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

McGill Electrical & Computer Engineering

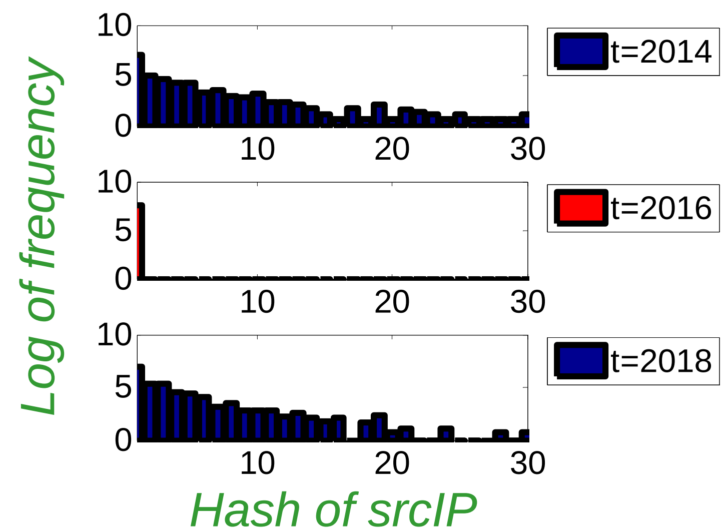
Introduction

- 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

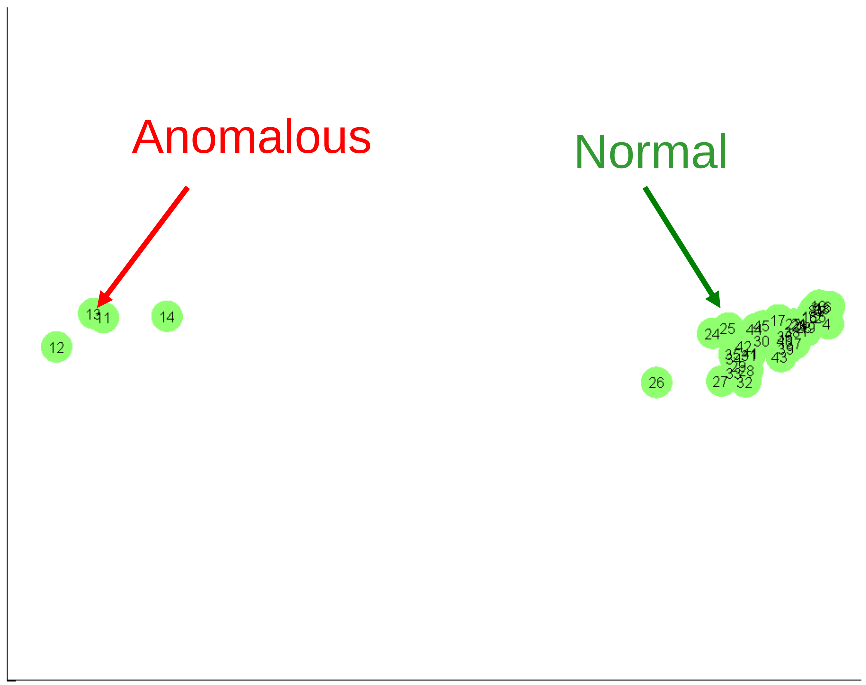
McGill

- 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 μ inside MVS



2-D Isomap of CHIN-LOSA backbone flow, *srcIP*

Kernel-based Online Anomaly Detection (KOAD): Introduction

McGill

- Sequence of multivariate measurements: $\{\mathbf{x}_t\}_{t=1:T}$
- Choose feature space with associated kernel:

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left\langle \varphi\left(\mathbf{x}_i\right), \varphi\left(\mathbf{x}_j\right) \right\rangle$$

where

$$\varphi : \mathbf{x} \in \mathbb{R}^n \rightarrow \varphi\left(\mathbf{x}\right) \in H^\infty$$

- Then feature vectors corresponding to normal traffic measurements should *cluster*

Kernel-based Online Anomaly Detection (KOAD): Dictionary

McGill

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

- Feature vector $\varphi\left(\mathbf{x}_t\right)$ is said to be *approximately linearly independent* on $\left\{ \varphi\left(\tilde{\mathbf{x}}_j\right) \right\}_{j=1}^{M}$ if [Engel 04]:

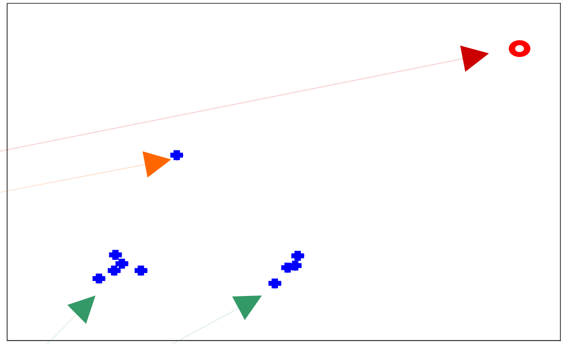
$$\delta_t = \min_{a} \left\| \sum_{j=1}^{m_{t-1}} a_j \phi(\tilde{\mathbf{x}}_j) - \phi(\mathbf{x}_t) \right\|^2 > \nu \qquad (1)$$

Dictionary approximation

Threshold

Kernel-based Online Anomaly Detection (KOAD): Algorithm

- At timestep t with arriving input vector $\mathbf{x}_t$:
  - Evaluate δ_t according to (1), compare with ν_1 and ν_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 (normal)

Camera 1

Camera 2

Camera 3

Camera 4

Camera 5

Camera 6

Sample Images (anomalous)

Camera 1

Camera 2

Camera 3

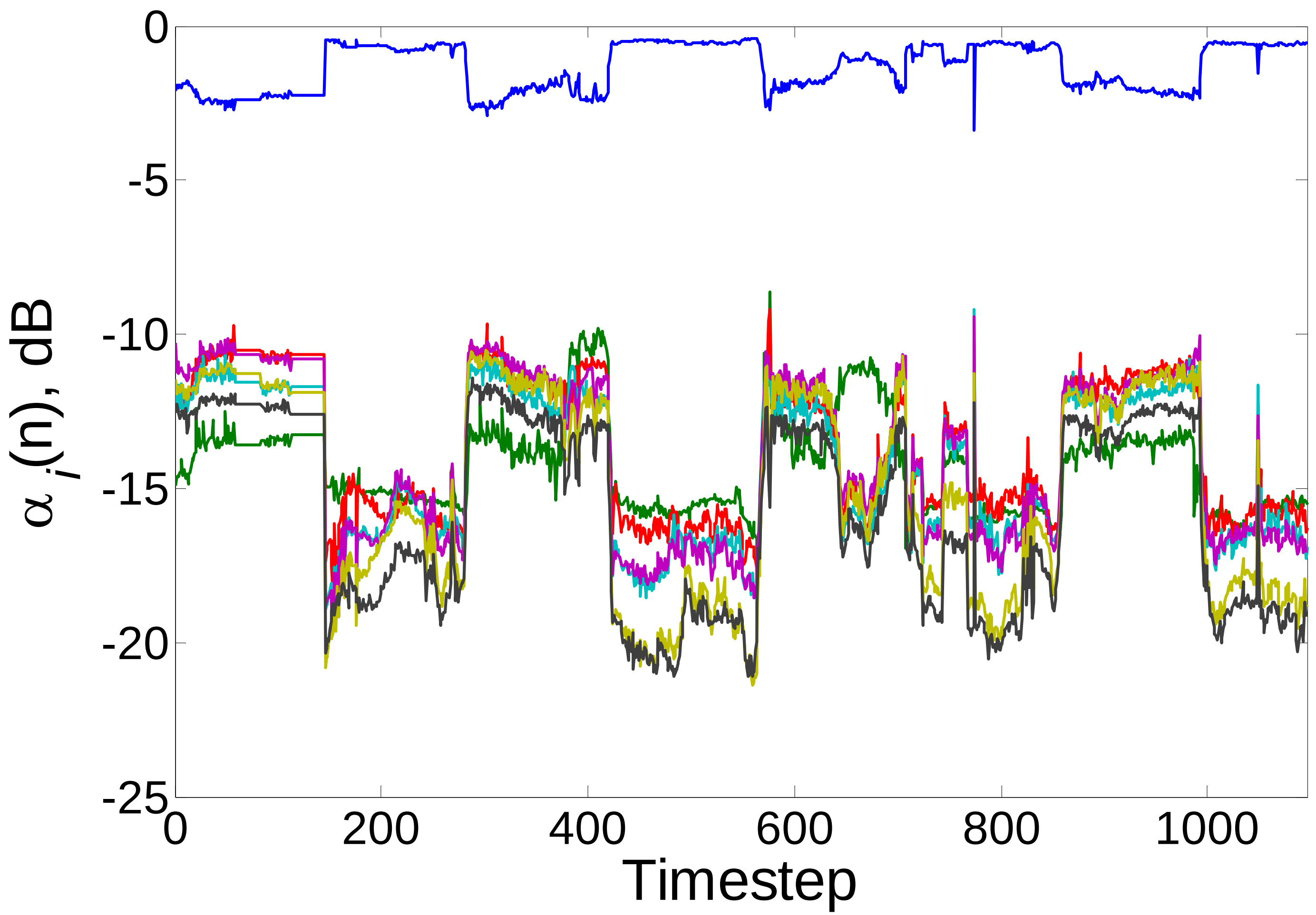
Camera 4

Camera 5

Camera 6

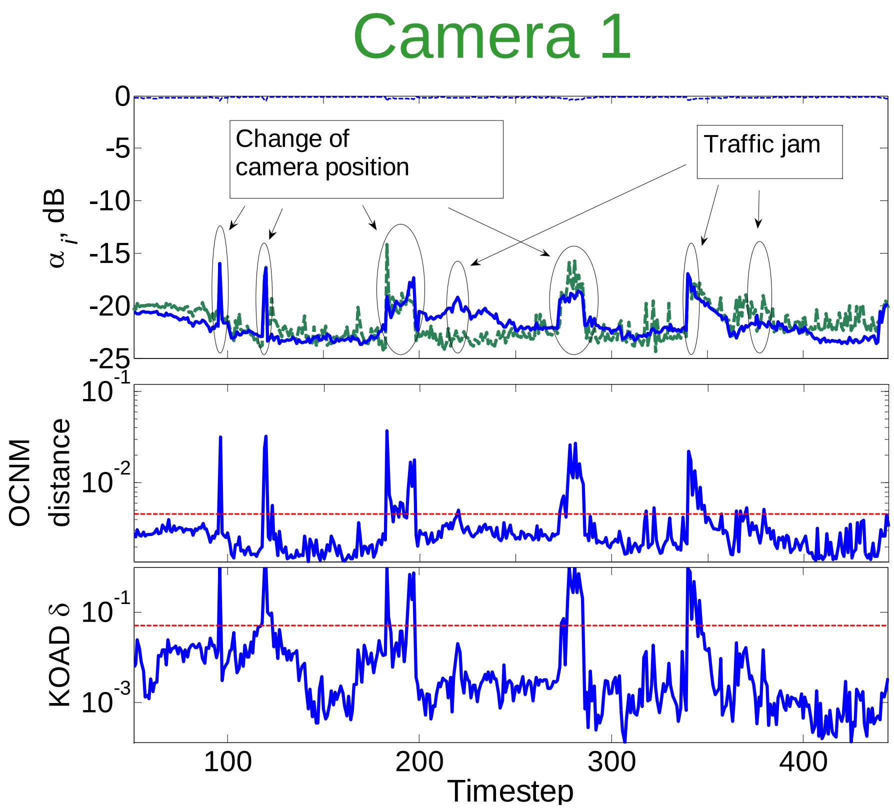
Feature Extraction: Discrete Wavelet Transform

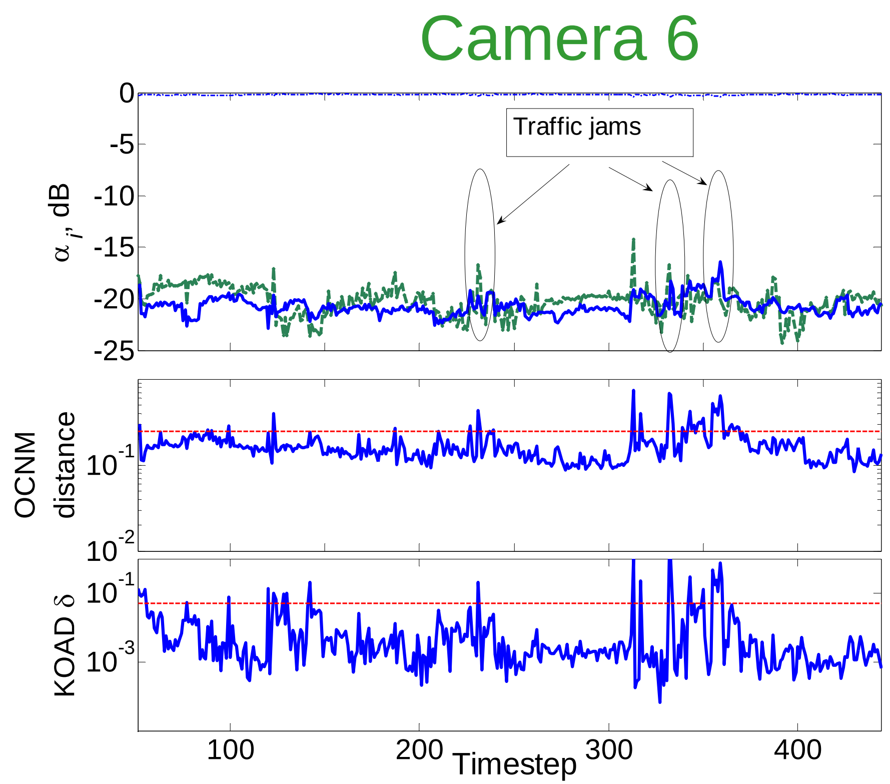
McGill



At each timestep, at each camera, get 6-D *wavelet feature vector*

Transports Quebec Results

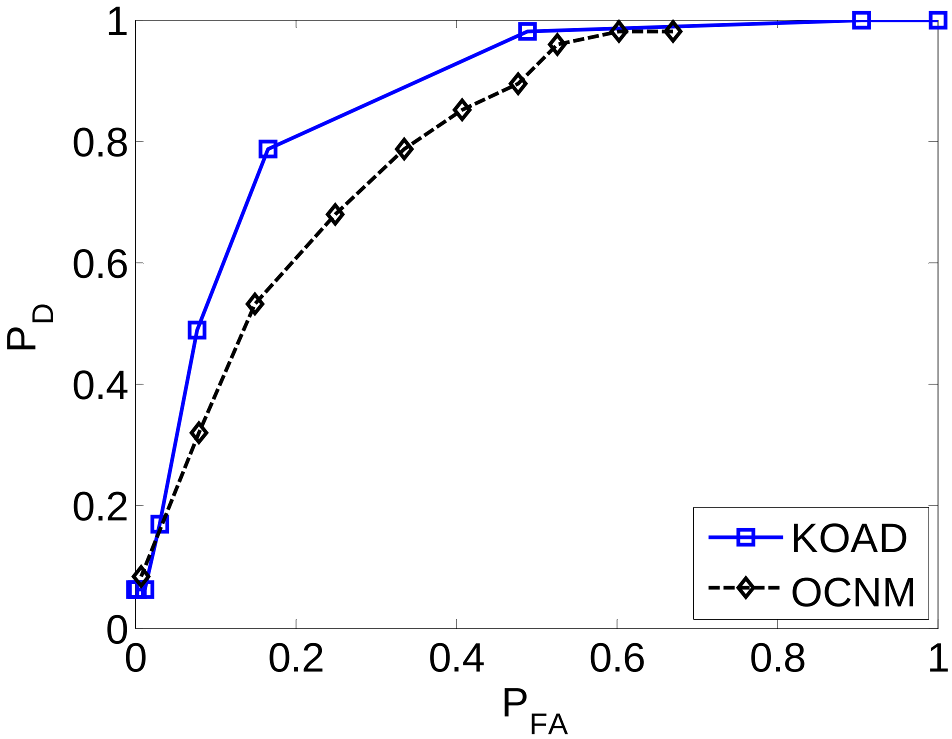




Use n = 3 out of c = 6 voting at central monitoring unit

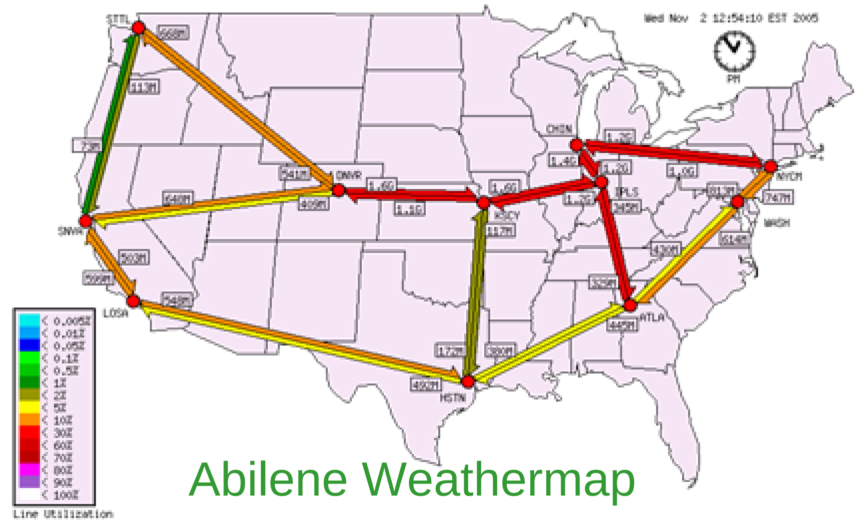
Transports Quebec ROC

- KOAD: Gaussian kernel, with varying standard deviation for the kernel function
- **OCNM**: identify 5%-50% of outliers



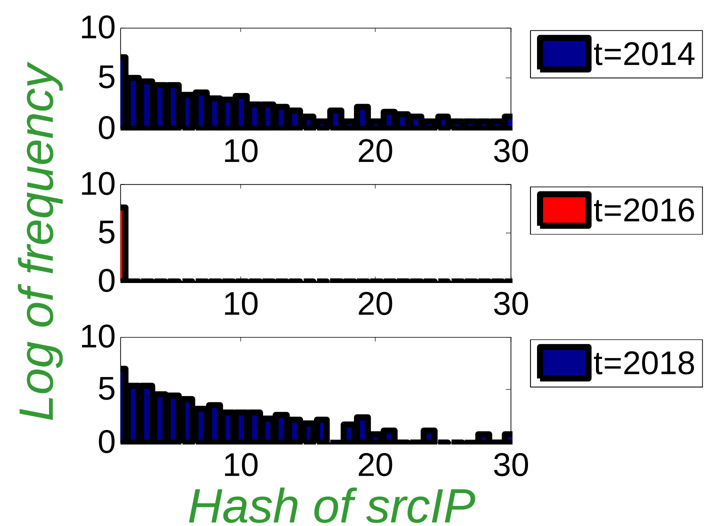
Dataset 2: Abilene

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

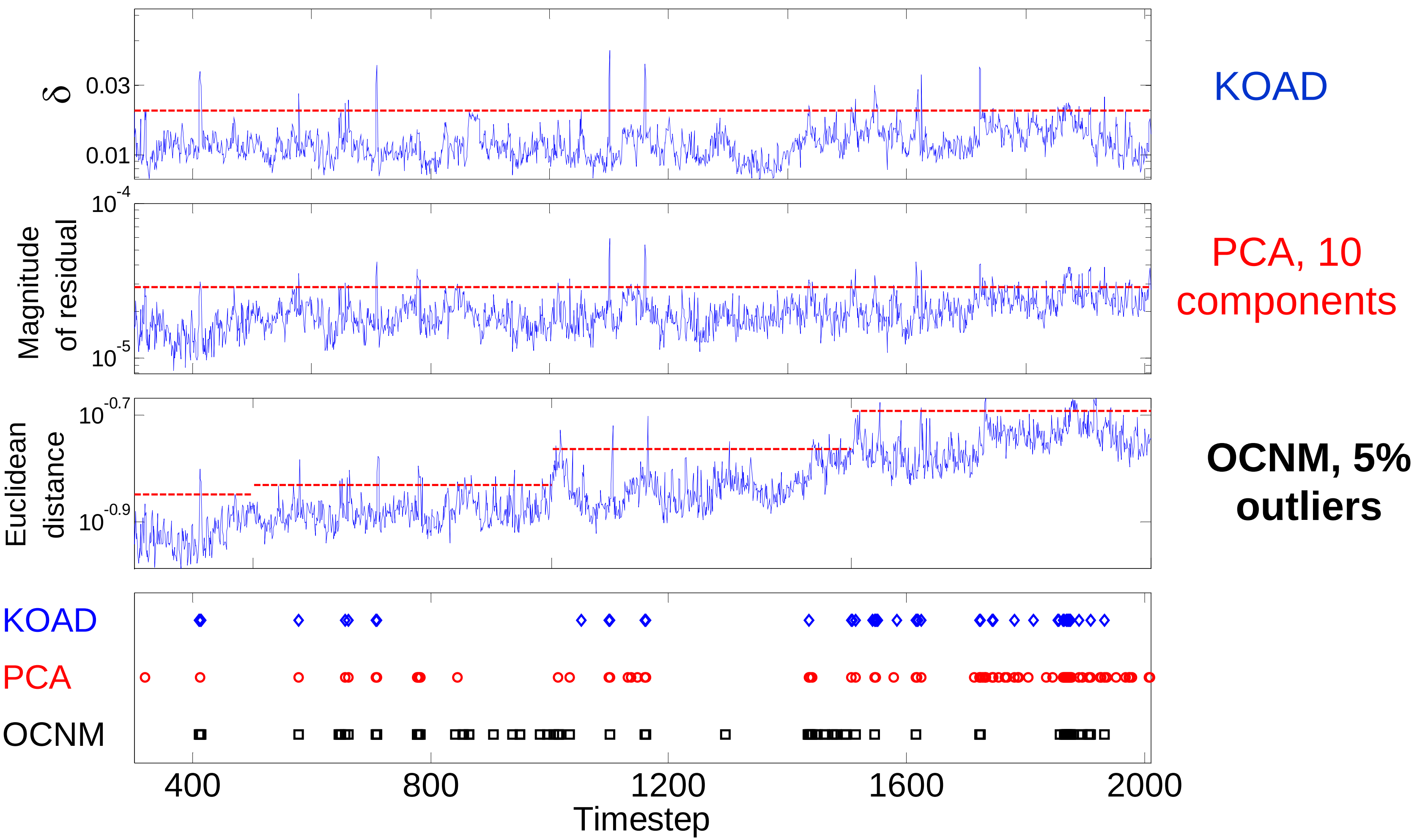


Abilene Weathermap

- 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



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, References

Acknowledgements:

Thanks to Sergio Restrepo for processing Transports Quebec data and to Anukool Lakhina for providing Abilene dataset.

References:

[Ahmed 07]

T. Ahmed, M. Coates, and A. Lakhina, “Multivariate online anomaly detection using kernel recursive least squares,” in *Proc. IEEE Infocom*, Anchorage, AK, May 2007, to appear.

[Engel 04]

Y. Engel, S. Mannor, and R. Meir, “The kernel recursive least squares algorithm,” *IEEE Trans. Signal Proc.*, vol. 52, no. 8, pp. 2275–2285, Aug. 2004.

[Lakhina 05]

A. Lakhina, M. Crovella and C. Diot, “Mining anomalies using traffic feature distributions,” in *Proc. ACM SIGCOMM*, Philadelphia, PA, Aug. 2005.

[Muñoz 06]

A. Muñoz and J. Moguerza, “Estimation of high-density regions using one-class neighbor machines,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, num 3, pp 476--480, Mar. 2006.