Defeating the PCAP Problem: Making a Mountain into a Molehill

Leigh Metcalf PhD, Daniel Ruef, Dillon Lareau, Angela Horneman

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213
Who am I?

• PhD in Theoretical Mathematics, specializing in Algebraic Topology and Dynamical Systems
• Researcher at CERT for 9 years
• Primary Author of Applied Mathematics and Cybersecurity
• Co-EiC of the new journal ACM Digital Threats: Research and Practice
• My work is challenging assumptions made in Cybersecurity and evidence based analysis
Analysis Pipeline…

…is not a spiffy User Interface.

It was developed to drop files for spiffy user interfaces to read.
What Can Analysis Pipeline Do?

- Narrow down where to look for bad guys
- Keep Inventory
- Network Profiling – Keep state
  - For example, suppose a system suddenly starts accepting web requests and it has never done that before
- It will only put actionable alerts into Security Information Management Systems
  - Can emit alerts in JSON for easy consumption
- Has thresholds for anomalies… custom thresholds even
- Can help you focus your analysis
- Analysis Pipeline uses SiLK concepts
Analysis Pipeline

• On an order of magnitude better than an IPS in certain cases
  • For example, adding an IP to SNORT means writing a new rule
  • Adding an IP to the Analysis Pipeline monitoring is simple and efficient and doesn’t change the inherent configuration
• For watchlists, it is incredibly efficient in processing speed and resource utilization. It costs almost nothing to add an indicator to a watchlist
• Works well in tandem with IPS
What Can’t Analysis Pipeline Do

• Still isn’t a cute User Interface
• And isn’t a SIM
• It’s only limitation is the amount of memory and disk on your system
• It also isn’t an AI/ML tool. You can use it to label data for AI/ML but it isn’t a flow passthrough tool to label all data.
Analysis Pipeline

**FLOWS**

**FILTERS**
- Collect notable/desirable traffic for further analysis

**INTERNAL FILTERS** (Stateless Updaters)
- Make lists of values from flows deemed important

**EVALUATIONS**
- Independent state for each check comprising an eval
- Output Lists
- Bulk of time and memory costs
Analysis Pipeline Configs

The configurations consist of two parts:

Filter – The step that determines if the record should be passed to an evaluation

Examples: dIP, sIP, Ports, Protocol, Domain names, etc

Evaluation – The step that does the work

Examples: statistics, thresh holds, moving time windows

Evaluation also handles the alerts
Case Study #1

1) Only worry about tcp traffic
2) Identify IP addresses that send more than 1MB in 5 minutes
Case Study #1

FILTER tcpFilter
    PROTOCOL == 6
END FILTER

EVALUATION CaseStudy1
    ALERT EVERYTHING
    FILTER tcpFilter
    CHECK THRESHOLD
        SUM BYTES > 1000000
        TIME_WINDOW 5 MINUTES
    END CHECK
END EVALUATION
Complexifying…

Identify web servers

• Sends more than 20,000 bytes in 10 minutes
• Talks to more than 15 different DIPs in 15 minutes
FILTER ourNetwork

SIP == 193.0.0.0/8 #arbitrarily chosen network

SPORT IN LIST [80, 8080] #Yes, there are more ports, but 80 and 8080 keep it simple

END FILTER
Configuration Continued

EVALUATION findWebServers

SEVERITY 2
OUTPUT_TIMEOUT 1 DAY
OUTPUT_LIST SIP
webServerList
CLEAR ALWAYS
ALERT EVERYTHING
ALERT 1 TIMES 1 HOUR
FILTER ourNetwork

FOREACH SIP

CHECK THRESHOLD
DISTINCT DIP > 15
TIME_WINDOW 15 MINUTES
END CHECK
CHECK THRESHOLD
SUM BYTES > 20000
TIME_WINDOW 10 MINUTES
END CHECK

END EVALUATION
Case Study #2 DDoS Detection

What is DDoS?

What are the characteristics of flows that participate in a DDoS attack?

What is the threshold event that causes us to say a DDoS attack is occurring?

Are there exceptions to the threshold?

Is this behavior ever expected and if so, what do we whitelist?
DDoS DNS Reflection

Rather than finding if we’re the victim of a DDoS

...Let’s look at if we’re the inadvertent part of one.

DNS Reflection attack detection
DDoS DNS Reflection

FILTER dnsFlows

   PROTOCOL IN_LIST [6,17]
   ANY PORT == 53
   ANY IP IN_LIST "whiteListDns.set"

END FILTER
DDoS DNS Reflection

EVALUATION dnsThreshold

FILTER dnsFlows

CHECK RATIO

OUTGOING 27 TO 1

END CHECK

FOREACH SIP DIP

CHECK THRESHOLD

RECORD COUNT > 100

TIME_WINDOW 5 MINUTES

END CHECK

OUTPUT_TIMEOUT 5 MINUTES

ALERT ALWAYS

CLEAR ALWAYS

END EVALUATION
Another Example

EVALUATION beacon

CLEAR NEVER
SEVERITY 3
ALERT JUST_NEW_THIS_TIME
ALERT 1 TIMES 60 SECONDS
FILTER all
CHECK BEACON

COUNT 5 CHECK_TOLERANCE 5 PERCENT
TIME_WINDOW 300 SECONDS

END CHECK

END EVALUATION
Analysis Pipeline

These Examples look at Network Flow as the input.

Analysis Pipeline also takes input from YAF (IPFIX-based file format) and can operate on that as well. In fact, it works on all flat IPFIX Data

Analyzing:

• User agent strings
• Domain names
• URLs…
• ….anything you can put in IPFIX…
Pipeline Can Also..

- Find system misconfigurations
- Find DNS Parking and Fast Flux
- Watchlists
- Find Sensor Outages
- Passive FTP detection
- Network Profiling
- Baseball Statistics
- ....
Analysis Pipeline Primitives

- Arithmetic Basics (Count / Sum / Average / # Distinct)
- Statistics Basics (Mean / Standard Deviation)
- Moving Average Windows
- Beacon Detection
- Persistence of Value Detection
- …And many more!
Analysis Pipeline

Analysis Pipeline:

SiLK:
https://tools.netsa.cert.org/silk/index.html

YAF:
https://tools.netsa.cert.org/yaf/index.html

Email for help, requests, other: netsa-help@cert.org
Thank you for your time!

Questions/Comments?