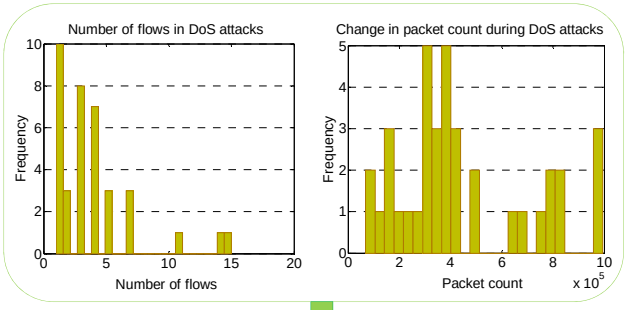
A Unified Approach to Network Anomaly Detection

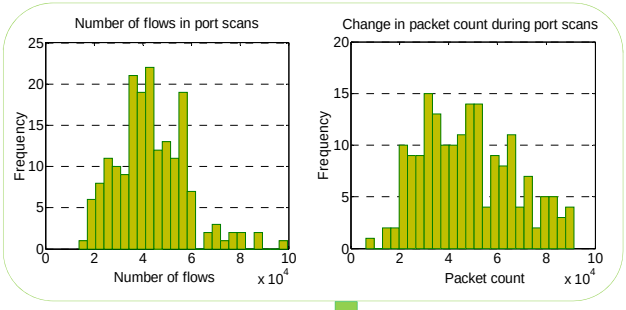
Tara Babaie
Sanjay Chawla
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

Current state of the art : detect one class of anomalies at the cost of others

Characterization of DoS attacks and port scans - Abilene network



DoS anomalies cause high temporal variation in the flows



port scan attacks cause spatial correlation across flows

DoS and port scan attacks are emblematic of two types of deviations: change across flows and through time

M-SSA: A Spatio-Temporal Approach

Traffic $\approx \mathcal{F}$([Spatial Component] , [Temporal Component]) + Noise

The count (or entropy) time series of : [sourceIP] → s(t) & [destIP] → d(t)

Embedding a manifold of arbitrary dimension *l* into the time dimension of both series

Joint-Trajectory Matrix $\mathbf{Y} = \left[\begin{array}{cccc|cccc} s_1 & s_2 & \cdots & s_{n-l+1} & d_1 & d_2 & \cdots & d_{n-l+1} \\ s_2 & s_3 & \cdots & s_{n-l+2} & d_2 & d_3 & \cdots & d_{n-l+2} \\ \vdots & \vdots & \cdots & s_{n-l+3} & \vdots & \vdots & \cdots & d_{n-l+3} \\ s_l & s_{l+1} & \cdots & s_n & d_l & d_{l+1} & \cdots & d_n \end{array}\right]$

SVD

$$\mathbf{C}_{s,d,\ell} = \frac{1}{\ell}[\mathbf{S} \quad \mathbf{D}]'[\mathbf{S} \quad \mathbf{D}]$$

$$\mathbf{C}_{s,d,\ell} U_{s,d,\ell} = \Sigma_{s,d,\ell} U_{s,d,\ell}$$

Sub-Space Technique: $\mathbf{Y} = \mathbf{X} + \mathbf{E}$

De-hankelization →

$s(t) = s(t)_{\{\text{Trend + Oscillations}\}} + \text{noise}_s(t)$

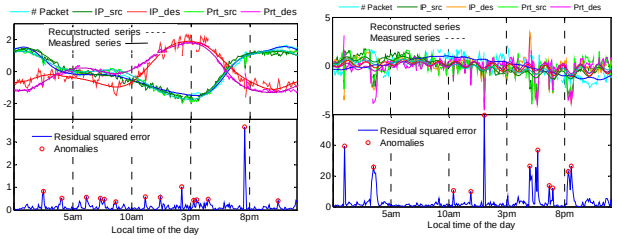
$d(t) = d(t)_{\{\text{Trend + Oscillations}\}} + \text{noise}_d(t)$

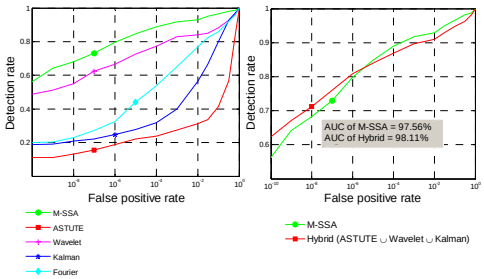
$(\text{noise}_s(t) , \text{noise}_d(t))$ → statistics-test for anomalies

Evaluation: detection capability and performance

- Datasets
  - **Real world data sets**
    Abilene Network: http://www.internet2.edu/
    Wide Network: http://www.wide.ad.jp/project/wg/mawi.html
  - **Synthetic data set**
    A simulation model which captures several distinctive characteristics of anomalies:
    1) The distribution of time between anomalies
    2) The distribution of anomaly Duration
    3) The distribution o anomaly type

- Main Result
  1) M-SSA, for the Abilene data, identifies 100% of DoS attacks and over 95% port scans. Similarly on the MAWI data set the detection rate was 100% for DoS attacks and over 90% for port scans
  2) Wavelets, Kalman and Fourier have high detection rates only for DoS attacks while ASTUTE performs exceedingly well only for port scan anomalies





ROC curves: M-SSA has a better detection rate than alternative techniques.

Future Work

- Distributed M-SSA– For more information contact tahereh.babaie@nicta.com.au

NICTA