IET Information Security

Research Article

IET Journals
The Institution of Engineering and Technology

Detecting anomalous traffic in the controlled network based on cross entropy and support vector machine

ISSN 1751-8709
Received on 28th November 2017
Revised 3rd May 2018
Accepted on 27th September 2018
E-First on 21st November 2018
doi: 10.1049/iet-ifs.2018.5186
www.ietdl.org

Weijie Han[1,2], Jingfeng Xue[1] ✉, Hui Yan[2]

[1]*School of Computer, Beijing Institute of Technology, Beijing, People's Republic of China*
[2]*School of Space Information, Space Engineering University, Beijing, People's Republic of China*
✉ *E-mail: xuejf@bit.edu.cn*

Abstract: Network anomaly detection is an effective way for analysing and detecting malicious attacks. However, the typical anomaly detection techniques cannot perform the desired effect in the controlled network just as in the general network. In the circumstance of the controlled network, the detection performance will be lowered due to its special characteristics including the stronger regularity, higher dimensionality and subtler fluctuation of its traffic. On the motivation, the study proposes a novel classifier framework based on cross entropy and support vector machine (SVM). The technique first subtracts the representative traffic characteristics from the network traffic and defines a 7-tuple feature vector for the controlled network by extending the traditional 5-tuple representation of the usual network. Then the probability distributions and cross entropies of the 7 tuples are calculated during the defined statistical window so as to generate the 7-tuple cross-entropy feature vector for profiling the network traffic fluctuation in the controlled network. Finally, the multi-class SVM classifier is trained by importing the 7-tuple cross-entropy feature vectors. Experimental results show that the proposed classifier can achieve higher detection rates and is more suitable to be used in the controlled network than the typical detection techniques.

Nomenclature

S_{ip}	source address of the network session
S_{port}	source port of the network session
D_{ip}	destination address of the network session
D_{port}	destination port of the network session
In	number of source nodes that is connecting with the observed node in the network session
Out	number of destination nodes that the observed node is connected within the network session
Vel	corresponding traffic rates of the different connections in the network session

1 Introduction

With the widespread use of the Internet, the potential risks due to the network attacks have become an urgent issue to be solved as soon as possible. To this end, researchers have carried out various kinds of anomalies detection methods to discover the attacks underlying in the huge network traffic [1]. According to the different techniques for detecting the anomalies, the conventional malicious traffic detection methods can be classified into three categories: the statistics-analysis based methods, the machine learning-based methods, and the signal-processing based methods. As the practical application results show, these approaches have solved the security challenges to some extent. However, these conventional anomalous traffic detection approaches cannot show the same satisfactory performance just as in the usual network environment when they are applied in the domain of the controlled network.

Compared with the traditional general-used network environment, the controlled network environment [2] has several different characteristics as follows:

(i) The controlled network always carries out critical tasks and demands higher security requirement. Usually, the network should go through strict security tests before practical execution so as to ensure its stability, safety, and reliability.
(ii) Both its access and operation process is controlled according to strict security mechanisms. On one hand, some constraints are always imposed on its access and utilisation including (i) which network terminals can access the network? (ii) how to set the IP addresses and open ports of the computers that are connected to the network? (iii) which application protocols can run on the network? On the other hand, its running status is also monitored in real-time so as to ensure its safety.

Due to its special characteristics, some new challenges will emerge when the common-used anomalous traffic detection techniques are applied to the controlled network including:

(i) The statistics-analysis based approach [3–5] cannot detect and discover the anomalous traffic occurring in a long interval or with subtle variations in the controlled network due to its relatively stable traffic characteristics.
(ii) The machine-learning based approach [6–8] may encounter the following problems: firstly, it is difficult to select distribution functions and solve the parameters due to lack of samples; secondly, it is a time-consuming task to detect the anomalies when the dimensionality of the traffic and the size of its features expand dramatically.
(iii) The signal-processing based approach [9–11] cannot recognise the small characteristic variations underlying in the traffic and obtain satisfactory classification results due to the slight fluctuation of the traffic.

In general, the controlled network traffic shows some different characteristics from the traditional general-used network, which include the stronger regularity, higher dimensionality, and subtler fluctuation. In order to solve the aforementioned problems, the paper proposes a novel detection technique by introducing cross entropy and support vector machine (SVM) because the metric of cross entropy can reflect the variation of the traffic characteristics more obviously [12] and the SVM classifier can classify the small-size, high-dimensionality and linearly non-separable samples effectively [13]. Following this line of thinking, we build a 7-tuple feature vector for profiling the controlled-network traffic, and then calculate the corresponding cross-entropy feature vector and input it to the multi-class SVM classifier to detect the anomalous traffic in the controlled network.

The contributions of this paper are as follows:

(i) In order to capture the subtle fluctuation underlying in the network traffic, we extend the traditional 5-tuple representation and build a novel 7-tuple representation for profiling its characteristics. Then, we generate the 7-tuple feature vector by calculating its probability distributions and cross entropies to reflect the characteristics of the traffic fluctuation.
(ii) In order to detect the anomalous traffic hidden in the controlled network, we build a classifier framework based on cross entropy and SVM. In this framework, the above 7-tuple cross-entropy feature vector is input to a multi-class SVM classifier for training and testing. In practical, the performance of the detection process depends on the 7-tuple cross-entropy feature vector.
(iii) In order to prove the effectiveness of the proposed approach, we conduct some experiments for comparing the approach with the typical methods on a representative dataset. The experimental results show that the proposed approach can solve the detection challenges faced by the controlled network efficiently and perform higher detection rate.

This paper is organised as follows: Section 2 analyses the motivation for the proposed approach, Section 3 describes the related work, Section 4 gives some main notations and necessary preliminaries, Section 5 defines the classifier model, Section 6 describes the classifier framework, Section 7 describes the classifying process, Section 8 evaluates the proposed approach by conducting comprehensive experiments, and finally Section 9 concludes the paper.

2 Motivation

With the proliferation of information techniques in the society, the Internet has become the main target for cyber attackers driven by the economic benefits. Therefore, the detection of network attacks has become the highest priority for the research community. Anomaly detection is an effective data analysis approach aiming at detecting anomalous or abnormal patterns from the seemingly normal dataset. Equipped with these advantages, anomaly detection has become an interesting research area and been widely applied in a large number of domains such as fraud detection, sensor networks, industrial damage and intrusion detection [14, 15].

In the domain of network intrusion, anomaly detection is an effective and useful tool which can detect the typical network attacks such as PROBE (collecting information about a targeted network for reconnaissance purpose), DOS (denial of service), R2L (remote guess password) and U2R (buffer overflow attack) [16]. Many network intrusion detection systems have been proposed in the literature [17]. However, the conventional anomaly detection techniques will reveal some drawbacks in the controlled network environment because of its special characteristics including the stronger regularity, higher dimensionality and subtler fluctuation of its traffic as mentioned earlier.

Motivated by the consideration, we propose to detect the anomalies in the controlled network based on cross entropy and SVM. Our proposed approach is augmented based on the following roadmap:

(i) Firstly, we are extending the traditional 5-tuple representation and build a novel 7-tuple representative feature vector for capturing the subtle fluctuation underlying in the network traffic. The additional elements include the numbers of the source and destination nodes in the network session along with the traffic rate of different connections. By supplementing these additional elements, the behavioural characteristics of the controlled network can be profiled more accurately.
(ii) Secondly, we construct a novel classifier model based on the 7-tuple cross-entropy feature vector and SVM technique. The classifier can detect the anomalies in the controlled network effectively because the support vector machine technique can show excellent performance when classifying the small-size, high-dimensionality and linearly non-separable samples.

3 Related work

This section discusses the previous research in the field of detecting anomalous traffic. According to the type of analysis theory, we divide these approaches into three categories: the statistics-analysis based approach, the machine-learning based approach, and the signal-processing based approach.

(i) *The statistics-analysis based approach:* Swarnkar and Hubballi [3] presented a method for detecting abnormal behaviour by obtaining the current data packet in the flag field information and calculating the possibility of the packet based on the polynomial Naïve Bayesian classifier for each network packet depth analysis.

Li and Li [4] enhanced the weak classifier based on Naïve Bayesian by introducing the Adaboost iterative algorithm to enhance the training speed and detection accuracy of the network intrusion detection system and reduce the false alarm rate.

Ahirwar *et al.* [5] combined the Naïve Bayesian network with the radial basis function (RBF) neural network to improve the detection accuracy. This technique combined statistics-based Naïve Bayes approach and weighted RBF network approach to determine the network traffic class.

ii) *The machine-learning based approach:* Catania *et al.* [6] presented a method to prepare the training sample-set for the SVM classifier by the filtering function of the SNORT software and improved the accuracy of the classifier greatly.

Ji *et al.* [7] presented an SVM-based predictive model to detect abnormal network behaviours and it was a two-level detection method. The anomaly detection was performed with the rules generated by CART. Then, SVM was applied to a predictive model capable of identifying exact attack types.

Tao and Zhoujin [8] improved the local least squares SVM by selecting and predicting the closer set of training samples to the dataset, reducing the computational complexity of the high-dimensional transpose matrix in the training phase.

iii) *The signal-processing based approach*: Novakov *et al.* [9] presented a novel anomaly detection approach based on a hybrid PCA-Haar wavelet analysis methodology. The hybrid approach used PCA to describe the data and Haar wavelet filtering for analysis. This approach utilised the effectiveness of PCA and wavelet algorithms in detecting network anomalies.

Jiang *et al.* [10] presented a wavelet-based adaptive approach to detect anomalies in network traffic. This approach used wavelet packet transform and continuous wavelet transform to perform the adaptive anomaly detection. The key points included the anomaly characteristics extraction and the further anomaly information obtaining by wavelet packet transformation.

Salagean and Firoiu [11] presented a detection mechanism of network traffic anomaly based on analytical discrete wavelet transform and high-order statistical analysis. The signal processing technique built a set of features based on different metrics to describe the network traffic information and could detect a wide range of anomalies.

The above three typical kinds of approaches have performed satisfactory detection performance to some extent in the usual network environment based on statistical analysis, machine learning and signal processing. But their detection performances will be wakened in the controlled network due to both the representative characteristics of the controlled network and the essential aspects of the detection approaches. Therefore, the aim of the approach proposed in this paper is to solve the above challenges by introducing cross entropy and SVM.

4 Notations and preliminaries

4.1 Notations

In our proposed approach, we define a 7-tuple feature vector for profiling the characteristics of the controlled network traffic.

Therefore, we first give an introduction of the main notations for the 7-tuple feature vector which is illustrated in the Nomenclature section.

4.2 Preliminaries

(i) *The 5-tuple representation of the network traffic:* The 5-tuple representation is the conventional representation of the traffic template for the network traffic, which consists of the most typical and important features [18]. The 5-tuple representation includes S_{ip}, S_{port}, D_{ip}, D_{port} and the transport layer protocol. Because the transport layer protocol is always restricted to TCP/UDP, it is not necessary for us to put the same attention on it like the other four in detecting network anomalies. For profiling the traffic in the controlled network precisely, we will make some extensions based on the conventional 5-tuple representation.

(ii) *Information entropy:* Information entropy [19] is a concept describing how much information there is in an event. In general, the higher the uncertainty degree of the event is, the more information it contains. In other words, we can think that the information means a decrease in uncertainty or entropy. This concept has been applied in the field of malicious traffic detection widely [20]. Information and its relationship to entropy can be modelled by the following formula:

$$R = H(x) - H(x|y) \tag{1}$$

The conditional entropy $H(x|y)$ is called equivocation, which measures the average ambiguity of the received signal. $H(x|y)$ means uncertainty or entropy. $H(x)$ represents the information. R represents the received signal.

(iii) *The Naïve Bayes classifier:* The Naïve Bayes classifier is a kind of statistically probabilistic classifier based on the Bayes' theorem [21]. The classifying process can be simplified greatly by the assumptions that strong independences exist between the features of the samples.

Despite the fact that its strong independence assumption is unrealistic, the Naïve Bayes classifier has performed quite well in many complicated domains which include text categorisation, automatic medical diagnosis, and anomalies detection so on. In some cases, it even outperforms the other machine learning approaches such as boosted trees and random forests [22].

(iv) *The SVM classifier:* The SVM classifier is one typical kind of machine learning technique for classification and regression analysis. Because a set of training examples are needed in the classifying process, it is called the supervised learning model. By introducing the kernel functions, SVM can perform both linear classification and non-linear classification tasks by mapping the input samples into higher dimensional feature spaces [23].

By adopting appropriate kernel functions and mapping the dataset with multi-classes into different feature spaces correspondingly, SVM can solve the multi-class classification problems efficiently [24]. This advantage provides one effective way for us to detect various kinds of malicious traffic in the controlled network.

(v) *The wavelet analysis classifier:* Because of its inherent time–frequency characteristic, the network traffic can be considered as a signal and be decomposed into different components at different frequencies [25]. According to the different ranges of frequency, the traffic is always split into three components: the low-frequency component, the mid-frequency component and the high-frequency component by wavelet transform. In general, the low-frequency component reflects the long-period behaviours of the traffic (a few days for example), the mid-frequency component reflects the daily fluctuations in the traffic, and the high-frequency component reflects the short-term fluctuations. By establishing appropriate thresholds for the wavelet signals, the wavelet analysis technique has been widely applied in the field of anomalous traffic detection [26].

5 Classifier model

In our proposed approach, we construct a multi-class classifier based on cross entropy and SVM. In order to gain a clear understanding of the classifier model, we give a set of definitions of the related concepts in this section.

Definition 1: The 7-tuple representation of the traffic.

The 7-tuple representation is defined to profile the characteristics of the traffic in the controlled network by extending the conventional 5-tuple representation. The 7-tuple vector is described in the following formula:

$$V = \left\{S_{\text{ip}}, S_{\text{port}}, D_{\text{ip}}, D_{\text{port}}, In, Out, Vel\right\} \tag{2}$$

Definition 2: The statistical window.

A statistical window is a time unit (i.e. an interval) in which we select m connections for making a statistic analysis of the 7 tuples.

Definition 3: The probability distribution of the 7 tuples.

The probability distribution of the 7 tuples is calculated as follows: the defined statistical window is first divided into two adjacent sub-windows, then the probability distribution is calculated by summing up its frequency during each sub-window. The probabilities are expressed as

$$\text{Prob}_{i-1} = \frac{d_i}{K}, \quad \text{Prob}_{i-2} = \frac{d'_i}{K'} \tag{3}$$

Here d_i and d'_i are the number of times that the ith element of the 7 tuples occurs in each of the two sub-windows, respectively. $K = \sum_{j=1}^{N} d_j$ and $K' = \sum_{j=1}^{N} d'_j$ $(N = 7)$ are the total number of times that all of the elements of the 7 tuples occur in each of the two sub-windows, respectively. Prob_{i-1} and Prob_{i-2} are the probability distributions of the ith element in the two sub-windows, respectively. If there are m connections in one statistical window, the size of its sub-window is $m/2$.

Definition 4: Cross entropy.

The cross-entropy [27] is defined as follows:

$$L_{0.5}(P, Q) = -2\log\sum_{i=1}^{N}(p_i q_i)^{1/2} \tag{4}$$

Here P and Q are two discrete probability distributions. After we input the distributions from Definition 3 into formula (4), we can calculate the cross entropy of the ith element of the 7 tuples in the statistical window as follows:

$$L_i = -2\log\sum_{j=1}^{N}(p_j q_j)^{1/2} \tag{5}$$

Definition 5: The feature vector of the network traffic.

The feature vector of the network traffic is composed of the cross entropies of the 7 tuples. The feature vector of the kth statistical window is denoted as follows:

$$R_t = \left[L^k_{S_{\text{ip}}}, L^k_{D_{\text{ip}}}, L^k_{S_{\text{port}}}, L^k_{D_{\text{port}}}, L^k_{\text{In}}, L^k_{\text{Out}}, L^k_{\text{Vel}}\right] \tag{6}$$

Here k denotes the current statistical window.

As an example, $L^{\,k}_{S_{\text{ip}}}$ is calculated as follows:

$$\text{(i)} \quad L^{\,k}_{S_{\text{ip}}} = -2\log\sqrt{\left[A^{\text{Before}}_{S_{\text{ip}}}\right]\cdot\left[A^{\text{After}}_{S_{\text{ip}}}\right]^{\text{T}}} \tag{7}$$

Here $A^{\text{Before}}_{S_{\text{ip}}}$ and $A^{\text{After}}_{S_{\text{ip}}}$ denote the probability distributions of the source IP address during the first and second sub-windows of one statistical window, respectively

$$\text{(ii)} \quad A^{\text{Before}}_{S_{\text{ip}}} = \left(S^1_{\text{ip}}, S^2_{\text{ip}}, \ldots, S^n_{\text{ip}}\right) \tag{8}$$

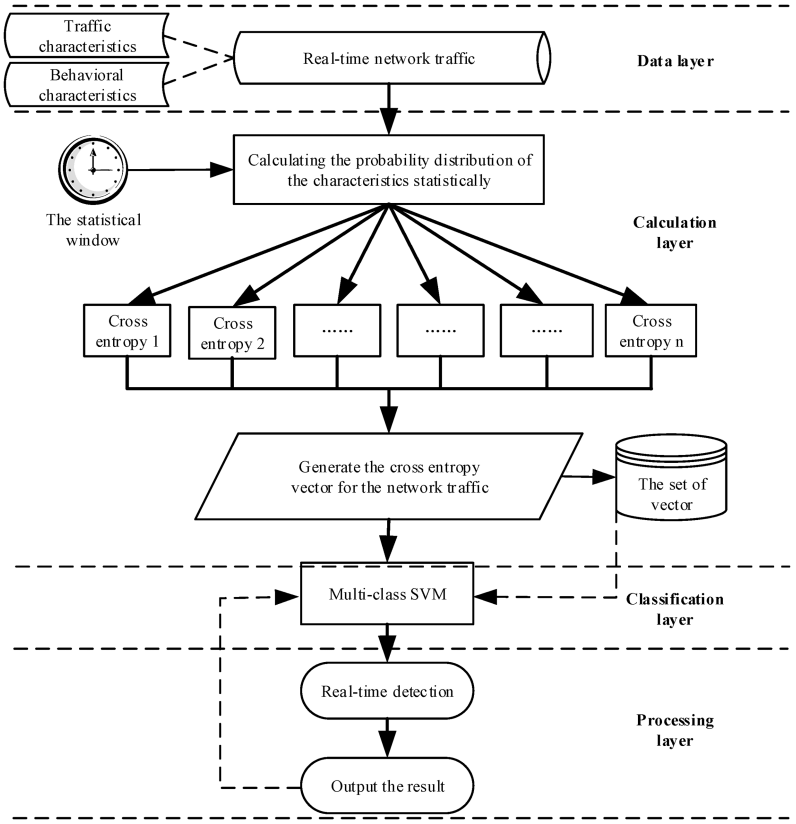

Fig. 1 *Classifier framework based on cross entropy and SVM*

Here n denotes the number of times that different S_{ip} occurs in the first sub-window. S_{ip}^{i} is calculated as follows:

$$S_{ip}^{n} = \frac{\text{Times}_{S_{ip}^{i}}}{\sum_{i=1}^{n} \text{Times}_{S_{ip}^{i}}} \quad (i = 1, \ldots, n) \tag{9}$$

Here $\text{Times}_{S_{ip}^{i}}$ denotes the number of times that the ith S_{ip} occurs in the first sub-window.

iii) $\boldsymbol{A}_{S_{ip}}^{\text{After}}$ is calculated similarly as $\boldsymbol{A}_{S_{ip}}^{\text{Before}}$ for the second sub-window.

iv) If the sizes of the vectors $\boldsymbol{A}_{S_{ip}}^{\text{Before}}$ and $\boldsymbol{A}_{S_{ip}}^{\text{After}}$ are not equal, then the shorter vector is filled to the same length as the longer one with additional zeros.

Definition 6: The multi-class SVM classifier.

In the case of a multi-label classifying problem, we propose a one-map-one training way and construct one label for every two different labels at first. Therefore, we will need $N = M(M-1)/2$ SVM classifiers for M labels.

6 Classifier framework

The classifier framework consists of four main layers: the data layer, the calculation layer, the classification layer, and processing layer from top to bottom as shown in Fig. 1.

(i) The data layer concerns the data source which consists of the traffic characteristics and behavioural characteristics. The layer captures the traffic packets from the network, subtracts and generates the 7-tuple dataset for the calculation layer.

(ii) The calculation layer computes the probability distribution and cross-entropy of the 7 tuples after receiving the 7-tuple dataset from the data layer. Then, the calculation layer generates the traffic feature vectors and supplies the classifier layer with the training set.

(iii) Multi-class classifiers are trained based on the traffic feature vectors from the calculation layer in the classification layer.

(iv) After being trained, the classifier is applied to real-time detection. Furthermore, the parameters of the classifier will be optimised dynamically according to the classification results.

7 Classifying process

7.1 Data pre-processing

In the controlled network environment, the variation of some traffic characteristics is usually subtle. In order to eliminate the influence between different characteristics due to their different attribute metrics, we use a normalised processing method called z-score [28] to transform the values of the traffic feature vector so that the values fall between 0 and 1.

The normalised process is denoted as follows:

$$l'_t = \frac{l_t - \bar{l}}{S}, \quad t = t_1, t_2, \ldots, t_n \tag{10}$$

where $\bar{l} = (1/n)\sum_{t=t_1}^{t_n} l_t$ and $S = \sqrt{(1/(n-1))\sum_{t=t_1}^{t_n}(l_t - \bar{l})}$. n is the number of statistical windows, and l_t denotes the value of the cross-entropy feature vector in a statistical window.

The normalisation process aims to reduce the variance of the cross entropy of the different dimensions of the traffic feature vector and improve the classification accuracy of the SVM.

One example of the feature vector matrix after normalisation processing is shown in the following formula:

$$A = \begin{bmatrix} 0.12 & 0.59 & 0.43 & 0.78 & 0.91 & 0.64 & 0.78 \\ 0.17 & 0.97 & 0.34 & 0.12 & 0.34 & 0.52 & 0.16 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.23 & 0.54 & 0.76 & 0.44 & 0.23 & 0.83 & 0.56 \end{bmatrix} \tag{11}$$

Each row of the matrix represents a traffic feature vector in a statistical window, and each column of the matrix represents the

normalised cross entropy of one feature during the entire statistical period.

7.2 Classification process based on cross entropy and SVM

Algorithm 1 (see Fig. 2) presents the main steps for classification based on cross entropy and SVM.

The classifying process is summarised as follows:

(i) Firstly, the 7-tuple representation should be constructed by subtracting characteristics from the traffic, and then the probability distributions of the 7 tuples are obtained by making a statistical computing.
(ii) Secondly, the cross entropies of the 7 tuples are calculated based on their probability distributions. Then we get the 7-tuple feature vector.
(iii) Thirdly, the 7-tuple feature vectors are normalised and labelled with their corresponding labels. Then the feature vectors are input into the SVMs as the training sample set.
(iv) Fourthly, the SVMs are trained based on the input feature vectors. The number of SVMs depends on the number of labels. We choose $N = M(M-1)/2$ SVMs to be trained in case of M labels.
(v) Finally, the classifier based on cross entropy and SVM is put into real-time detection. The traffic will finally be labelled by using a voting mechanism over the multi-class SVMs.

The key points of the algorithm include generating the 7-tuple feature vector and training the multi-class SVM classifier.

8 Evaluation

In this section, we conducted comprehensive experiments to evaluate the performance of our classifier. The evaluation is performed by 10-fold cross-validation.

8.1 Dataset

In the experiment, we utilise KDDCUPP99 dataset [29] to validate our approach which is a classical dataset used for modelling anomalies detection in the network and appropriate for the controlled network environment in spite of some limitations [30, 31]. Typically, KDDCUPP99 dataset consists of 41 features and can be divided into three categories: basic feature, content feature and time feature [32].

In order to profile the characteristic of the controlled network traffic exactly, we select partial data from the KDDCUPP99 dataset which consists of 574,760 traffic records. Our selected dataset is composed of five types of traffic which include NORMAL, PROBE, DOS, R2L and U2R. Every record is labelled and marked as numbers according to the alphabetic orders of their type names. The composition and proportion of the dataset are as shown in Table 1. As we can see, the composition and proportion of various types of traffics in the dataset can reflect the characteristics of the controlled network, and its data dimensionality can meet the requirement of data validation [33].

1: **Input:** Train Dataset and Test Dataset
2: **Output:** Anomalous or normal status sets of the Test Dataset
3: **Procedure** ClassifierTrain(TrainDataset, SizeOfWindow, Labels, SVMs)
4: $Len \leftarrow TrainDataset.Size$
5: $Num \leftarrow \lceil Len / SizeOfWindow \rceil$
6: //Construct the set of the 7-tuple cross-entropy feature vector
7: **for** $p \leftarrow 1$ to Num do
8: //Compute the probability distributions of the 7 tuples
9: $A_{Slp}^{Before}[p] = Cal\_Slp\_Bef\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
10: $A_{Slp}^{After}[p] = Cal\_Slp\_Aft\left(TrainDataset[(p-\frac{1}{2})\times SizeOfWindow, p\times SizeOfWindow]\right)$
11: $A_{Dlp}^{Before}[p] = Cal\_Dlp\_Bef\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
12: $A_{Dlp}^{After}[p] = Cal\_Dlp\_Aft\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
13: $A_{SPort}^{Before}[p] = Cal\_SPort\_Bef\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
14: $A_{SPort}^{After}[p] = Cal\_SPort\_Aft\left(TrainDataset[(p-\frac{1}{2})\times SizeOfWindow, p\times SizeOfWindow]\right)$
15: $A_{DPort}^{Before}[p] = Cal\_DPort\_Bef\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
16: $A_{DPort}^{After}[p] = Cal\_DPort\_Aft\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
17: $A_{In}^{Before}[p] = Cal\_In\_Bef\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
18: $A_{In}^{After}[p] = Cal\_In\_Aft\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
19: $A_{Out}^{Before}[p] = Cal\_Out\_Bef\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
20: $A_{Out}^{After}[p] = Cal\_Out\_Aft\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
22: $A_{Vel}^{Before}[p] = Cal\_Vel\_Bef\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
23: $A_{Vel}^{After}[p] = Cal\_Vel\_Aft\left(TrainDataset[(p-1)\times SizeOfWindow, (p-\frac{1}{2})\times SizeOfWindow]\right)$
24: //Compute the cross entropies of the 7 tuples
25: $L_{S_{ip}}[p] = -2lb\sqrt{\left[A_{S_{ip}}^{Before}[p]\right]\cdot\left[A_{S_{ip}}^{After}[p]\right]^T}$
26: $L_{D_{ip}}[p] = -2lb\sqrt{\left[A_{D_{ip}}^{Before}[p]\right]\cdot\left[A_{D_{ip}}^{After}[p]\right]^T}$
27: $L_{S_{port}}[p] = -2lb\sqrt{\left[A_{S_{port}}^{Before}[p]\right]\cdot\left[A_{S_{port}}^{After}[p]\right]^T}$
28: $L_{D_{port}}[p] = -2lb\sqrt{\left[A_{D_{port}}^{Before}[p]\right]\cdot\left[A_{D_{port}}^{After}[p]\right]^T}$
29: $L_{In}[p] = -2lb\sqrt{\left[A_{In}^{Before}[p]\right]\cdot\left[A_{In}^{After}[p]\right]^T}$
30: $L_{Out}[p] = -2lb\sqrt{\left[A_{Out}^{Before}[p]\right]\cdot\left[A_{Out}^{After}[p]\right]^T}$
31: $L_{Vel}[p] = -2lb\sqrt{\left[A_{Vel}^{Before}[p]\right]\cdot\left[A_{Vel}^{After}[p]\right]^T}$
32: //Construct the 7-tuple cross-entropy vector
33: $L[p] = \left[L_{S_{ip}}[p], L_{D_{ip}}[p], L_{S_{port}}[p], L_{D_{port}}[p], L_{In}[p], L_{Out}[p], L_{Vel}[p]\right]$
34: **end for**
35: //Train SVM based on the vector set L
35: for $r \leftarrow 1$ to $Labels(Labels-1)/2$ do
36: $SVMs[r] \leftarrow SVM\_Train(L)$
37: end for
38: return $SVMs[]$
39: **end Procedure**
40: **Procedure** ClassifierDetect(Dataset, Status, SVMs[])
41: **for** $i \leftarrow 1$ to $Dataset.Size$ **do**
42: $Status[i] \leftarrow SVMs(Dataset[i]).Voting()$
43: **switch** ($Status[i]$) {
44: case "DDos": return DDos Anomaly; break;
45: case "PROBE": return PROBE Anomaly; break;
46: case "R2L": return R2L Anomaly; break;
47: case "U2R": return U2R Anomaly; break;
48: default: return;
49: }
50: **end switch**
51: **end for**
52: **return** Status[]
53: **end Procedure**

Fig. 2 *Algorithm 1: Classifying process based on cross entropy and SVM*

Table 1 Composition and proportion of the dataset

Type of traffic	Marked label	Number of traffics	Proportion, %
NORMAL	2	558,860	97.23
PROBE	3	560	0.097
DOS	1	11,490	1.999
R2L	4	280	0.049
U2R	5	3570	0.62

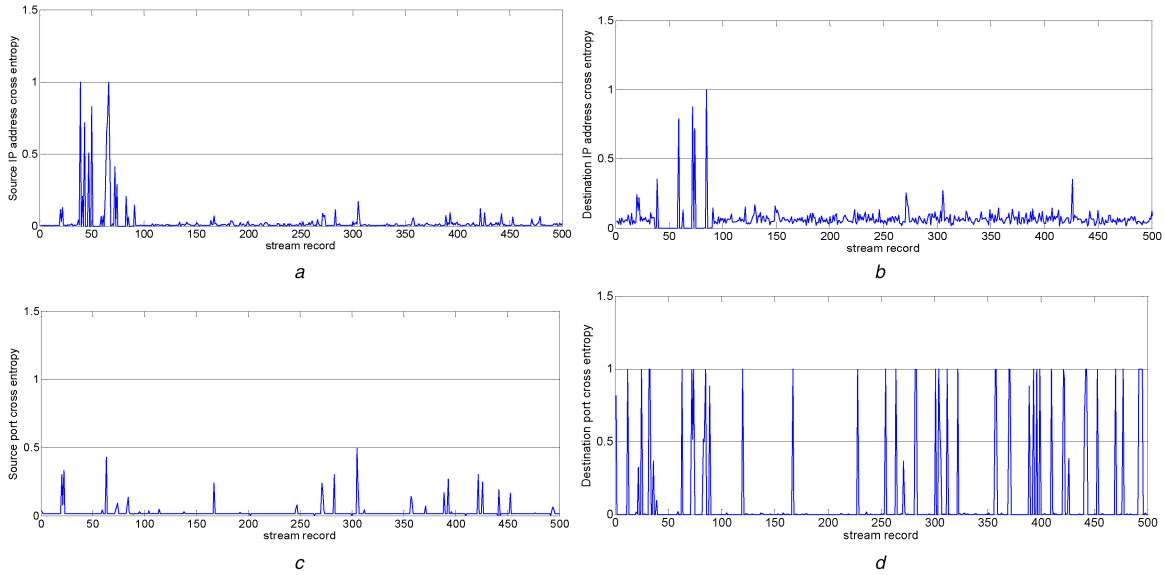

Fig. 3 *Variation of the cross-entropy values*
(a) Cross entropy of S_{ip}, ***(b)*** Cross entropy of D_{ip}, ***(c)*** Cross entropy of S_{port}, ***(d)*** Cross entropy of D_{port}

8.2 Evaluation criterion

To evaluate the experiment results, we define and use the following measurements according to the popular metric methods:

(i) *TPR (true positive rate):* TPR measures the rate of positive instances (i.e. anomalous traffic) which are classified correctly.
(ii) *FPR (false positive rate):* FPR measures the rate of negative instances (i.e. normal traffic) which are classified incorrectly.
(iii) *The detection rate:* The detection rate measures the number of correctly classified instances, either positive or negative, divided by the total number of instances.
(iv) *ROC (receiver operating characteristic) curve:* The ROC curve is a graph produced by plotting the fraction of TPR versus the fraction of FPR for a binary classifier as its discrimination threshold varies.

8.3 Results and discussion

(i) *Comparison between the cross-entropy feature and its original feature of the traffic.* We compute 5747 7-tuple cross-entropy vectors from the dataset as we set the size of the statistical window to 100. The variations of the cross-entropy array of the four features including S_{ip}, S_{port}, D_{ip} and D_{port} are shown in Fig. 3.

As we can see, the cross entropies of the four features approach 0 when the network traffic tends to be normal, and the cross entropy will fluctuate sharply when the anomaly occurs. As shown in Fig. 3*a*, the cross entropy of S_{ip} increases rapidly from 50 to 70 because a DOS attack occurs at that point which makes the numbers of S_{ip} and the value of its cross entropy increase simultaneously. In Fig. 3*b*, the cross entropy of D_{ip} increases rapidly between 70 and 90 because of the occurrence of a PROBE attack during this period.

From Table 1 and Fig. 3, we can conclude that the change of the cross entropy of one traffic feature is more obvious than the change of its original proportion when the anomalous traffic occurs in the controlled network. This gives us a preferable advantage to train the classifier by using the cross-entropy vector instead of the original feature vector as well as taking advantage of the outstanding performances of SVM in case of classifying the small-size, high-dimensionality and linearly non-separable sample-set.

(ii) *Comparison of the detection performance under different-size statistical windows*. The size of the statistical window will play influence to some extent on the detection performance indirectly. If we modify the size of the statistical window, the 7-tuple cross entropy vector of the traffic features will change correspondingly. Then the detection performance may be influenced negatively unless we select the appropriate size for the statistical window. Fig. 4 demonstrates a comparison among three different sizes of the statistical window.

As we can see, the detection performance reaches to the best when the size of the statistical window equals to 100.

(iii) *Detection rate comparison between the 7-tuple feature vector and the traditional 5-tuple feature vector*. In order to evaluate the performance of the additional three elements including In, Out and Vel, we compare the performance of a 7-tuple feature vector against the traditional 5-tuple one. Because the transport layer protocol is always restricted to TCP/UDP, the 5-tuple feature vector actually includes four elements (i.e. S_{ip}, S_{port}, D_{ip} and D_{port}). Table 2 and Fig. 5 illustrate the comparison comprehensively.

Judging from the experimental results between the two different-size tuple feature vectors, the detection rate can be improved by from 2.9 to 11.6%. The detection performance can be improved obviously by adding the three additional elements. Among the four typical attacks, the detection rate of PROBE is increased, most obviously because the three additional elements can reflect the characteristics of PROBE more accurately. On the contrary, the detection rate of DOS did not increase in an apparent way because the 5-tuple feature vector can already reflect the characteristic of DOS largely and the three additional elements do not add more valuable representative information.

(iv) *Comparison with classifiers based on wavelet analysis, Naïve Bayes and SVM*. In order to evaluate the performance of our method, we make a comparison with the other three classifiers based on wavelet analysis, Naïve Bayes and SVM in which wavelet analysis classifier represents the signal-processing based approach, Naïve Bayes represents the statistics-analysis based approach, and SVM represents the machine-learning based approach. The performance comparison of these four classification techniques is shown in Table 3 and Fig. 6.

As we can see from Table 3 and Fig. 6, the wavelet analysis technique, unfortunately, obtains the worst classifying performance because this technique cannot detect the subtle traffic variation underlying in the controlled network environment. The classifying performance of the Naïve Bayes technique is relatively better than the wavelet analysis technique. However, the Naïve Bayes classifier still cannot give a satisfactory performance because of the lack of a sufficient amount of training samples with precise labels and the explicit probability distribution functions of the traffic in the controlled network. Additionally, we evaluated the performance of the SVM classifier with the 7-tuple feature vector, instead of the cross-entropy vector. Its performance is better than the Naïve Bayes approach and the Wavelet analysis approach. In comparison to the above three techniques, the classifier based on cross entropy and SVM shows the best performance because the cross-entropy feature vector can reflect the traffic variation in a more apparent way and the SVM classifier outperforms the others when classifying small-size, high-dimensionality and linearly non-separable samples in the controlled network. However, the detection rate of R2L and PROBE is not as satisfactory as the other types of anomalies, mainly due to their smaller proportions in the dataset.

(v) *Summary of the detection approach based on cross entropy and SVM*. Judging from the above experiment results, the approach proposed by this paper is more efficient for detecting anomalous traffic in the controlled network. The characteristics of the approach can be summarised as follows:

(a) Compared to the statistics-analysis based approach and the signal-processing based approach, the new approach can discover and profile the subtle fluctuation underlying the traffic in the controlled network in a more apparent way and more effectively.
(b) Compared to the machine-learning based approach, it can classify small-size, high-dimensionality and linearly non-separable samples set in the controlled network more precisely.
(c) Compared to the aforementioned three detection approaches, the new approach is more suitable for the controlled network due to its unique characteristics (strong regularity, high dimensionality, and subtle fluctuation).

9 Conclusion

Because of the regularity, high dimensionality and subtle fluctuation of the traffic in the controlled network, the typical anomaly detection techniques cannot discover the subtle variations in the traffic and detect the anomalies effectively. On this motivation, we propose a novel classifier based on cross entropy and SVM. Compared to the typical representation of the traffic characteristic, the 7-tuple cross-entropy feature vector built on the basis of the typical 5-tuple representation can reflect the variation underlying the controlled network traffic in a more apparent way. Additionally, the SVM classifier can shed more light on the subtle variations in the traffic based on the cross-entropy input and its advantage in case of classifying the small-size, high-dimensionality and linearly non-separable sample set in the controlled network.

Currently, we are carrying out deeper research to optimise the process, in order to increase the accuracy and efficiency for detecting unknown anomalous traffic. Furthermore, we will put more emphasis on strengthening the classification algorithm in our future work.

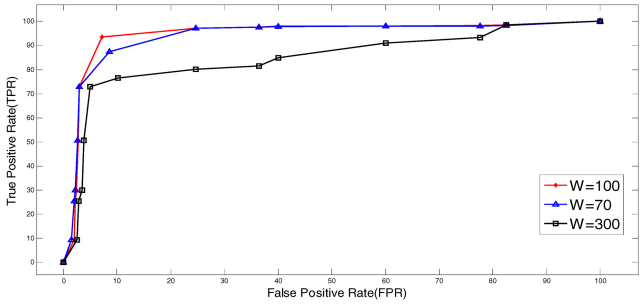

Fig. 4 *ROC curve of various statistical windows*

Table 2 Performance comparison between 7-tuple and 5-tuple feature vectors

Anomaly	Feature Detection rate		
	5-tuple, %	7-tuple, %	Increase, %
PROBE	74.1	85.7	11.6
DOS	93.3	96.2	2.9
R2L	63.7	72.9	9.2
U2R	85.8	92.0	6.2

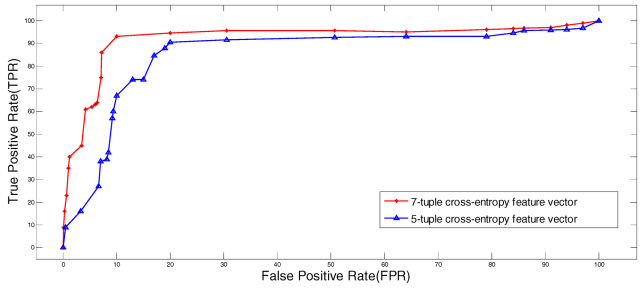

Fig. 5 *Performance comparison between the 7-tuple and the 5-tuple feature vectors*

Table 3 Detection rate of the four classifiers

Anomaly	Classifier Detection rate			
	Wavelet, %	Naïve Bayes, %	SVM, %	CE-SVM, %
PROBE	68.6	78.5	81.5	85.7
DOS	76.8	89.2	92.6	96.2
R2L	59.4	64.8	66.2	72.9
U2R	79.6	83.9	85.3	92.0

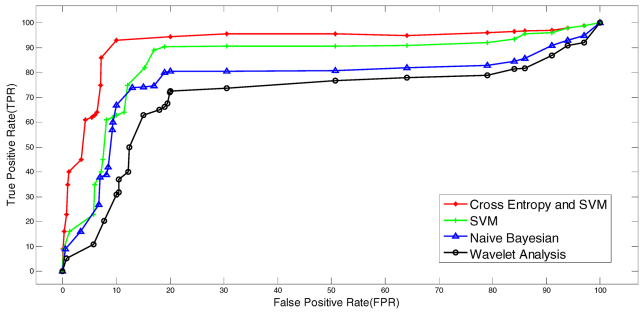

Fig. 6 *Comparison of the four classifiers*

10 Acknowledgments

This work was supported by the Key Technology Research Project of the General Armaments Department (grant no. 2015ZD004025) and the National Key Research & Development Program 2016 (2016YFB0801304).

11 References

[1] Ahmed, M., Mahmood, A.N., Hu, J.: 'A survey of network anomaly detection techniques', *J. Netw. Comput. Appl.*, 2016, **60**, pp. 19–31

[2] Kaur, R., Nagpal, E.S., Chamotra, S.: 'Malicious traffic detection in a private organizational network using honeynet system'. Annual IEEE India Conf., New Delhi, India, December 2015, pp. 1–6

[3] Swarnkar, M., Hubballi, N.: 'OCPAD: one class naive Bayes classifier for payload based anomaly detection', *Expert Syst. Appl.*, 2016, **64**, pp. 330–339

[4] Li, W., Li, Q.X.: 'Using naive Bayes with AdaBoost to enhance network anomaly intrusion detection'. IEEE Int. Conf. on Intelligent Networks & Intelligent Systems, Shenyang, China, November 2010, pp. 486–489

[5] Ahirwar, D.K., Saxena, S.K., Sisodia, M.S.: 'Anomaly detection by Naive Bayes & RBF network', *Int. J. Adv. Res. Comput. Sci. Electron. Eng.*, 2012, **1**, (1), pp. 14–18

[6] Catania, C.A., Bromberg, F., Garino, , *et al.*: 'An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection', *Expert Syst. Appl.*, 2010, **39**, (2), pp. 1822–1829

[7] Ji, S.-Y., Choi, S., Jeong, D.H.: 'Designing a two-level monitoring method to detect network abnormal behaviors'. IEEE Int. Conf. on Information Reuse and Integration, Redwood City, USA, August 2014, pp. 703–709

[8] Tao, P., Zhoujin, T.: 'A small scale forecasting algorithm for network traffic based on relevant local least squares support vector machine regression model', *Appl. Math. Inf. Sci.*, 2015, **9**, (2), pp. 653–659

[9] Novakov, S., Lung, C.-H., Lambadaris, I., *et al.*: 'Studies in applying PCA and wavelet algorithms for network traffic anomaly detection'. 14th IEEE Int. Conf. on High Performance Switching and Routing, Taipei, China, July 2013, pp. 185–190

[10] Jiang, D., Zhang, P., Xu, Z., *et al.*: 'A wavelet-based detection approach to traffic anomalies'. 7th IEEE Int. Conf. on Computational Intelligence and Security, Hainan, China, December 2011, pp. 993–997

[11] Salagean, M., Firoiu, I.: 'Anomaly detection of network traffic based on analytical discrete wavelet transform'. 8th IEEE Int. Conf. on Communications, Bucharest, Romanian, June 2010, pp. 49–52

[12] Yan, R., Zheng, Q.: 'Using cross entropy to detect and classify network anomalous traffic', *J. Xi'an Jiaotong Univ.*, 2010, **44**, (6), pp. 10–15

[13] Jing, N., Yang, M., Cheng, S., *et al.*: 'An efficient SVM-based method for multi-class network traffic classification'. 30th IEEE Int. Conf. on Performance Computing and Communications, Orlando, USA, November 2011, pp. 1–8

[14] Ahmed, M., Anwar, A., Mahmood, A.N., *et al.*: 'An investigation of performance analysis of anomaly detection techniques for big data in SCADA systems', *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, 2015, **15**, (3), pp. 1–16

[15] Mahmood, A.N., Hu, J., Tari, Z., *et al.*: 'Critical infrastructure protection: resource efficient sampling to improve detection of less frequent patterns in network traffic', *J. Netw. Comput. Appl.*, 2010, **33**, (4), pp. 491–502

[16] Yao, H., Liu, Y., Fang, C.: 'An abnormal network traffic detection algorithm based on big data analysis', *Int. J. Comput. Commun. Control*, 2016, **11**, (1), pp. 567–579

[17] Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K.: 'Network anomaly detection: methods, systems and tools', *IEEE Commun. Surv. Tutor.*, 2014, **16**, (1), pp. 303–336

[18] Peng, L., Zhang, H., Yang, B., *et al.*: 'Traffic labeller: collecting internet traffic samples with accurate application information', *China Commun.*, 2014, **11**, (1), pp. 69–78

[19] Shannon, C.E.: 'A mathematical theory of communication', *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, 2001, **5**, (1), pp. 3–55

[20] Berezinski, P., Jasiul, B., Szpyrka, M.: 'An entropy-based network anomaly detection method', *Entropy*, 2015, **17**, (4), pp. 2367–2408

[21] Langley, P., Iba, W., Thompson, K.: 'An analysis of Bayesian classifiers'. 10th National Conf. on Artificial Intelligence, San Jose, USA, July 1992, pp. 223–228

[22] Rich, C., Alexandru, N.-M.: 'An empirical comparison of supervised learning algorithms'. 23rd ACM Int. Conf. on Machine Learning, Pittsburgh, USA, June 2006, pp. 161–168

[23] Cortes, C., Vannik, V.: 'Support vector networks', *Mach. Learn.*, 1995, **20**, (3), pp. 273–297

[24] Liu, B., Xiao, Y., Cao, L.: 'SVM-based multi-state-mapping approach for multi-class classification', *Knowl.-Based Syst.*, 2017, **129**, pp. 79–96

[25] Barford, P., Kline, J., Plonka, D., *et al.*: 'A signal analysis of network traffic anomalies'. 2nd ACM SIGCOMM Workshop on Internet Measurement, Marseille, France, November 2002, pp. 71–82

[26] Callegari, C., Giordano, S., Pagano, M., *et al.*: 'WAVE-CUSUM: improving CUSUM performance in network anomaly detection by means of wavelet analysis', *Comput. Secur.*, 2012, **31**, (5), pp. 727–735

[27] Nobre, R.H., Rodrigues, F.A.A., Marques, R.C.P., *et al.*: 'SAR image segmentation with Renyi's entropy', *IEEE Signal Process. Lett.*, 2016, **23**, (11), pp. 1551–1555

[28] Ko, Y.C., Fujita, H., Li, T.A.R.: 'An evidential analysis of Altman Z-score for financial predictions: case study on solar energy companies', *Appl. Soft Comput.*, 2017, **52**, pp. 748–759

[29] 'KDD Cup 1999'. Available at http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, accessed 28 December 2016

[30] Sommer, R., Paxson, V.: 'Outside the closed world: on using machine learning for network intrusion detection'. IEEE Symp. on Security and Privacy, Berkeley/Oakland, USA, May 2010, pp. 305–316

[31] Creech, G., Hu, J.: 'Generation of a new IDS test dataset: time to retire the KDD collection'. IEEE Wireless Communications and Networking Conf.: Services & Applications, Shanghai, China, April 2013, pp. 4487–4492

[32] Tavallaee, M., Bagheri, E., Lu, W., *et al.*: 'A detailed analysis of the KDD CUP 99 data set'. IEEE Symp. on Computational Intelligence for Security and Defense Applications, Ottawa, Canada, July 2009, pp. 53–58

[33] Diana, N.E., Sabiq, A.: 'Cognitive-affective emotion classification: comparing features extraction algorithm classified by multi-class support vector machine', *Int. J. Comput. Commun. Eng.*, 2016, **5**, (5), pp. 350–357