Intrusion Patterns Recognition in Computer Networks

Ali Farzan, Naser Razavi, Mohammad Ali Balafar and Farshad Arvin

Abstract - One of the hottest research areas in recent years is detecting network intrusion patterns in computer networks. Because of dynamic nature of intrusion patterns in networks, intelligently inspecting the behavior of networks and detecting anomalies are mostly desirable. KDD-Cup99 pattern database are used as a standard source of network packets in our research. K-mean, Bayesian method and Support Network Machine (SVM) are used as anomaly detectors. Results show the superiority of SVM over other two methods regarding the accuracy of classifying patterns into normal packets and suspicious ones. It can be concluded that using high dimensional pattern recognition methods have reasonable competence in detecting attack patterns in computer networks.

Keywords: **k-mean; Bayesian; SVM; Network Intrusion Detection**

I. INTRODUCTION

With the development of complicated networks and specially Internet technology, security of the networks has become one of the major issues in designing the networks [1]. Availability of comprehensive and rich information sources for the various ways of destructive attacks motivates more hackers to use simple operations in performing fatal attacks [2-5]. It is supposed that the amount of hacking attacks is growing 10 times per year [6] and this makes the security of computer networks a critical topic.

Traditional methods for enforcing security in networks such as VPN, firewall or encryption methods suffer from their static nature and cannot be adapted to the dynamic nature of the attacks.

Manuscript received July 26, 2011; revised August 9, 2011. This work was supported by the Islamic Azad University, Shabestar Branch, Iran under Grant 51954900129001.

Ali Farzan is with the Islamic Azad University, Shabestar Branch and University of Putra Malaysia (Corresponding author to provide phone: +601-76737629; e-mail: alifarzanam@gmail.com).

Naser Razavi is with the Islamic Azad University, Shabestar Branch (email: razavi@iust.ac.ir).

Mohammad Ali Balafar is with the Islamic Azad University, Shabestar Branch (email: balafarila@yahoo.com).

Farshad Arvin is with the Islamic Azad University, Shabestar Branch (email: farshadarvin@yahoo.com).

That is, the attack data packets, often don't follow a pre specified well known pattern format. Rather, regarding the type of attack and the severity of attacker, its format varies. This dynamic nature of attack types motivates researchers to develop new methods in detecting intrusion packets [1-2, 7-8]. A network intrusion detection system has the responsibility of monitoring traffic on the network, modeling the normal an abnormal behavior of it and regarding this model, to issue an alarm when detecting any data packet which matches the abnormal state of the model [8].

Three different classification methods are used in this paper to classify data packets into normal or abnormal ones. According to our sample data set, the abnormal packets are also divided into 4 different groups [9]. This categorization has been done based on the type of attacks as:

- Denial of Service Attack (**DoS**): is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine.
- User to Root Attack (**U2R**): is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system.
- Remote to Local Attack (**R2L**): occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine.
- Probing Attack: is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls.

For each packet of dada, there are 41 various characteristics of them in the database which are used as

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

features in classifying packets into 5 different groups (1 group for normal packets and 4 groups for attacks).

Three various classifiers are adopted in designing the intrusion detection systems or packet classifiers. K-mean as a clustering method, Bayesian classifier as a statistical method and finally, SVM as a kernel based method are three methods with different natures which are used in this paper.

A total of 9354 randomly selected data packets are used in this study. All of these patterns are labeled with their appropriate groups and so can be used as training or testing samples.

II. CLASSIFICATION BASED ON K-MEAN

Suppose $X = \{x_i \in \mathbb{R}^p | i = 1 \dots n\}$ denotes the set of n observations of p–dimensional patterns and the goal is to classify those n observations into $K < n$ classes, $C_k \quad 1 < k < K$. A classifying rule (denoted by R is a many-to-one mapping such as $R(x_i) = C_k \quad (1 < i < n, 1 < k < K)$. The K-means aims to minimize the overall objective function

$$\sum_{k=1}^{K} \sum_{R(x_i)=C_k} \|x_i - \mu_k\|^2$$

With respect to the classification rule R, where μ_k is the means (or centroid) of patterns from cluster k. This paper, assumes $k = 5$ and therefore, there are five cluster means, μ_1 to μ_5. Beginning with an initial assignment of patterns to clusters or an initial assignment of cluster means the K-means algorithm iterates through the following two steps:

Step 1: Reassign each observation to the cluster whose mean is closest to that observation.

$$R(x_i) = C_k \quad \leftrightarrow \quad k = arg\ \min_l \|x_i - \mu_l\|^2$$

Step 2: Recalculate the new cluster means.

The convergence is reached if the cluster means do not change. An observation x is therefore, classified to C_1 with cluster mean μ_1 if and only if

$$\|x_i - \mu_1\|^2 < \|x_i - \mu_2\|^2$$

This method is conducted on all of sample data packets with randomly initialized cluster means. Results are shown in Table 1.

Table 1
Clustering Results for K-mean

Assigned to Cluster	1	2	3	4	5
Dos	57	4923	53	1717	150
R2L	302	0	48	13	133
Probe	0	0	31	96	1
Normal	1568	4	99	2	145
U2R	0	0	0	5	7

Packets of each group assigned into 5 clusters

Regarding these results, each cluster can be labeled as Table 2:

Table 2
Labeling Cluster Numbers

Cluster No	Label
1	Normal
2	Dos
3	No Label
4	Probe
5	R2L

Each cluster is labeled according to the type of maximally assigned packets

It can be revealed that 2634 packets are misclassified which is equal to the error rate of 28.16%.

III. CLASSIFICATION BASED ON BAYESIAN METHOD

Statistical methods have been used widely for estimating or modeling the histogram or probability density function of random variables [10]. Bayes classifier is used as our decision rule in assigning packet labels. For a given pattern with feature vector of x, Bayes rule is used to calculate a posteriori probability of assigning label w to it

$$p(w|x) = \frac{p(x|w) \quad p(w)}{p(x)}$$

Expectation Maximization (EM) is used to compute MAP. It is constituted of two iterative steps, *Expectation* and *Maximization*. In *expectation step (E-Step)* given the current estimate of distribution parameters, the conditional probability of w is calculated using above equation. In *maximization step (M-Step)* based on the last classification performed in *expectation step (E-Step)*, it calculates new values of distribution parameters as well as a priori probability.

Given the normal distribution of intensities for each of brain tissues, there are two following steps

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

E-Step:

$$p(w_i|x_j) = \frac{G(x_j, \mu_i^m, \sigma_i^m).p(w_i)^m}{\sum_k G(x_j, \mu_k^m, \sigma_k^m).p(w_k)^m}$$

M-Step:

$$\mu_i^{m+1} = \frac{\sum_j x_j.p(w_i|x_j)}{\sum_j p(w_i|x_j)}$$

$$(\sigma_i^{m+1})^2 = \frac{\sum_j (x_j - \mu_i^{m+1})^2.p(w_i|x_j)}{\sum_j p(w_i|x_j)}$$

$$p(w_i)^{m+1} = \frac{\sum_j p(w_i|x_j)}{\sum_i \sum_j p(w_i|x_j)}$$

Results of adopting this method in our data are illustrated in Table 3.

Table 3
Classification Results for Bayesian Method

Label	True Positive Rate	False Positive Rate
Dos	0.965	0.001
R2l	0.935	0.029
Probe	0.922	0.014
Normal	0.864	0.002
u2r	0.75	0.014

True positive rate shows the percentage of correctly classified packets with appropriate label. False positive rate shows the percentage of misclassified packets into the appropriate label.

A total of 534 packets are misclassified which is equal to the error rate of 5.71%.

IV. CLASSIFICATION BASED ON SVM

Support Vector Machine is a supervised learning method. In brief, given some known examples $(x^k, y^k)_{k=1,2,\ldots,n}$ where $x^k \in R^p$ is the observed pattern of features and $y^k \in \{1, -1\}$ is the appropriate class label, linear SVM aims to find the best hyperplanes which can separate patterns of classes from each other. The optimal hyperplanes are the ones for which the margin between groups is maximal, and at the same time the number of misclassified patterns is minimal. To find such hyperplanes, the following constrained optimization problem must be solved

$$min\left(\frac{1}{2}w^T w + c\sum_{k=1}^{n}\xi^k\right)$$

Subject to $y^k(w^T.x^k + b) \geq 1 - \xi^k$

c, is a parameter controlling the tradeoff between maximizing margin and minimizing misclassified patterns. ξ^k is a positive slack variable allowing some of the patterns lie in the wrong side of the margin. Suppose the classification function is $y = sign(x.w + b)$ where w determines the orientation of the hyperplane and b is the offset from the origin. The vector w which maximizes the margin can be written as a linear combination of some patterns. These patterns are called "Support Vectors". It is obvious that classification is done based on dot products of patterns which provide a linear classifier. By replacing the dot product with a kernel evaluation such as RBF kernel, one can design a nonlinear SVM classifier.

Results of applying SVM classifier over our data is depicted in Table 4.

The number of misclassified packets is 276 which lead to the error rate of 2.95%.

Table 4
Classification Results for SVM

Assigned to Cluster	Dos	R2L	Probe	Normal	U2R
Dos	6882	0	1	17	0
R2L	3	480	2	11	0
Probe	1	4	123	0	0
Normal	8	227	1	1582	0
U2R	0	1	0	0	11

Packets of each group assigned into 5 groups

V. CONCLUSION and RESULTS

To simplify investigating the results of all three classifiers and comparing them versus each other, Table 5 shows all the results.

It seems that kernel based methods outperforms other two types. The capability of kernel methods in transforming original feature space into new one which is more appropriate to classification purposes, can be considered as a reason for SVM's highest accuracy.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table 5
Classification Results for all three methods

	K-mean		Bayesian		SVM	
Correctly Classified Packets	6720	71.84%	8829	94.2874%	9078	97.0494%
Incorrectly Classified Packets	2634	28.16%	525	5.7126%	276	2.9506%

SVM performs better than two other methods.

REFERENCES

[1] Markatos EP, Xinidis K, Anagnostakis KG, editors. Design and Implementation of a High-Performance Network Intrusion Prevention System2010.
[2] Carl G, Brooks RR, Rai S. Wavelet based Denial-of-Service detection. Computers & Security. 2006;25(8):600-15.
[3] Schneier B, Gross AH, Callas JD. Method and system for dynamic network intrusion monitoring, detection and response. Google Patents; 2011.
[4] Siris VA, Papagalou F. Application of anomaly detection algorithms for detecting SYN flooding attacks. Computer communications. 2006;29(9):1433-42.
[5] Sommer R, Paxson V, editors. Outside the closed world: On using machine learning for network intrusion detection2010: IEEE.
[6] Xue M, Zhu C, editors. Applied Research on Data Mining Algorithm in Network Intrusion Detection2009: IEEE.
[7] Chen RC, Cheng KF, Hsieh CF. Using rough set and support vector machine for network intrusion detection. Arxiv preprint arXiv:10040567. 2010.
[8] Sun S, Wang YZ, editors. A Weighted Support Vector Clustering Algorithm and its Application in Network Intrusion Detection2009: IEEE.
[9] Tavallaee M, Bagheri E, Lu W, Ghorbani AA, editors. A detailed analysis of the kdd cup 99 data set2009.
[10] Broadhurst RE, Stough J, Pizer SM, Chaney EL, editors. A statistical appearance model based on intensity quantile histograms2006: IEEE.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)