Anomaly Internet Network Traffic Detection by Kernel Principle Component Classifier*

Hanghang Tong[1], Chongrong Li[2], Jingrui He[1], Jiajian Chen[1], Quang-Anh Tran[2], Haixin Duan[2], and Xing Li[2]

[1] Department of Automation, Tsinghua University, Beijing 100084, China
{walkstar98,hejingrui98}@mails.tsinghua.edu.cn
[2] Network Research Center of Tsinghua University, Beijing 100084, China
{licr,qa,dhx,xing}@cernet.edu.cn

Abstract. As a crucial issue in computer network security, anomaly detection is receiving more and more attention from both application and theoretical point of view. In this paper, a novel anomaly detection scheme is proposed. It can detect anomaly network traffic which has extreme large value on some original feature by the major component, or does not follow the correlation structure of normal traffic by the minor component. By introducing kernel trick, the nonlinearity of network traffic can be well addressed. To save the processing time, a simplified version is also proposed, where only major component is adopted. Experimental results validate the effectiveness of the proposed scheme.

1 Introduction

Intrusion detection has received great attention from researches in the past years [5], [6]. It has been widely recognized that a malicious intrusion or unauthorized use could cause severe damages [11], [13]. According to [2], an intrusion can be defined as "any set of action that attempts to comprise the integrity, confidentiality or availability of information resources". Existing intrusion detection methods can be classified into two categories [13]: misuse detection [8], [9] and anomaly detection [11], [13]. Compared with misuse detection, anomaly detection does not need any prior knowledge of attack, and therefore can detect novel attack types. With the development of Internet, it has become a crucial issue from both application and theoretical point of view [5], [6].

Almost, if not all, existing anomaly detection methods are based on the following assumption [5], [6], [11], [13]: the network traffic with attack has different statistical character compared with that without attack. The main difficulties of anomaly detection lie in two aspects. On one hand, there are a lot of different kinds of attacks. For example, the authors in [4] identified five cases where anomalies present in attack traffic, including user behavior, bug exploits, response anomalies, bugs in the attack and evasion. The statistical nature of attack traffic might vary dramatically throughout different attack types. On the other hand, even for the normal traffic, its statistical nature is very complex. Two of the most important discoveries of the statistics of

* This work is supported by National Fundamental Research Development (973) under the contract 2003CB314805.

J. Wang, X. Liao, and Z. Yi (Eds.): ISNN 2005, LNCS 3498, pp. 476–481, 2005.

anomaly Internet traffic over the last ten years are that Internet traffic exhibits self-similarity [3], [7], [12] (in many situations, also referred as long-range dependence) and non-linearity [1], [12]. The diversity of attack types and the complex statistical nature of network traffic make highly accurate anomaly detection very difficult.

In the past years, many methods have been proposed for anomaly network traffic detection. Statistical-based techniques build a norm profile and make use of statistical tests to perform anomaly detection [5]. A representative work in this category is based on chi-square [15]. More recently, researchers have applied machine learning technique to anomaly detection (See [6] for a detailed review). For example, the authors in [13] proposed using One-Class Support Vector Machine (OCSVM); the authors in [11] proposed a robust principle component classifier (PCC) for anomaly detection and the experimental results on KDD-99 dataset showed its superiority over existing methods. However, as a linear dimension reduction method, PCC cannot capture the non-linearity of network traffic, and therefore, its effectiveness might be compromised.

To deal with the non-linear nature of Internet traffic, in this paper, we introduce kernel trick into principle component classier (PCC) and propose a novel anomaly network traffic detection scheme, namely kernel principle component classifier (KPCC). Like PCC, KPCC can detect anomaly network traffic which has extreme large value on some original feature by the major component, or does not follow the correlation structure of normal traffic by the minor component. As a non-linear dimensionality reduction method, Kernel PCA ensures that the non-linear nature of network traffic can be well addressed. However, PCC cannot achieve this goal. We also proposed a simplified version of KPCC (SKPCC), in which only major component is used. In SKPCC, the processing time for solving eigen-problem can be greatly reduced since only the first several eigen-values and eigen-vectors are needed, which is a desirable property for real applications. Experimental results on DARPA 1999 dataset demonstrate the effectiveness of the proposed methods.

The rest of this paper is organized as follows: in Section 2, we present our Kernel Principle Component Classifier in detail; experimental results are given in Section 3; finally, we conclude the paper in Section 4.

2 Kernel Principle Component Classifier

2.1 Notation

Let $\{X_i, i=1,\cdots,N\} \in R^m$ the training set and $X_{test} \in R^m$ a testing sample;

Let $K=(k(x_i,x_j))_{i,j}$ the dot product matrix defined on training set by a certain type of kernel [10], [14]. Let $\{(\lambda_i, e_i); i=1,\cdots,N\}$ be the eigen-spectrum of K, where is λ_i is the i^{th} largest eigen-value and e_i is the corresponding eigen-vector;

Let $Y_i=[y_i^1,\cdots,y_i^p]^T$ be the principle component of X_i $(i=1,\cdots,N)$, where $y_j (j=1,\cdots,p)$ is the j^{th} principle component. Similarly, Let $Y_{test}=[y_{test}^1,\cdots,y_{test}^p]^T$ be the principle component of X_{test};

Let $\lambda_j (j = 1, \cdots, q)$ and be the major eigen-value, and $C_{maj}(X_i) = \sum_{j=1}^{q} \frac{y_i^j}{\lambda_j}$ the major component of X_i. Let $\lambda_{p-r+j} (j = 1, \cdots, r)$ and be the minor eigen-value, and $C_{min}(X_i) = \sum_{j=p-r+1}^{r} \frac{y_i^j}{\lambda_j}$ the minor component of X_i.

2.2 Architecture

Like in [13], our system contains three modules as shown in Fig. 1:

- The collection module sniffs network traffic to 1) form the training set $\{X_i, i = 1, \cdots, N\}$, where all traffic are normal and 2) prepare a testing sample X_{test}.
- The training module generates KPCC or SKPCC by the training set.
- The Testing module determines whether or not the testing sample X_{test} is attack.

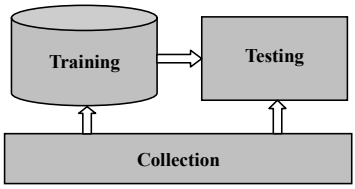

Fig. 1. Architecture of the proposed system

2.3 Algorithm

The network traffic is likely to be an attack if its statistical nature is different from those normal ones. In KPCC, it indicates that such traffic whether 1) has extremely large value on some of its original feature; or 2) it does not the correlation structure of normal traffic. While the former character can be capture by the major component, the latter can be described by the minor component [11]. Based on the above observation, KPCC can be designed as follow:

KPCC for anomaly traffic detection

- If $C_{maj}(X_{test}) = \sum_{j=1}^{q} \frac{y_{test}^j}{\lambda_j} > c_1$, or $C_{min}(X_{test}) = \sum_{j=p-r+1}^{r} \frac{y_{test}^j}{\lambda_j} > c_2$, X_{test} is an attack;
- Else X_{test} is a normal traffic.

Where c_1 and c_2 parameters, which can be determined by the specified false alarm rate (α_1 and α_2) from the training set:

$\alpha_1 = P\left(C_{maj}(X_i) > c_1 \mid X_i \text{ is normal}\right)$; and $\alpha_2 = P\left(C_{min}(X_i) > c_2 \mid X_i \text{ is normal}\right)$.

If the dot product matrix $K = (k(x_i, x_j))_{i,j}$ is replaced by the correlation matrix of $\{X_i, i = 1, \cdots, N\} \in R^m$, KPCC is degraded into PCC. However, unlike in PCC, the non-linearity within network traffic can be well described by kernel trick in KPCC.

Note that, in both PCC and KPCC, we need to get the complete eigen-spectrum to get the minor component. When the dimension is high, which is often the case for a real problem, the processing time might be much high. To address this issue, we also propose a simplified version of KPCC (SKPCC), in which only major component is used:

SKPCC for anomaly traffic detection

- If $C_{maj}(X_{test}) = \sum_{j=1}^{q} \frac{y_{test}^{j}}{\lambda_j} > c_1$, X_{test} is an attack;
- Else X_{test} is a normal traffic.

3 Experimental Results

DARPA 1999 data (http://www.ll.mit.edu/IST/ideval) is used to evaluate the performance of KPCC and SKPCC. The data in the first week is attack free, while that in second week contains various types of attack as listed in Table 1. As in [13], we use the *inside-tcpdump* dataset in the first week as the training set, and that in the second week as the testing set. To perform a fair comparison, the same types of statistics in [13] are adopted as listed in Table 2, which are generated by TCPSTAT (http://www.frenchfries.net/paul/TCPSTAT).

Table 1. Attack information in the second week

Day	Attack	Destination	Start Time	End Time
Tue	Portsweep	172.16.114.50	21:44:15	22:11:11
	Mailbomb	172.16.112.50	03:25:10	03:35:06
	Ipsweep	172.16.112.0/23	02:05:13	02:29:14
Wed	Satan	172.16.114.50	01:02:09	01:04:38
	Mailbomb	172.16.112.50	02:44:13	02:54:08
	Ipsweep	172.16.112.0/23	09:17:04	09:29:13
Thu	Satan	172.16.114.50	22:33:20	22:35:37
	Portsweep	172.16.114.50	23:50:07	00:07:31
	Neptune	172.16.114.207	00:04:12	00:07:37
Fri	Ipsweep	172.16.112.0/23	05:36:06	05:39:33
	Neptune	172.16.114.50	00:20:11	00:23:36
	Portsweep	172.16.112.50	06:13:02	06:25:06

Table 2. Traffic statistics used in KPCC/SKPCC

TCPSTAT Output Options	Traffic Statistics
%C	# of ICMP packets
%T	# of TCP packets
%U	# of UDP packets
%a	Mean of packet size
%d	Deviation of packet size

There are a set of parameters and operations that need to be set in KPCC and SKPCC:

- p is set so that the major eigen-value accounts for 50% of the total variance;
- r is set so that the minor eigen-value accounts for 0.02% of the total variance;
- The duration to generate the traffic statistics is set to be 300s;
- RBF kernel is adopted to formulate the dot product matrix $K = (k(x_i, x_j))_{i,j}$, that is, $k(x_i, x_j) = \exp(\frac{\|x_i - x_j\|^2}{\sigma})$, where $\sigma = 1$;
- $\alpha_1 = \alpha_2$; $\alpha_1 + \alpha_2$ is the total specified false alarm rate.

The receiver operating characteristic (ROC) curve on the testing set by KPCC and SKPCC is plotted in Fig. 2. It is compared with that by OCSVM [13], and PCC [11]. The recall and precision by KPCC are given in Table 3, with specified total false alarm rate 10%.

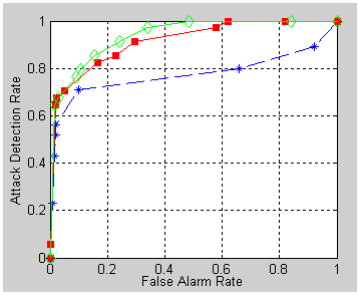

Fig. 2. ROC curve on the testing set by different anomaly detection schemes

Table 3. Recall and Precision of KPCC (10% total specified false alarm rate)

Actual \ Predicted	Attack	Normal	Recall
Attack	*TP=29*	*FN=5*	85.3%
Normal	*FP=193*	*TN=1058*	84.6%
Precision	13.1%	99.5%	

4 Conclusion

In this paper, we have proposed a novel anomaly network traffic detection scheme, namely kernel principle component classifier (KPCC). KPCC can capture two kinds of anomaly traffic, which have some extremely large values on some original feature, or do not follow the correlation structure of normal traffic. By introducing kernel trick into existing PCC, KPCC can well address the non-linearity of network traffic. To save the processing time, a simplified version of KPCC is also proposed, where only major component is used. Experimental results demonstrate the effectiveness of the proposed scheme. Future work includes: 1) investigating other kinds of kernel in

KPCC (SKPCC); 2) exploring other traffic statistics; 3) exploring the relationship between KPCC (SKPCC) and spectral methods.

References

1. Hansegawa, M., Wu, G., Mizuno, M.: Applications of Nonlinear Prediction Methods to the Internet Traffic. The 2001 IEEE International Symposium on Circuits and Systems, (2001) 169-172
2. Heady, R., Luger, G., Maccabe, A., Servilla, M.: The Architecture of a Network Level Intrusion Detection System. Tech Report, University of New Mexico, (1990)
3. Leland, W.E., Taqqu, M.S., Willinger, W., Wilson, D.: On the Self-similar Nature of Ethernet Traffic. IEEE/ACM Tran. on Networking, (1994) 1-15
4. Mahoney, M., Chan, P.K.: Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks. SIGKDD, (2002) 376-385
5. Markou, M., Singh, S.: Novelty Detection: A Review Part1: Statistical Approaches. Signal Processing, (2003)
6. Markou, M., Singh, S.: Novelty Detection: A Review Part2: Neural Network-based Approaches. Signal Processing, (2003)
7. Ostring, S., Sirisena, H.: The Influence of Long-rang Dependence on Traffic Prediction. IEEE ICC, (2001) 1000-1005
8. Paxson, V.B.: A System for Detecting Network Intruders in Real-Time. Lawrence Berkley National Laboratory Proceedings, 7'th USENIX Security Symposium, (1998)
9. Roesch, M.: Snort - Lightweight Intrusion Detection for Networks. Proceedings of USENIX Lisa'99, (1999)
10. Scholkopf, B., Smola, A.J., Muller, K.R.: Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation, (1998) 1299-1319
11. Shyu, M.L., Chen, S.C., Sarinnapakorn, K., Chang L.W.: A Novel Abnormal Detection Scheme Based on Principle Component classifier. ICDM (2003)
12. Tong, H., Li, C., He, J.: A Boosting-Based Framework for Self-similar and Non-linear Interet Traffic Prediction. ISNN (2004) 931-936
13. Tran, Q.A., Duan, H., and Li, X.: One-Class Support Vector Machine for Anomaly Network Traffic Detection. APAN (2004)
14. Vapnik, V.N.: An Overview of Statistical Learning Theory. IEEE Trans on Neural Networks, (1999) 988-999
15. Ye, N., Chen, Q.: An Anomaly Detection Technique Based on a Chi-Square Statistic for Detecting Intrusions into Information Systems. Quality and Reliability Eng Int'l, (2001) 105-112