Machine Learning Approaches to Network Anomaly Detection

Tarem Ahmed, Boris Oreshkin and Mark Coates
tarem.ahmed@mail.mcgill.ca, boris.oreshkin@mail.mcgill.ca, coates@ece.mcgill.ca

USENIX SysML, Cambridge, MA

April 10, 2007

Research supported by Canadian National Science and Engineering Research Council (NSERC) through the Agile All-Photonics Research Network (AAPN) and MITACS projects
Introduction

- Network data are
  - voluminous
  - high-dimensional
  - high-rate

- What is a network anomaly?
  - rare event
  - short-lived

- ML-based network anomaly detection methods
  more general than
  - model-based
  - signature-based
Our Methodology

- Show applicability of **ML approaches** to network anomaly detection

- **Two example algorithms:**
  - One-Class Neighbor Machine (OCNM) [Muñoz 06]
  - Kernel-based Online Anomaly Detection (KOAD) [Ahmed 07]

- **Two example datasets:**
  - Transports Quebec
  - Abilene
One-Class Neighbor Machine (OCNM)

- Region of normality should correspond to a Minimum Volume Set (MVS)
- OCNM for estimating MVS proposed in [Muñoz 06],
- Requires choice of sparsity measure, $g$. Example: $k$-th nearest-neighbour distance
- Sorts list of $g$, identifies fraction $\mu$ inside MVS

2-D Isomap of CHIN-LOSAnormal backbone flow, srcIP
Kernel-based Online Anomaly Detection (KOAD): Introduction

- Sequence of multivariate measurements: \( \{x_t\}_{t=1:T} \)
- Choose feature space with associated kernel:

\[
k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle
\]

where

\[
\varphi : x \in \mathbb{R}^n \rightarrow \varphi(x) \in H^\infty
\]

- Then feature vectors corresponding to normal traffic measurements should \textit{cluster}
Kernel-based Online Anomaly Detection (KOAD): Dictionary

- Should be possible to describe *region of normality* in feature space using sparse *dictionary*, 
  \[ D = \{ \tilde{x}_j \}_{j=1}^M \]

- Feature vector \( \varphi(x_t) \) is said to be *approximately linearly independent* on \( \{ \varphi(\tilde{x}_j) \}_{j=1}^M \) if [Engel 04]:
  \[ \delta_t = \min_a \left\| \sum_{j=1}^{m_t-1} a_j \varphi(\tilde{x}_j) - \varphi(x_t) \right\|^2 > \nu \]  
  (1)

Dictionary approximation

Threshold
Kernel-based Online Anomaly Detection (KOAD): Algorithm

- At timestep $t$ with arriving input vector $\mathbf{x}_t$:
  - Evaluate $\delta_t$ according to (1), compare with $\nu_1$ and $\nu_2$ where $\nu_1 < \nu_2$
  - If $\delta_t > \nu_2$, infer $\mathbf{x}_t$ far from normality: Red1
  - If $\delta_t > \nu_1$, raise Orange, resolve $l$ timesteps later, after “usefulness” test
  - If $\delta_t < \nu_1$, infer $\mathbf{x}_t$ close to normality: Green

- Delete obsolete elements, use exponential forgetting
- For details of KOAD see [Ahmed 07]
Dataset 1: Transports Quebec

Cameras monitored
Sample Images (anomalous)
Feature Extraction: Discrete Wavelet Transform

At each timestep, at each camera, get 6-D wavelet feature vector
Use $n = 3$ out of $c = 6$ voting at central monitoring unit
- **KOAD**: Gaussian kernel, with varying standard deviation for the kernel function
- **OCNM**: identify 5%-50% of outliers
Dataset 2: Abilene

- **Data collection**
  - 11 core routers, 121 *backbone flows*
  - 4 main pkt header fields collected: (*srcIP, dstIP, srcPort, dstPort*)

- **Data processing**
  - Construct histogram of headers
  - Calculate header *entropies* for each backbone flow, at each timestep
  - Variations in entropies (distributions) reveal many anomalies [Lakhina 2005]
Abilene Results

KOAD
PCA, 10 components
OCNM, 5% outliers
Conclusions and Future Work

- Preliminary results indicate potential of ML approaches
- Parameters set using supervised learning
- Computations must be distributed
- **Online**: complexity must be independent of time
Acknowledgements:
Thanks to Sergio Restrepo for processing Transports Quebec data and to Anukool Lakhina for providing Abilene dataset.

References:
[Ahmed 07]

[Engel 04]

[Lakhina 05]

[Muñoz 06]