Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

Model-Driven Situational Awareness (TAGs)

EMILIE PURVINE

Asymmetric Resilient Cybersecurity Initiative Roadshow

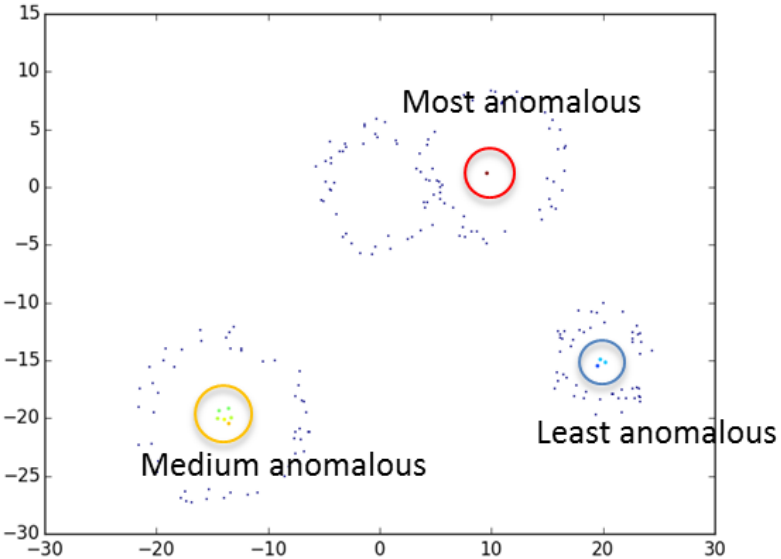
May 12, 2017

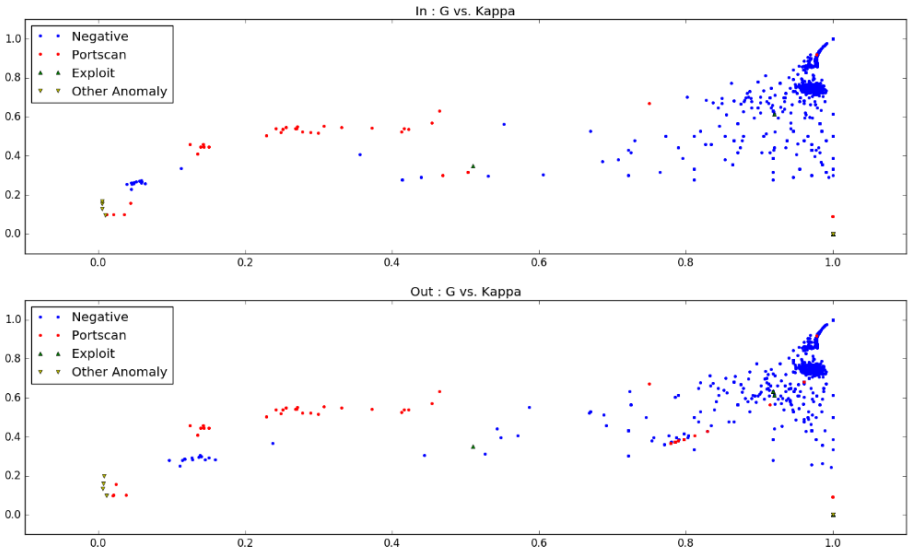
Need

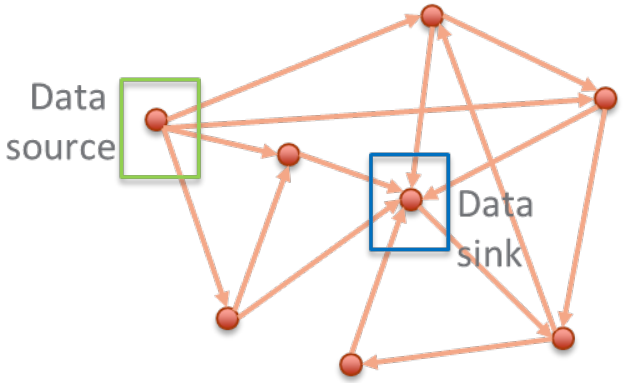
- How to get the most coherent picture of the state of our cyber system?
  - Each sensor returns data that can be analyzed with machine learning or other techniques – draw separate conclusions based on different data and perform correcting actions
- But current practice is not adequate, because…
  - Separate data sources and conclusions not being merged into a robust decision making capability
  - Network is treated as a graph, but there is so much more higher order and behavioral structure to take advantage of
  - Signature-based or machine learning methods require ground truth and training data which is often not available

Approach

- Study the shape of data to ***discover deviations from running normal behavior*** which is re-learned constantly
- ***Understand evolution*** of directed systems
- Cluster ***behavioral*** measures
- Example data sources:
  - Network flow
  - Firewall logs
  - Authentication records

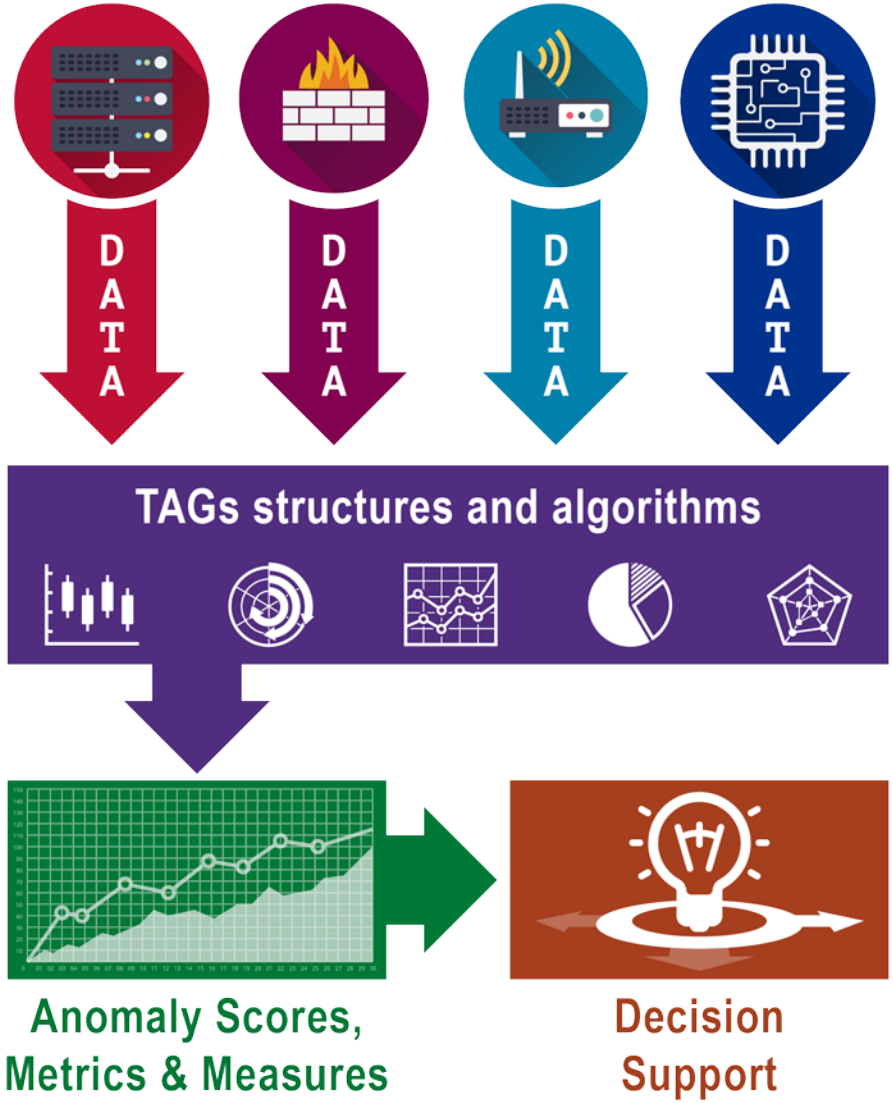






Benefit

- No training needed!
  - Discovers deviations from normal
  - Baseline tracked through time to account for time-varying properties of the system
- Extensible – multiple data types can augment current knowledge of the system
  - Data agnostic approaches
  - Flexible to additional domain expert knowledge
- Transparent algorithms provide repeatable methods and high confidence results



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Emilie Purvine
Principal Investigator
Topological Analysis of Graphs

emilie.purvine@pnnl.gov

Asymmetric Resilient
Cybersecurity Initiative

cybersecurity.pnnl.gov

Demo Materials

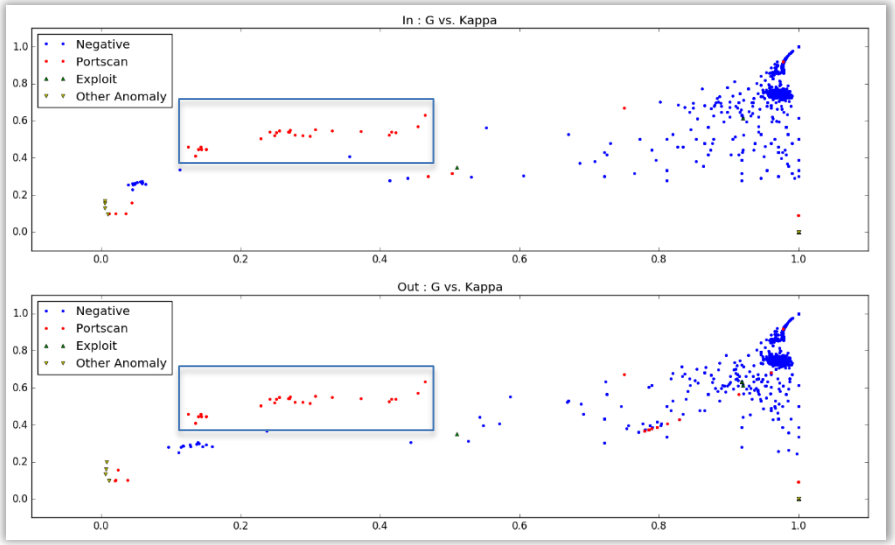
Video Demo – Setup

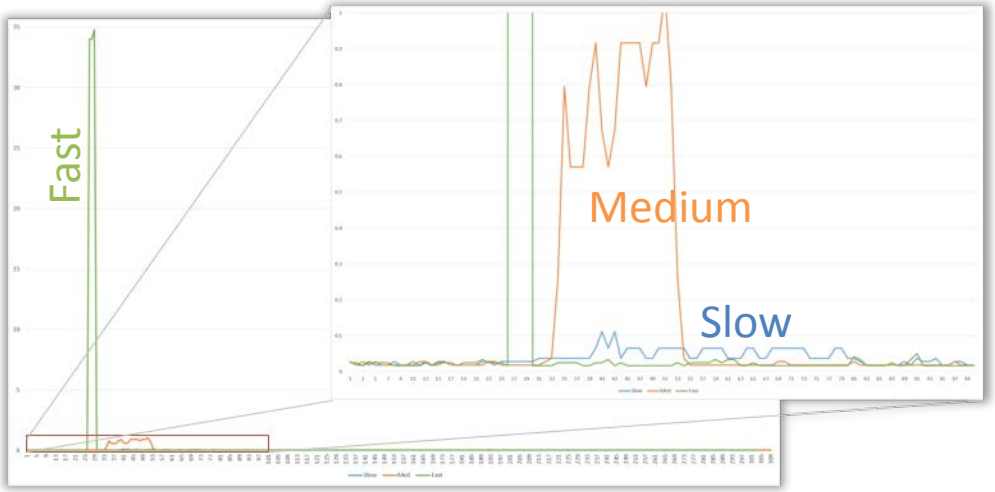
- Experiment
  - 7 VMs networked together
    - Web server, adversaries, and workstations
    - Scripts built to simulate browsing external web and the “corporate” site
  - Initial time period of network recon – nmap port scans
  - Some quiet period as attacker is “planning”
  - Remote buffer overflow exploit at the end to gain privileged shell, and change the root password
- Environment
  - MiiRCS environment ingested previously recorded NetFlow
  - Calculates running anomaly score using one of our TAGs algorithms
    - X axis = time
    - Y axis = anomaly score (this will adapt to the current values in the plot)
- Video is sped up to run in 30 seconds, actual time period is 52 minutes

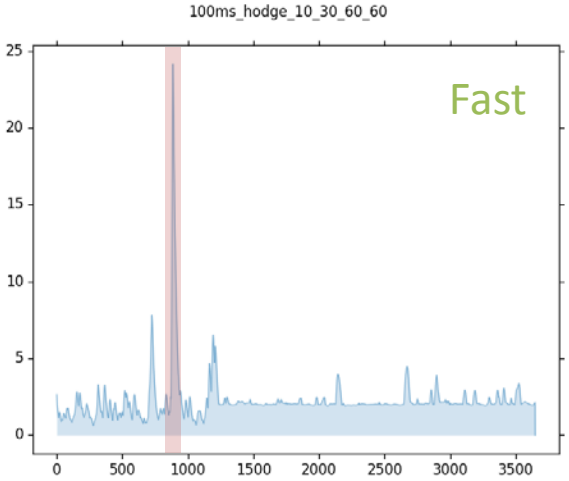
Video Demo

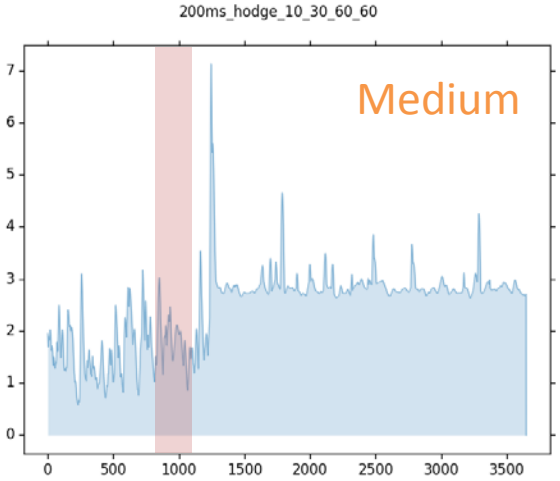
- 4:06 - 4:13 AM
  - netdiscover scan for local network discovery
- 4:14, 4:18, 4:20 AM
  - nmap scans to ping and list targets (no flow sent)
- 4:30, 4:31, 4:34, 4:35 AM
  - nmap TCP SYN scan with OS detection, version detection
- 4:39 AM
  - nikto scan to identify comprehensive vulnerabilities
- 4:40 – 4:50 AM
  - Some time would pass for the adversary to develop and test an exploit, then return to use it.
- 4:50 - 4:52 AM
  - The adversary returns, but does not find an open machine to exploit. Given their previous lack of success at IP hijacking, they opt for speed and deploy the exploit (in under two minutes!). A **remote buffer overflow** is used against the secure web server to gain a privileged shell, then change the root password, allowing the adversary access at any time.
  - /usr/local/src/OpenF*** 0x6b 192.168.0.109 443

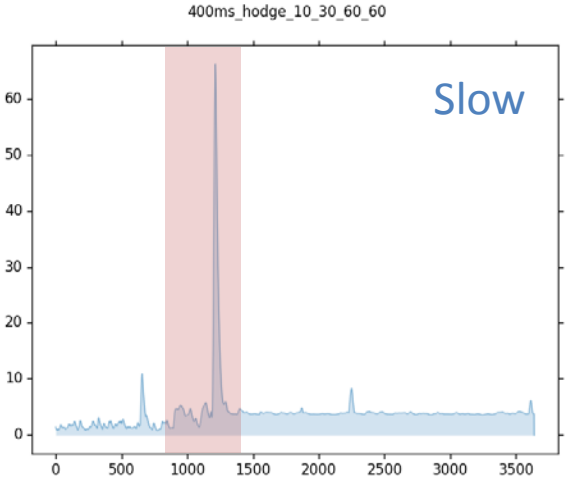
Port Scans


In : G vs. Kappa
Negative
Portscan
Exploit
Other Anomaly
Out : G vs. Kappa
Negative
Portscan
Exploit
Other Anomaly


Fast
Medium
Slow


100ms_hodge_10_30_60_60
Fast

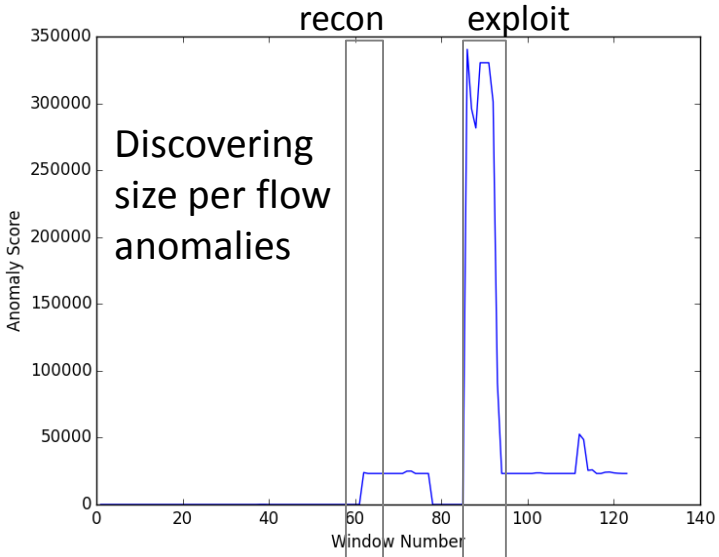

200ms_hodge_10_30_60_60
Medium

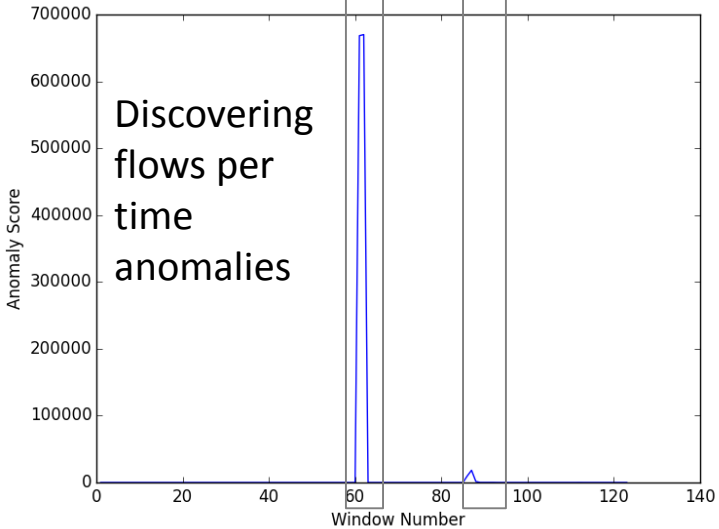

400ms_hodge_10_30_60_60
Slow

Kill Chain

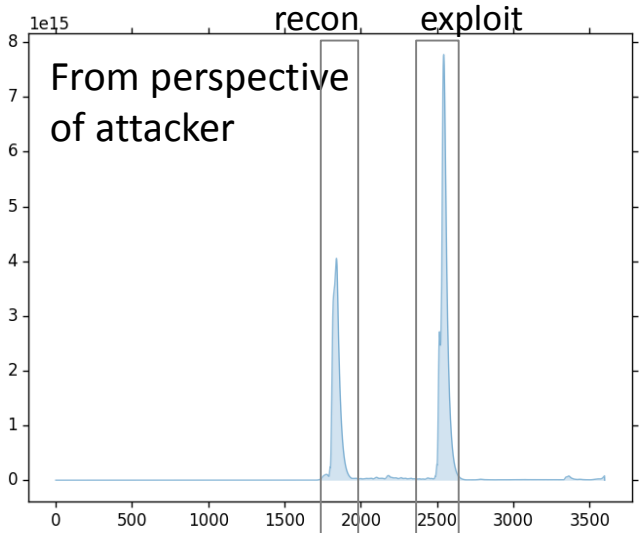
Recon = nmap -sS -v -A -PN 192.168.0.1
Exploit = ./OpenF*** 0x6b 192.168.0.109 443

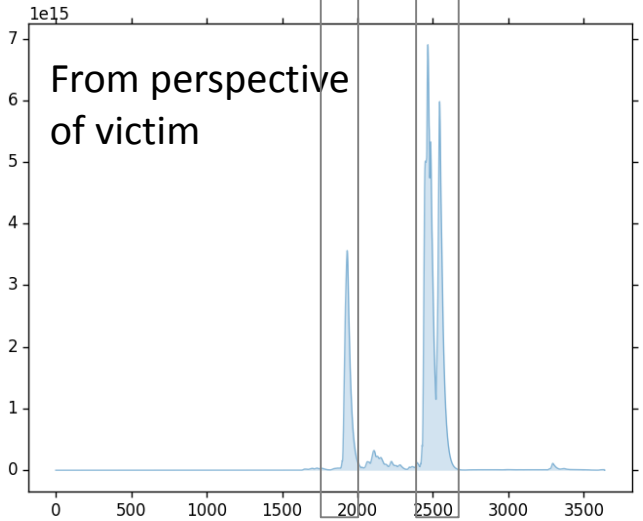
Persistent Homology





Hodge Decomposition

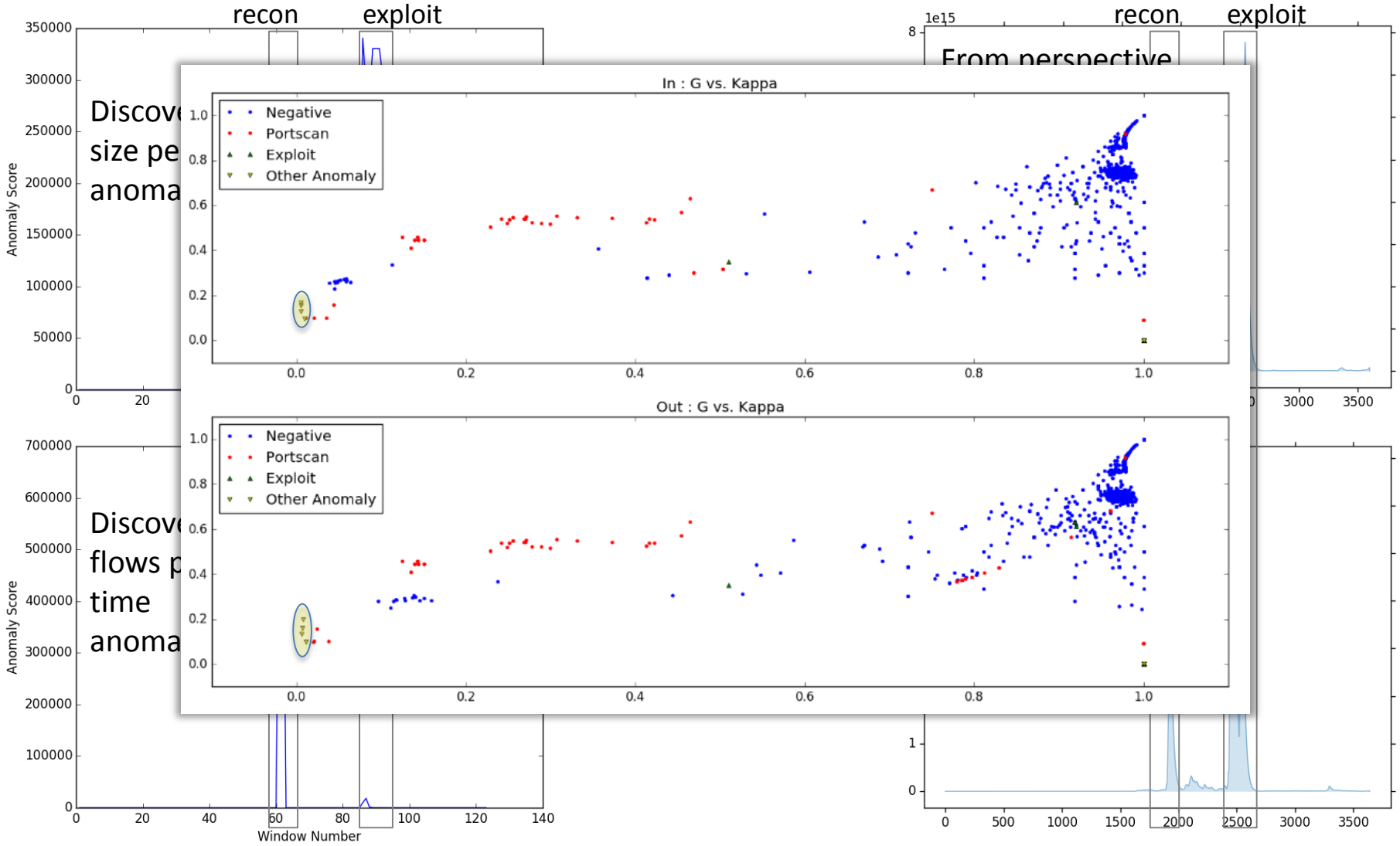




Kill Chain

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Recon = nmap -sS -v -A -PN 192.168.0.1
Exploit = ./OpenF*** 0x6b 192.168.0.109 443



Firewall Data – Intro

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Access to “live” PNNL firewall logs from network via our Living Laboratory initiative
- Batched into 1 hour blocks
- We present analysis of one hour 11/29/2016 Midnight to 1:00am
  - 15,046,500 records
  - Consider 1 minute windows, every 30 seconds. Baseline sampled 50% of windows from first 6 minutes

Description:

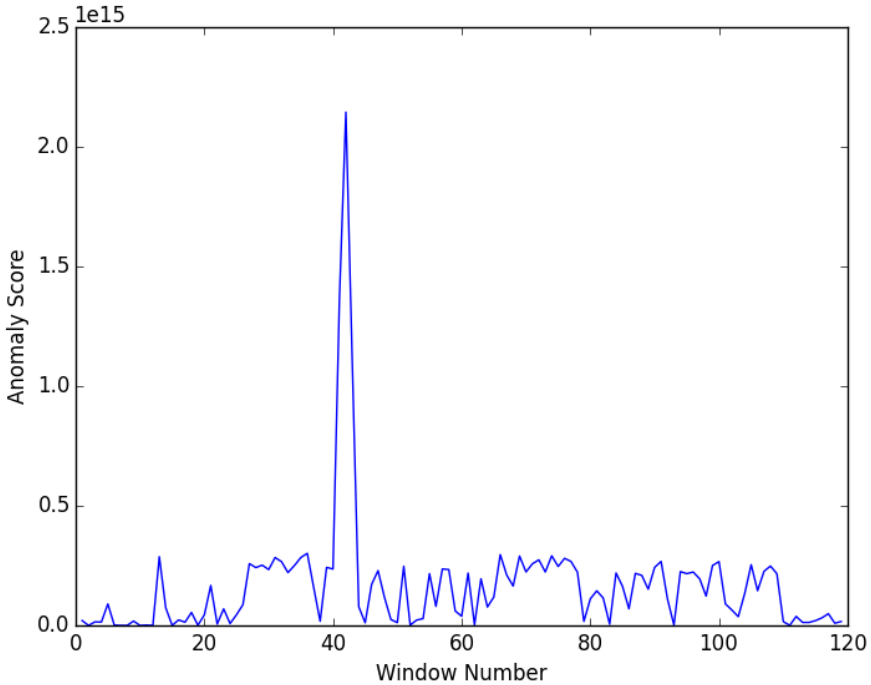
This is the data dictionary of PNNL Firewall data as it flows out of Splunk and into the CyberFit Data Warehouse.

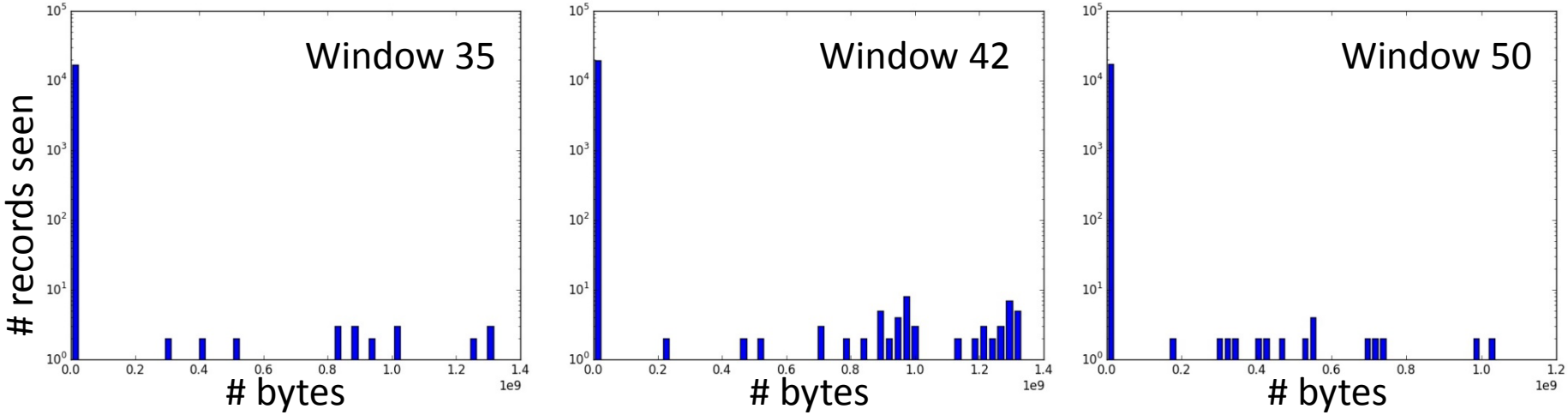
Data Definition:

Column Name	Type	Description
_TIME	DATETIME	Event Time
DATE_ZONE	INTEGER	Timezone offset? (Ex. -420)
DURATION	TIME	Duration of connection captured on teardown. HH:MM:SS
DVC	VARCHAR	IP address for the ASA firewall device
ID	BIGINT	Thought to be the unique connection id number. Probably unique to the DVC
LOG_LEVEL	INTEGER	Log Severity Level 1 - Alert Messages 2 - Critical Messages 3 - Error Messages 4 - Warning Messages 5 - Notification Messages 6 - Informational Messages 7 - Debugging Messages
MESSAGE_ID	INTEGER	Syslog Message Number
EVENTTYPE	VARCHAR	Type of event
TRANSPORT	VARCHAR	Transport protocol (UDP, TCP)
SRC_IP	VARCHAR	Source IP address
SRC_PORT	INTEGER	TCP/UDP source port number or equivalent
SRC_INTERFACE	VARCHAR	Identifies the interface or Zone the source is coming from
DEST_IP	VARCHAR	Destination IP address
DEST_PORT	INTEGER	TCP/UDP destination port number or equivalent
DEST_INTERFACE	VARCHAR	Identifies the interface or the Zone the destination for the connection
VENDOR_ACTION	VARCHAR	Identifies whether this event is a build up or teardown
BYTES	BIGINT	Total Bytes transmitted
REASON	VARCHAR	Reason for the connection teardown

Firewall Data

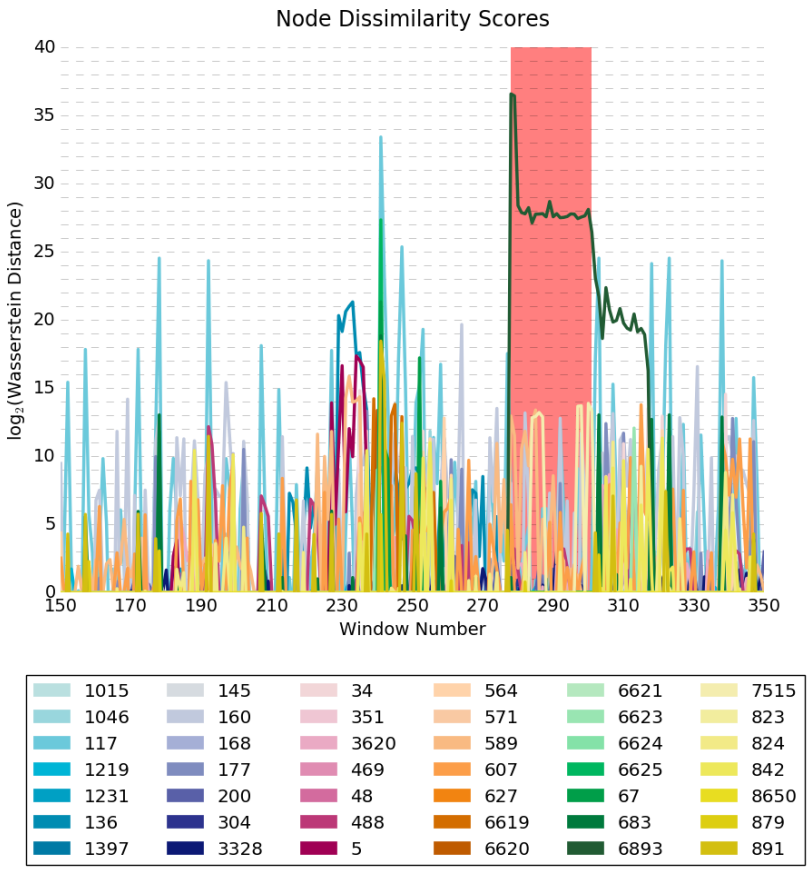
- Large anomaly score in windows 41 – 43 correspond to rise in traffic to port 53 with large byte counts
- Histograms of all destination port 53 firewall logs by # of bytes





BitTorrent Detection

- Network flow for a single building was captured[1]
  - BitTorrent traffic added after the fact by node 6893 during windows 278-301
- PH methods were able to detect an anomaly from 6893 during the correct time window
  - Used graph information derived from network flow in cited paper



[1]Christopher R. Harshaw, Robert A. Bridges, Michael D. Iannacone, Joel W. Reed, and John R. Goodall. 2016. GraphPrints: Towards a Graph Analytic Method for Network Anomaly Detection. In Proceedings of the 11th Annual Cyber and Information Security Research Conference (CISRC '16).