NOVETTA

CYBER ANALYTICS

Architecture Overview

Technical Brief

May 2016 | novetta.com |

NOVETTA

Novetta Cyber Analytics:
Technical Architecture Overview

INTRODUCTION

Novetta Cyber Analytics is an advanced network traffic capture, inspection and analytics application for medium to large enterprises and government agencies. Designed to ensure rapid detection of breaches, significantly reduce attacker dwell times, and enable immediate remediation to reduce further exposure, Novetta Cyber Analytics delivers the fastest and most precise network traffic analytics possible.

Most security applications monitor network and endpoint log activity, correlate actions, and produce automated alerts about events, many of which aren't relevant to protecting networks and data from advanced threats. In contrast, Novetta Cyber Analytics focuses on empowering IT security analysts by providing them with critical information to detect and research threats and the ability to rapidly query and analyze data to stop attacks. Instead of providing analysts with alerts about known attacker and malware behaviors that can change in an instant, Novetta Cyber Analytics gives analysts all the tools and data they need to dynamically identity and counter intruders in real time.

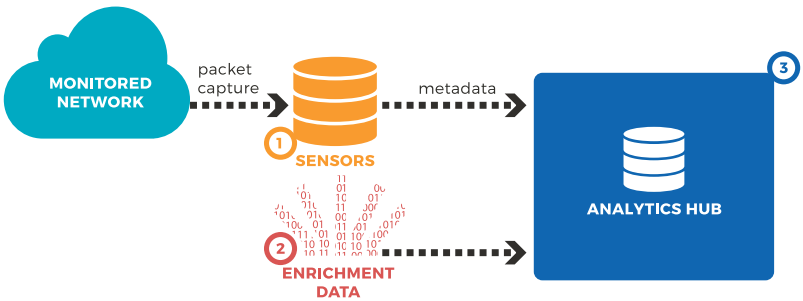
Novetta Cyber Analytics is made up of three main categories of technology:

1. **SENSORS** Novetta sensors perform full lossless packet capture, traffic metadata extraction, compression and storage of packet data, and capture and processing of data from existing legacy sensors and other network traffic data sources.

2. **ENRICHMENT DATA** Geolocation, DNS, domain and threat intelligence information, and customer black lists and white lists, add greater context to network traffic data to enable the identification of suspicious activities that would otherwise go undetected.

3. **ANALYTICS HUB** processes, combines and centrally stores metadata and enrichment data, performs powerful and rapid ad hoc and automated queries, and retrieves full packet capture data from sensors as required.

This white paper provides an overview of the technical architecture of Novetta Cyber Analytics.



Novetta Cyber Analytics Three Main Elements

CAPTURE AND PROCESS ALL NETWORK TRAFFIC

Novetta sensor software continuously performs out-of-band, lossless packet capture of all observed network traffic. The software is a collection of Novetta-authored code and open source libraries that executes a wide range of functions, from receiving and processing network packets off the network to extracting metadata, and compressing and storing packets for future access. The software runs on the Red Hat Enterprise Linux operating system on commodity rack-mounted server hardware, and passively captures a copy of all network traffic at a choice of up to 1 or 10 Gbps (clustering enables up to 100 Gbps). Once data is captured, customizable traffic sessionization logic is then applied that identifies, combines and consolidates packets to make host-to-host conversations understandable by analysts. Novetta Cyber Analytics also identifies traffic protocols and service types within captured sessions to make it easier for the software to extract metadata and for analysts to review traffic data and perform analytics.

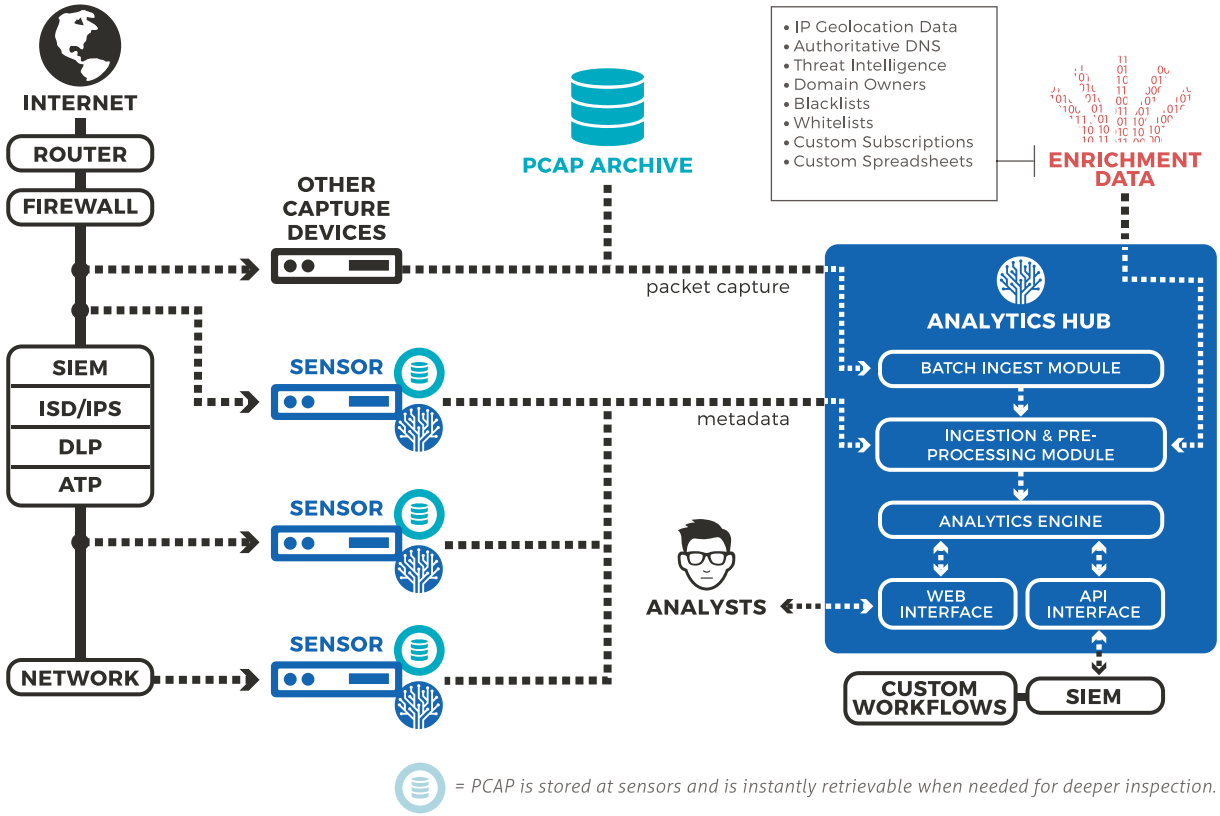
Novetta sensor software reviews the captured packets to extract IP addresses, ports, hosts, query strings, cookies and other data attributes that most concisely characterize what is happening within the traffic. Average metadata extraction ratios for Novetta sensor software are 100:1, such that for every 100 TB of recorded packet capture, 1 TB of metadata is generated. This is substantially less than the 10:1 or less average ratios of other security solutions, which require far more resources to store and maintain metadata, and can take too long to perform queries or potentially overwhelm analysts with too much data.

Novetta Cyber Analytics is intentionally selective when performing metadata extraction. The sensor software includes knowledge provided by Novetta's team and other leading security analysts who have spent many years detecting the subtlest and most prolific threat actors. Through this work, they have identified the data attributes most useful for detecting critical tactics, techniques and procedures (TTPs). To optimize analyst and system response times, only those attributes are included in Novetta Cyber Analytics metadata.

Once metadata is extracted, Novetta sensor software creates an index that links captured packets to corresponding metadata to make it easy for analysts to rapidly retrieve raw network traffic as needed. The software then compresses and stores the packet capture data and sends the metadata to the analytics hub.

Novetta Cyber Analytics employs a first-in first-out buffer approach and a variety of storage options for packet capture data. Organizations choose the type and amount of commodity storage that best accommodates their overall business and security requirements. Storage options include onboard storage on sensor hardware, direct-attached storage, storage area network or network-attached storage.

In addition to the traffic captured by Novetta sensors, organizations can also leverage existing full packet capture repositories. The batch ingest module, which is part of the analytics hub, processes data from these repositories and extracts metadata.



ADD RICHER CONTEXT

Novetta Cyber Analytics further enhances analysis of packet capture metadata by adding enrichment data. The data is inserted into the analytics hub to be used alongside the metadata as an integral part of analysis. A range of data types can be added, from IP geolocation, DNS and threat intelligence, to white lists and black lists and other information obtained from third party sources or provided by customers. Specific examples of enrichment data include:

- **IP GEOLOCATION** - Estimated locations (including longitudes and latitudes) of routable IP addresses.
- **AUTHORITATIVE DNS** - Known Internet domains and their links to IP addresses.
- **THREAT INTELLIGENCE** - Malicious IP addresses, domains and traffic types; IP addresses and domains known to be infected with malware; command and control bots and other threats.
- **DOMAIN OWNERS** - Companies and organizations that are registrants for public domains.

The data services above are included with Novetta Cyber Analytics. Organizations can also add their

own enrichment data, such as the data types listed below:

- **CUSTOMER BLACK / WHITE LISTS** - IP addresses and domains an organization considers to be good or bad that they want to utilize as contextual data or regularly search to review activities.

- **CUSTOM SUBSCRIPTIONS** - Some companies want to add their own subscription services, such as other threat intelligence services, to the data repository.

- **LISTS OF KNOWN BAD OR INNOCENT IP ADDRESSES** - Spreadsheets containing lists of IP addresses and other search criteria analysts want to review on a regular basis.

PERFORM FAST AND EFFECTIVE NETWORK TRAFFIC ANALYTICS

The Novetta Cyber Analytics hub, as shown in the diagram on page 3, includes a batch ingest module, data ingestion and pre-processing layer, data repository, analytics engine, web interface, and API interface. The analytics hub software runs on the Red Hat Enterprise Linux operating system on a cluster of commodity rack-mounted servers. The number of servers an installation requires depends on the volume of data and number of users to be supported. The hub performs the following activities:

Processing Operations

The batch ingest module, discussed above, collects and processes data from packet capture archives, extracts metadata for use in the data repository, and stores packet capture data as required.

The data ingestion and pre-processing nodes receive metadata from multiple Novetta sensors throughout the network, perform classic extract, transform and load operations on the metadata, and send it to the data repository.

Storage and Retrieval

The data repository includes a collection of database worker nodes, managed by a database leader node. The repository is built on a massively parallel processing, columnar database that centrally stores all metadata and the enrichment data tied to it. The database is able to rapidly respond to all ad hoc and automated queries because of its column-oriented database structure. Column-oriented databases run dramatically faster on metadata queries of network traffic than row-oriented structures common in conventional database systems. Speed differences are especially pronounced when the system is aggregating values of fields across many records, a common operation for Novetta Cyber Analytics.

In contrast to Novetta's approach, other network traffic capture systems employ a distributed metadata architecture and row-based database management system for storage and retrieval of data. Instead of consolidating and querying metadata in one location, each sensor node within these systems stores a portion of the metadata and queries must be run on each node. Query results from each node are returned to a central server where they are stitched together to provide analysts with results. Not only are query operations much slower with these systems than with Novetta Cyber Analytics, but also results can

be inaccurate because queries cannot run on all of the data at once.

Data Access and Review

Novetta Cyber Analytics provides two ways for users to access and retrieve data. The web interface node, host to the software's web application, provides access to metadata and analytics, and retrieves full packet capture data as required for investigations. The application programming interface (API) node allows other applications to invoke system functions similar to those available in the web application over a RESTful web API or Python API.

Rapid Queries and Research

The analytics engine receives and manages concurrent execution of queries and rapidly returns results. Queries are initiated from the Novetta Cyber Analytics web application or from security information and event management (SIEM) tools and other software applications over an API interface.

Incident responders, security analysts and cyber hunters can use the flexible query builder, presented over the web or API interface, to create ad hoc queries manually, or set them to run automatically at time intervals. The categories of supported analytics include:

- Summary reports, generated from simple queries, provide analysts with details about volumetric traffic over days, weeks or months for review, and comparison with average or normal activities e.g. total volume of traffic exchanged between hosts, volume of traffic from the network to specific countries, services (HTTP, TLS, DNS, etc.) used, server ports used.
- Investigative analytics are initiated when suspicious behavior is detected on the network, often by SIEMs, which usually consists of tightly constrained queries that deliver information about activity in a specific area of the network, e.g. traffic between two IP addresses in the past 48 hours.
- Threat detection analytics employ complex queries in a free flowing manner to proactively look for network activities that could signify intrusions, e.g. traffic on ports using the incorrect service. For examples of actual queries, see Threat Detection Queries Explained below.

Novetta Cyber Analytics includes more than 130 prebuilt queries* developed by Novetta's team based on the TTPs used by sophisticated threat actors and malware families. Many prebuilt queries, such as those to detect non-standard keyboard use, distant administrators, protocol abuse and beaconing, detect attackers attempting to gain access to and travel within secured networks, exploit vulnerabilities, and hide from traditional signature-based real-time scanning and blocking applications.

Analysts can quickly click through from metadata and query results and view packet data of interest. Metadata and packets can also be exported for use in an investigative tool of choice, e.g. Wireshark. Specified ranges of records can be downloaded as a single file to make it easier for analysts to review data involved in complex investigations.

**See the Top Ten Analytics brief at novetta.com/cyber-analytics.*

THREAT DETECTION QUERIES EXPLAINED

For proactive threat detection, analysts can use the query builder to create their own searches or customize many of the 130 prebuilt queries included in the software. Below are examples of two of the prebuilt queries included in Novetta Cyber Analytics.

INTERACTIVE SESSIONS

Interactive sessions, including reverse shell TCP sessions, are created when attackers are interacting with a network resource while trying to hide within the network noise. This query looks for low and slow traffic streams with long durations on uncommon ports or using unknown services, which indicate persistent interaction between a client and server, to uncover stealthy remote interaction traffic (e.g. using a keyboard or mouse to control a compromised host).

Query to Detect Interactive Sessions:

Find all sessions...

In one or more network traffic collections (datasets),
For given start and end date/times,
Where...
- *One IP (e.g. client) contacts another IP (e.g. server),*
- *Over TCP protocol,*
- *With a session duration of >= 60 seconds (configurable)*
- *With a total byte transfer of >= 1KB (configurable)*
- *With packet size (bytes/packet) for the server or client of < 300 (configurable)*
- *And the transfer rate (bytes/second) for the server or client is < 300 (configurable)*

MALWARE BEACON DETECTION

Beaconing is one of the first network-related indications of a malware infection, but previously unseen malware samples are hard to detect because they have no known signatures and often come from previously unseen attacker infrastructure. This query helps analysts detect infections early by identifying malware beaconing behavior originating from an organization's network.

Query to Detect Beaconing:

Find all sessions...

In one or more network traffic collections (datasets),
For given start and end date/times,
Where...
- *One IP (e.g. client) contacts another IP (e.g. server),*
- *Every [Second, Minute, Hour, Day, Week, Month] interval,*
- *[<, <=, =, >=, >] N times per interval,*
- *And there is a % persistence factor to allow for failed signals*
- *Excluding potential false positives like DNS TTL expirations*

TAGGING FOR ENHANCED ANALYTICS

Tagging capabilities within Novetta Cyber Analytics allow analysts to manually add freeform text and color-coded labels to data elements, create automated filters that apply tags to data as it flows in, or apply tags in bulk to existing data. Tags are a good way to filter out large categories of network traffic and make it easier for analysts to find suspicious or malicious traffic. Tags can be easily shared by all team members to improve collaboration or be used as search input, e.g. to look for all traffic to or from a sensitive subnet from all IP addresses outside of the United States. Below are the most common reasons analysts use tags.

- Make it easier to understand the traffic details by labeling IP addresses with their function within the network infrastructure, e.g. firewall, proxy server, wireless access point or web server.
- Flag items for further investigation by tagging elements with generic terms, such as adding a "suspicious" tag to an IP address or a "review later" label to an unrecognized element.
- Tie elements together as part of a specific investigation or threat type by adding case numbers to external IP addresses involved in a known incident or with ties to a threat attack group or an industry-specific threat already being tracked by analysts.

client ip		client port
56980340|192.168.2.107	Case123 HR	4608
56980340|192.168.2.107	Case123 HR	4609
56980340|192.168.2.107	Case123 HR	4610
56980340|192.168.2.107	Case123 HR	4611
56980340|192.168.2.118	Recruiting	4616
56980340|192.168.2.118	Recruiting	4617
56980340|192.168.2.118	Recruiting	4618
56980340|192.168.2.118	Recruiting	4619
56980340|192.168.2.120	Finance	3856
56980340|192.168.2.108	PCI_Subnet	4343
56980340|192.168.2.108	PCI_Subnet	4344
56980340|192.168.2.109	Suspicious	4814
56980340|192.168.2.109	Suspicious	4815

Tagged IP Address Examples

DELIVERING ADVANCED NETWORK INTELLIGENCE

Novetta Cyber Analytics performs full packet capture, metadata extraction and data enrichment to provide analysts with the most comprehensive, scalable network traffic analytics solution available. The software retrieves relevant and useful packet capture data, based on more subtle and complex criteria than other tools allow, and returns query results in seconds.

The true power of Novetta Cyber Analytics is delivered when advanced analytics intersect with rapid retrieval of full packet capture data to empower analysts to rapidly analyze network traffic, quickly retrieve the specific data they need to identify breach locations, decrease attacker dwell times, and immediately remediate and prevent exposures.

FOR MORE INFORMATION, PLEASE VISIT

novetta.com/cyber-analytics