Wednesday, August 17, 2022
2:30 PM – 2:50 PM
Ingesting Sensor Feeds from Multiple Enclaves to a Central Platform

Daniel Haas
Director of Grown, Cyber Mission Sector
Peraton

Abstract:
Evolving into the future fighting force, the Army must utilize data as the critical component in determining the outcome of military operations. Success in the operating environment hinges on the ability of Army decision-makers to have real-time information available to realize decision dominance. Supporting the need for real-time data means the Army must operate on a common data fabric leveraging speed, security, and simplicity at echelon while understanding the dynamics and complexity of its globally distributed systems. The resulting benefit is an information repository with agility, scalability, and integrity while operating in contested operating environments.

Starting with the fundamental question of "what decision(s) is our data supporting?" should drive how the Army organizes to become a data-centered fighting force. A common data fabric reduces data silos from stove-piped legacy systems and enables the integration, management, and distribution of trusted and secure data in real-time to geographically dispersed users in any location regardless of enclave or classification levels. However, this only solves how to make data visible, accessible, secure, and interoperable. Data must be parsed, classified, aggregated, tagged, stored, and secured for users to recognize the content, context, and applicability of data and then link it to other data elements for exploitation and meaningful reporting. Applied in Army cyberspace operations, data ingested from DODIN-A creates searchable repositories, regardless of the domain, to enable knowledge of incidents for comparison against Army prioritized threats/vulnerabilities with relevant data.

High volume feeds are accomplished via API's allowing real-time ingestion and bulk uploads from Army sources which includes historical data in cloud-based containers or cleansed datasets from isolated networks. Data inputs are fed into and processed in seconds in a natural language processor (NLP) analysis engine providing the performance and scalability required to cover large amounts of data input.

Defenders can customize these NLP engine-identified objects of information based on past and future threats. The objects are tagged to create data sets of information for analysts and threat hunters at any skill level. Identified observables automatically move to the attribution layer of the fabric and are enriched using the previously identified NLP objects through organizational, third-party, free, open, paid, and customer-provided threat intelligence sources. Next, a secondary NLP engine begins processing the combined event and enriched data.

This secondary NLP process is key to being able to increase the accuracy of the enriched threat data. Data used for analytics is stored in a separate graph container. This container holds data for analytics and metrics, both visually and as data for correlation with previous and future events. The service also directs data required for queries into a searchable database. At the time of the first observation by an analyst or threat hunter, the data fabric has already gathered threat research information, reducing time-intensive research and correlation allowing for optimized operations. A new approach to security

against modern and evolving threats requires new thinking. By breaking through the data silos, leveling the playing field, and enriching data at machine speed, this new approach transforms organizations' security posture.