A Scalable Network Event Detection Framework for Darknet Traffic

Max Gao Ricky K.P. Mok KC Claffy

UC San Diego/CAIDA

Introduction

- *Darknet traffic* is widely used to study internet-wide network events (e.g. censorship, malware outbreaks, DDoS impacts, etc.)
- Larger network telescopes collect more traffic. Processing it longitudinally at packet and flow resolutions is challenging.

Goal

- Evaluate the applications of ML-based approaches on detecting events of interest from time series derived from raw *darknet traffic*

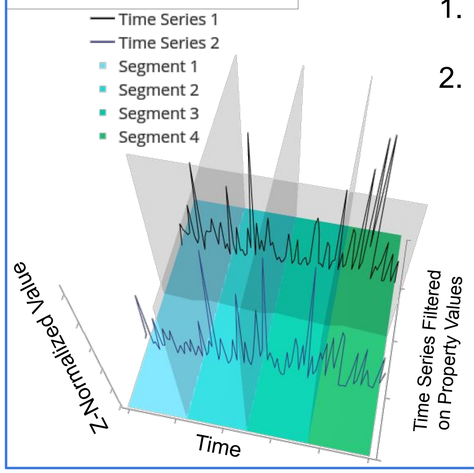
Approach

Property Set

- Origin ASN (pfx2AS)
- Country of Origin (NetAcuity)
- IP Protocol Number
- TCP/UDP Destination Port
- ICMP Type & Code
- Spoofing Inference

Metric Set

- Packets / minute
- Bytes / minute
- # of unique source IPs / minute
- # of unique source ASNs / minute
- # of unique destination IPs / minute

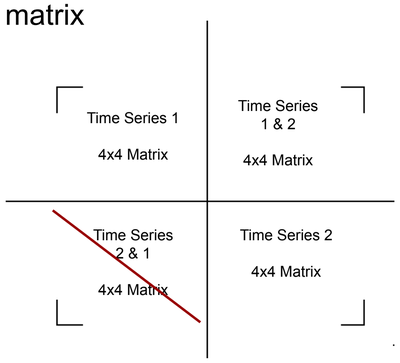


Distance Matrix Computation

1. Select *S*>=1 time series of length *T*, each filtered by a property value
2. Partition each time series by segment length *b*, yielding *S*N* total segments
3. Apply *Dynamic Time Warping* on each unique pair of segments for a *SNxSN* symmetric matrix

Evaluation

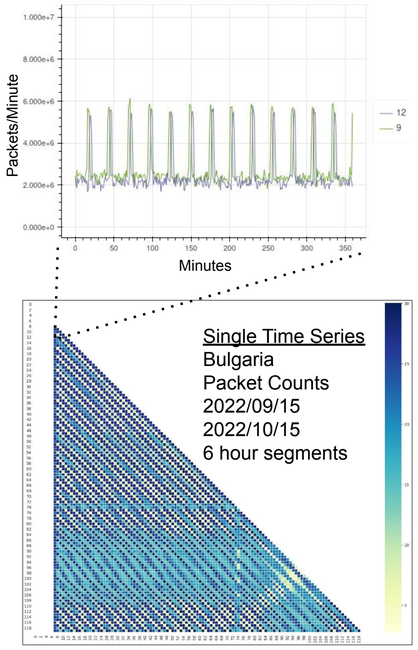
1. Apply Hierarchical Agglomerative Clustering to extract clusters of dissimilarity scores
2. Extract statistics for further analysis from subsets of the matrix



Preliminary Results

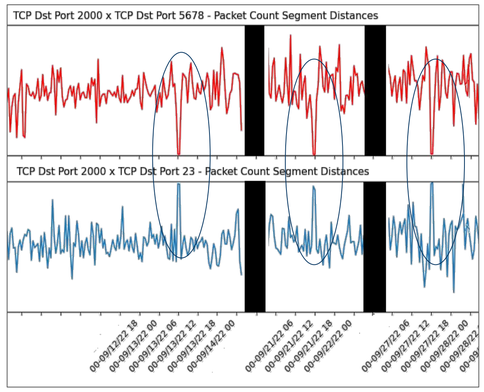
Type I Events

- Detection of temporal relationships in dissimilarity scores between a single time series' segment pairs



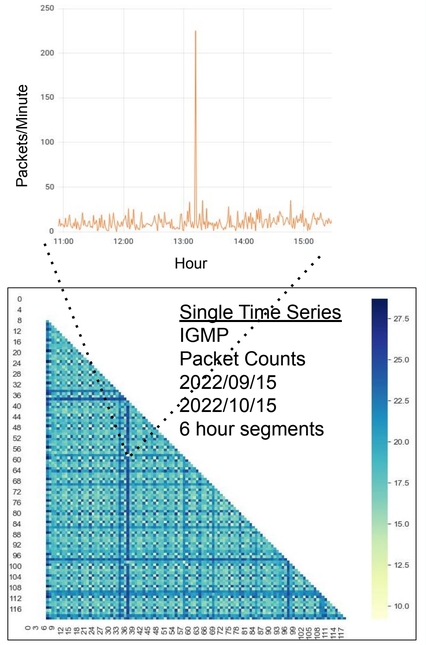
Type II Events

- Detection of temporally correlated dissimilarity scores in segments across multiple time series



Type III Events

- Detection of outlying dissimilarity scores within a set of time series segments



Discussion & Future Work

- The use cases we investigated are applicable for retrospective analysis and anomaly detection within darknet traffic
- In our future work, we plan to:
  - Comprehensively evaluate a large number of time series inputs
  - Automatically detect events, serving as a trigger for reactive measurements
  - Define events of interest, cross-evaluation with other internet datasets

SDSC UC San Diego

This work is based on research sponsored by the National Science Foundation grants CNS-2120399, CNS-2131987, CNS-1730661, and the Amateur Radio Digital Communications (ARDC) Foundation. The views are those of the authors and do not necessarily represent endorsements, either expressed or implied, of the sponsors.