Enabling Network Visibility and Security through Tensor Analysis

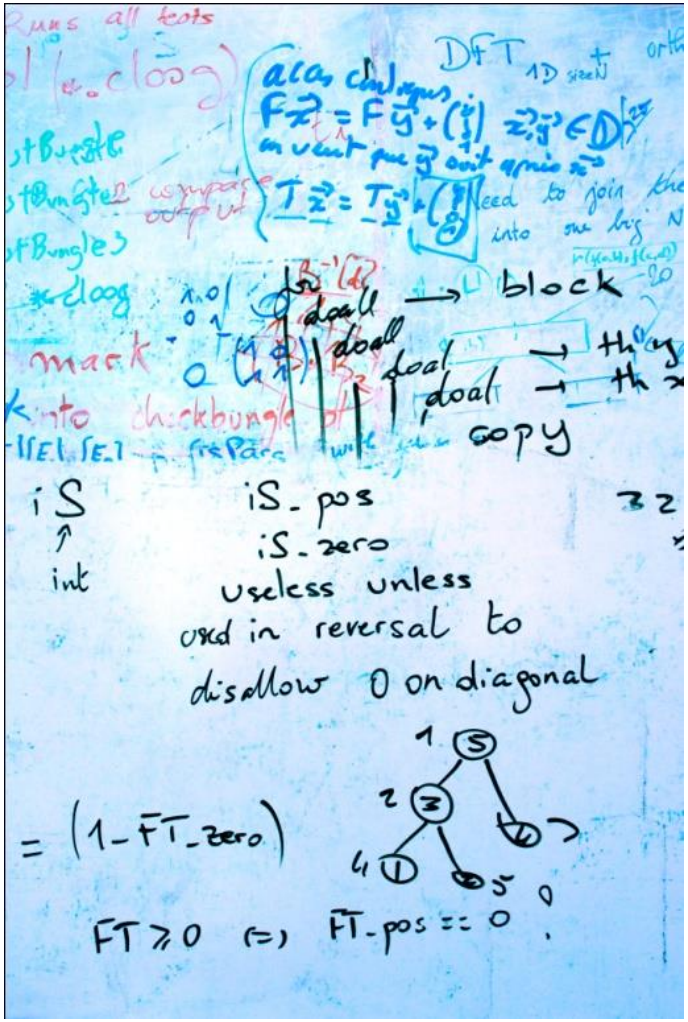
INDIS 2017

Muthu Baskaran
David Bruns-Smith[#]
Thomas Henretty
James Ezick
Richard Lethin

Reservoir Labs

[#] UC Berkeley (Work done at Reservoir Labs)

12 November 2017



www.reservoir.com

Value Add to the Security Ecosystem

We have tools that excel at monitoring flows

- Signature-based tools that provide real-time alerts
- More advanced metadata collection tools enable deeper offline analysis

Challenges

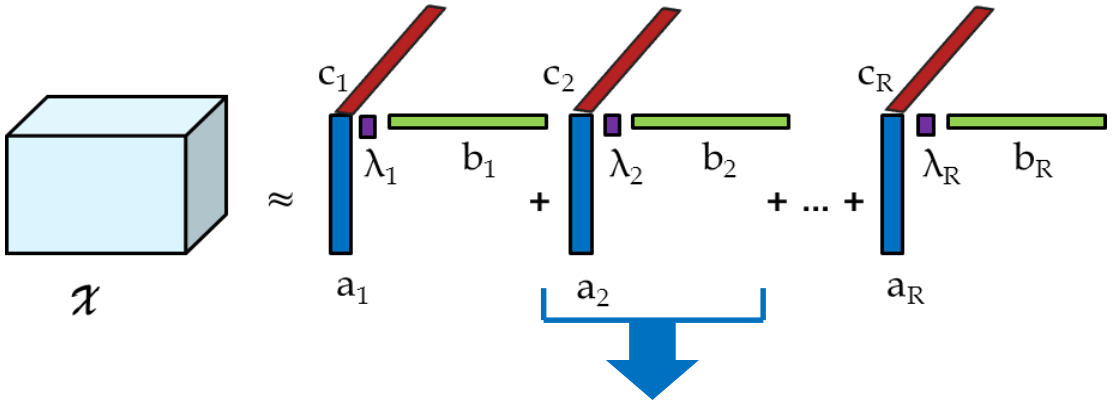
- Rules can be anticipated and evaded
- Metadata analysis is a forensic activity
- Both require a known starting point

Question

- Can we supplement traditional incident-focused approaches to threat discovery with an approach that feeds metadata to a pattern-focused analytic?

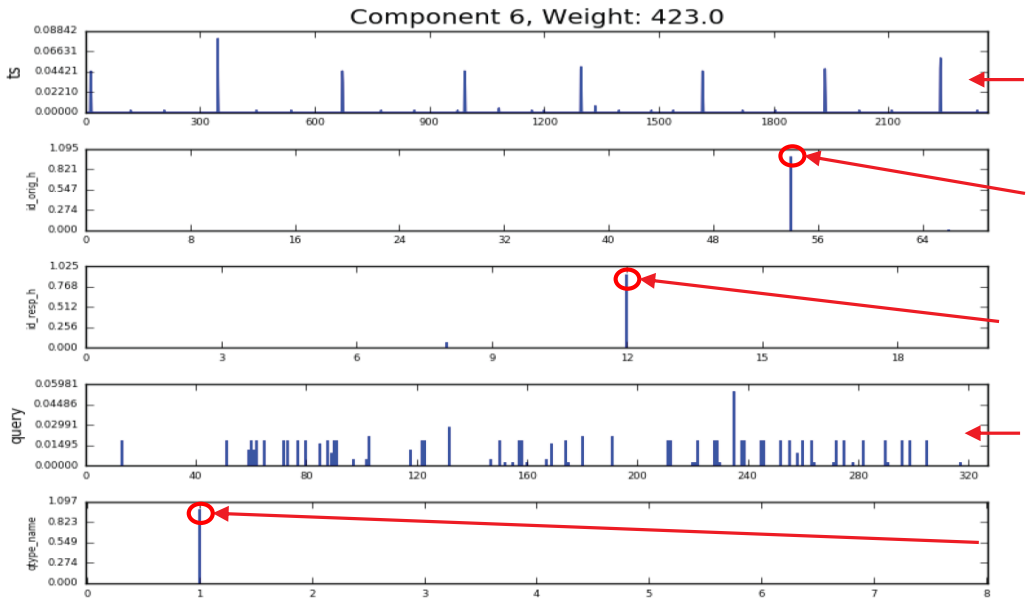
Tensor Analysis and Tensor Decompositions

A New Paradigm for Network Analysis



Unsupervised learning

Detects unknown unknowns



Regular time intervals

Single user PC

Network primary DNS server

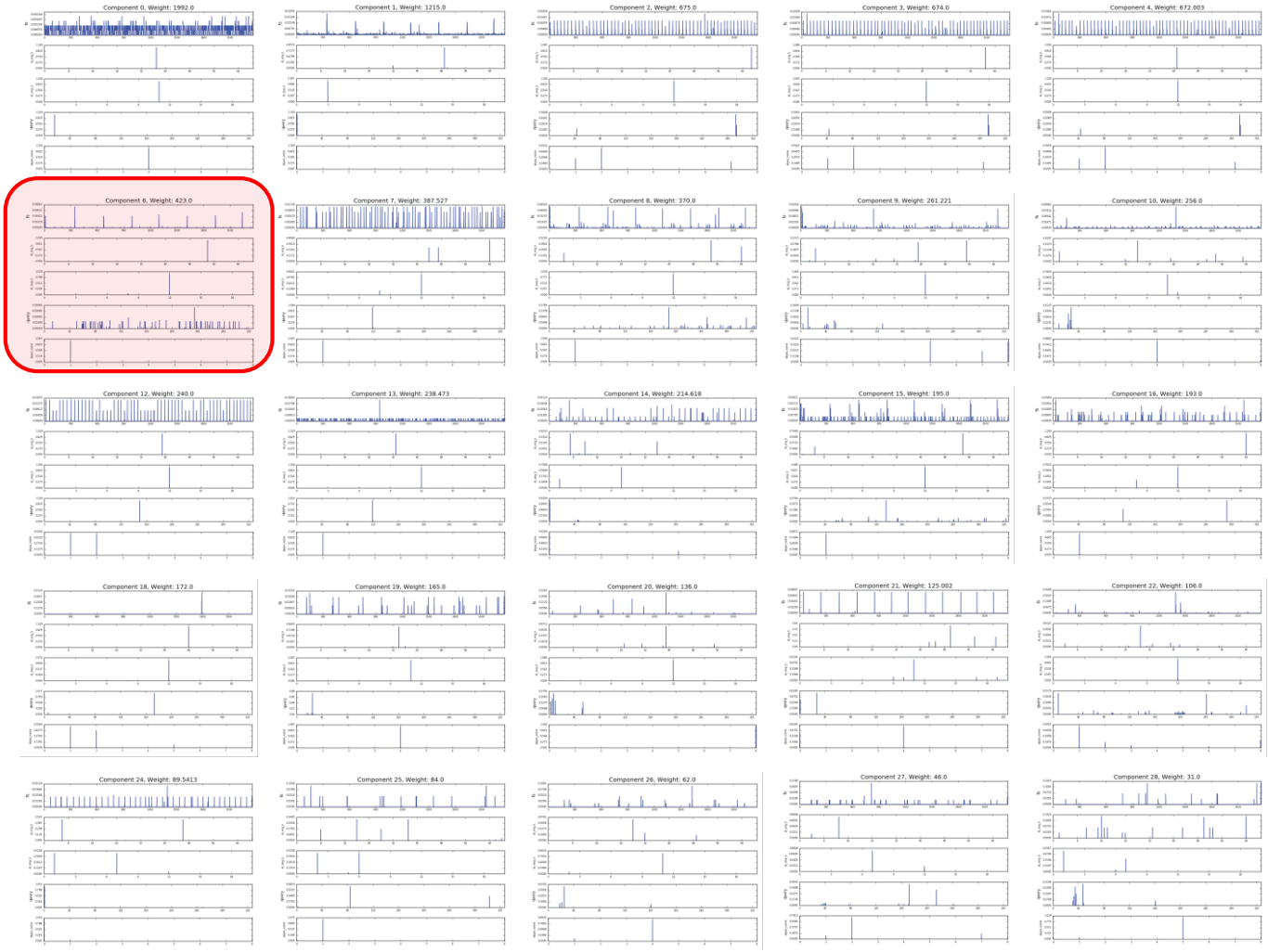
The New York Times

IPv4 traffic

Captures coherent patterns of activity spanning multiple dimensions

Tensor Analysis and Tensor Decompositions

A New Paradigm for Network Analysis



Lesser cognitive load for analyst

Looks at fewer components (indicating activities of interest)

Use the patterns of activity to guide further investigation

CANDID and ENSIGN: Context and Overview

CANDID: A tool for network security and traffic analysis

- Provides comprehensive and contextual insights into the network
  - Malicious and obfuscated network threats
  - Network state and network performance indicators
- Reduces the cognitive load of network analysts

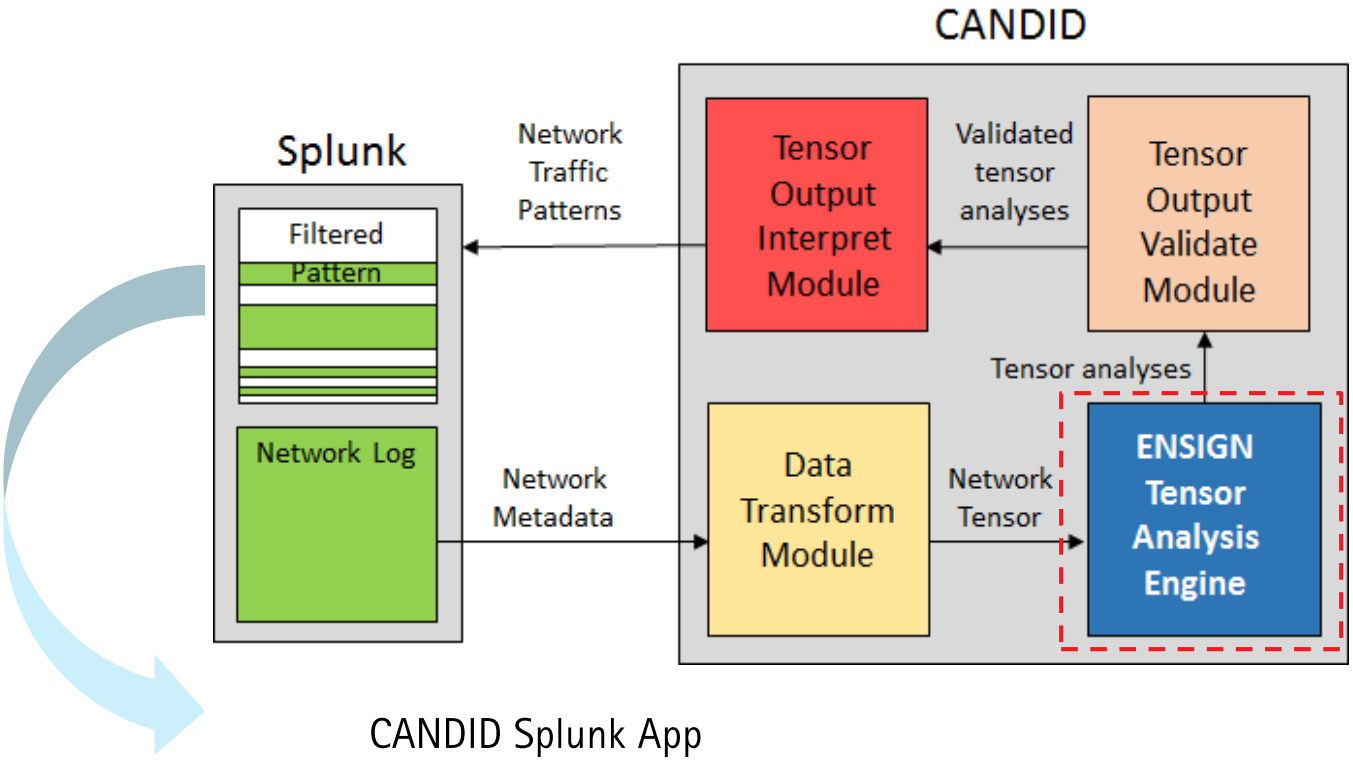
ENSIGN: High-performance tensor analysis engine driving CANDID

- Tensor Toolbox with advanced mathematical methods for data analysis
- High performance, rich capability, easy usability

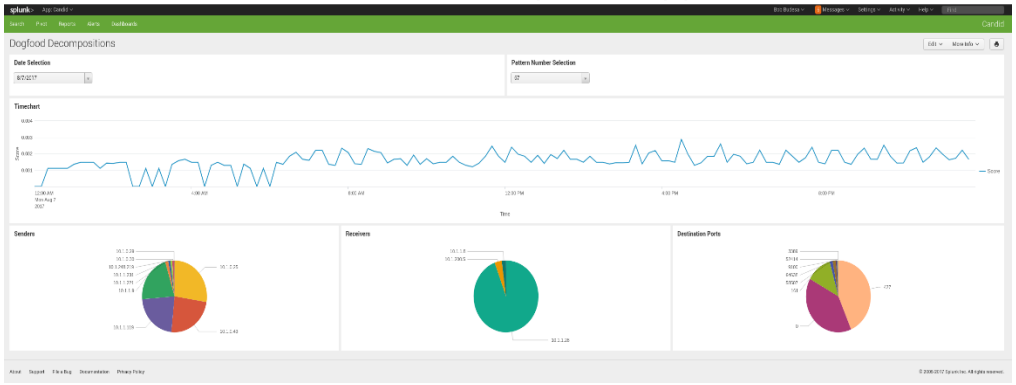
Successfully used in diverse operational environments

- Security Operations Center (SOC) for the SCinet network at SC16
- Reservoir Labs' Local Area Network (LAN)

CANDID : Tool workflow

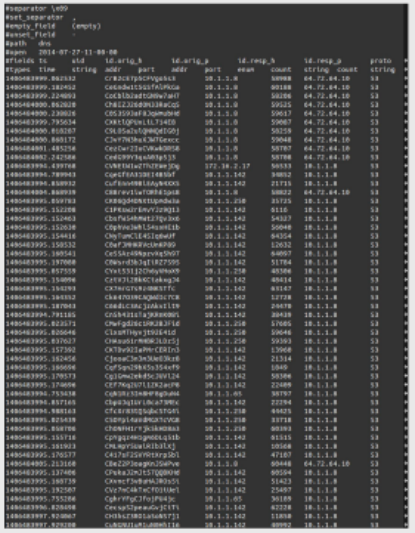


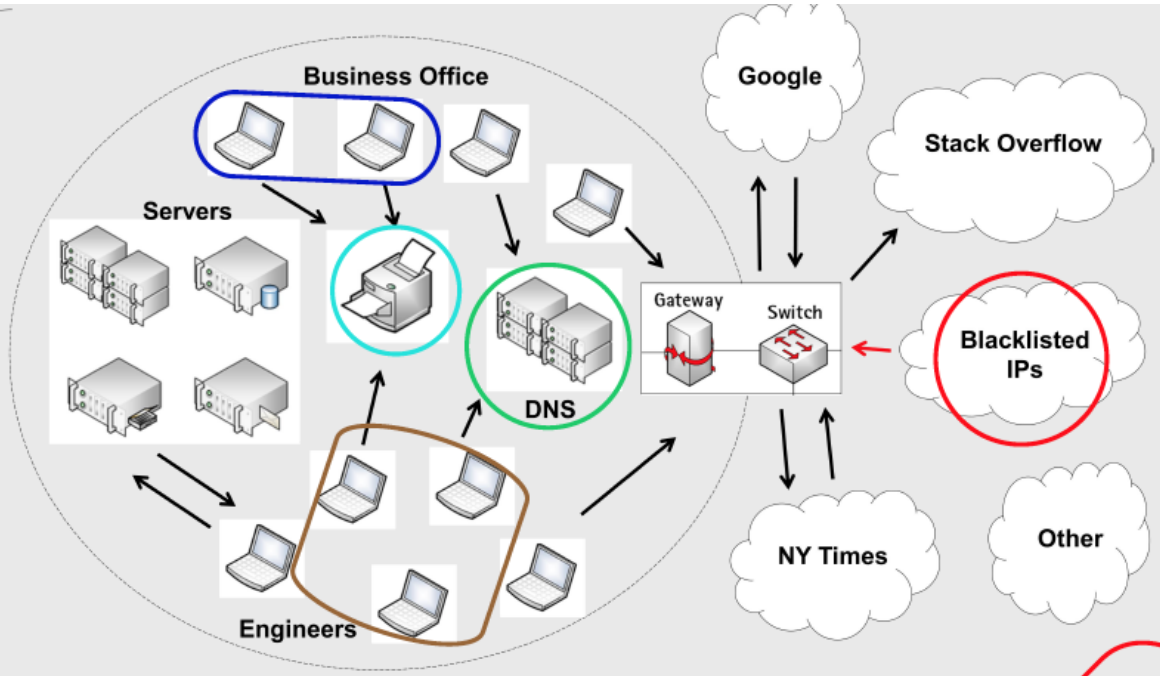
CANDID Splunk App

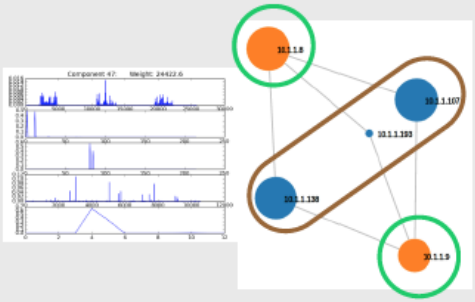


CANDID on Reservoir Network Traffic

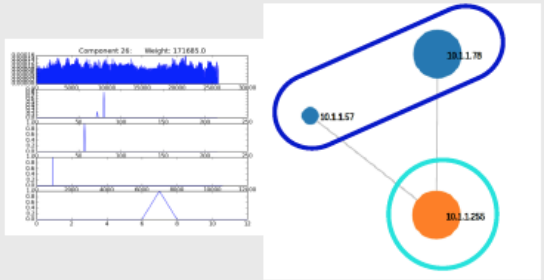
Raw Data from R-Scope®
(Reservoir's Network Security Monitoring Appliance)





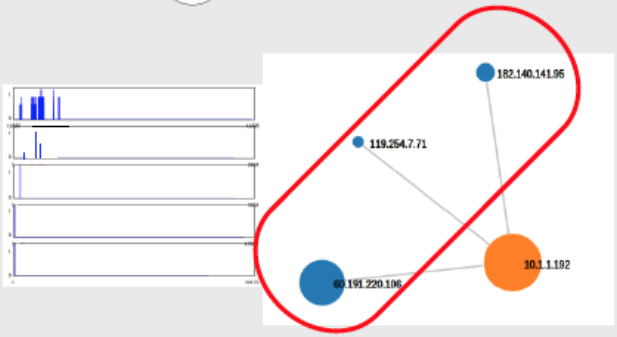


- Time: 9 am to 5 pm every day
- Senders: Intern's computers
- Receivers: DNS servers
- Requests: Google, Stack Overflow, various yale.edu websites



- Time: Constant and regular
- Senders: Two business computers
- Receiver: Broadcast address
- Request: Faulty printer's address

(*Request denied so repeated constantly*)



- Time: Between 2 am and 5 am in the morning
- Senders: Blacklisted Chinese IP addresses
- Receiver: Reservoir code repository
- Access denied

CANDID on SCinet Network

SCinet 2016 Analysis Highlights

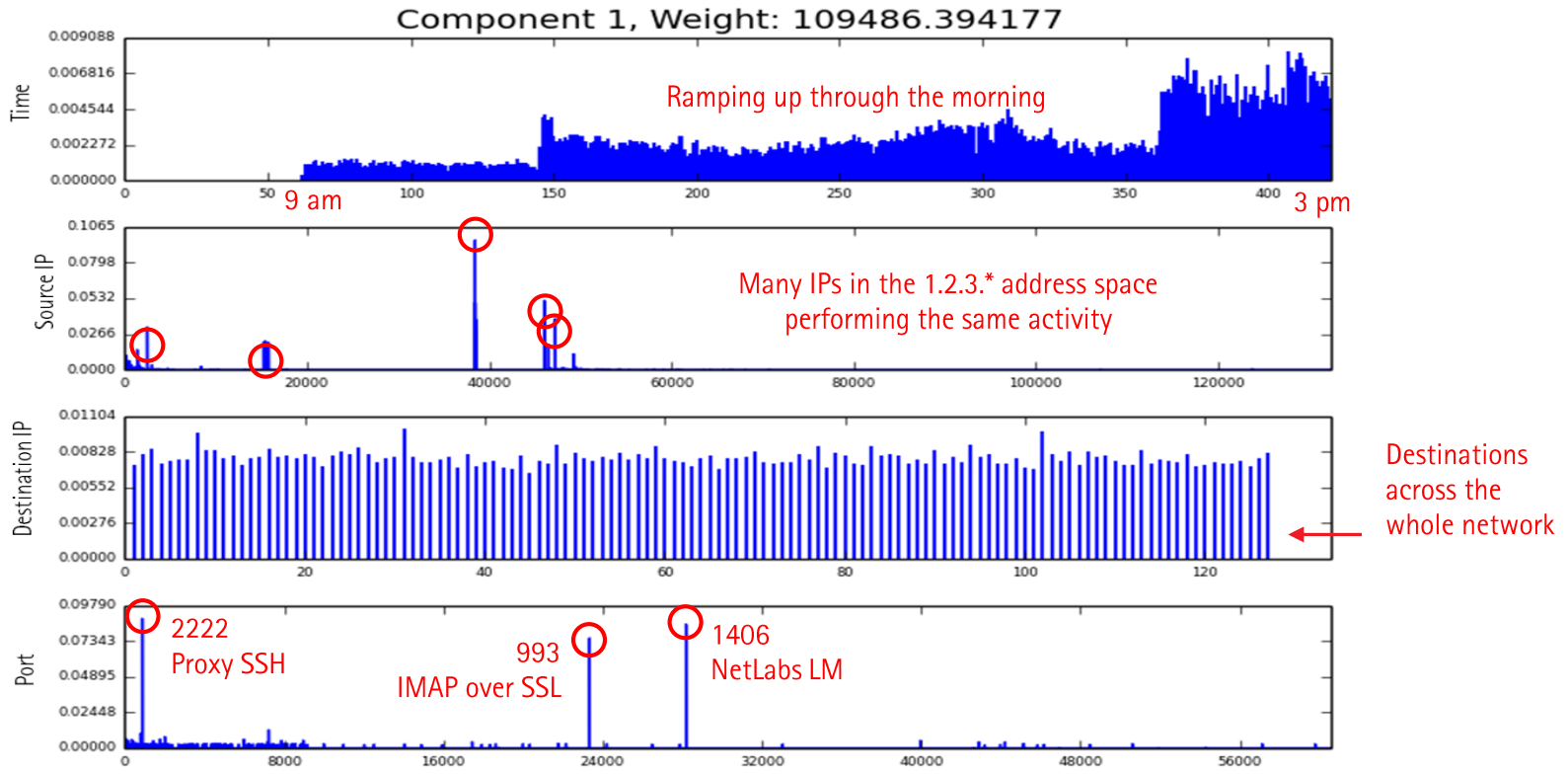
- Network Research Exhibition (NRE)
- Metadata collected from 6 R-Scope boxes
- HPE Apollo 2000 running ENSIGN
- Data collected within specific time windows (8-24 hours)
- Data filtered by internal/external source/destination
- Binning applied by time interval, subnet, etc.
- Typical tensor ~106 non-zeros with >99% sparsity
- Typical 100 component decomposition required ~5 minutes
- Post-decomposition exploration with Splunk

Tensors formed from selected, binned metadata pulled from filtered R-Scope (Bro) logs

HPE Apollo 2000
12 Cores, 256 GB RAM utilized per tensor

Case Study #1: External Scanners

Indicator of a reconnaissance phase of an attack

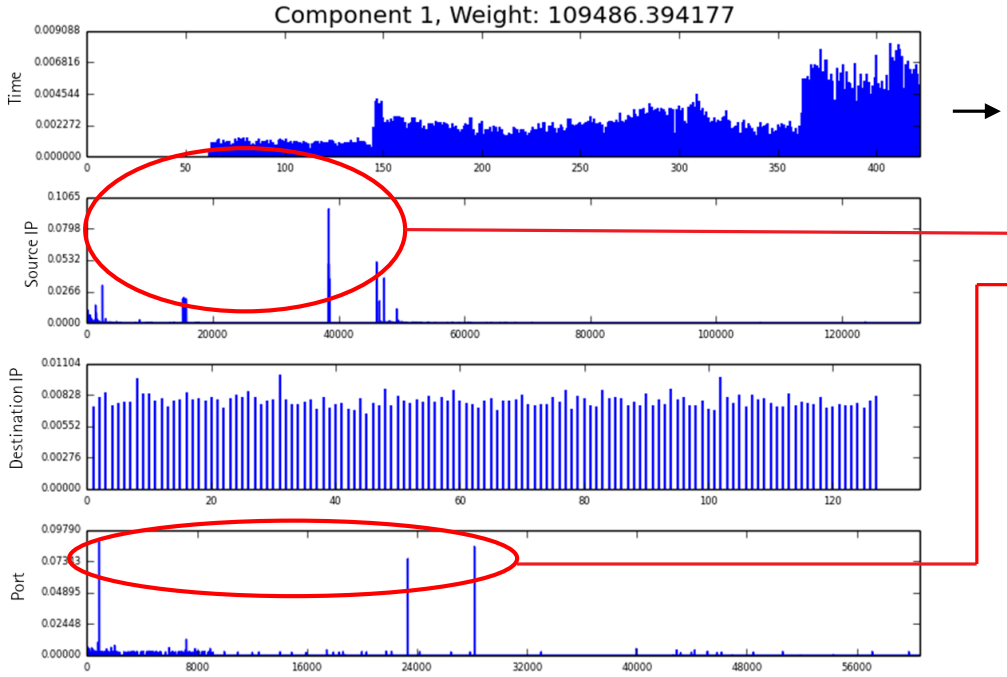


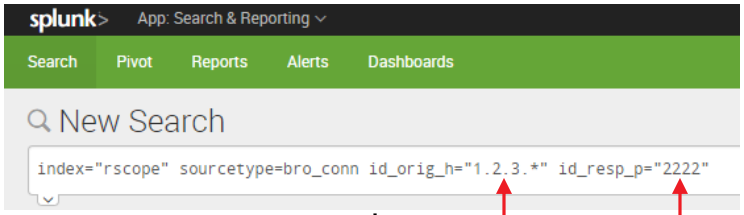
Distributed network mapping and port scanning with likelihood of hostile intent

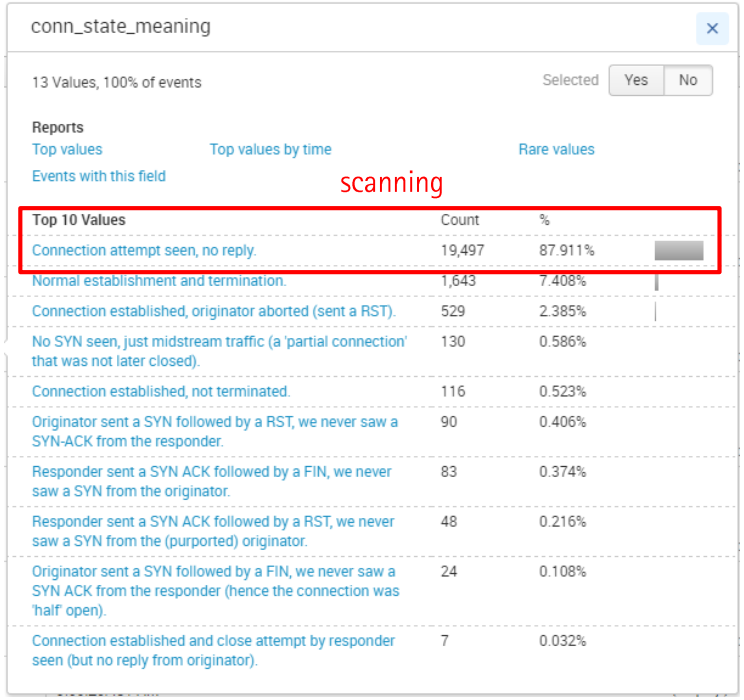
- Coordinated attempt by multiple external actors to find hosts on SCinet with particular services

Case Study #1: External Scanners

Confirmation of hostile intent





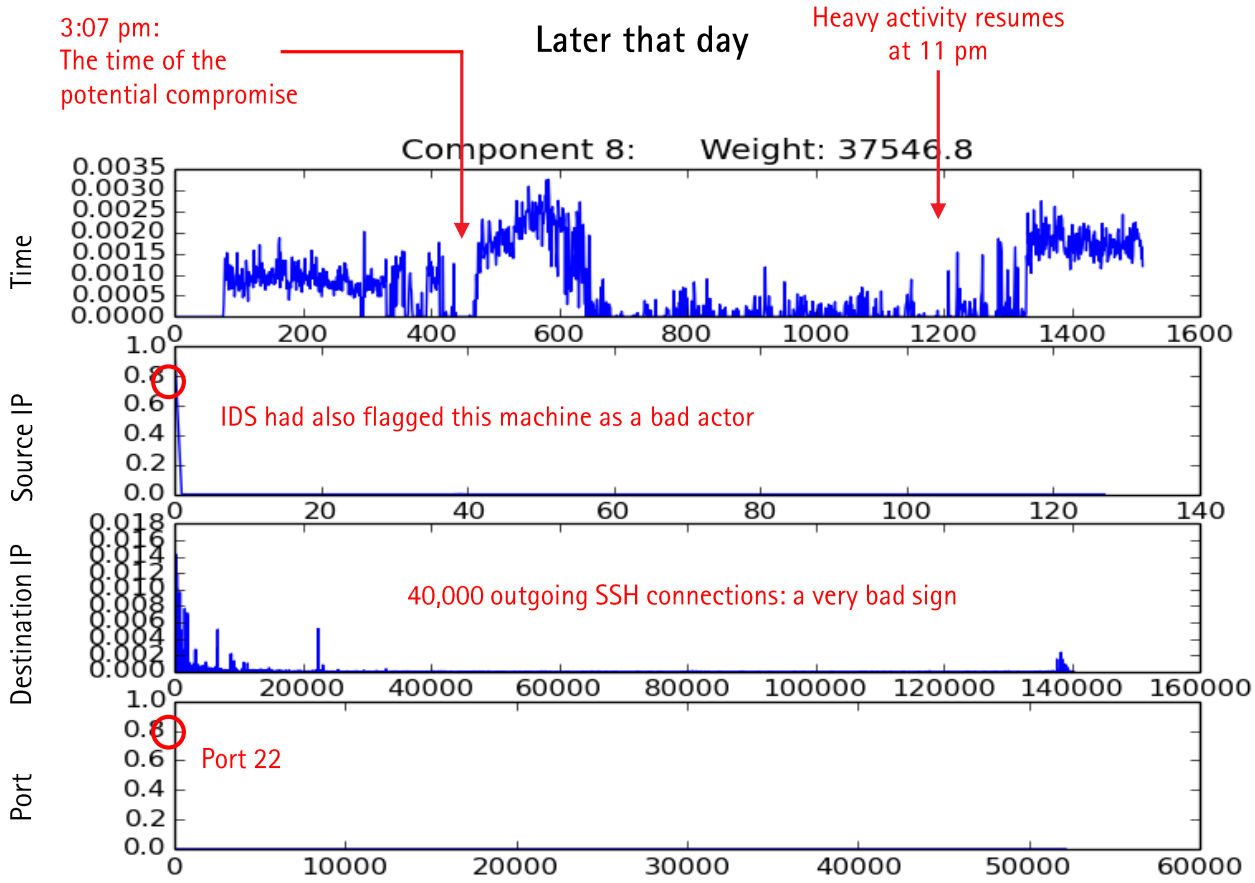


Further Investigation in Splunk

- Filtered search for successful connections
  - "focused" Splunk query guided by the component
- Confirmed the scanning

Case Study #1: External Scanners

Confirmation of evolution of an attack

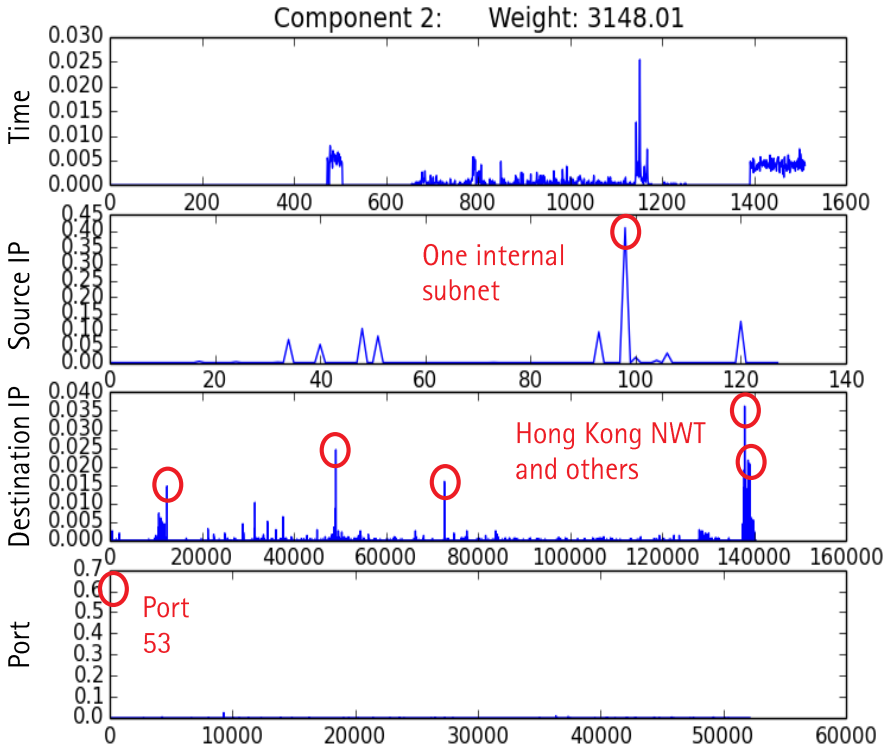


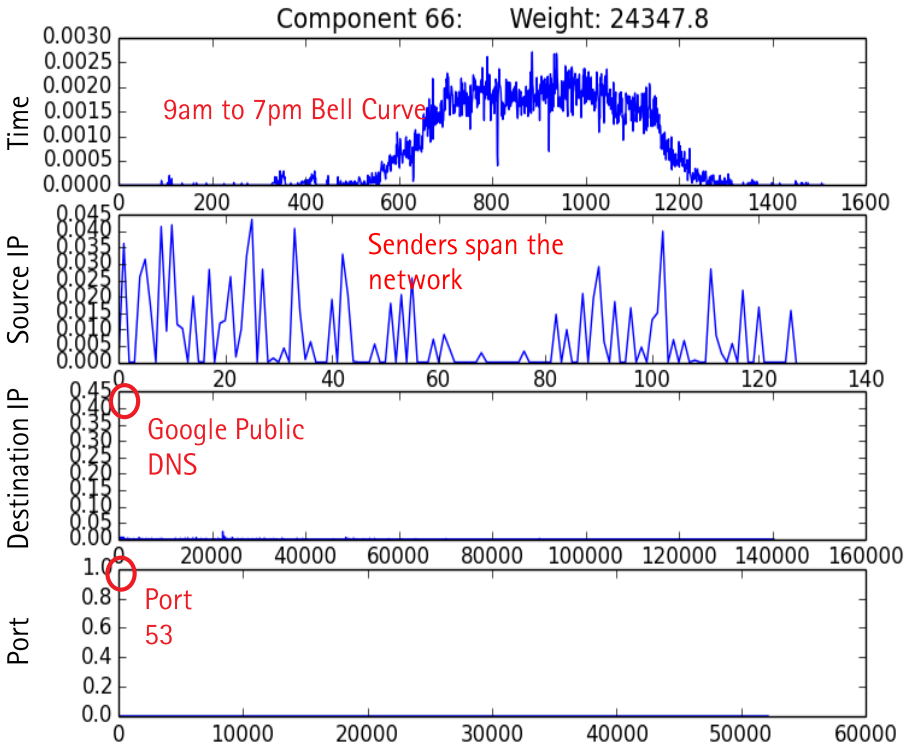
Confirmed the evolution of an attack using a later component

- Outgoing SSH connections from a compromised host

Case Study #2: Suspected Data Exfiltration

Isolating Suspicious DNS Traffic





Suspicious DNS Traffic

- Irregular spikes in time
- Suspicious destinations

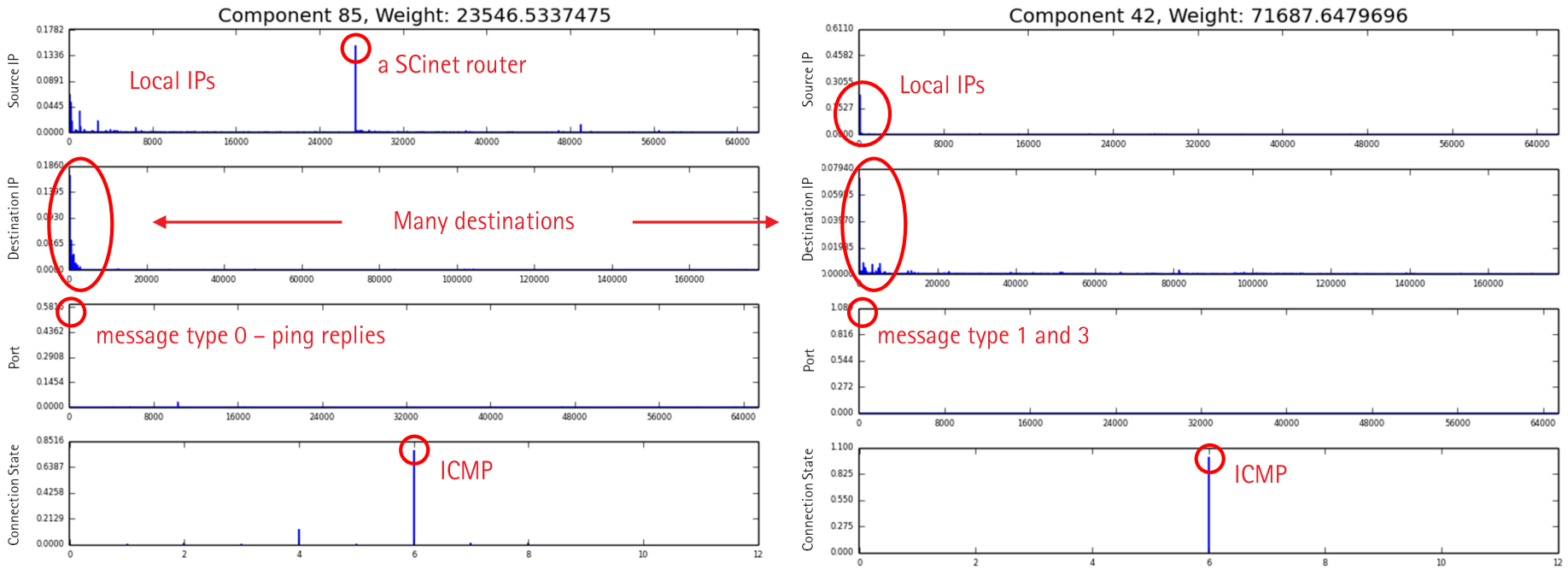
Typical DNS Traffic

- Highly regular traffic between 9 am and 7 pm – the running hours of the conference
- Valid DNS server destination

Case Study #3: ICMP Tunneling

Anomalous ICMP is difficult to distinguish through

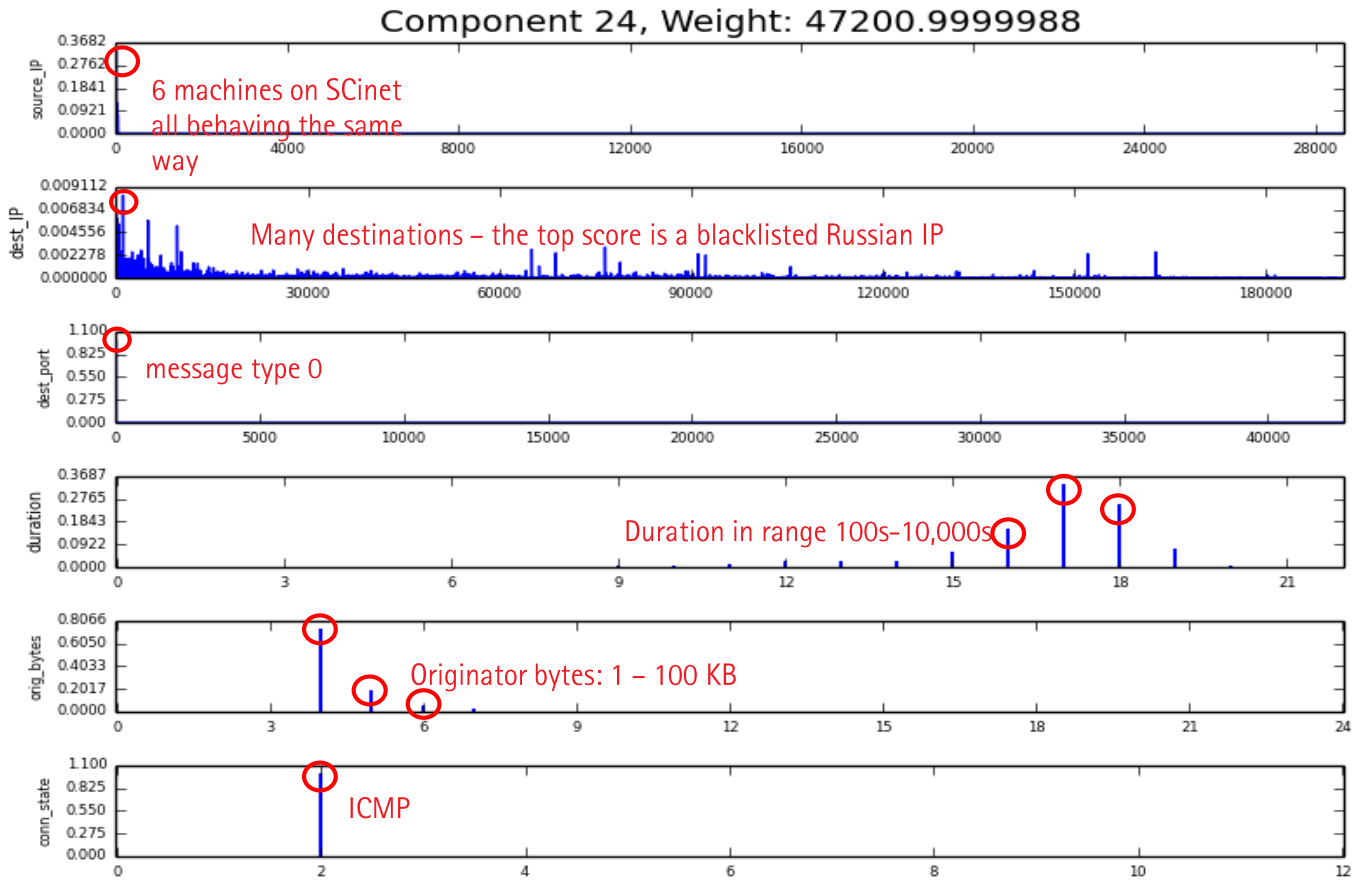
- Decompositions with fewer metadata attributes
  - IP addresses, port, connection state/time



Case Study #3: ICMP Tunneling

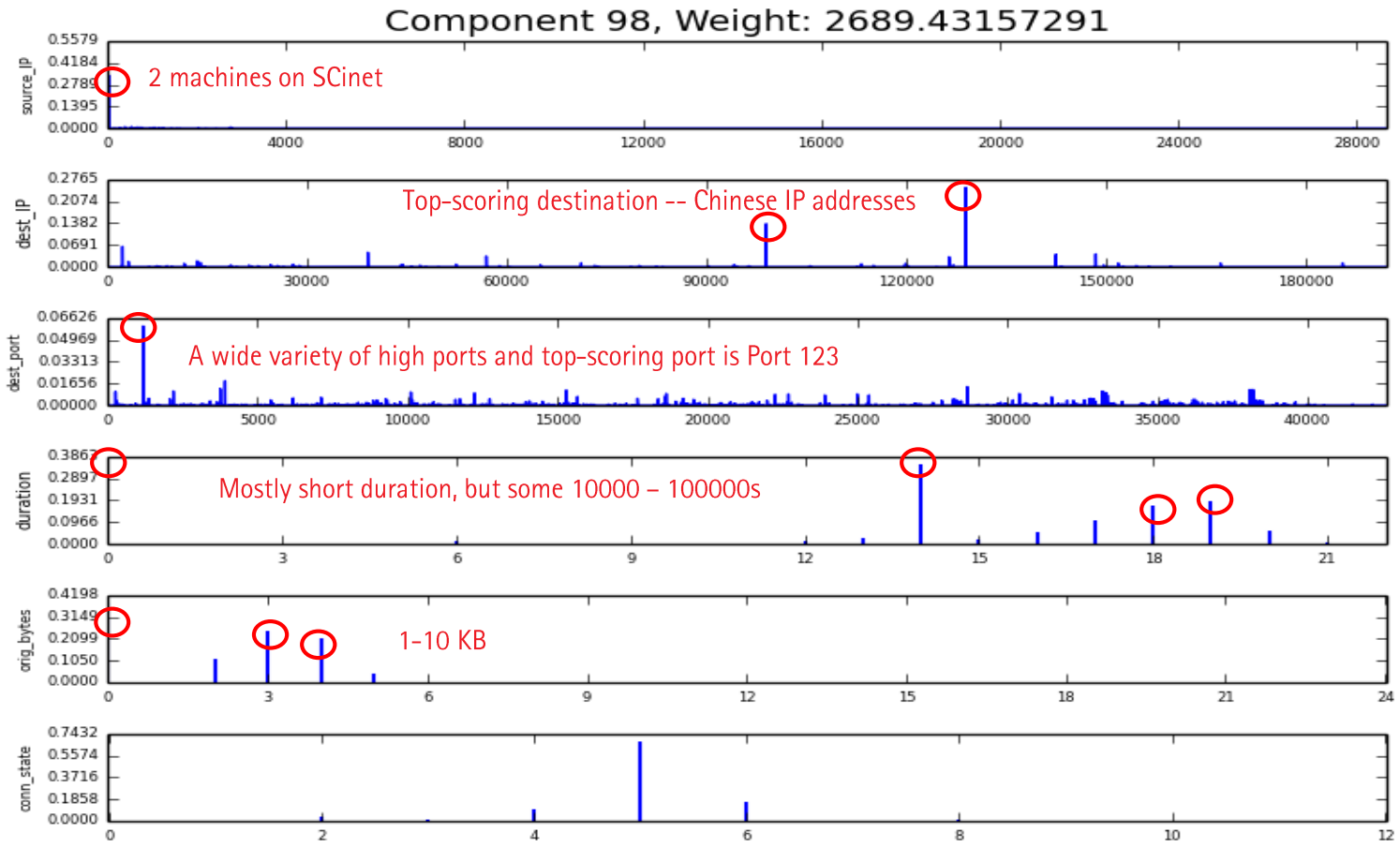
Anomalous ICMP easily distinguished through

- Decompositions with additional key metadata attributes
  - adding connection duration and number of bytes to analysis



Case Study #4: NTP Amplification Attack

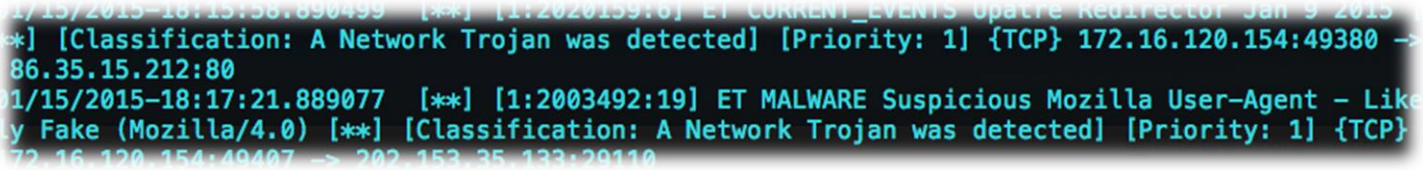
Another successful use case of decompositions with more metadata attributes



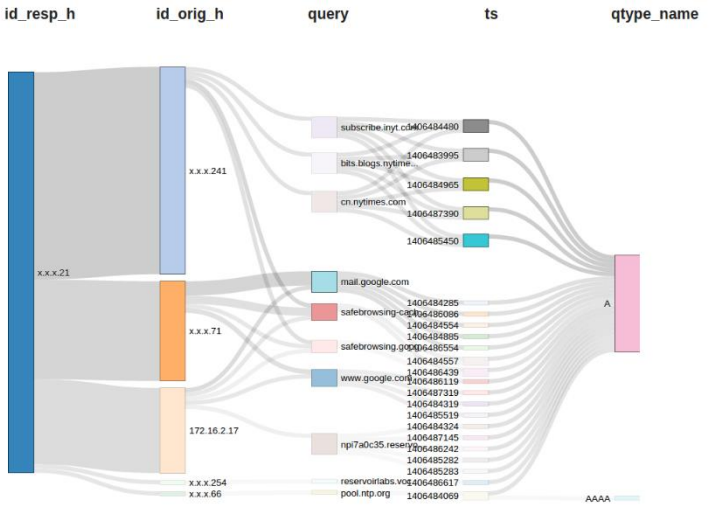
Using CANDID and ENSIGN Tensor Analysis…

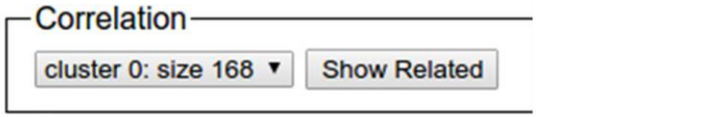
We have uncovered and visualized patterns indicative of:

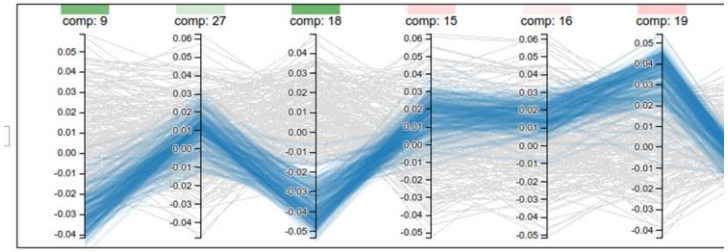
- Distributed port scans evolving to machine takeover
- Distributed denial of service attacks
- DNS-based data exfiltration/insider threat
- SSH password guessing (apart from scanning)
- Network policy violations
- Exploitation of application-specific port vulnerabilities
- Patterns of traffic indicative of scans for printers or IoT devices
- Broken or misconfigured network services
- Selective, persistent use of cryptographic methods in point-to-point communication



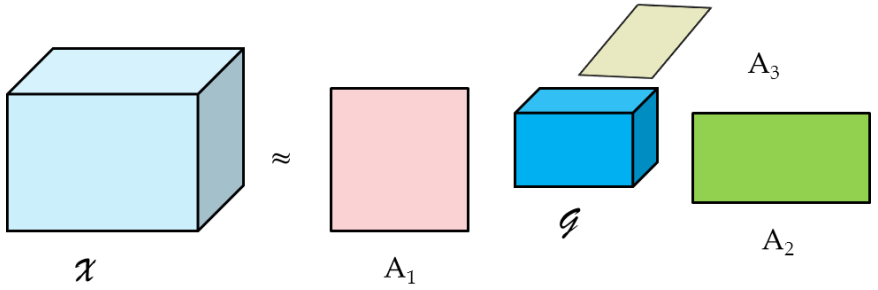
More Features in the Pipeline



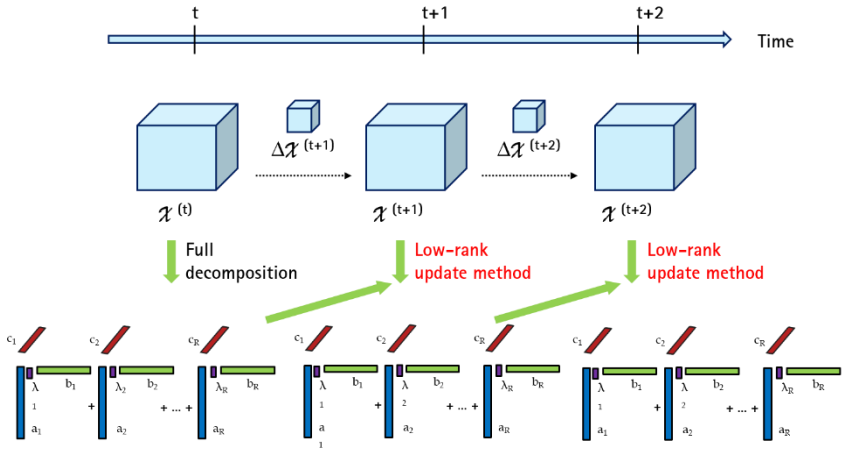




Specialized visualization and
ML-driven post-processing methods



Alternate and advanced methods for tensor decompositions



Support for streaming updates

Conclusion

Contact **Reservoir** Labs

- https://www.reservoir.com

Contact the Speaker

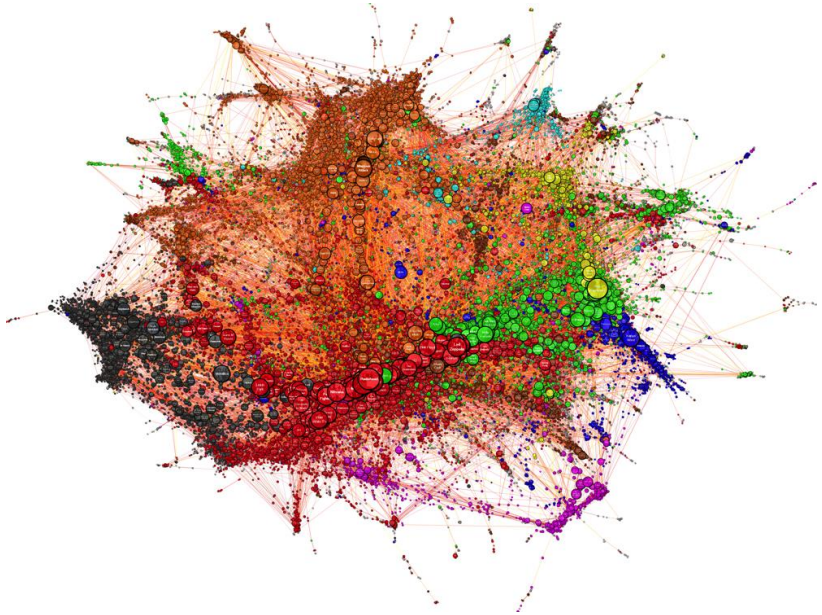
- baskaran@reservoir.com

Meet us at SCinet NRE Demo 2017

Other Recent Papers

- *Cyber Security Through Multidimensional Data Decompositions*
  D. Bruns-Smith, M. Baskaran, T. Henretty, J. Ezick, R. Lethin, in CYBERSEC, Apr 2016

- *Memory-efficient Parallel Tensor Decompositions*
  M. Baskaran, T. Henretty, D. Bruns-Smith, M. H. Langston, J. Ezick, R. Lethin, in IEEE HPEC, Sep 2017
  (Best Paper Award)



Tensor decompositions provide a fast, scalable linear algebra based solution to finding patterns in linked metadata