Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

LINEBACKER

LINE-speed Bio-inspired Analysis and Characterization for Event Recognition

A Biosequence-based approach in the discovery of evolving threats

Christopher Oehmen
christopher.oehmen.pnnl.gov | (509) 375-2038

OVERVIEW

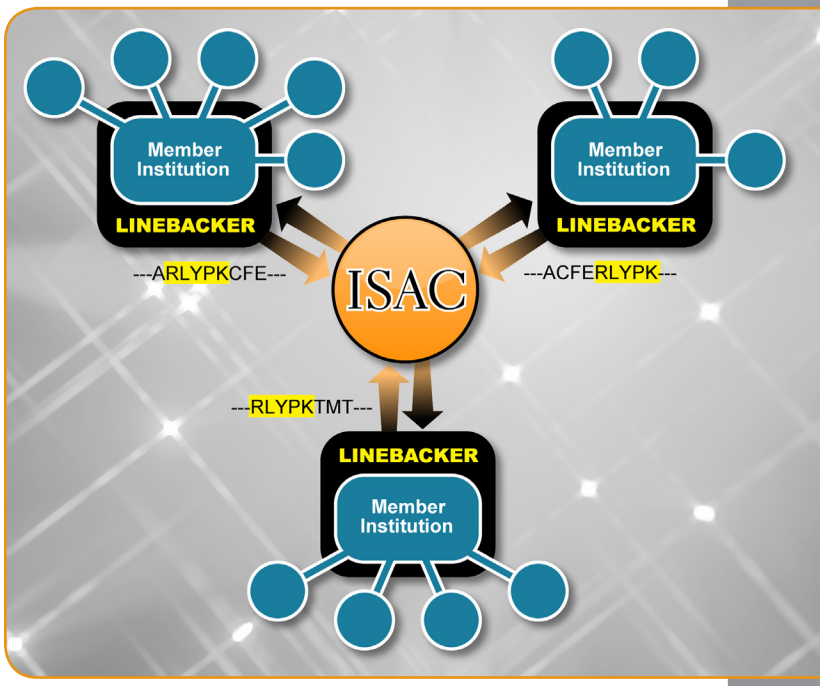
LINEBACkER allows cyber security analysts to quickly discover and analyze behaviors of interest in network traffic to enhance situational awareness, enable timely responses, and facilitate rapid forensic and attribution analysis. In a collaborative, operational setting, netflow data can be converted on site in near real-time and then shared with collaborators in obfuscated form. This allows for finding attacks and anomalies faster without exposing sensitive data.

CHALLENGE

Our reliance on cyber systems permeates virtually every aspect of national infrastructure. From banking, finance and industry to education and research, from national defense to power generation and delivery, secure computer networks are the lifeblood for maintaining critical infrastructure, information, and the US strategic advantage over our adversaries. The volume of network traffic data generated has outpaced our ability to effectively analyze it fast enough to prevent many forms of network-based attacks. In most cases new forms of attacks cannot be detected with current methods. We need a method to drastically reduce the amount of data to be analyzed, to quickly characterize an attack, and to identify previously unseen types of attacks before they're executed. Network analysts need the ability to discover malicious traffic in computer networks, and share their insight or a signature of the threat with others, without jeopardizing sensitive or institutional data.

APPROACH

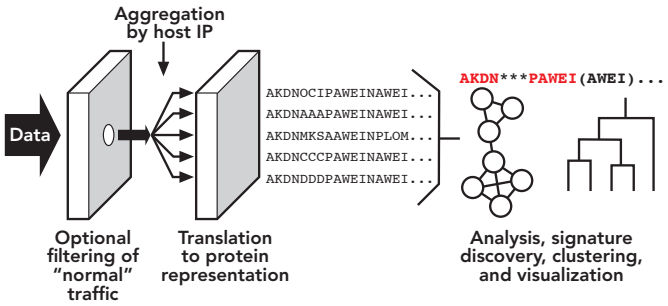
The LINEBACkER tool allows analysts to share signatures without sharing data. This is especially beneficial when sensitive data is involved and sharing threat signatures across



multiple organizations is necessary. Simply, LINEBACkER applies the MLSTONES methodology to the problem of discovering malicious sequences of traffic in computer networks. MLSTONES leverages technologies and methods from biology and DNA research, and have effectively mapped a solution to flexibly represent and identify signatures and express them in a biology-based language that cannot be "translated" back to the original data.

We've translated several biology and bioinformatics concepts onto cyber defense data. Specifically we've created a methodology that uses the concepts of protein identification and families, inheritance, and function to apply to a number of cyber based data types. The MLSTONES process creates cyber "proteins" and then create a single representation of an entire family of entities thus reducing the amount of data to analyze by several orders of magnitudes.

The Department of Homeland Security sponsored the production of this material under DOE Contract Number DE-AC05-76RL01830 for the management and operation of Pacific Northwest National Laboratory.



Example of the Analysis Process for Discovering New Signaatures and Trends in Highly Variable Network Traffic Data Using Bioinformatics-inspired String Matching

REQUIREMENTS

1. Bidirectional netflow collection software (e.g., NFSEN)
2. Connectivity to the REN-ISAC data sharing infrastructure
3. Installation of PNNL-developed software
   a) Shipped as source code for transparency
   b) Code needs access to the netflow data mentioned above
   c) We do not anticipate that any additional storage devices will be required beyond what is commonly used to accumulate netflow data

NOTES: LINEBACkER will convert netflow data to an obfuscated format where host systems are not identifiable. This obfuscated data will be collected at the member institution, processed, and shipped to REN-ISAC periodically.

We can also infer the function of a "cyber protein" by its relationship to other similar proteins. This is the same process used in biology to discover similar proteins. This helps to identify completely new (zero-day) cyber threats. We apply high-performance biosequence analysis that enables inexact string matching of streaming network traffic; our approach is robust when there is more than one form of threat and supports "family resemblance" attribution. The tool characterizes baseline behavior, converts raw netflow data to bio-representation, constructs a family tree of cyber event types, and creates visual interface to deploy in a client setting, against specific threats or suspicions. The translation of network behavior is accomplished at the site of collection and it is the translated representation that is shared among collaborating agencies. This translation is lossy and cannot be reverse engineered. Thus no identifying network data is ever sent away from its host institution.

COMPETITIVE ADVANTAGE

Most conventional identification strategies are based some form of exact matching, which fails in the face of a high degree of diversity. Our approach maps the problem into one where we can apply the mathematics and statistics of bioinformatics to look for the defining characteristic of a "family of behaviors" and use this for identification and potentially for attribution. This provides the advantage to recognize previously unseen events (i.e., zero-day), simply by their similarity to something that has been seen.

LINEBACkER has the ability to discover malicious network activity through sequence analysis across the U.S. research and engineering computing infrastructure; our goal is to scale the computational processes to handle line-speed analysis at a variety of institutions. Additionally it provides a secure mechanism for collaboration of potential threat vectors without releasing sensitive information. By translating at the source no identifiable information is sent to other institutions but "cyber proteins" can be shared and identifications of potential threats can be shared quickly.

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*
U.S. DEPARTMENT OF ENERGY