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

Jayro Santiago-Paz, Deni Torres-Roman, Angel Figueroa-Ypiña.
Cinvestav, Campus Guadalajara

November 2014

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

Outline

1st International Electronic Conference on Entropy and Its Applications

3 - 21 November 2014

Outline

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

Outline

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

Outline

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

Outline

1st International Electronic Conference on Entropy and Its Applications

3 - 21 November 2014

Outline

1st International Electronic Conference on Entropy and Its Applications

3 - 21 November 2014

Network Intrusion Detection Systems (NIDS)

1. **Signature-NIDS**. Use a database with attack signatures.
2. **Anomaly-NIDS**. Classify the traffic in normal and abnormal to decide if an attack has occurred. Uses network features such as destination and source IP Addresses and Port, packet size, number of flows, and amount of packets between hosts.

A class of Anomaly-NIDS is the **entropy-based approach,** which:

- Provide more information about the structure of anomalies than traditional traffic volume analysis.
- Capture the degree of dispersal or concentration of the distributions for different traffic features.

ınic
and

3 - 21 November 2014

Statement problem

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, this region R_N will serve to detect anomalies.
- If ψ was captured while network attack occurred, the defined region R_A characterizes the anomaly

onic
and Its Applications
3 - 21 November 2014

Approach

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.

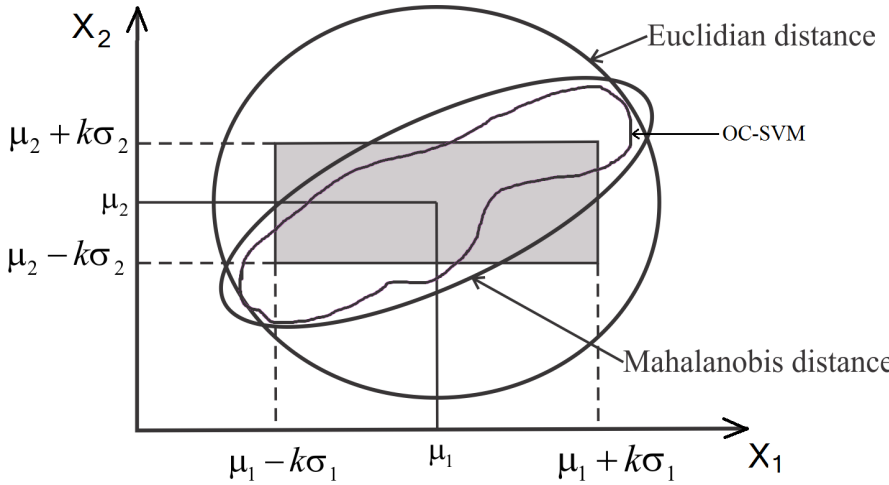


Figure 1: Different regions based on different methods and metrics.

Entropy

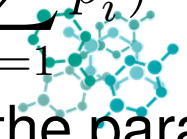
Let X be a r.v. that take values of the set $\{x_1, x_2, ..., x_M\}$, $p_i := P(X = x_i)$ the probability of occurrence of x_i.

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

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

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

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.



Mahalanobis distance

$$d^2 = (\mathbf{x} - \bar{\mathbf{x}})^T \boldsymbol{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}). \tag{4}$$

An unbiased sample covariance matrix is

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

where the sample mean is

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

OC-SVM

$$\min_{w \in F, b \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}^N} \frac{1}{2} \|w\|^2 + \frac{1}{\nu N} \sum_i^N \xi_i - b \tag{7}$$

Decision function

$$f(\mathbf{x}) = sgn\left(\sum_i^N \alpha_i k(\mathbf{x}_i, \mathbf{x}) - b \right), \tag{8}$$

Mahalanobis Kernel

$$k(\mathbf{x}, \mathbf{y}) = exp(-\frac{\eta}{p}(\mathbf{x} - \mathbf{y})' \mathbf{S}^{-1}(\mathbf{x} - \mathbf{y})), \tag{9}$$

where p is the number of features, η is a control parameter of

Training

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

MD method

- $LT = (\frac{(m-1)^2}{m})\beta_{[\alpha,p/2,(m-p-1)/2]}$, where $\beta_{[\alpha,p/2,(m-p-1)/2]}$ represents a beta distribution.
- The mean vector $\bar{\mathbf{x}} = \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_p\}$.
- The matrix equation $\mathbf{S}\boldsymbol{\gamma} = \boldsymbol{\lambda}\boldsymbol{\gamma}$ is solved.
- $\{LT, \bar{\mathbf{x}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}\}$.

Training

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

OC-SVM method

- The equation (7) is solved using two different kernel functions (Radial Basis Function (RBF) and Mahalanobis kernel(MK)).
- $\{x_i = sv_i, \alpha_i, b\}$, where $x_i = sv_i$ is the i-support vector, α_i, b are constants that solve the equation (7).

Detection

$$\mathbf{h_i} = \left(\bar{H}(X_i^1), \bar{H}(X_i^2), \cdots, \bar{H}(X_i^p)\right). \tag{10}$$

- The decision function for MD region is given by (4), if $d_i^2 \leq LT$ then $i-$slot is considered "normal" otherwise is a potential anomaly.
- The decision function for OC-SVM is (8), if the function is $+1$ then $\mathbf{h_i}$ is considered "normal" otherwise is a potential anomaly.

onic
and
Its Applications
3 - 21 November 2014

Classification

If the vector (10) 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.

Datasets

LAN

- Normal (β_1).
- port scan (ψ_1).
- Blaster worm (ψ_2).
- Sasser worm (ψ_3).
- Welchia worm(ψ_4).

MIT-DARPA

- Normal (β_2).
- Smurf worm (ψ_5).
- Neptune worm (ψ_6).
- Pod worm(ψ_7).
- port sweep (ψ_8).

Conference on Entropy and Its Applications
3 - 21 November 2014

Anomaly detection

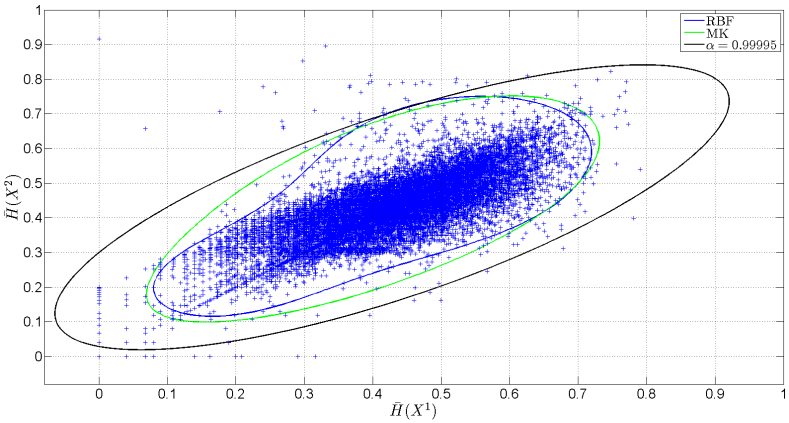


Figure 2: Estimated entropy of IP addresses from LAN traces.

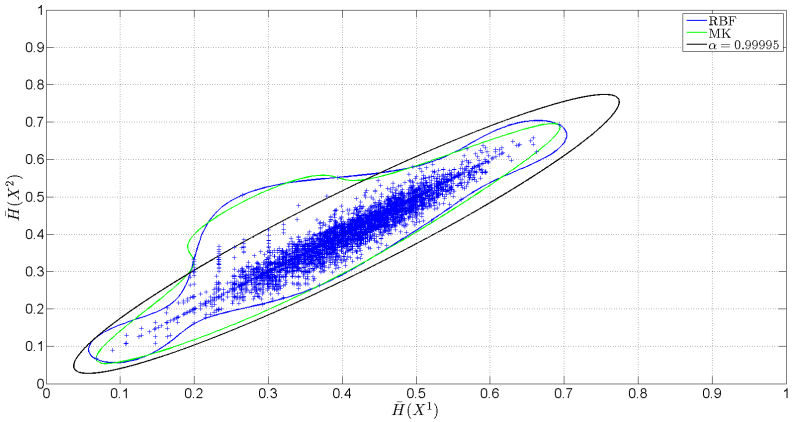
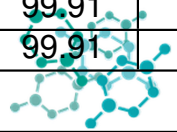


Figure 3: Estimated entropy of IP addresses from MIT DARPA traces.

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

	β_1	ψ_1	ψ_2	ψ_3	ψ_4	β_2	ψ_5	ψ_6	ψ_7	ψ_8
$\boldsymbol{H}_{Ip}$										
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



Classification

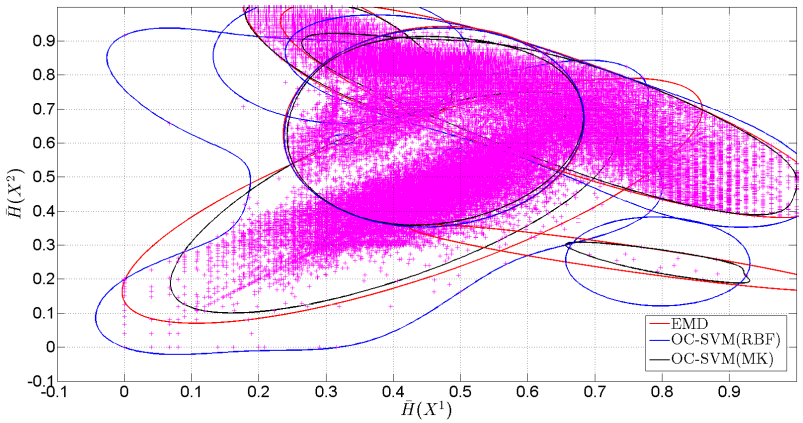


Figure 4: Worm attack regions from LAN traces in the 2D space ($L = 32$).

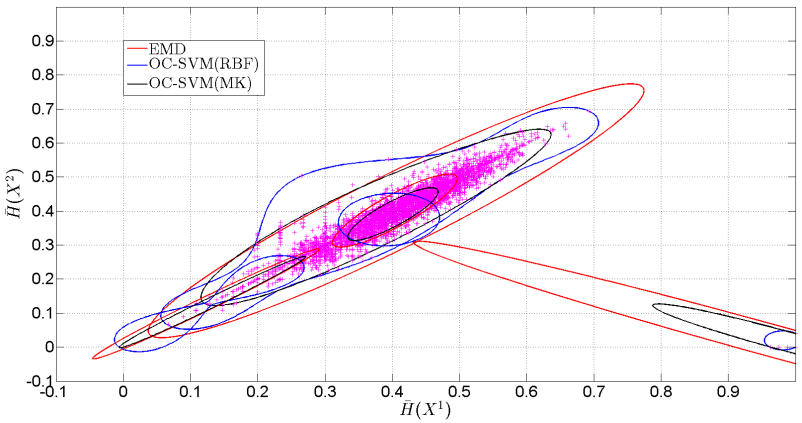


Figure 5: Worm attack regions from MIT-DARPA traces in the 2D space ($L = 32$).

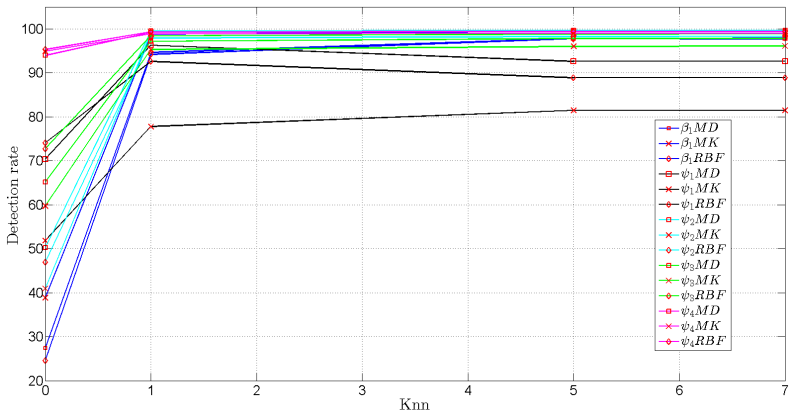


Figure 6: Classification of LAN traces using Tsallis entropy.

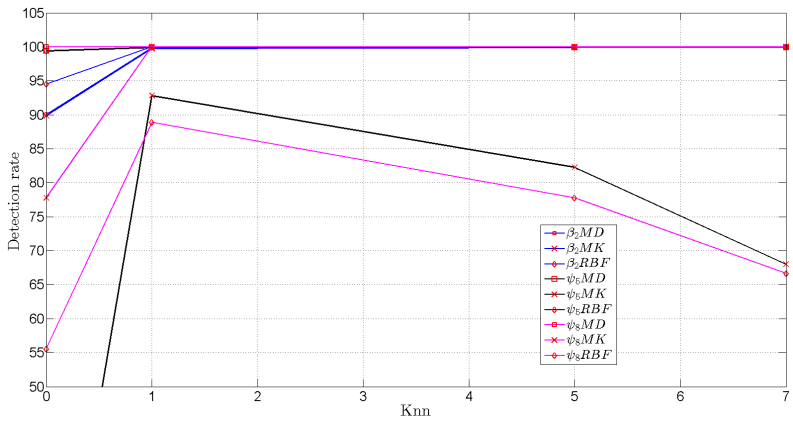


Figure 7: Classification of MIT-DARPA traces using Tsallis entropy.

1st International Electronic Conference on Entropy and Its Applications

3 - 21 November 2014

Conclusions

- Ellipsoidal regions based on Mahalanobis distance and the computation of $\{\bar{\mathbf{x}}, \gamma, \boldsymbol{\lambda}, LT\}$ 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.
- Using the Knn method the classification is improved, however, it has a delay of k-slots to perform the classification.
- Using a PC with Intel Core i7 3.4 Ghz and 16G of RAM, a C-implementation of the proposed method using MD and including the decision function took computation times of not more than $5\mu s$.

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