SC19 Network Research Exhibition: Publishable Submission

Automated Tensor Analysis for Deep Network Visibility*

Thomas S Henretty and Muthu M Baskaran, Reservoir Labs Inc.
henretty@reservoir.com, baskaran@reservoir.com

Abstract

We will demonstrate a usable and scalable network security workflow based on ENSIGN, a high-performance data analytics tool based on tensor decompositions, that can analyze huge volumes of network data and provide actionable insights into the network. The enhanced workflow provided by ENSIGN assists in identifying actors who craft their actions to subvert signature-based detection methods and automates much of the labor intensive forensic process of connecting isolated incidents into a coherent attack profile. This approach complements traditional workflows that focus on highlighting individual suspicious activities.

ENSIGN uses advanced tensor decomposition algorithms to decompose network data with multiple metadata attributes into components that capture multimodal network patterns and behaviors. This enables easier identification of anomalies and suspicious patterns and simpler analysis of large, complex patterns. We will apply ENSIGN over the network security logs available through the SCinet network stack to provide deep visibility into network behaviors/trends including, but not limited to, port scans, network mapping attempts, scans targeting specific services, SSH brute forcing, NTP amplification attacks, DNS amplification DDoS attacks, and obfuscated data exfiltration using DNS and ICMP tunneling.

Goals

At SC15, SC16, SC17, and SC18 NRE, we successfully demonstrated ENSIGN [1] on offline network data feeds provided by R-Scope network appliances [2]. Specifically, we demonstrated how ENSIGN separated normal and off-normal traffic patterns in a way that led to the discovery of indicators consistent with, and in some cases prior to, human analyst discovery (e.g., a distributed takeover attack on a vendor booth and a suspected ICMP-based data exfiltration) [3]. At SC17 NRE, we also introduced an initial version of a streaming analysis capability to improve the timeliness of previously demonstrated offline analysis of network metadata [4]. At SC18 NRE, we demonstrated timely detection of alarming behaviors and the capability to provide actionable insights into network logs through the use of network security utility tools built on top of ENSIGN.

At SC19 NRE, our overall objective is to further demonstrate the effectiveness of ENSIGN in an operational cyber security setup in extracting anomalous patterns of network traffic, detect alarming behaviors, and provide actionable insights into network data. With mature development of ENSIGN and associated tools over the last year, our specific research objectives at SC19 NRE are the following:

- Demonstrate **highly-scalable static tensor analysis** to analyze large batches of offline network data collected and stored in the form of Zeek logs, especially, conn.log, dns.log, files.log, and http.log.
- Demonstrate the use of **cyber utility tools** built on top of ENSIGN to report suspicious behaviors and attacks disrupting the operation and performance of SCinet.
- Demonstrate **advanced streaming analysis** capability in ENSIGN to improve the timeliness of detection of malicious and anomalous behaviors.
- Demonstrate the results of **advanced anomaly detection** techniques in ENSIGN to identify and isolate anomalous network behaviors/patterns.

* This material is partially based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Award Number DE-SC0017081.

Planned Experiments at SCinet

We divide the experiments and activities that we plan to do at SCinet into two major categories, namely, operational and research. The operational activities will be closely tied to the NOC security activities. The research activities will be optional and will be done based on the availability of personnel and network resources that are free after being used for the operational activities.

Operational Activities

Static network data analysis

We will use highly-scalable static tensor analysis methods [5] to analyze large batches of offline network data collected and stored in the form of Zeek logs, especially, `conn.log`, `dns.log`, `files.log`, and `http.log`. These logs are expected to be exported to a Splunk data store hosted in the TACC Chameleon Cloud. We have multiple tensor decomposition methods that are suited for analyzing network data. These methods have varied degrees of tradeoff between "interpretability" of the tensor analysis output and time needed to complete tensor analysis. We will run two of these static tensor methods that we have extensively tested, in parallel, to provide a more robust analysis of the network data.

Expected outcome:

- Separate normal and off-normal traffic patterns
- Discovery of indicators related to suspicious traffic patterns (port scans, network mapping attempts, scans targeting specific services, SSH brute forcing, NTP amplification attacks, DNS amplification DDoS attacks, and more)
- Discovery of indicators of obfuscated behaviors with potential malicious intent (data exfiltration using DNS, ICMP tunneling, and more)

Streaming network data analysis

We will use advanced streaming tensor decompositions [6] (a prototype of which we tested at SC17 and SC18 SCinet) to analyze small batches of network logs (as opposed to waiting for accumulation of large batches of network logs) as soon as they are made available in the Splunk data store hosted in the TACC Chameleon Cloud. We will start with "hourly" batches for the streaming analysis and we will dynamically adjust the batch size as needed during the experiments.

Expected outcome:

- Identify new network patterns, if any, that appear in the new data stream
- Identify how already-seen patterns have changed or evolved after the new data stream
- Discovery of indicators of suspicious traffic patterns and obfuscated behaviors close to their onset

Using and testing cyber utility tools

We have developed cyber utility tools to operate on top of the ENSIGN tensor analysis engine with the primary objective of facilitating the usability of ENSIGN. Specifically, we have developed utility tools to query, process, and present the mathematically sophisticated raw tensor analysis output in a network analyst friendly format. We have also developed tools that will parse the tensor output to check for the occurence of specific (most common) network patterns such as beaconing, network mapping, and port scanning.

We will use, test, and dynamically enhance these tools as needed by running them on the output generated by the static and streaming tensor decompositions (described above).

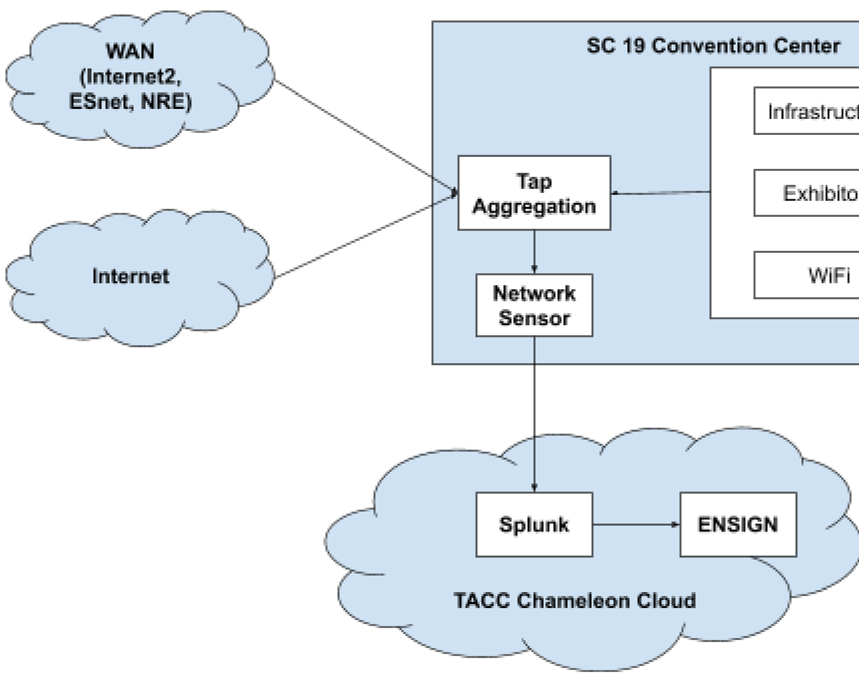
Research activities

Advanced anomaly detection

As a research activity, we plan to experiment with advanced anomaly detection approaches around ENSIGN for which we have a working prototype implementation. These methods are aimed at easily separating off-normal traffic patterns from normal traffic patterns. Specifically, we plan to exercise our technique for automated clustering and classification of the tensor analysis output based on topic modeling [7]. This automation technique is expected to substantially increase the usability of the tool by reducing the cognitive load of the analysts. This activity needs computational resources that may contend with the resources needed for the operational activities and hence it is planned as an optional activity.

Network Resources

We will install and operate ENSIGN from a compute node in the TACC Chameleon Cloud. The primary source that will feed ENSIGN with the network logs from SCinet network security stack will be a Splunk instance hosted in the TACC Chameleon Cloud. ENSIGN will access the Zeek logs through the Splunk Python API. The diagram describing the network workflow of ENSIGN is shown below.



References

[1] ENSIGN, https://www.reservoir.com/product/ensign-cyber

[2] R-Scope, https://www.reservoir.com/product/r-scope

[3] M. Baskaran et. al, "Enhancing Network Visibility and Security through Tensor Analysis." in FGCS 2019

[4] J. Ezick et. al, "Eliminating Barriers to Automated Tensor Analysis for Large-scale Flows," in FloCon 2018

[5] M. Baskaran et. al, "Fast and Scalable Distributed Tensor Decompositions." in IEEE HPEC 2019

[6] P. Letourneau et. al, "Computationally Efficient CP Tensor Decomposition Update Framework for Emerging Component Discovery in Streaming Data," in IEEE HPEC 2018

[7] T. Henretty, et al., "Topic Modeling for Analysis of Big Data Tensor Decompositions," in SPIE Disruptive Technologies in Information Science 2018