage. Then the instruction detection sustained through several major DARPA projects and other Government programs. Later in the beginning of the 1990s it became a hot research topic and commercial IDRS started to emerge [10] when the internet era arrived.

Unlike a traditional network, which only passively transforms data packets, an active network allows the network node to execute mobile code carried in packets. The proposed IDRS combines distributed monitoring techniques (through individual host and LAN monitors) and data mining methods to analyze the threats of the incoming data and respond to the intrusions effectively (which is achieved by an intrusion detection center).

The KDDCup99 datasets includes very high training samples. This dataset is a common standard for evaluation of intrusion detection techniques. This dataset contains a number of connection records where each connection is a sequence of packets containing values of 41 features. The attack types in this dataset drop into four main categories: denial of service (DoS), probe, user to root (U2R), and remote to local (R2L). Feature values in this data set are as continuous, discrete and symbolic. Range of values for some of these features is very large and diverse. [11]

An emerging trend that addresses these problems is the deployment of intrusion detection systems. These systems are designed for detecting threatening situations which occur in spite of other security measures and follow two main paradigms: anomaly detection and misuse detection systems. They are planned to be

used in conjunction with additional security measures enforcing the security policy. They constitute a second line of defense sitting behind the locks on the doors and windows like motion detector in building. [12]

3. METHODOLOGY

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

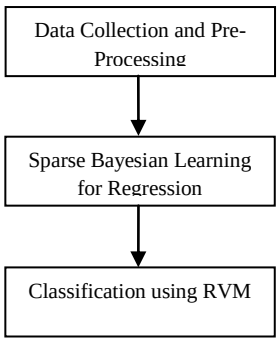

Figure 1. Steps involved in the proposed method

3.1 Data Collection and Preprocessing

The proposed IDS are experimented using the data taken from KDD Cup '99, which is widely accepted as a benchmark dataset and referred by many researchers. "10% of KDD Cup'99" from KDD Cup '99 data set was chosen to evaluate rules and testing data sets to detect intrusion. The entire KDD Cup '99 data set contains 41 features. Connections are labeled as normal or attacks fall into 4 main categories [13].

- DOS:- Denial Of Service
- Probe:- e.g. port scanning
- U2R:- an unauthorized access to root privileges,
- R2L:- an unauthorized remote login to machine.

Both training and testing data are divided into following three protocol types such as UDP, TCP or ICMP in order to test and train the data separately. The repeating number of remaining data has been deleted. The data to be deleted were chosen mostly from "normal" labeled data from the dataset. Still there were some attacks remaining in testing data set that were not in training data set and it can be tested using RVM classification.

3.2 Sparse Bayesian Learning for Regression

Given a data set of input-target pairs: $\{X_n, t_n\}_n^N = 1$ considering scalar-valued target functions only, then follow the standard probabilistic formulation and assume that the targets are samples from the model with additive noise:

$$t_n = y(X_n; W) + \epsilon_n$$

Where ϵn are independent samples from some noise process which is further assumed to be mean-zero Gaussian by variance. [17]

To extend the probit (or the logistic) model to include non-linear transformations of the input features, the link can be applied to a non-linear function of x: $P(y = 1|x) = \varphi(\mu(x, \beta))$

Here, consider classifiers only where this non-linear function is of the form $\mu(x, \beta) = \beta^T h(x)$ this includes:

- Linear classifiers; $h(x) = [1, x_1, .., x_d]$, in which case βis a $(d + 1)^{-1}$dimensional vector.
- Non-linearclassifiers; $h(x) = [1, \emptyset_1(x), \dots., \emptyset_k]^T$ where the $\emptyset_i(.)$ are nonlinear functions. Here, the dimensionality ofβis k+1.
- Kernel classifiers; $h(x) = [1, k(x, x^{(1)}), k(x, x^{(n)})]^T$ where K(.,.) is some (symmetric) kernel function. Here,β is (n+1)- dimensional. This is used in SVM and RVM approaches.

The more important characteristic of the probit link to exploit is that it has a simple interpretation in terms of hidden (or latent) variables.

3.3 Relevance Vector Machine

Relevance vector machine is simply a specialization of a spares Bayesian model which utilizes the same data dependent kernel basis. The main feature of RVM is that the inferred predictors are exceedingly sparse in that they contain relatively few ''relevance vectors", they are good generalization performance. For this self-contained paper, RVM for regression is introduced concisely here Supposing the mapping relationship is multiple-input-single out (MISO), sampled a dataset of N input vectors $\{X_n\}_{n=1}^n$ alongwith N corresponding scalar-valued target $\{t_n\}_{n=1}^N$, and assuming that the outputs are independent, identically distributed (IID) observations. Some observations could be assumed to contain mean-zero Gaussian noise with variance observations and some observations could be assumed to contain mean-zero Gaussian noise with variance in engineering point of view.

$$\sigma^2 : p(\epsilon_n|\sigma^2) = N(0, \sigma^2)$$

$$t = y(X : W) + \epsilon = \Phi W + \epsilon$$

Where $t = [t_1, \dots\dots., t_N]^T$, $W = [w_1, \dots\dots, w_m]^T$ is the weight vector and where U is the $N \times M$ design matrix, wherein its element is $\Phi_{nm} = k(X_m, X_n)$. In fact, the sparse Bayesian learning framework has the ability to utilize arbitrary basis functions, such as symmlet wavelet kernel, Gaussian kernel, Haar wavelet kernel and splines kernel,

The classical approach is used to estimate t which is to maximize likelihood or to minimizing ''least-squares" of the measured training dataset in order to estimate of W and r, however, it would lead to over-fitting.

$$p(t|W, \sigma^2) = (2\pi\sigma^2)^{-\frac{\alpha}{2}} exp\left\{-\frac{1}{2\sigma^2}\|t - \Phi W\|^2\right\}$$

To control the complexity of model and avoid over-fitting, a zero mean Gaussian prior probability distribution is defined over every ω_i with varianceσ_1^{-1}, the prior of W is written as:

$$p(W| \propto) = (2\pi)^{-M/2} \prod_{m=1}^{M} \alpha_m^{1/2} exp\left\{-\frac{\alpha_m \omega_m^2}{2}\right\}$$

Where hyper parameters vector $\alpha = [\alpha_0, \alpha_1, \alpha_2, \dots. \alpha^N]^T$ controls how far from zero each weight is allowed to deviate. For completion of hierarchical prior, hyper priors over $\alpha : p(x)$ and the inverse noise variance $\sigma^2 : p(\sigma^2)$ are specified as as Gamma distributions within a; b; c; d sets to some uninformative value$(e.g., a = b = c = d = 10^{-4})$.

$$p(\propto) = \Gamma(\propto |a, b) = \prod_{n=0}^{N} \frac{b^a}{\Gamma(a)} \propto^{a-1} e^{-b\propto}$$

$$p(\sigma^2) = \Gamma(\sigma^2|c, d) = \frac{d^c}{\Gamma(c)} \sigma^{-2(c+1)} e^{-b/\sigma^2}$$

Consequently, using Bayes' posterior inference, the posterior distribution over W is also conveniently Gaussian:

$p(W|t, \propto, \sigma^2) \sim N(\mu, \sum)$. Where the posterior mean μ and covariance ∑ are as follows:

$$\mu = \sigma^{-2} \sum \Phi^T t$$

$$\sum = (\sigma^{-2} \Phi^T \Phi + A)^{-1}$$

Where $A = diag(\propto_0, \propto_1, \dots, \propto_N)$

Sparse Bayesian learning can then be formulated as maximization of hyper parameter posterior,$p(\propto, \sigma^2|t)\ \alpha\ p(t|\alpha, \sigma^2)p(\alpha)p(\sigma^2)$. In practice, due to $\propto, \sigma^2 > 0$,to avoid adding positive constraints in the optimization problem, their logarithm are considered, then uniform hyper priors are defined over a logarithmic scale, which ultimately raises spare solutions. So the MAP of hyper parameter needs only to maximize the marginal likelihood $p(t|\propto, \sigma^2)$ known as type-II Maximum Likelihood procedure. The RVM marginal likelihood, also called evidence by MacKay is given by:

$$L(\propto, \sigma^2) = p(t|\alpha, \sigma^2) = \int p(t|X, W, \sigma^2)p(W|\alpha)\, dW \sim N(0, c)$$
$$= -\frac{1}{2}[Nlog2\pi + log|C| + t^T C^{-1} t]$$

Where the covariance is $C = \sigma^2 I + \Phi A^{-t} \Phi^T$

Obtained in this optimization process, the value of $\propto_{MP}$ and $\propto^2_{MP}$as the substitution of the$\propto$ and σ^2, the posterior mean and variance can be computed, and then a mean final approximator at unseen data X^*could be gained with:

$$\mu_* = \mu^T \Phi(X^*)$$

If the sparse Bayesian learning framework utilize Gaussian kernel $k_0(X_m, X)$ basis function, cross validation on the validation set is used to get good unified kernel width.

4. Experimental Results

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

- TCP.
- UDP and
- ICMP.

Totally 600 database are collected from the KDD cup99, these database are used for classification. Here two type of classification are done, that are RVM and SVM

4.1 Performance Measures

The performance measure which are used to evaluate the proposed RVM against SVM is

Correctly detected attack and False-alarm rate

4.2 Correctly Detected Attack

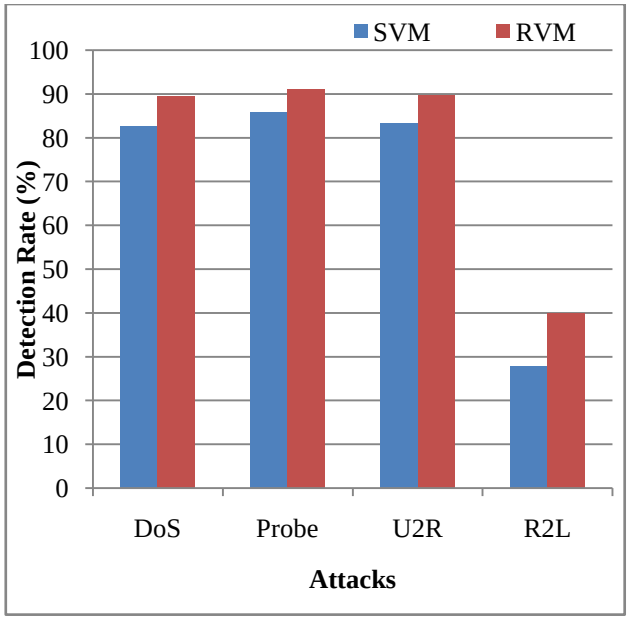

Figure 2. Comparison of Detection Rate

The results of correctly detection rate for different types of attacks are shown in figure 2. From the results it is observed that the TCP have a high accuracy in the RVM classification against SVM classification, after the TCP, UDP have a high accuracy than ICMP in the Classification of RVM and SVM.

4.3 False Alarm rate

False alarm rate indicates the percentage of normal data which is wrongly recognized as attack. By using this False Alarm find out the overall accuracy in the classification

The results of false alarm rate for different types of attacks are shown in figure 3. From the figure that the RVM is observed which have higher accuracy than the SVM. Thus the experimental results proved that the proposed RVM obtains better results.

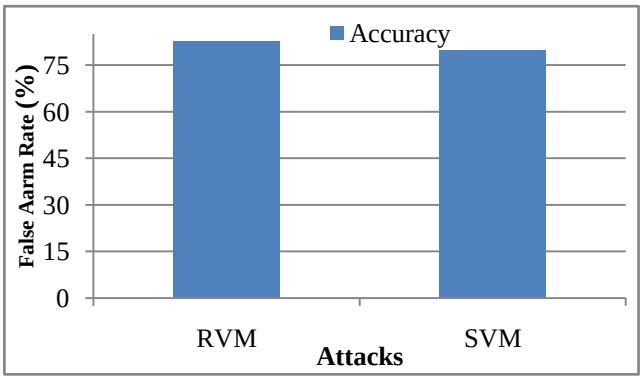

Figure 3. Overall Accuracy

5. Conclusion

At present, security inside the network communication is of a main concern. Being the information the data which are considered as the precious asset of an organization, trying to provide security in opposition to the intruders is very important. Intrusion detection system tries to recognize security attacks of intruders by investigating several data records observed in processes on the network. In this paper, the unity-based normalization is proposed to standardize data and Relevance Vector Machine (RVM) is used for efficient classification. The experiment is carried out in MATLAB and 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 SVM. As a future work, various training algorithms are employed to improve its performance

6.References

[1] Avolio, Frederick M. "A multidimensional approach to internet security." netWorkervol. 2.2, April 1998, pp. 15-22.

[2] Vivek K. Kshirsagar, Sonali M. Tidke & Swati Vishnu, "Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview", International Journal of Computer Science and Informatics ISSN (PRINT): 2231 – 5292, Vol-1, Iss-4, 2012

[3] D. Anderson, T. Frivold and A. Valdes, "Next-generation intrusion detection expert system (NIDES): a summary", Technical Report SRI-CSL-95-07. Computer Science Laboratory, SRI International, Menlo Park, CA, 1995.

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

[5] S. Freeman, A. Bivens, J. Branch and B. Szymanski, "Host-based intrusion detection using user signatures", Proceedings of the Research Conference. RPI, Troy, NY, 2002.

[6] K. Ilgun, R.A. Kemmerer and P.A. Porras, "State transition analysis: A rule-based intrusion detection approach", IEEE Trans. Software Eng, Vol. 21, No. 3, Pp. 181–199, 1995.

[7] D. Marchette, "A statistical method for profiling network traffic", Proceedings of the First USENIX Workshop on

[8] Intrusion Detection and Network Monitoring, Santa Clara, CA, Pp. 119–128, 1999.

[9] R.G. Bace, "Intrusion Detection", Macmillan Technical Publishing, 2000.

[10] F. Chuan, P. Jianfeng, Q. Haiyan, and W. R. Jerzy, "Alert fusion for a computer host based intrusion detection system," in Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems, 2007, pp. 433-440.

[11] W. Rensheng and V. N. Jeffrey, "Search strategy optimization for intruder detection," IEEE Sensors Journal, Vol. 7, 2007, pp. 315-316.

[12] J. G. Tront and R. C. Marchany, "Internet security: Intrusion detection and prevention in mobile systems," in Proceedings of the 40th Annual Hawaii International Conference on System Sciences, 2007, pp. 162-162.

[13] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi and Lilly Suriani Affendey, "Intrusion Detection Using Data Mining Techniques", IEEE, 2010. ISBN: 978-1-4244-5651-2/10

[14] N.S.Chandolikar1 & V.D.Nandavadekar2, "Comparative Analysis Of Two Algorithms For Intrusion Attack Classification Using Kdd Cup Dataset", International Journal of Computer Science and Engineering (IJCSE) Vol.1, Issue 1 Aug 2012 81-88 © IASE.

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.