Outlier Identification in Temporal Network Data

Kyle Boyer
Erik Sturcke

Motivations

In network security:

- Tools can flag potentially malicious traffic.
- Much of it can be blocked automatically.
- Often an analyst needs to evaluate anomalous behavior and if other corrective actions needs to be taken.

Goals and Research Questions

How can vis enhance anomaly detection and evaluation of temporal networking data?

Goals and Research Questions

How can we help analysts discover potentially malicious behavior?

- How can we give analysts context to evaluate?
- How can we scale both visually and data volume?

Data Abstraction

Network data from VAST Challenge 2013

- 1000+ hosts (servers & workstations)
- 2 weeks of network flow data (8GB)
- 1 week of IPS data (1.8GB)
- 2 weeks of host health data (2.8GB)

Data Abstraction

Host attributes

IP	categorical
Hostname	categorical
Type	categorical

Data Abstraction

Network flow attributes

Start time	quantitative (interval)
IP, port, protocol	categorical
Payload size	quantitative (ratio)
Duration	quantitative (ratio)

Data Abstraction

IPS attributes

Time/IP/Port	*(same as network flow)*
Priority	ordinal
Operation	categorical
Message	categorical

Data Abstraction

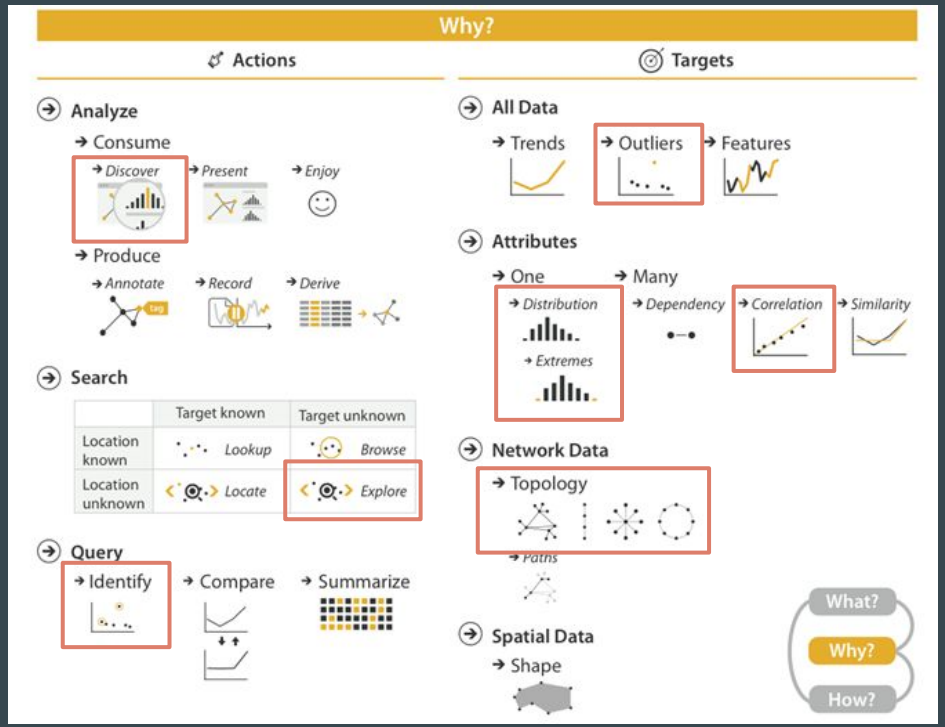
Health attributes

Hostname	categorical
Time	quantitative (interval)
Status	ordinal (+unknown)
Disk/mem/load	quantitative (ratio)

Dataset Type

Temporal network data

Task Abstraction

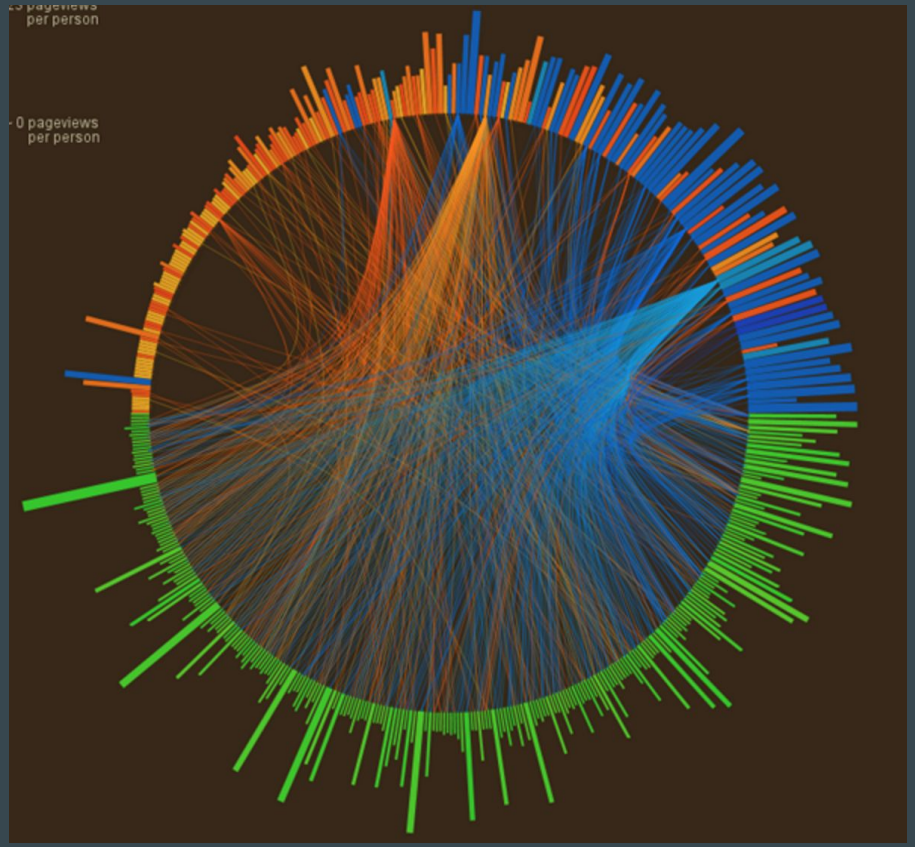


Proposed Methods: Overview Attributes

Important overview attributes

Host	Type
Network flow	Payload size, connection counts
IPS	Priority, operation
Health	Status

Proposed Methods: Chord Diagram (Hierarchical Tree)



Timeline and Action Plans

- Oct 24 — Survey and sanitize data
- | — Data server
- Nov 7 — Literature review
- | — Primary visualization
- Nov 14 — Dynamic queries
- Dec 5 — Details on demand
- Dec 7 — Presentation

Expected Outcomes and Contributions

A vis tool to assist a network security analyst in detecting and evaluating anomalous behavior.