Peraton | LABS

WHITEPAPER

CREATING A FRAMEWORK FOR CYBER RECONNAISSANCE

HOW TO ELIMINATE BIASED ANALYSIS TO UNCOVER THE TRUTH IN INTELLIGENCE DATA

The number of devices in the cyber landscape is expanding at an alarming rate. Newly invented devices, firmware and waveforms are eagerly adopted by business and consumer users while older technologies are never completely discarded. Users are empowered to actively add applications, customize interfaces or make other changes as they desire. This combination of accelerated expansion and ongoing change presents a compelling problem for organizations working to understand or build intelligence about the cyber landscape.

This document describes the challenges presented by the threat environment and current tools, and introduces how Peraton Labs' cyber reconnaissance framework can enhance cyber data collection and analysis across the operational theater.

CYBER INTELLIGENCE CHALLENGES

The vast amounts of cyber information—data about devices, networks and the people who own and operate them—present processing and analysis challenges for federal agencies tasked with addressing national intelligence issues or critical infrastructure cyber defense. Data overload and the bias introduced at all levels of cyber collection by specific tools, the availability of information, perceived intent, threat attribution and overconfidence of analyses, make it difficult for agencies to quantify a true understanding of their cyber information at any given time. The urgent nature of cyber data collection, often demanding tightly correlated collection to uncover relationships between information, creates an additional challenge.

Preventing or initiating cyberattacks requires getting the right intelligence and information to the right people at the right time, and it all starts with gaining a clear picture of the cyber landscape. Utilizing disparate information sources, operators can build a knowledge base of cyber intelligence with analysis of specific targets or any information collected "in" or "through" the cyber landscape.[1]

Yet, the resulting intelligence picture is often ambiguous, because the still-evolving cyber intelligence discipline is not as mature as imagery intelligence and signals intelligence that were established decades ago. This shortcoming becomes clear when former Defense Secretary Donald Rumsfeld's 2002 explanation of the limits of intelligence reports using "known and unknowns" (figure 1) is applied to cyber intelligence: when we start with IPv6 as a known-unknown, it is reasonable to anticipate the next variant of a known threat from a known actor (known-known), but difficult to anticipate the size of the North Korean offensive cyber program (known-unknown), and even more challenging to quantify the utilization of internet protocol version 6 (IPv6) for global military operations (unknown-unknown).

Cyber intelligence analysis is further challenged because the data collection sources for cyber network defense are predominantly informed by network- and host-based agents that provide voluminous, segregated, ambiguous and non-standard data—placing a heavy cognitive burden on the operators who integrate and process the data. This strain introduces perceptual and cognitive biases to the overall understanding of the defensive landscape.[2]

The gap between data collection and intelligence presents both a challenge and an opportunity to change the future of cyber intelligence collection and analysis. Agencies need to obtain high-confidence analysis and more certain conclusions about data. They require a means to eliminate biased analysis. Data normalization and integration is not enough to address this challenge and existing tools fail to address the problem of how data gets collected in the first place, so high confidence in these disparate tools is misplaced.

PERATON LABS' CYBER RECONNAISSANCE FRAMEWORK–HIGH CONFIDENCE ANALYSIS AND MORE CERTAIN CONCLUSIONS

To help the government obtain high-confidence analysis and more certain conclusions about data, Peraton Labs has invested in the development of the cyber reconnaissance framework. This proof-of-principle implementation uses a data-driven, automated cyber collection pipeline that identifies information gaps, then strategizes and executes techniques to collect relevant data to fill those gaps. The result is a unified model for cyber intelligence gathering that supports operational priorities, informs decision makers, and dynamically adapts to available resources and current intelligence needs.

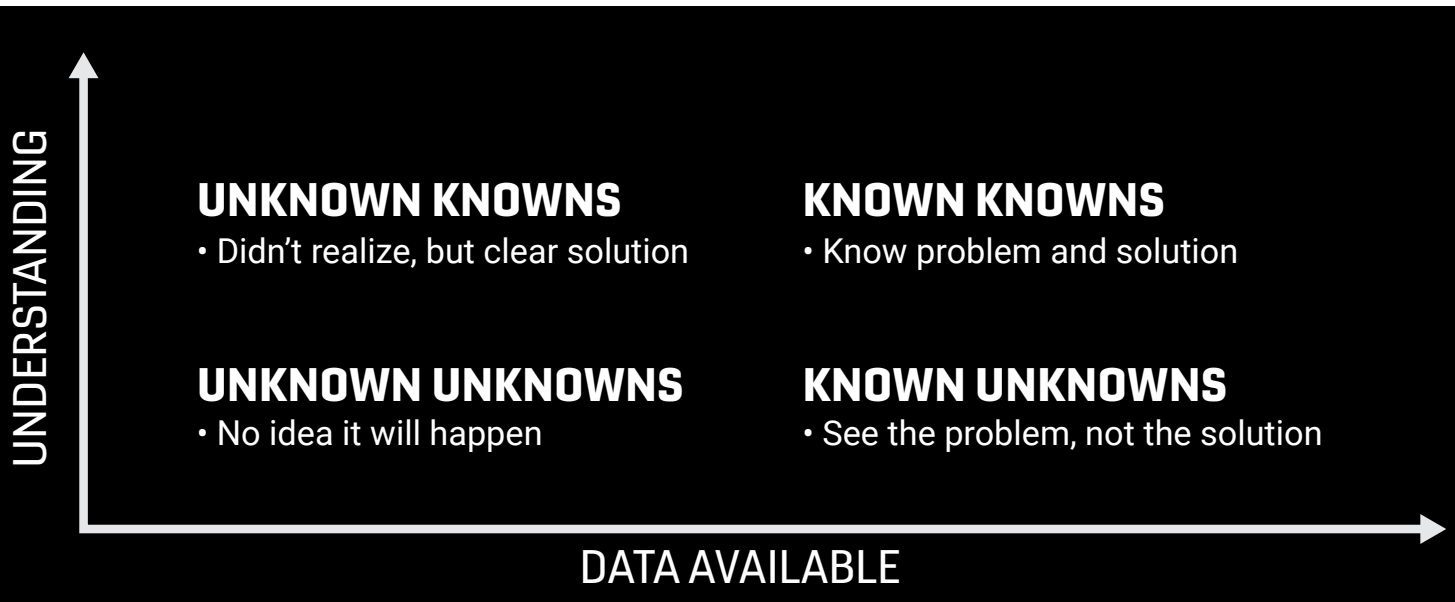

Figure 1: Rumsfeld's epistemology—Known and unknowns

Our framework mitigates the information gaps in cyber intelligence and its application within cyber network defense. It goes a step beyond the integration of disparate data to focus on how and when data was collected, the confidence in the collection methodology, the utility of available resources and the value to the questions that need to be answered. As such, the framework can be used for a variety of mission applications, including several specifically relevant to cyber network defense.

How it works

The cyber reconnaissance framework identifies coordinated vulnerabilities and attacks across disparate intelligence domains and helps anticipate future attacks. The framework can also identify current attack activities and expand them with attribution to actor and intent. If investigation of previous attacks provides insight, the framework can deliver the dynamic adaptation of the intelligence posture with the introduction of new historical information.

The architecture

The cyber reconnaissance framework is comprised of three main high-level architectural elements: probability store, collection planner and collection orchestrator (figure 2).

Probability store: This topological database stores the post-processed and confidence-weighted intelligence collected by the cyber reconnaissance framework. It also provides natural and intuitive access languages which simplify questions about complex relationships and interdependencies in intelligence data. The probability store provides the fundamental schema for a multi-domain common operating picture, which underpins the logic and reasoning of the framework's other subcomponents.

Collection planner: Using the stored common intelligence picture from the probability store, the collection planner computes a collection of information-theoretic metrics that identify and prioritize information gaps. Collection strategies are subsequently planned for each gap and tasked for collection. Specific to cyber network defense information priorities, the collection planner focuses on maintaining information critical to the attribution of identified threatening activity (e.g., network ownership by organization), as well as tasking the defense posture of both agents and tools to use operational resources to target specific threats. It also adjusts for any duplication or excess.

Collector orchestration: By optimizing the execution of the collection strategy across naturally limited resources (i.e., compute, storage, licensing), the collector orchestration element automatically works across a distributed collection of agentless endpoints and all tiers of the infrastructure relevant to a mission. In a typical cyber network defense mission, collector orchestration would deploy the internal blue host and network agents, the gray endpoints for penetration testing and situational awareness, and a collection of global endpoints for intelligence fusion. In essence, enterprise defenses can be managed by the cyber reconnaissance framework's subcomponents. Moreover, the collector orchestration can accommodate the integration of new tools, techniques and agents with the addition of a simple module script automatically deployed by the orchestration framework.

THE CYBER RECONNAISSANCE FRAMEWORK IN ACTION

1. Evaluation of global technology transitions: To validate the fundamental value of the framework as a different approach to national cyber intelligence, Peraton Labs evaluated global technology transitions. The use case scenario focused on the global attribution of network ownership and the dynamic changes in national network technologies—away from IPv4 and to IPv6. To inform motive and intent in these transitions, the team sought to attribute the locality and sectors responsible for the largest dynamic change.

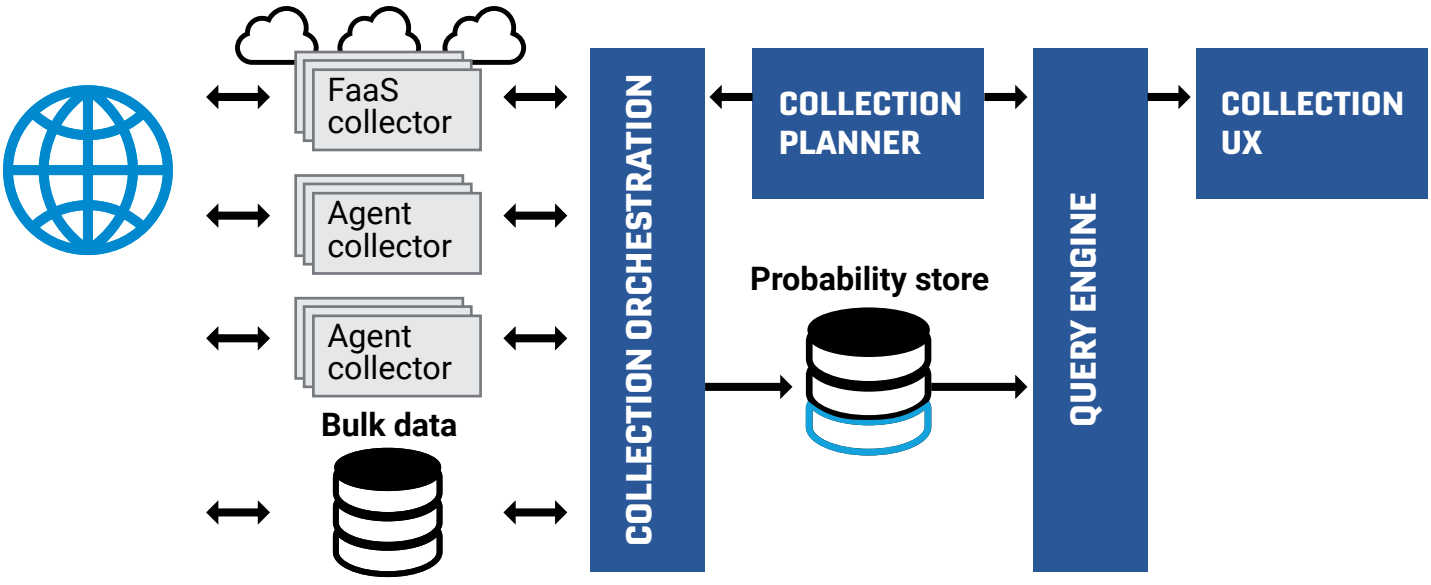

Figure 2: The cyber reconnaissance framework architecture

The exercise started with the question, “What are the most specific networks (those networks that are the customer edge of the tier-3 internet infrastructure) that we know to be in China?” This was asked using the domain query language: “all netblocks delegated, with high confidence, to the country whose country code is “CN”, where “netblock” and “country” are nodes and “delegated” is a relationship.” The resulting response (figure 3) shows there are a collection of both IPv4 and IPv6 networks known to be associated with China with high confidence.

The information provided by our framework led to a series of additional questions regarding how China has adopted IPv6 as a networking technology:

- How quickly has the adoption occurred?
- Does the adoption execute in rolling updates or on demand?

The questions were answered by binning (aggregating by time) the allocation of IPv6, which is attributable to China with high confidence for the time since IPv6 was first accepted as an internet standard. The result (figure 4) shows a strong surge of IPv6 adoption in the last quarter of 2015, lasting approximately one year and subsiding by the first quarter of 2017. Peraton Labs used this information to then identify which organizations and industries are responsible for the 2015 - 2016 uptick in IPv6 adoption.

The cyber reconnaissance framework traced the organizational ownership of the networks, starting with identification of the background data (figure 5). This led to an understanding that the surge in IPv6 utilization was primarily from tier-3 network providers.

This use case demonstrates our framework’s value in reducing the cognitive burden on operators. CyberRecon can deeply inform the case study questions without requiring operators to understand the tools used to acquire the intelligence. Less cognitive burden reduces operator bias in data collection and interpretation of conclusions. The basic structure leveraged three implemented metrics and collectors, which can be expanded for a broader concept of operations and intelligence priorities.

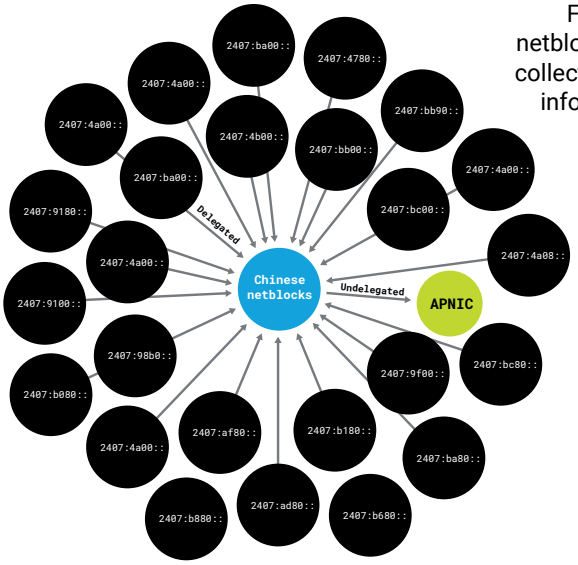

Figure 3: Chinese netblocks information collected from routing information registry delegated

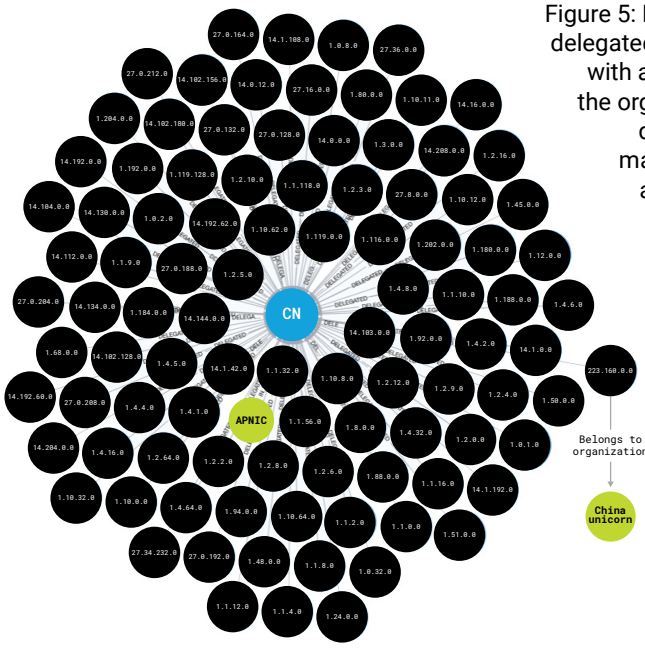

Figure 5: Netblocks delegated to China with a focus on the organization delegating malicious IP addresses

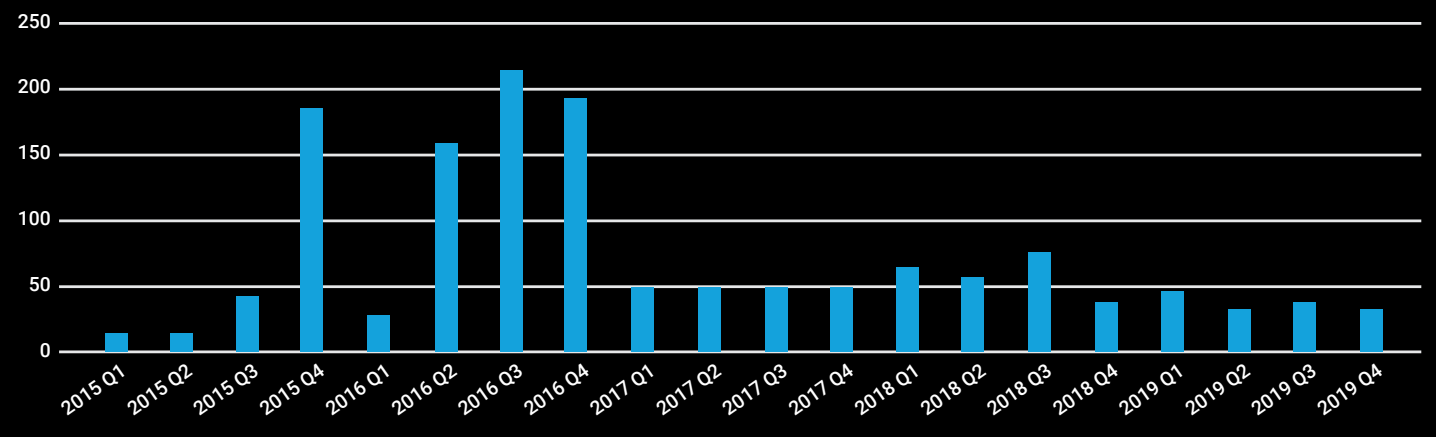

Figure 4: Generated histogram of Chinese IPv6 delegations

2. Attributed cyber defense: A critical function for any defensive cyber operation is pinpointing a cyber threat's network origin. In our second use case we confirmed that the framework and approach can attribute a potential cyber threat to a critical operational network. Unlike a tool, our framework integrates data and subsequent intelligence derived from multiple cyber network defense and non-cyber network defense sensors.

The use case emulated classic cyber defensive operations such as CERT and hunt. The environment was described as a complex operational network in which several host- and network-based sensors were deployed. The cyber reconnaissance framework was positioned as the controlling system that orchestrated and managed the disparate sensors, and processed their output in a single, unified, common intelligence picture. The fused common intelligence picture also included information from the framework's other processing modules, including network attribution intelligence, IP geolocation and organizational delegation.

To begin, a cyber network defense operator, using the framework's domain query language, described a behavior of concern and requested to be notified if it was observed within the defensive area of operations. The behavior of concern was described as an intentional, credentials-based brute-force attack on any one of a class of devices. When our framework identified a behavior of concern matching the description—a credentials-based brute-force attack on any one of a class of devices—the operator received a snapshot of the observed behavior (figure 6), which was assembled from several cyber defense network and intelligence modules feeding the cyber reconnaissance framework.

The snapshot of the attack led to additional questions to gain a better understanding of the threat, including: "Was there any evidence of service enumeration of this device prior to the supposed brute force attempt?" In response to this question, our framework provided the operator with evidence of service enumeration, as well as the metadata associated with the offender of the service enumeration, which indicated a foreign adversary-owned virtual private server. The resulting threat description was assembled into a single snapshot and flagged for interdiction (figure 7).

In addition to pinpointing the cyber threat's network origin, our framework can answer questions about operational efficiency. If there are too many operational cyber defense tools in use, operational availability and resource utilization can be impacted which can degrade the defensive mission. For this use case, the operator asked, "Which cyber defense tools have provided the most information value in all identified threats?" In response, the framework provided a list of tools, along with cost and licensing considerations, delivering a seamless data-driven approach to evaluating the defensive posture and improving the operating environment.

AN ENDURING SOLUTION FOR THE CYBER DOMAIN

The volume and velocity of cyber threats requires new thinking and solutions such as the Peraton Labs' cyber reconnaissance framework. Its raises the level of efficiency in cyber-relevant intelligence because analysts can perform their core functions and develop a common intelligence picture without tool bias.

As a technical framework, it accelerates integration across cyber intelligence systems, research and operations by automating data collection and feeding the process for observation, orientation, decision and action. This goes a long way in the development of the cyber discipline so that its conclusions balance well with the more established imagery and signals intelligence.

Finally, the cyber reconnaissance framework introduces a partner-oriented approach to addressing the critical issues in cyber intelligence. With access to the solution, the expertise behind it and the resources to operate it, agencies can optimize their cyber intelligence activities to mitigate threats and achieve their missions.

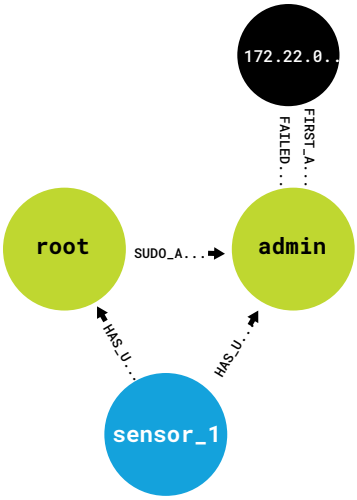

Figure 6: Detected brute force attack

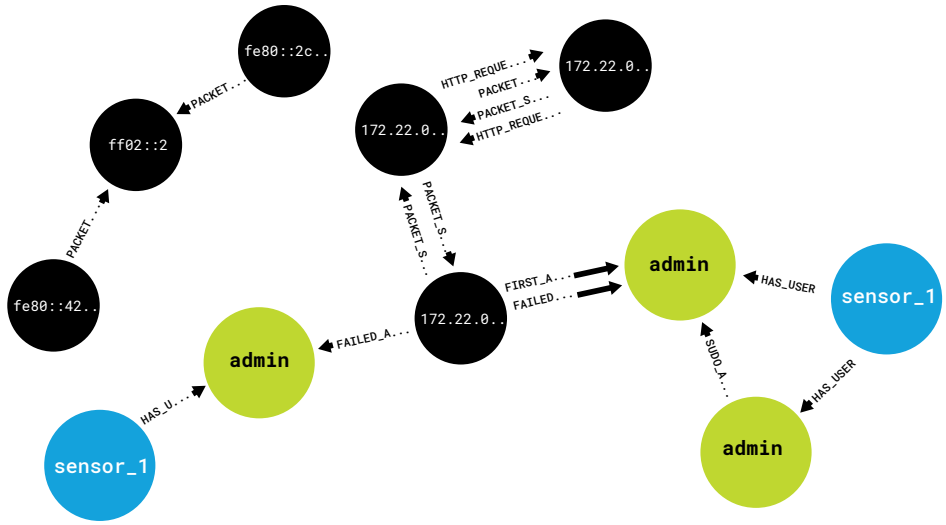

Figure 7: Threat description snapshot of service enumeration and subsequent brute force attack in the defensive operating environment

Peraton | LABS

LEARN MORE AT

PERATONLABS.COM

150 Mount Airy Road
Basking Ridge, NJ 07920