

Parallel Processing in Netflow Data Fusion

- George Saylor
- Michael Rash
 - G2, Inc.
- Flocon, 2010



Agenda

- Setting the stage – a description of factors
- Environment
- Parallel processing to facilitate:
 - Fusing large netflow dataset with other large datasets
 - Joining netflow and other data to enrichment data such as IP reputation and GeoIP lists
 - Incident investigation on long time ranges using fused datasets



Agenda cont'd

- Specific use cases
 - Social Networks for Computers and Networks
 - Long Term Data Exfiltration
 - Comparing netflow, system characteristics, and time to determine impact
 - DNS Fast Flux detection and associated botnet detection
 - SPAM behavior on a given network



Setting the stage

- Targeting very large environments (ISP)
 - > than 2 trillion netflow records per year
 - Existing large siloed data stores of:
 - Large netflow database
 - Assets – with history
 - System characteristics – with history
 - DNS queries/responses
 - Large store of PCAP data



Setting the Stage cont'd

- Goals

- Fuse this data together
- Add enrichment data
 - IP reputation data/Black Lists/White Lists
 - Malware domains
 - Google Safe Browsing API
 - Spamhaus
 - Dshield
 - GeoIP



Setting the Stage Cont...

- Goals Cont...
 - Unified query interface (instead of scripting)
 - Ability to run against big datasets (without hurting anyone else)
 - Iterative discovery by pivoting on results of previous queries
 - Large scale summary and computation
 - Beacon detection over long time ranges
 - Correlating with other indicators
 - System Level Social Network Analysis



Environment

- Selected a computing cluster that was optimized for:
 - Very large joins
 - Large data searches
 - Fast indexed access to data
 - Index building does not affect end users
 - Index builds on one system, distributes to queried nodes
 - Ability to partition on values and ranges



Use Case: Social Network Growth

- Social Networks for Computers and Networks
 - Given an IP address or addresses, or network range characterize relationships between those and other computers or networks
- Scenario:
 - An IDS alert provides some indication that a given host is compromised
 - We may want to know who this machine talks to
 - Whether the machine's "friends list" (other systems) is growing rapidly



Use Case: Data Exfiltration Discovery

- Social Network prioritization based on bytes transferred
- Scenario:
 - An IDS alert provides some indication that a given host is compromised.
 - Cross reference attacker's IP with the GeoIP and other reputation databases.
 - Construct the social network of other systems this IP communicates with.
 - Sort by bytes transferred – the top systems in this list may be candidates for data exfiltration nodes.



Use Case: SPAM Node and C2 Discovery

- SPAM nodes and mail server error code thresholds
- Scenario:
 - IDS signatures for thresholding large numbers of recipient unknown error codes sent back from mail servers to local systems.
 - Build system level social network concentrating on command and control communications and other observables.
 - Try to identify command and control connections – may become obvious with overlapping external IP's in SPAM node communications



Use Case: DNS Fast Flux

- Storing large number of DNS requests and responses
- Scenario:
 - Use thresholding operation for the number of IP addresses given in response to DNS requests for the same hostname.
 - Also use low TTL values in DNS responses as a search metric.
 - Look for “unusual” DNS hostnames that use multiple numbers and strange patterns (e.g. “f4n3upyhqj.com”).



Use Case: Complex Fusion

- Mixing together numerous data sources to gain better situational awareness
- Scenario:
 - Retrieve all flows where:
 - For a given IP list (derived from previous exercises for example).
 - Platform is Windows XP and the patch level is below a minimum level or there is no netflow record to the Windows Update service to indicate that it has been patched.
 - Exhibits communications to disreputable hosts and/or suspicious geo locations.
 - Compute this via a single interface.



Questions?

george.saylor@g2-inc.com

michael.rash@g2-inc.com

