OPEN ACCESS

1st International Electronic
Conference on Entropy and
Its Applications
3 - 21 November 2014

www.sciforum.net/conference/ecea-1

Conference Proceedings Paper – Entropy

Detection and Classification of Anomalies in Network Traffic Using Generalized Entropies and OC-SVM with Mahalanobis Kernel

Jayro Santiago-Paz [1,⋆], Deni Torres-Roman [1], Angel Figueroa-Ypiña[1]

[1] CINVESTAV, Campus Guadalajara, Av. del Bosque 1145, Col. El Bajio, Zapopan, Mexico

* Author to whom correspondence should be addressed; jsantiago@gdl.cinvestav.mx, +52(33) 3777 3600, +52(33) 3777 3609.

Received: 12 September 2014 / Accepted: 26 October 2014 / Published: 3 November 2014

Abstract: Network anomaly detection and classification is an important open issue of network security. Several approaches and systems based on different mathematical tools have been studied and developed. Among them, the Anomaly-Network Intrusion Detection System (A-NIDS), this monitors network traffic and compares it against an established baseline of “normal” traffic profile. Then, it is necessary to characterize the “normal” Internet traffic. This paper presents an approach for anomaly detection and classification based on: the entropy of selected features (including Shannon, Renyi and Tsallis entropies), the construction of regions from entropy data employing the Mahalanobis distance (MD), and One Class Support Vector Machine (OC-SVM) with different kernels (RBF and particularity Mahalanobis) for “normal” and abnormal traffic. Regular and non-regular regions built from “normal” traffic profiles, allow the anomaly detection; whilst the classification is performed under the assumption that regions corresponding to the attack classes have been characterized previously. Although, this approach allows the use of as many features as required, only four well known significant features were selected in our case. To evaluate our approach two different data sets were used: one set of real traffic obtained from an Academic LAN, and the other a subset of the 1998 MIT-DARPA set. The selected features sets computed in our experiments provide detection rates up to 99.90% with “normal” traffic and up to 99.83% with anomalous traffic and false alarm rate of 0.086%. Experimental results show that certain values of the q parameter of the generalized entropies and the use of OC-SVM improves the detection rate of some attack classes, due to a better fit of the region to the data. Besides, our results show that MD allows to obtain high detection rates with an efficient computation time, while OC-SVM achieved detection rates slightly higher but more expensive computationally.

Keywords: Generalized entropies; Network traffic; Anomaly detection; OC-SVM; Mahalanobis kernel, Mahalanobis distance

PACS classifications: one, two, three

1. Introduction

The detection and prevention of attacks and malicious activities have led to the development of technologies and devices, designed to provide a certain degree of security. One of first technologies to counter attacks launched against computer networks were the Network Intrusion Detection Systems (NIDS). NIDS are classified into two groups: Signature-NIDS (S-NIDS use a database with attack signatures) and Anomaly-NIDS (A-NIDS use the principle of classify the traffic in normal and abnormal to decide if an attack has occurred).

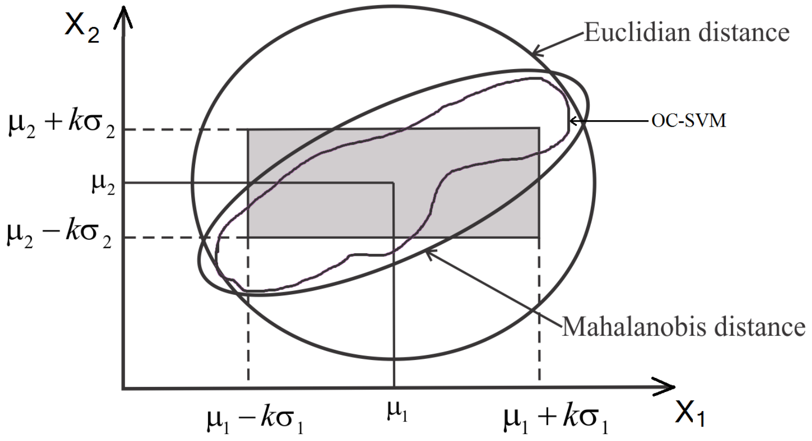
A-NIDS, also known in the literature as behavioral-based, apply various processes modeling in order to detect security events. They try to establish what a "normal profile" or anomaly-free profile for system or network behavior is, using the network features or variables e. g: destination and source IP Addresses and Port, packet size, number of flows, and amount of packets, between hosts to identify deviations from a "normal" behavior.

For anomaly detection [1] can be employed some traffic variables directly or functions of these variables, e.g, the entropy. Entropy-based approaches for anomaly detection are appealing since they provide more information about the structure of anomalies than traditional traffic volume analysis [2]. The entropy is used to capture the degree of dispersal or concentration of the distributions for different traffic features, see for example [3], [4]. The attractiveness of entropy metrics stems from their capability of condensing an entire feature distribution into a single number and at the same time retaining important information about the overall state of the distribution. A sequence of packets from network traffic is captured, network features are selected and the entropy of these network features are calculated. With the estimated values of entropy the anomalies detection is performed, for this, a profile with "normal" traffic is generated, the data that deviate from this profile will be considered anomalies. In the work [5], by constructing a matrix $\boldsymbol{H}$ which contains entropy estimates of the training trace ψ, an ellipsoidal region through the Mahalanobis distance was defined, where the greatest distance Mahalanobis found was used to generate the ellipsoidal regions. However, this method requires training traces which have been previously refined because the exclusion of outliers is not performed, and therefore, the generated region can not properly represent the "normal" profile.

An improvement to the previous work [5] was proposed in [6] where the proposed algorithm uses the Mahalanobis distance to the exclusion of outliers, and an ellipsoidal regions were generated by calculating the parameters $\{\bar{\mathrm{x}}, \gamma, \boldsymbol{\lambda}, LT\}$, where $\bar{\mathrm{x}}$ is the mean vector of the matrix $\boldsymbol{H}$, γ, $\boldsymbol{\lambda}$ are the eigenvectors and eigenvalues of the covariance matrix of $\boldsymbol{H}$, and LT is the limit of Mahalanobis distance for $\boldsymbol{H}$ [7]. In both works, network traffic behavior was characterized by regular ellipsoidal regions.

This paper proposes to define non-regular regions from training traces, i.e. "normal" traffic, through OC-SVM which contains parameters that adjust the region to the training traces. In other works, for example [9] and [10], was used the RBF kernel. But, this work proposes to use the Mahalanobis kernel, which in general showed higher detection rates than others. Figure 1 shows different defined regions for the case of two variables.

Figure 1. Different regions based on different methods and metrics.



The paper is organized as follows: section 2 introduces the theoretical background including different entropy estimators, distance metrics, and OC-SVM. Section 3 formulates the problems and the proposed methods associated with the definition of a region in the space $\mathbb{R}^p$ that characterizes the entropy behavior of the p intrinsic variables associated with the trace ψ are presented. In Section 4 the experiments to define regions and detect and classify anomalies, employing two different types of datasets are presented. Finally, in section 6, the conclusions of this paper are outlined.

2. Mathematical Background

2.1. Entropy estimators

Let X be a r.v. that take values of the set $\{x_1, x_2, ..., x_M\}$, $p_i := P(X = x_i)$ the probability of occurrence of x_i, and M the cardinality of the finite set, hence the Shannon entropy is:

$$\hat{H}^S(P) = -\sum_{i=1}^{M} p_i \; logp_i. \tag{1}$$

Based on the Shannon entropy [11], Rényi [12] and Tsallis [13] defined generalized entropies, which are related with the q-deformed algebra

$$\hat{H}^R(P,q) = \frac{1}{1-q} log(\sum_{i=1}^{M} p_i^q) \tag{2}$$

and

$$\hat{H}^T(P,q) = \frac{1}{q-1}(1 - \sum_{i=1}^{M} p_i^q), \tag{3}$$

where P is a probability distribution and the parameter q is used to make less or more sensitive the entropy to certain events within the distribution. When $q \to 1$ the generalized entropies are reduced to Shannon entropy. In order to compare the changes of entropy at different times, the entropy were normalized.

2.2. Feature Space

Let $X_t^i,\ i = 1, 2, ..., p$ be features or random variables of some phenomenon under study. When the phenomenon is observed during a time period T, N observations are collected. These observations can be studied one by one or by group. In our case, the N observations are partitioned into m sequences or windows of length L.

For each sequence or time windows a functional $f(\bullet)$ is applied. As our purpose is the study of network traffic and the randomness of the features, we will employ the entropy as $f(\bullet)$.

Let $X_j,\ \ j = 1, 2, ..., N \in \mathbb{R}^p$ be the vectors associated at p features, and $H_i,\ \ i = 1, .., m,$ the entropies associated at X_j in each sequence. So, we have $\boldsymbol{X}_{N\times p}$ a matrix representing the observations and $\boldsymbol{H}_{m\times p}$ the matrix of the entropy of the m sequences.

$$\boldsymbol{X}_{N\times p} = \begin{pmatrix} X_1^1 & X_1^2 & \cdots & X_1^p \\ X_2^1 & X_2^2 & \cdots & X_2^p \\ \vdots & \vdots & \vdots & \vdots \\ X_N^1 & X_N^2 & \cdots & X_N^p \end{pmatrix} \overset{f(\bullet)}{\rightarrow} \boldsymbol{H}_{m\times p} = \begin{pmatrix} \bar{H}(X_1^1) & \bar{H}(X_1^2) & \cdots & \bar{H}(X_1^p) \\ \bar{H}(X_2^1) & \bar{H}(X_2^2) & \cdots & \bar{H}(X_2^p) \\ \vdots & \vdots & \vdots & \vdots \\ \bar{H}(X_m^1) & \bar{H}(X_m^2) & \cdots & \bar{H}(X_m^p) \end{pmatrix} \tag{4}$$

A row of the matrix $\boldsymbol{H}_{m\times p}$ represents a point in the $p-$dimensional feature space and the m points generates a cloud of points, they characterizes the behavior of p variables of the phenomenon under study. The estimations of entropy are normalized, $\bar{H}(X_i^p) \in [0, 1]$, to perform a comparison between the variables.

2.3. Mahalanobis distance

The Mahalanobis distance is defined as [14]: $d^2 = (\mathbf{x} - \mu)^T C^{-1}(\mathbf{x} - \mu)$, where $\mathbf{x} \in \mathbb{R}^p$ is the sample vector, $\mu \in \mathbb{R}^p$ denote the theoretical mean vector, and $C \in \mathbb{R}^{p\times p}$ denote the theoretical covariance matrix.

An unbiased sample covariance matrix is

$$\mathrm{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})', \tag{5}$$

where the sample mean is

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i. \tag{6}$$

Thus, Mahalanobis distance using (6) and (5) is given by:

$$d^2 = (\mathbf{x} - \bar{\mathbf{x}})^T \boldsymbol{S}^{-\mathbf{1}}(\mathbf{x} - \bar{\mathbf{x}}). \tag{7}$$

2.4. One Class Support Vector Machine and Mahalanobis Kernel

OC-SVM maps input data $\mathbf{x}_1, ..., \mathbf{x}_N \in A$ (a certain set) into a high dimensional space F (via Kernel $k(\mathbf{x}, \mathbf{y})$) and finds the maximal margin hyperplane which best separates the training data from the origin. Theoretical fundamentals of SVM and OC-SVM were established in [16], [17], [18]. To separate the data from the origin, the following quadratic program must be solved [15]

$$\min_{w \in F, b \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}^N} \frac{1}{2} \|w\|^2 + \frac{1}{\nu N} \sum_{i}^{N} \xi_i - b \tag{8}$$

subject to $(w \cdot \varphi(\mathbf{x}_i)) \geq b - \xi_i;\ \xi_i \geq 0,\ \nu \in (0, 1]$, where w is the normal vector, φ be a map function $A \to F$, b is the bias, ξ_i are nonzero slack variables, ν is the outliers parameter control and $k(\mathbf{x}, \mathbf{y}) = (\varphi(\mathbf{x}), \varphi(\mathbf{y}))$. Besides, the decision function is given by $f(\mathbf{x}) = sgn\left((w \cdot \varphi(\mathbf{x}_i)) - b\right).$

By applying the kernel function and Lagrangian multiplier (α_i) to the original quadratic program, the solution of Eq.(8) creates a decision function:

$$f(\mathbf{x}) = sgn\left(\sum_{i}^{N} \alpha_i k(\mathbf{x}_i, \mathbf{x}) - b\right), \tag{9}$$

where $w = \sum_i \alpha_i \varphi(\mathbf{x}_i)$ and $\sum_i \alpha_i = 1$.

In this work we used the Mahalanobis kernel (MK) that is defined as: $K(\mathbf{x}, \mathbf{y}) = exp(-(\mathbf{x}-\mathbf{y})' C(\mathbf{x}-\mathbf{y}))$, where $\mathbf{C}$ is a positive definite matrix. The Mahalanobis kernel is an extension of the Radial Basis Function kernel (RBF). Namely, by setting $\mathbf{C} = \eta \mathbf{I}$, where $\eta > 0$ is a parameter for decision boundary control and $\mathbf{I}$ is the unit matrix, we obtain the RBF kernel: $exp(-\eta \|\mathbf{x} - \mathbf{y}\|^2)$. The Mahalanobis kernel approximation [20] is:

$$k(\mathbf{x}, \mathbf{y}) = exp(-\frac{\eta}{p}(\mathbf{x} - \mathbf{y})' \mathbf{S}^{-1}(\mathbf{x} - \mathbf{y})), \tag{10}$$

where p is the number of variables, and $\mathbf{S}$ is defined in (5).

3. Problem Statement

Let ψ be an Internet traffic data trace and p the number of random variables X_i representing the traffic features. Using entropy of these traffic features we can find a region that characterize the feature behavior of the trace in the feature space. If ψ was obtained during "normal" network behavior, i.e. it is free of anomalies, this region R_N will serve to detect anomalies, since deviations from "normal" behavior occur at one or more variables when an anomaly is present, thus any behavior that is out of this region may represent an anomaly. On the other hand, if ψ was captured while network attack occurred, the defined region R_A characterizes the anomaly. Our goal is to build a well-defined region with a given set of observations. So, the problem is to find a region in the feature space $\mathbb{R}^p$ that characterizes the behavior of the entropy of the p intrinsic variables associated with a class characterized by the trace ψ.

Our approach for define the “normal” R_N or abnormal region R_A in the space is to use Mahalanobis distance to define regular regions (i.e. hyperellipsoids) and OC-SVM which allows finding a non-regular region based on the support vectors.

3.1. Algorithm for the construction of decision regions

The method comprises three stages: training, detection and classification. Training stage: different regions in the feature space are defined using the Internet traffic data trace; The outliers exclusion is performed with the definition of the α parameter like in [6]). Detection stage: the anomaly detection using the “normal” regions R_N and decision functions generated previously is accomplished. Finally, the classification of known anomalies through pre-defined regions R_A is performed.

3.1.1. Training stage

The data trace ψ is divided into m non-overlapping slots of L number of packets. Normalized entropy estimates of each one p variable in every $i-$slot of size L is obtained through the relative frequency $\hat{p_i} = \frac{n_i}{L}$, where n_i is the number of times that the element x_i appears in the $i-$slot using 2 and 3 , then the matrix $\boldsymbol{H} \in \mathbb{R}^{m \times p}$ is built

$$\boldsymbol{H}_{m \times p} = \begin{pmatrix} \bar{H}(X_1^1) & \bar{H}(X_1^2) & \cdots & \bar{H}(X_1^p) \\ \bar{H}(X_2^1) & \bar{H}(X_2^2) & \cdots & \bar{H}(X_2^p) \\ \vdots & \vdots & \vdots & \vdots \\ \bar{H}(X_m^1) & \bar{H}(X_m^2) & \cdots & \bar{H}(X_m^p) \end{pmatrix}, \tag{11}$$

where $\bar{H}(X_i^p)$ represents the normalized entropy estimation of the p variable of each $i-$slot obtained from ψ. Different regions can be defined through the matrix $\boldsymbol{H}$, in this paper, Mahalanobis distance and OC-SVM are used.

Construction of regions based on the Mahalanobis distance method (MD):

1. To perform the exclusion of outliers the limit for Mahalanobis distance LT is calculated through $LT = (\frac{(m-1)^2}{m})\beta_{[\alpha,p/2,(m-p-1)/2]}$ [7], [8], where $\beta_{[\alpha,p/2,(m-p-1)/2]}$ represents a beta distribution with level of confidence α and parameters $p/2$ and $(m-p-1)/2$, m is the number of rows and p the number of columns of matrix $\boldsymbol{H}$.

2. The mean vector $\bar{\mathbf{x}} = \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_p\}$; where the $i-$element is the mean of the $i-$column of matrix $\boldsymbol{H}$, see eq.(6).

3. In [6] was pointed out that regular regions based on the Mahalanobis distance achieve a better fit than those based on euclidean-distance, so the covariance matrix S of matrix $\boldsymbol{H}$ is calculated by (5). As the matrix S is positive definite and Hermitian, all of its eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ are real and positive, and its eigenvectors $\gamma_1, \gamma_2, ..., \gamma_p$ form a set of orthogonal basis vectors that span the $p-$dimensional vector space.

4. The matrix equation $\mathbf{S}\boldsymbol{\gamma} = \boldsymbol{\lambda}\boldsymbol{\gamma}$ is solved according to an specific or known algorithm to obtain the eigenvalues λ_i and eigenvectors γ_i of $\mathbf{S}$.

5. Finally, we can define an hyperellipsoidal region R_N based on MD, which characterizes the matrix $\boldsymbol{H}$ by means of $\{LT, \bar{\mathrm{x}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}\}$.

Construction of regions based on the OC-SVM method:

1. The equation (8) is solved using two different kernel functions (Radial Basis Function (RBF) and Mahalanobis kernel(MK)) by the sequential Minimal Optimization Algorithm (SMO) [19], considering as input data the matrix obtained in the previous step.
2. The vector obtained in the previous step is: $\{x_i = sv_i, \alpha_i, b\}$, where $x_i = sv_i$ is the i-support vector, α_i, b are constants that solve the equation (8).

3.1.2. Detection stage

1. In the current traffic an $i-$slot of size L packets is captured, the p features or variables associated to every packet are extracted and their entropies estimated, with these values the input vector $\mathbf{h_i}$ for decision functions is built:

$$\mathbf{h_i} = \left(\bar{H}(X_i^1), \bar{H}(X_i^2), \cdots, \bar{H}(X_i^p)\right). \tag{12}$$

2. The decision function for MD region is given by (7), if $d_i^2 \leq LT$ then $i-$slot is considered “normal” otherwise is a potential anomaly.
3. The decision function for OC-SVM is (9), if the function is $+1$ then $\mathbf{h_i}$ is considered “normal” otherwise is a potential anomaly.

3.1.3. Classification stage

Regions R_A for anomaly traffic are built employing the same algorithm described in the training stage. If the vector (12) is out of the “normal” region, i.e $\mathbf{h_i} \notin R_N$, but $\mathbf{h_i} \in R_A$ the abnormal behavior, then it will be classified. Here $\mathbf{h_i}$ is evaluated with all decision functions defined in the training stage. The classification is refined using the k-nearest neighbors algorithm to insure that a point belongs to a specific class.

4. Experiments and results

4.1. Data sets

We evaluate our approach by analyzing its performance over two different experimental databases. The first one is from an Academic LAN [21], and it is composed of traffic information collected during seven days. A trace contain “normal” traffic (β_1) and four traces are formed with traffic considered “normal” plus traffic generated by four real attacks: port scan (ψ_1), and the worms: Blaster (ψ_2), Sasser (ψ_3) and Welchia (ψ_4). The second one is a sub-set of the 1998 MIT-DARPA [22] (public set benchmark for testing NIDS), and it is composed of one training trace (β_2) that was collected during five days of "normal" behavior of the network and four traces contains the traffic generated by Smurf (ψ_5), Neptune (ψ_6), Pod (ψ_7) and portsweep (ψ_8) attacks.

4.2. Detection of anomalies in network traffic

As it was pointed out, anomaly-free traces were divided into m non-overlapping slots of size L (in our case $L = 32$) packets. This size was chosen because the duration of some attacks in the test traces were around to 30 packets, assuring at least one slot with malicious traffic. The attacks that are used in this work (scans and denial of service) generate deviations from the typical behavior of the IP and Ports addresses, respectively. So, they were defined for this experiment $p = 4$ random variables $X^r,\ r = 1, ..., 4$; that represent four traffic features: X^1 source IP addresses, X^2 destination IP addresses, X^3 source port addresses and X^4 destination port addresses. In each i-slot the normalized entropy estimates from each p variable were obtained and the vectors $\mathbf{h}_{\mathbf{X}^\mathbf{p}} = \left(\bar{H}(X_1^p), \bar{H}(X_2^p), ..., \bar{H}(X_m^p)\right)$ were constructed. The values of q parameter of the generalized entropies used in the experiments are $\{0.01, 0.1, 0.5, 1.5, 2, 10, 20\}$.

Next, five matrices were formed: $\boldsymbol{H}_{Ip} \in \mathbb{R}^{m\times 2}$ containing the entropy estimates of source and destination IP addresses, $\boldsymbol{H}_{Pt} \in \mathbb{R}^{m\times 2}$ containing estimates of entropy of the source and destination port addresses, $\boldsymbol{H}_{IpDPt} \in \mathbb{R}^{m\times 3}$ containing estimates of entropy of the source and destination IP addresses and entropy estimates of destination port address, $\boldsymbol{H}_{IpSPt} \in \mathbb{R}^{m\times 3}$ containing estimates of entropy of the source and destination IP addresses and entropy estimates of source port address, finally $\boldsymbol{H}_{IpPt} \in \mathbb{R}^{m\times 4}$ where each column contains estimates of entropy of each variable considered in this paper. For example

$$\boldsymbol{H}_{IpPt} = (\mathbf{h}_{\mathbf{X}^1}{}' \ \mathbf{h}_{\mathbf{X}^2}{}' \ \mathbf{h}_{\mathbf{X}^3}{}' \ \mathbf{h}_{\mathbf{X}^4}{}') \tag{13}$$

The ellipsoidal regions in the feature space were obtained with these matrices, non-regular regions were found through Oc-SVM with Radial Basis Function (RBF) and Mahalanobis Kernel (MK). Parameters η and ν of OC-SVM were selected in the 5-fold cross-validation process. For implementation of OC-SVM the LIBSVM library [23] was used. The found regions are used to detect anomalies in network traffic. Therefore, traces containing traffic generated by different anomalies were used. Each test trace was divided into slots of size $L = 32$ packets and the estimates of entropy for each variable considered in this work were obtained. To each i-slot the Mahalanobis distance was computed by (7). Likewise, each i-slot was analyzed with OC-SVM decision function (9) for both kernels used in this paper, and thus it was determined whether it belongs to the non-regular region or not.

Results for anomaly detection of the LAN and MIT-DARPA traces using Tsallis entropy with $q = 0.01$ and the features considered in this work through the ellipsoidal and non-regular regions are displayed in Table 1. The detection rate for the attack ψ_6 is 0 or 100, because it is contained in only one slot.

4.3. Classification of worms attack

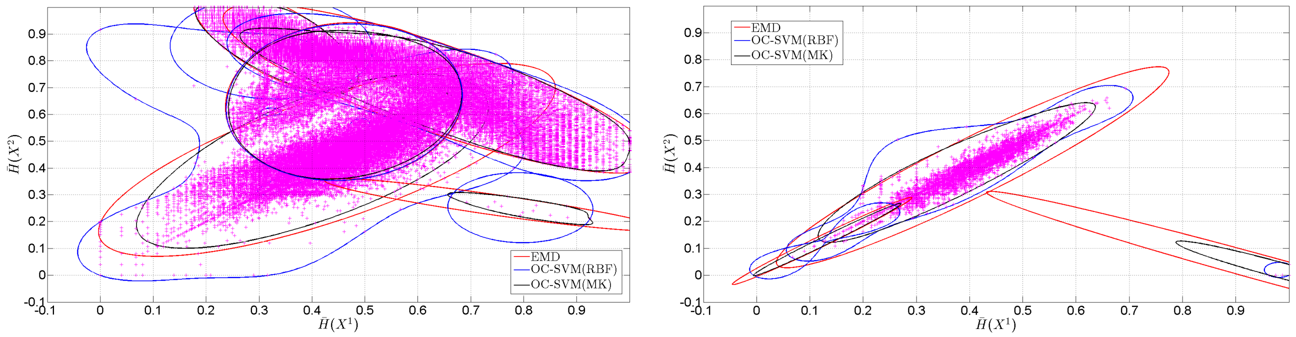
Each anomaly-traffic traces were divided into m non-overlapping slots of size $L = 32$ packets. The four traffic feature were extracted from each packet and for each i-slot, $i = 1, ..., m$, of the traces the estimation of entropy $\bar{H}(X_i^r)$ is obtained. Next, five matrices were formed. Each anomaly-traffic trace generates a region on the feature space, the ellipsoidal region was defined with Mahalanobis distance, and

Table 1. Detection rate using Tsallis entropy with $q = 0.01$.

Region	LAN					MIT-DARPA				
$\boldsymbol{H}_{Ip}$										
	β_1	ψ_1	ψ_2	ψ_3	ψ_4	β_2	ψ_5	ψ_6	ψ_7	ψ_8
MK	96.56	100	99.38	87.48	95.68	99.80	99.91	0.0	92.85	22.22
RBF	91.53	100	99.37	84.66	86.75	99.04	99.91	0.0	92.85	22.22
MD	99.27	100	99.53	81.89	97.23	99.99	99.91	0.0	0.0	0.0
$\boldsymbol{H}_{Pt}$										
MK	97.02	92.59	86.38	61.67	86.99	99.79	99.39	100	92.85	88.88
RBF	93.04	88.88	85.2	63.0	86.74	99.89	99.82	100	92.85	88.88
MD	99.3	77.77	86.89	63.37	1.35	99.89	0.0	100	0.0	100
$\boldsymbol{H}_{IpSPt}$										
MK	97.81	100	99.72	84.98	99.63	99.92	99.91	100	92.85	66.66
RBF	96.72	100	99.69	81.28	99.47	99.9	99.91	100	92.85	66.66
MD	99.04	100	99.62	81.52	98.74	99.89	99.91	100	0.0	44.44
$\boldsymbol{H}_{IpDPt}$										
MK	97.18	100	99.43	85.09	99.56	99.91	99.91	0.0	92.85	88.88
RBF	93.96	100	99.40	90.35	99.38	99.77	99.91	0.0	92.85	100
MD	99.05	100	99.6	82.67	98.4	99.88	99.91	0.0	0.0	100
$\boldsymbol{H}_{IpPt}$										
MK	97.75	100	99.64	78.84	99.56	99.90	99.91	100	92.85	100
RBF	97.45	100	99.61	87.99	99.71	99.89	99.91	100	92.85	100
MD	98.87	100	99.67	81.91	98.78	99.84	99.91	100	0.0	100

non-regular regions were found through OC-SVM with Radial Basis Function (RBF) and Mahalanobis Kernel (MK). Figure 2 shows the ellipses and non-regular regions defined on the feature space of IP addresses for each anomaly-traffic traces from LAN and MIT-DARPA traces.

Figure 2. Worm attack regions in the 2D space.

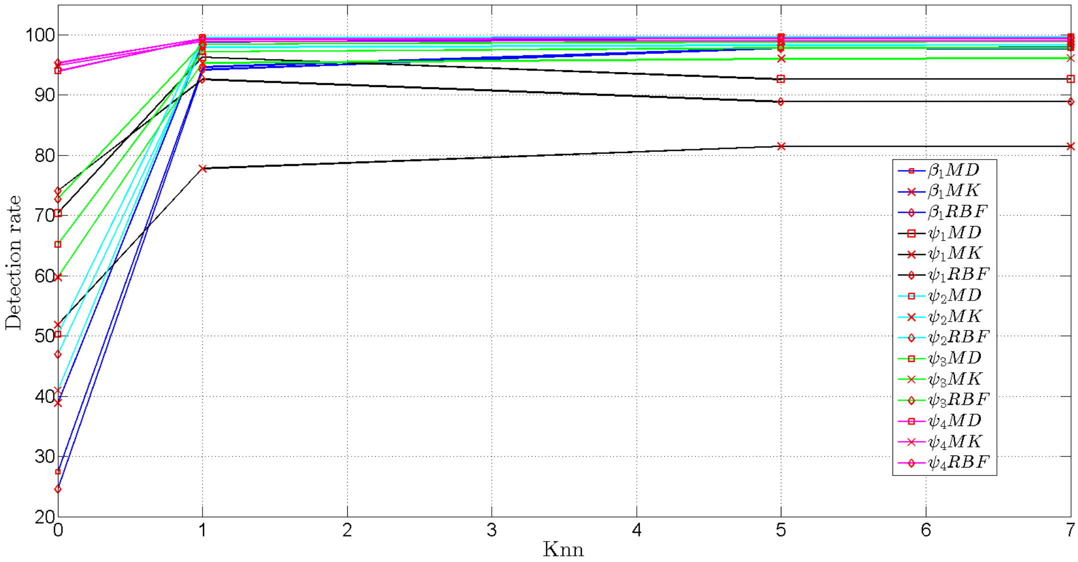


(a) Worm attack regions from LAN traces in the 2D space ($L = 32$).

(b) Worm attack regions from MIT-DARPA traces in the 2D space ($L = 32$).

In figure 3 the results of the classification of anomalies on LAN network traces using Tsallis entropy estimates with $q = 0.01$ of the Ips and ports variables and the different regions found with the proposed method are shown. These results were obtained using the knn approach as in [6].

Figure 3. Impact of the order of Knn on the detection rate.



5. Discussion of the experiment results

The effects of the number of features on the detection rate is showed in table (1). In OC-SVM method the Mahalanobis Kernel is better than RBF, since improves the detection rate in certain traces. Using the Knn method the classification is improved, however, it has a delay of k-slots to perform the classification.

Considering packet sizes of 60 bytes, in a 100Mbs network, to capture a slot of 32 packets, the time required is $\frac{32\times60\times8}{100Mbs} = 153.6\mu S$. The time of the training stage is not critical since this is done only once and offline. Using a PC with Intel Core i7 3.4 Ghz and 16G of RAM, a C-implementation of the here a proposed method using MD and including the decision function took computation times of not more than $5\mu s$. Therefore, the proposed method can be implemented in real time.

6. Conclusions

In this paper was proposed a method to detect and classify Internet traffic anomalies using: entropy of selected four features, Mahalanobis distance, and OC-SVM with two kernels RBF and particularly Mahalanobis Kernel. Regular and non-regular regions were built from normal traffic from training datas. Ellipsoidal regions based on Mahalanobis distance and the computation of $\{\bar{\mathbf{x}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, LT\}$ allow detection rate in the order of 98.81%. OC-SVM using Mahalanobis kernel achieve detection rate of 99.83% slightly higher than those using RBF kernel. Computation times in order of a few μs were obtained with ellipsoidal regions for detection of an anomaly. Consequently these results are very significant for real time implementation.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Chandola Varun, Banerjee Arindam, Kumar Vipin, Anomaly detection: A survey, ACM Computing Surveys, 09 2009, pp. 1-72
2. Anukool Lakhina, Mark Crovella, Christophe Diot, Mining Anomalies Using Traffic Feature Distributions, SIGCOMM 2005, pp. 217-228.
3. Arno Wagner, Bernhard Plattner, Entropy Based Worm and Anomaly Detection in Fast IP Networks, In Proc. of the 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise, WETICE 2005.
4. K. Xu, Z. Zhang, S. Bhattacharyya, Profiling Internet Backbone Traffic: Behavior Models and Applications, SIGCOMM 2005, pp. 22-26.
5. Santiago-Paz J., Torres-Roman D., Velarde-Alvarado P., Detecting Anomalies in Network Traffic Using Entropy and Mahalanobis distance, In Proc. of CONIELECOMP 2012.
6. Santiago-Paz J., Torres-Roman D., Characterization of Worm Attacks Using Entropy, Mahalanobis Distance and K-Nearest Neighbors, In Proc. of CONIELECOMP 2014.
7. R. Lee Mason, J. C. Young, Multivariate statistical process control with industrial applications, ASA-SIAM, 2002.
8. Tracy, N.D, Young, J.C., Mason, R.L, Multivariate Control Charts for Individual Observations, J. Quality Technology, 24, pp.88-95.
9. Kun-Lun Li, Hou-Kuan Huang, Sheng-Feng Tian, Wei Xu, Improving one-class SVM for Anomaly Detection, In Proc. of Second International Conference on Machine Learning and Cybernetics, November 2003.
10. Rui Zhang, Shaoyan Zhang, Yan Lan, Jianmin Jiang, Network Anomaly Detection Using One Class Support Vector Machine, In Proc. of International Multiconference of Engineers and Computer Scientists, IMECS 2008.
11. C.E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, Vol.27, 1948, pp. 379-423, 623-656.
12. A. Renyi, Probability Theory. North Holland, Amsterdan, 1970.
13. C. Tsallis, Possible Generalization of Boltzmann-Gibbs Statistics, Journal of Statistical Physics, Vol. 52, Nos. 1/2, 1988.
14. P.C. Mahalanobis, On the generalized distance in statistics, Proceedings of the National Institute of Science of India 12, 1936, pp.49-55.
15. B. Scholkopf et al., Estimating the Support of a High-Dimensional Distribution, Neural Computation, vol. 13, pp. 1443-1471, 2001.
16. B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classifiers. In Proc. of the 5th Annual ACM Workshop on Computational Learning Theory, pages 144-152, Pittsburgh, PA, 1992. ACM Press.
17. V. Vapnik, The Nature of Statistical Learning Theory. Springer, N.Y., 1995.
18. Scholkopf, C. J, C. Burges, A. J. Smola, Advances in Kernel Methods Support Vector Learning. MIT Press, Cambridge, MA, 1999.

19. J. C. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, 1998.
20. S. Abe, Training of support vector machines with Mahalanobis kernels, Lecture Notes in Computer Science, 3697 - Artificial Neural Networks: Formal Models and Their Applications, pp. 571-576, 2005.
21. P. Velarde-Alvarado, C. Vargas-Rosales, D. Torres-Roman, and A. Martinez-Herrera, Entropy-based profiles for intrusion detection in LAN traffic, Res. in Computing Science, vol. 40, pp. 119-130, 2008.
22. Lincoln Laboratory, MIT. DARPA Intrusion Detection Data.
23. C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology, Vol.2, pp.27:1-27:27, 2011.

7. Appendix: Results of detection and classification of LAN and MIT-DARPA traces.

Table 2 and 3 show the detection rate using the Shannon and Renyi entropy with $q = 0.01$ respectively and the features considered in this work, for the ellipsoidal and non-regular regions.

Table 2. Detection rate using Shannon entropy.

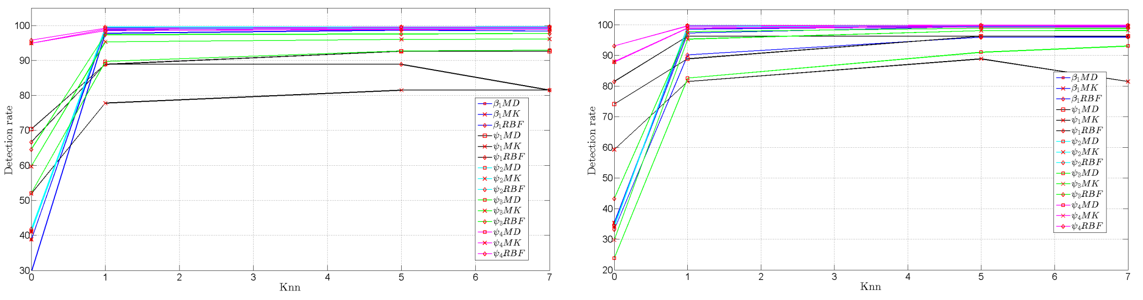
Region	**LAN**					**MIT-DARPA**				
					$\boldsymbol{H}_{Ip}$					
	β_1	ψ_1	ψ_2	ψ_3	ψ_4	β_2	ψ_5	ψ_6	ψ_7	ψ_8
MK	96.94	100	99.16	85.39	97.14	99.97	99.91	0.0	92.85	33.33
RBF	96.45	100	99.11	88.47	97.10	99.96	99.91	0.0	92.85	33.33
MD	98.77	100	99.08	65.86	95.79	99.93	99.91	0.0	92.85	33.33
					$\boldsymbol{H}_{Pt}$					
MK	97.36	77.77	82.34	66.74	89.23	99.71	99.82	100	92.85	100
RBF	97.33	74.07	81.90	66.77	89.28	99.92	99.82	100	92.85	100
MD	98.51	11.11	70.58	60.41	83.04	99.73	99.82	100	92.85	88.88
					$\boldsymbol{H}_{IpSPt}$					
MK	97.45	100	99.43	80.71	99.42	99.89	99.91	100	92.85	55.55
RBF	97.35	100	99.48	80.74	99.48	99.74	99.91	100	92.85	55.55
MD	99.56	100	99.16	66.25	97.20	99.84	99.91	0.0	92.85	55.55
					$\boldsymbol{H}_{IpDPt}$					
MK	97.01	100	99.26	85.02	99.28	99.91	99.91	0.0	92.85	88.88
RBF	97.50	100	99.25	87.88	99.47	99.91	99.91	0.0	92.85	88.88
MD	98.44	100	99.17	67.64	98.39	99.83	99.91	0.0	92.85	88.88
					$\boldsymbol{H}_{IpPt}$					
MK	97.02	100	99.39	82.21	99.54	99.96	99.91	100	92.85	100
RBF	97.36	100	99.43	82.66	99.63	99.77	99.91	100	92.85	100
MD	98.65	100	99.19	67.80	98.54	99.82	99.91	100	92.85	100

Table 3. Detection rate for Renyi entropy with $q = 0.01$.

Region	LAN					MIT-DARPA				
					$\boldsymbol{H}_{Ip}$					
	β_1	ψ_1	ψ_2	ψ_3	ψ_4	β_2	ψ_5	ψ_6	ψ_7	ψ_8
MK	96.74	100	99.22	86.96	95.93	99.98	99.91	0.0	92.85	22.22
RBF	96.77	100	99.17	85.08	94.02	99.95	99.91	0.0	92.85	22.22
MD	98.77	100	99.23	60.89	79.02	99.93	99.91	0.0	92.85	22.22
					$\boldsymbol{H}_{Pt}$					
MK	92.84	88.88	83.77	62.45	88.31	99.76	99.39	100	92.85	88.88
RBF	93.09	88.88	81.88	60.8	91.11	99.86	99.82	100	92.85	77.77
MD	98.38	14.81	70.14	60.03	86.74	99.58	99.39	100	92.85	88.88
					$\boldsymbol{H}_{IpSPt}$					
MK	97.93	100	99.58	70.59	98.84	99.92	99.91	100	92.85	55.55
RBF	96.72	100	99.69	81.28	99.47	99.87	99.91	100	92.85	66.66
MD	97.36	100	99.57	56.79	99.12	99.78	99.91	100	92.85	77.77
					$\boldsymbol{H}_{IpDPt}$					
MK	97.18	100	99.43	85.09	99.56	99.92	99.91	0.0	92.85	88.88
RBF	96.9	100	99.35	84.76	99.59	99.76	99.91	0.0	92.85	88.88
MD	98.65	100	99.32	62.78	98.85	99.77	99.91	0.0	92.85	100
					$\boldsymbol{H}_{IpPt}$					
MK	97.75	100	99.64	78.84	99.56	99.28	99.91	100	92.85	100
RBF	97.14	100	99.58	79.05	99.59	99.77	99.91	100	92.85	100
MD	98.65	100	99.38	63.83	98.88	99.42	99.91	100	92.85	100

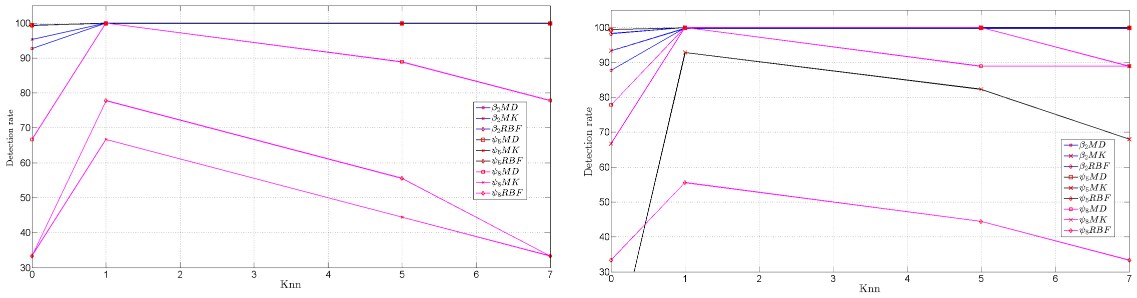
In figures (4(a)), (4(b)), (5(a)) and (5(b)) the results of the classification, and the impact of the order of Knn on the detection rate of LAN traces and ψ_5 y ψ_8 and β_2 from MIT-DARPA traces using entropy estimates of the Ips and ports variables and the different regions found with the proposed method are shown. These results were obtained with $q = 0.01$ for generalized entropies. The others attacks from MIT-DARPA traces are excluded because the estimates of entropy are zero or there is only one point.

Figure 4. Impact of the order of Knn on the classification of LAN traces using Shannon and Renyi entropies respectively.



(a) Classification using Shannon entropy. (b) Classification using Renyi entropy with $q = 0.01$.

Figure 5. Impact of the order of Knn on the classification of β_2, ψ_5 y ψ_8, using Shannon and Renyi entropies respectively.

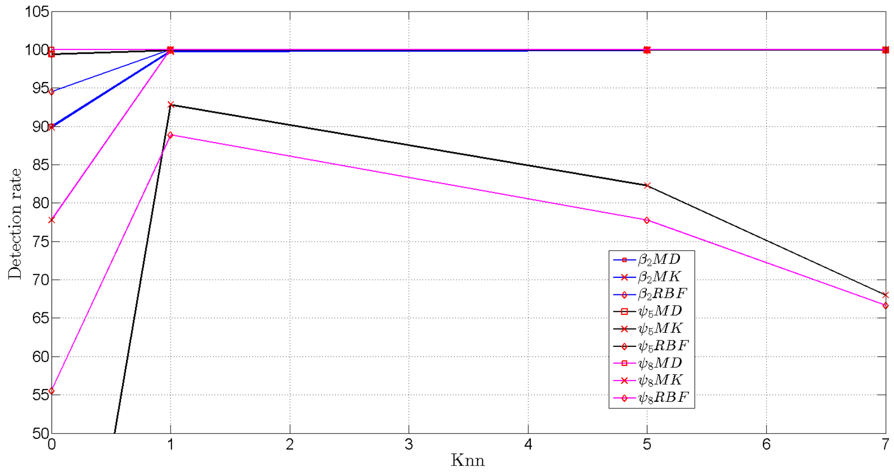


(a) Classification using Shannon entropy.

(b) Classification using Renyi entropy with $q = 0.01$.

In figure 6 the results of the classification of anomalies on MIT-DARPA network traces using Tsallis entropy estimates with $q = 0.01$ of the Ips and ports variables and the different regions found with the proposed method are shown.

Figure 6. Impact of the order of Knn on the detection rate with Tsallis entropy and $q = 0.01$.


Detection rate
Knn
$\beta_2 MD$
$\beta_2 MK$
$\beta_2 RBF$
$\psi_5 MD$
$\psi_5 MK$
$\psi_5 RBF$
$\psi_8 MD$
$\psi_8 MK$
$\psi_8 RBF$