Clustering and Monitoring Edge Behaviour in Enterprise Network Traffic

Chris Schon, N. Adams & M. Evangelou
christopher.schon.16@ucl.ac.uk

University College London & Imperial College London

September 2017

Outline

Outline

Introduction

- Cyber-security has become an increasing threat at an organisational and national level.
- The WannaCry ransomware attack affected 230,000 computers in over 150 countries. *Image: Malware Tech.*
- Traditional signature-based detection techniques can easily be circumvented by sophisticated intruders.

Outline

Aim

- Complement established methods with novel anomaly detection techniques.
- Provide a dashboard-like collection of statistics, interpretable to a network analyst, which give extra support for situational awareness and network cognisance.
- Previous efforts include quantifying 'normal' activity at a global level (*A. Valdes and K. Skinner, 2000*) and at a local level, such as monitoring specific device connections and clusters (*J. Neil et al., 2013*).

Outline

Data

- One week of NetFlow event records from Los Alamos National Laboratory (LANL).

Time	Conn. duration	Source comp.	Source port	Destination comp.	Destination port	Protocol	Packet Count	Byte count
1	0	C1065	389	C3799	N10451	6	10	532
1	0	C1423	N1136	C1707	N1	6	5	847
1	0	C1423	N1142	C1707	N1	6	5	847

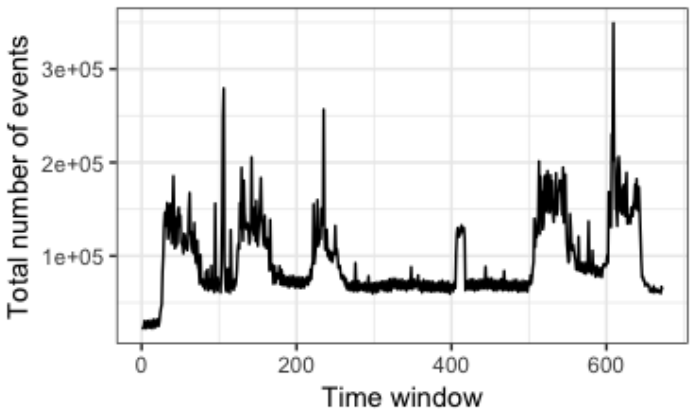


Figure 1 : Number of NetFlow events by time window.

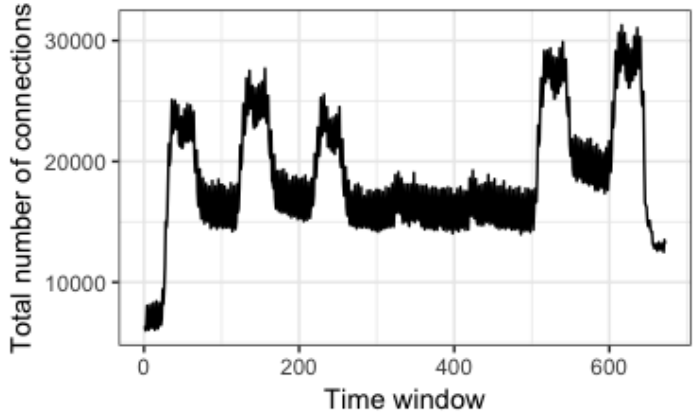


Figure 2 : Number of unique connections by time window.

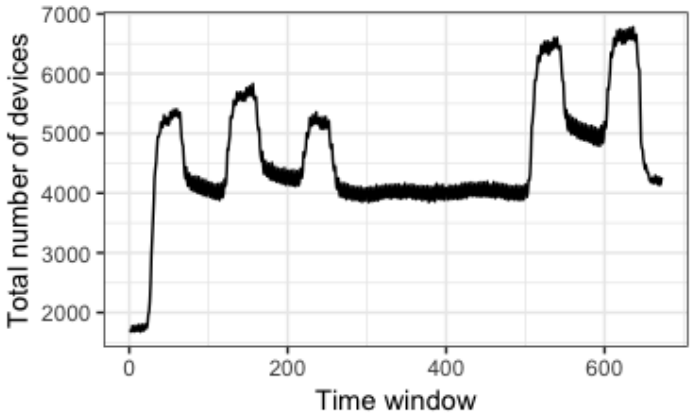


Figure 3 : Number of unique devices by time window.

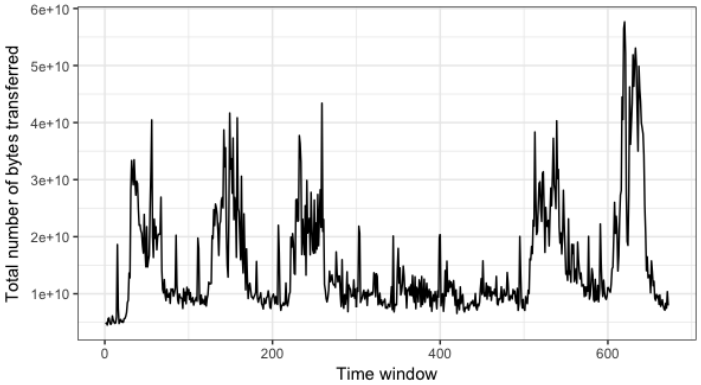
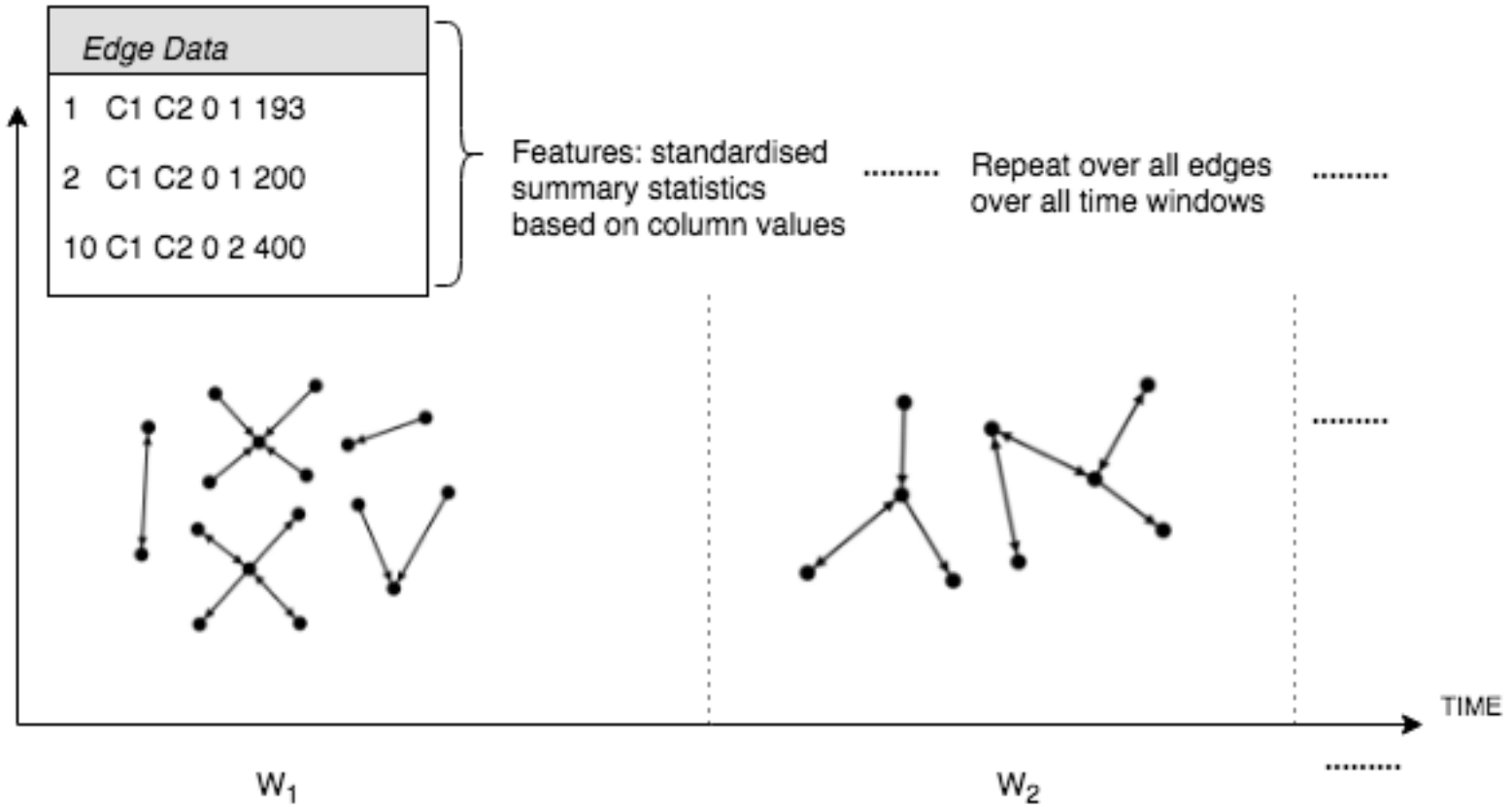


Figure 4 : Total bytes transferred by time window.

Outline

Methodology

- Our anomaly detection technique follows a three-step procedure:
  1. Gather features which describe the active connections (edges) in the network during a 15-minute time window.
  2. Group edges into clusters based on their feature vectors.
  3. Repeat this process over contiguous windows. Then, derive a series of informative indicators by examining the relationship of edges with the observed cluster structure.


Edge Data
1 C1 C2 0 1 193
2 C1 C2 0 1 200
10 C1 C2 0 2 400
Features: standardised summary statistics based on column values
.....
Repeat over all edges over all time windows
.....
.....
.....
TIME
.....
W1
W2

Features

Time related

- Mean, median, standard deviation and inter-quartile range (IQR) of connection *duration*.
- Mean and standard deviation of *inter-arrival times* of events.

Bytes related

- Mean, median, standard deviation and IQR of *bytes*.

Packet related

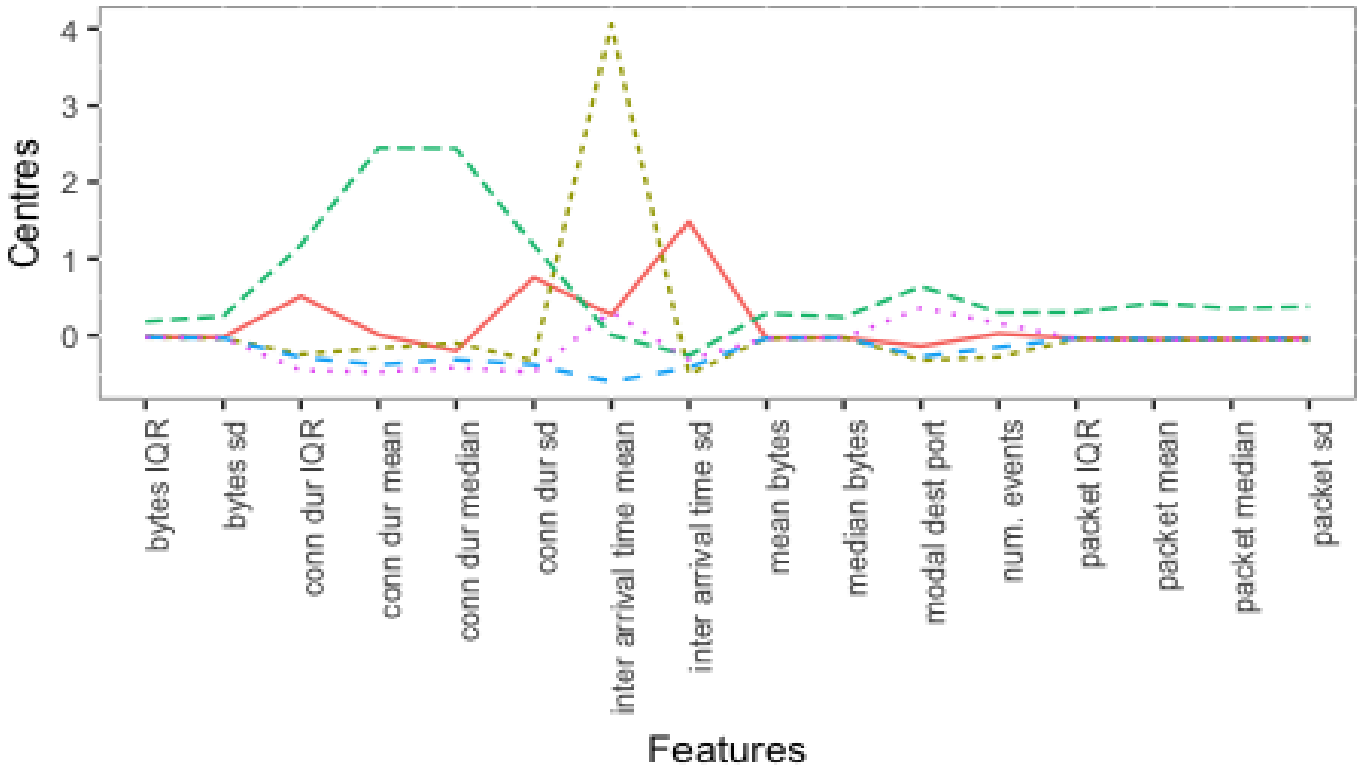
- Mean, median, standard deviation and IQR of *packets*.

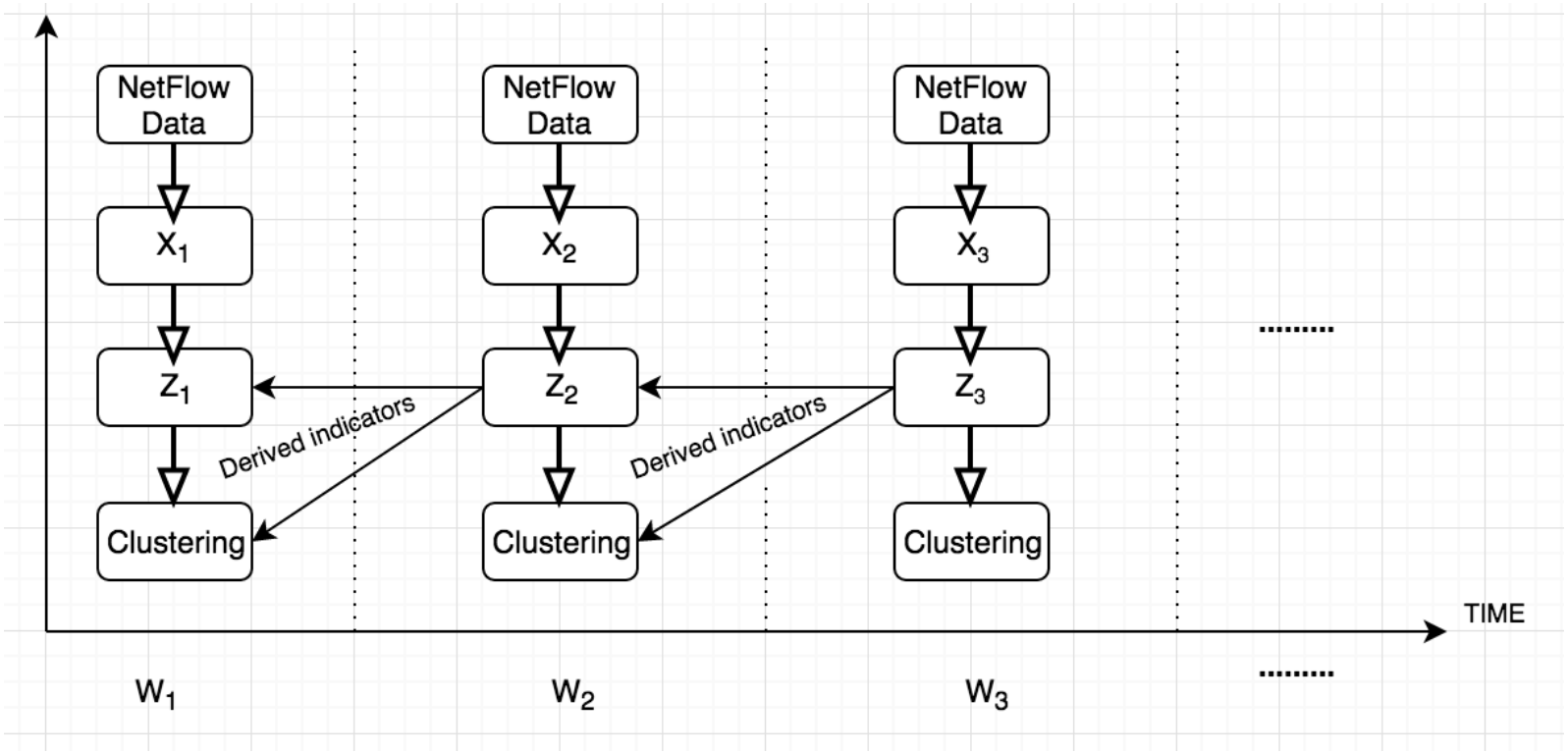
Other

- Count of the *number of events* that occured on the edge.
- Count of events on the *modal destination port* observed on the edge.

Clustering

- K-means clustering with 5 centroid vectors using standardised edge features over each of the chosen time windows.




NetFlow Data
NetFlow Data
NetFlow Data
X_1
X_2
X_3
Z_1
Z_2
Z_3
Derived indicators
Derived indicators
Clustering
Clustering
Clustering
TIME
W_1
W_2
W_3

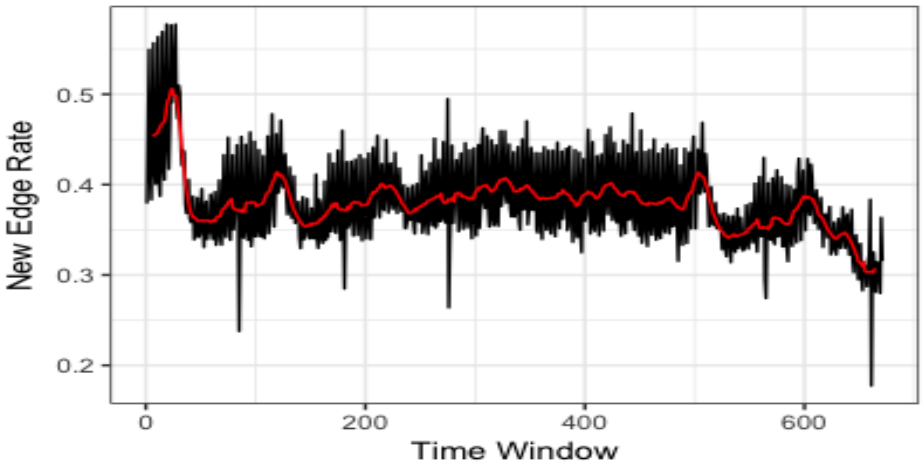
Derived Indicators

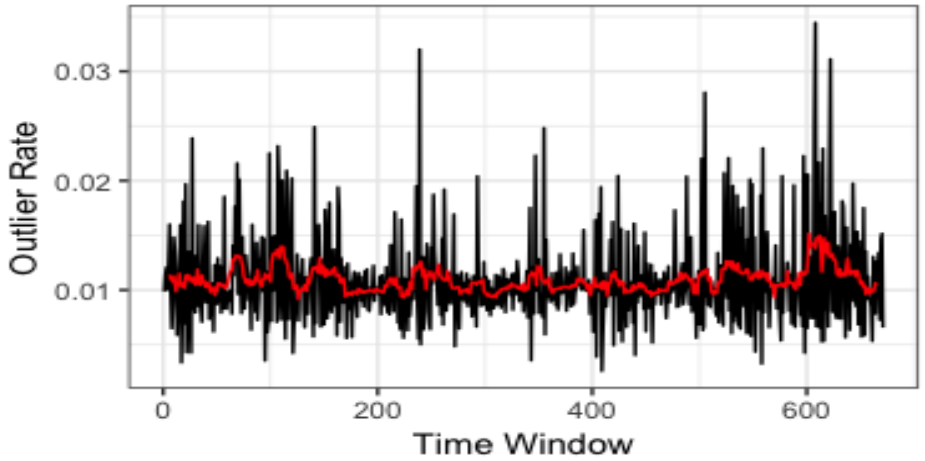
Indicator variables for each active edge of each time window are constructed. One each for:

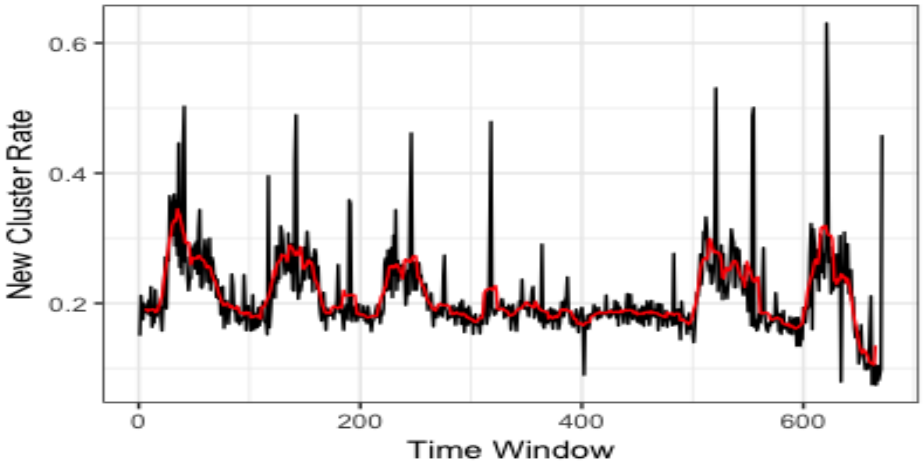
- Whether the edge was present in the previous time window.
- Whether the edge is an *outlier*.
- If the edge was present in the previous window, whether its cluster assignment has changed.

Where the 1 indicates a shift from the 'normal' behaviour.
By *burning in* the algorithm over a day's worth of data after the edge's first appearance, *control charts* can be constucted for the expected frequency of each indicator for each edge.

Global indicator value rates



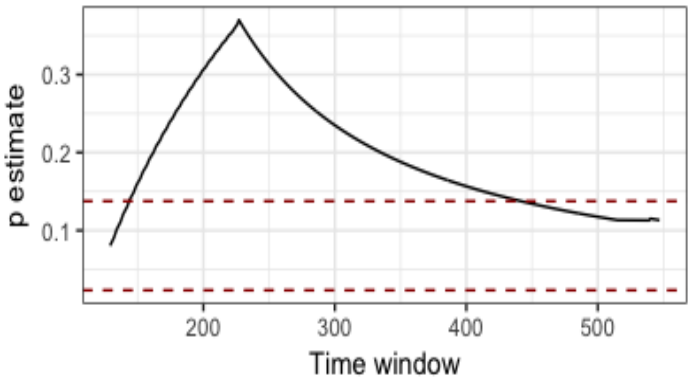


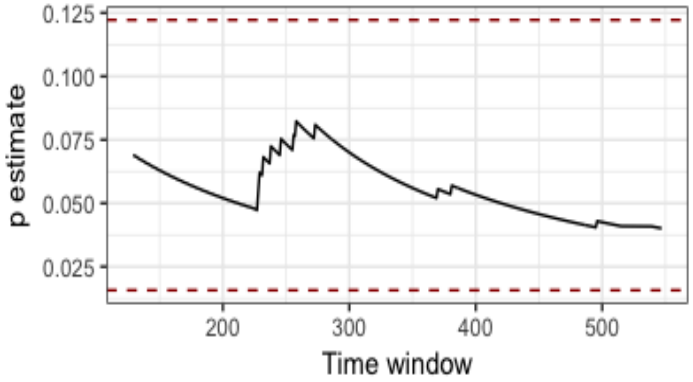


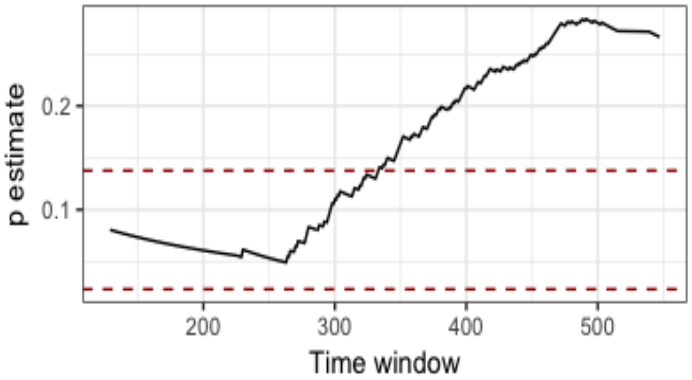
Cluster change rates show week-day seasonality. Outlier rates are noisy about 0.01.

Outline

Edge Control Charts



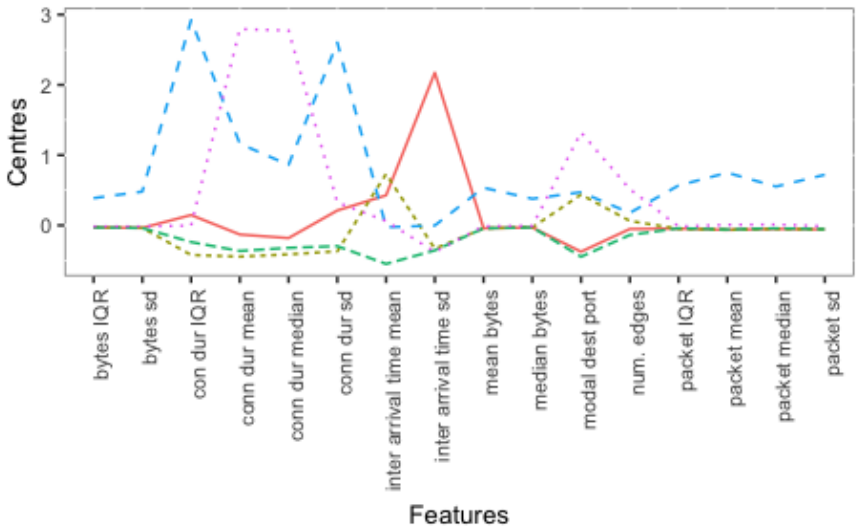


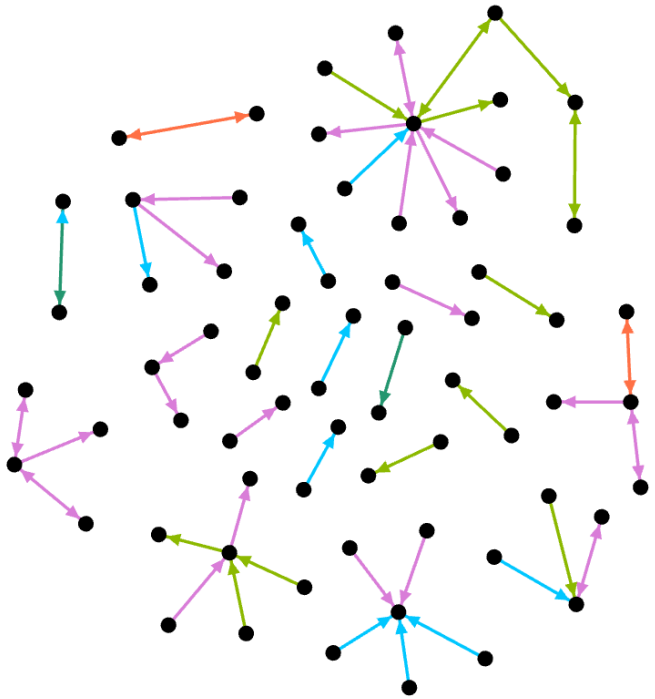


This edge shows changing new edge and outlier rates (upper and lower left), but constant cluster change rate (upper right).

Anomalous Edges

If the indicators for the edge sum to two (note: not three), then the edge is considered *anomalous*. Example window:





Outline

Conclusion

- We have developed a methodology that can monitor NetFlow traffic statistics in a real-time environment.
- We can construct different scales of data analysis: individual edges, specific subgraphs, and entire network edge-behaviour.
- This provides a foundation for a complementary enterprise network analysis tool for detecting abnormal edge behaviour.

Thank you

References

Valdes and K. Skinner. Adaptive, model-based monitoring for cyber attack detection. In *International Workshop on Recent Advances in Intrusion Detection*, pages 80–93. Springer, 2000.

J. Neil, C. Hash, A. Brugh, M. Fisk, and C. B. Storlie. Scan statistics for the online detection of locally anomalous subgraphs. *Technometrics*, 55(4):403–414, 2013.