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
Outline

1. Introduction
2. Statement problem
3. Mathematical background
4. Algorithm
5. Experiments
6. Conclusions
Outline

1. Introduction
2. Statement problem
3. Mathematical background
4. Algorithm
5. Experiments
6. Conclusions
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.
Let $\psi$ 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 $\psi$ was obtained during “normal” network behavior, this region $R_N$ will serve to detect anomalies.
- If $\psi$ was captured while network attack occurred, the defined region $R_A$ characterizes the anomaly.
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.
Entropies

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

\begin{align*}
\hat{H}^S(P) &= - \sum_{i=1}^{M} p_i \log p_i. \quad (1) \\
\hat{H}^R(P, q) &= \frac{1}{1 - q} \log(\sum_{i=1}^{M} p_i^q) \quad (2) \\
\hat{H}^T(P, q) &= \frac{1}{q - 1} (1 - \sum_{i=1}^{M} p_i^q) \quad (3)
\end{align*}

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

\[ d^2 = (x - \bar{x})^T S^{-1} (x - \bar{x}). \] (4)

An unbiased sample covariance matrix is

\[ S = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})', \] (5)

where the sample mean is

\[ \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i. \] (6)
OC-SVM

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

Decision function

\[
f(x) = sgn \left( \sum_{i}^{N} \alpha_i k(x_i, x) - b \right),
\]

Mahalanobis Kernel

\[
k(x, y) = exp\left(-\frac{\eta}{p} (x - y)' S^{-1} (x - y) \right),
\]

where \(p\) is the number of features, \(\eta\) is a control parameter of
Training

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

**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 \( \bar{x} = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\}. \)
- The matrix equation \( \mathbf{S}\gamma = \lambda\gamma \) is solved.
- \( \{LT, \bar{x}, \gamma, \lambda\}. \)
Training

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

**OC-SVM method**

- The equation (7) is solved using two different kernel functions (Radial Basis Function (RBF) and Mahalanobis kernel (MK)).

\[\{x_i = sv_i, \alpha_i, b\}, \text{ where } x_i = sv_i \text{ is the } i\text{-support vector, } \alpha_i, b \text{ are constants that solve the equation (7)}.\]
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 \( h_i \) is considered “normal” otherwise is a potential anomaly.
If the vector (10) is out of the “normal” region, i.e. $\mathbf{h}_i \notin \mathbb{R}^N$, but $\mathbf{h}_i \in \mathbb{R}^A$ 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 ($\beta_1$).
- port scan ($\psi_1$).
- Blaster worm ($\psi_2$).
- Sasser worm ($\psi_3$).
- Welchia worm ($\psi_4$).

MIT-DARPA
- Normal ($\beta_2$).
- Smurf worm ($\psi_5$).
- Neptune worm ($\psi_6$).
- Pod worm ($\psi_7$).
- port sweep ($\psi_8$).
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 $q = 0.01$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$\psi_3$</th>
<th>$\psi_4$</th>
<th>$\beta_2$</th>
<th>$\psi_5$</th>
<th>$\psi_6$</th>
<th>$\psi_7$</th>
<th>$\psi_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK</td>
<td>96.56</td>
<td>100</td>
<td>99.38</td>
<td>87.48</td>
<td>95.68</td>
<td>99.80</td>
<td>99.91</td>
<td>0.0</td>
<td>92.85</td>
<td>22.22</td>
</tr>
<tr>
<td>RBF</td>
<td>91.53</td>
<td>100</td>
<td>99.37</td>
<td>84.66</td>
<td>86.75</td>
<td>99.04</td>
<td>99.91</td>
<td>0.0</td>
<td>92.85</td>
<td>22.22</td>
</tr>
<tr>
<td>MD</td>
<td>99.27</td>
<td>100</td>
<td>99.53</td>
<td>81.89</td>
<td>97.23</td>
<td>99.99</td>
<td>99.91</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$\psi_3$</th>
<th>$\psi_4$</th>
<th>$\beta_2$</th>
<th>$\psi_5$</th>
<th>$\psi_6$</th>
<th>$\psi_7$</th>
<th>$\psi_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK</td>
<td>97.02</td>
<td>92.59</td>
<td>86.38</td>
<td>61.67</td>
<td>86.99</td>
<td>99.79</td>
<td>99.39</td>
<td>100</td>
<td>92.85</td>
<td>88.88</td>
</tr>
<tr>
<td>RBF</td>
<td>93.04</td>
<td>88.88</td>
<td>85.2</td>
<td>63.0</td>
<td>86.74</td>
<td>99.89</td>
<td>99.82</td>
<td>100</td>
<td>92.85</td>
<td>88.88</td>
</tr>
<tr>
<td>MD</td>
<td>99.3</td>
<td>77.77</td>
<td>86.89</td>
<td>63.37</td>
<td>1.35</td>
<td>99.89</td>
<td>0.0</td>
<td>100</td>
<td>0.0</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$\psi_3$</th>
<th>$\psi_4$</th>
<th>$\beta_2$</th>
<th>$\psi_5$</th>
<th>$\psi_6$</th>
<th>$\psi_7$</th>
<th>$\psi_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK</td>
<td>97.81</td>
<td>100</td>
<td>99.72</td>
<td>84.98</td>
<td>99.63</td>
<td>99.92</td>
<td>99.91</td>
<td>100</td>
<td>92.85</td>
<td>66.66</td>
</tr>
<tr>
<td>RBF</td>
<td>96.72</td>
<td>100</td>
<td>99.69</td>
<td>81.28</td>
<td>99.47</td>
<td>99.9</td>
<td>99.91</td>
<td>100</td>
<td>92.85</td>
<td>66.66</td>
</tr>
<tr>
<td>MD</td>
<td>99.04</td>
<td>100</td>
<td>99.62</td>
<td>81.52</td>
<td>98.74</td>
<td>99.89</td>
<td>99.91</td>
<td>100</td>
<td>0.0</td>
<td>44.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$\psi_3$</th>
<th>$\psi_4$</th>
<th>$\beta_2$</th>
<th>$\psi_5$</th>
<th>$\psi_6$</th>
<th>$\psi_7$</th>
<th>$\psi_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK</td>
<td>97.18</td>
<td>100</td>
<td>99.43</td>
<td>85.09</td>
<td>99.56</td>
<td>99.91</td>
<td>99.91</td>
<td>0.0</td>
<td>92.85</td>
<td>88.88</td>
</tr>
<tr>
<td>RBF</td>
<td>93.96</td>
<td>100</td>
<td>99.40</td>
<td>90.35</td>
<td>99.38</td>
<td>99.77</td>
<td>99.91</td>
<td>0.0</td>
<td>92.85</td>
<td>100</td>
</tr>
<tr>
<td>MD</td>
<td>99.05</td>
<td>100</td>
<td>99.6</td>
<td>82.67</td>
<td>98.4</td>
<td>99.88</td>
<td>99.91</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$\psi_3$</th>
<th>$\psi_4$</th>
<th>$\beta_2$</th>
<th>$\psi_5$</th>
<th>$\psi_6$</th>
<th>$\psi_7$</th>
<th>$\psi_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK</td>
<td>97.75</td>
<td>100</td>
<td>99.64</td>
<td>78.84</td>
<td>99.56</td>
<td>99.90</td>
<td>99.91</td>
<td>100</td>
<td>92.85</td>
<td>100</td>
</tr>
<tr>
<td>RBF</td>
<td>97.45</td>
<td>100</td>
<td>99.61</td>
<td>87.99</td>
<td>99.71</td>
<td>99.89</td>
<td>99.91</td>
<td>100</td>
<td>92.85</td>
<td>100</td>
</tr>
<tr>
<td>MD</td>
<td>98.87</td>
<td>100</td>
<td>99.67</td>
<td>81.91</td>
<td>98.78</td>
<td>99.84</td>
<td>99.91</td>
<td>100</td>
<td>0.0</td>
<td>100</td>
</tr>
</tbody>
</table>
**Classification**

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

**Figure 5:** Worm attack regions from MITDARPA 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.
Conclusions

- Ellipsoidal regions based on Mahalanobis distance and the computation of $\{\bar{x}, \gamma, \lambda, LT\}$ 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.