White Paper

Using entity context for visibility into devices, applications, and users

Corelight Entity Collection

Introduction

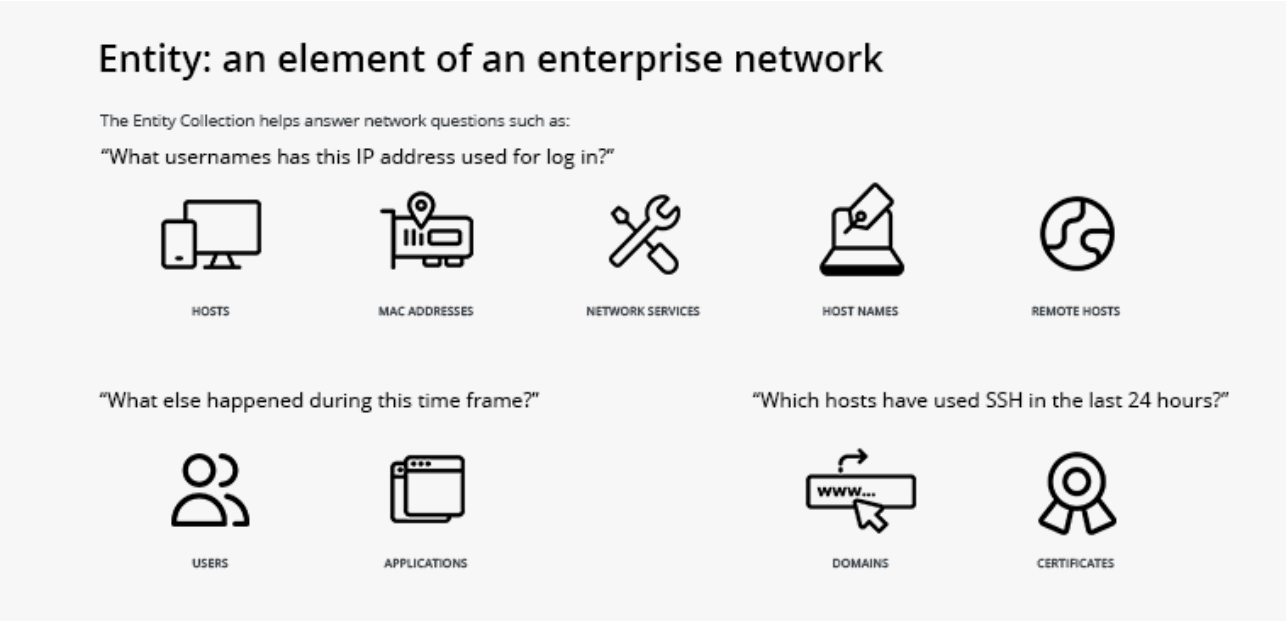
Understanding the context of an alert or incident during an investigation can be challenging when the information needed is difficult to find. To compensate, analysts often expend too much effort hand-crafting summaries and running SIEM searches that cover large spans of time. In addition, asset/inventory management is notoriously difficult in large enterprise environments—beyond most CMDBs lies an ever-changing inventory of unknown, unmanaged entities traversing the network. The Corelight Entity Collection improves entity visibility while accelerating incident response and threat hunting workflows.

Capabilities

With the Corelight Entity Collection, security teams gain powerful identification capabilities to more comprehensively map and defend their environment. The collection identifies and summarizes activity related to apps, devices, services, subnets, certs, hosts, and more to help teams close asset visibility gaps, raise operational awareness, and gain immediate context about related entities to accelerate investigation and hunts. The Corelight Entity Collection comprises the Known Entities Package, Application Identification Package, and the Local Subnets Package.

Known Entities Package

This package provides relevant summaries of information normally hand-crafted by analysts and security teams for threat hunting and incident response. Corelight defines an entity as an enterprise network element such as a system, server, user, domain, or certificate. These attributes are available in a set of interlinked logs that are summarized from the full Corelight log streams for fast searching. This log set includes entity information about everything on your networks, from IT devices (laptops, servers, phones, printers) to Industrial Control System (ICS) and Operational Technology (OT) devices (building automation, cameras, and industrial control systems).



The package simplifies extraction of the data from Corelight logs, synthesizing it together in a compact form. The known logs operate differently than most of Corelight's other data logs in that they focus on summarizing activity over 15-minute periods and, by default, only operate on your internal networks.

Log format differs by type, but all include a KUID (known UID) for showing all entity data for a specific time window and an option to display the last KUID to create a chain. Each log aggregates and summarizes data from multiple other sources.

Known Entity Logs:

Log	Description
known_hosts.log	Lists the hosts (in the local network) that were active during an observation interval
known_devices.log	Lists the hardware MAC address for the hosts observed
known_services.log	Lists the TCP and UDP services that hosts in the local network offer
known_names.log	Records the host names observed within the local network
known_certs.log	Lists the details of the certificates handed out by a server
known_domains.log	Lists the domains associated with hosts observed in an observation interval
known_users.log	Tracks user accounts (across remote login and authentication protocols) in the local network
known_remotes.log	Lists the remote entities involved in bidirectional communication with entities in the known_hosts.log

Use cases

There are four primary use cases for this new set of known entity logs:

Context

Context from the known entity logs allows an analyst to quickly find important context about an entity or set of entities related to an alert or investigation.

For example, imagine starting the incident response process with a single internal host IP from an alert:

As an analyst, I first want to understand all the information about that host, and I do that by asking questions:

- *Is a hostname associated with this IP? What is its MAC address?*
- *Have I seen authentication events that would tie this host to a user?*
- *Are services offered by this IP that provide a clue to the type of machine or its function in the network? How busy is the host on the network?*
- *Can I tell where this device is and how it might be related to the organization*

The Known Entities logs help quickly take a single host IP address and find its associated entities. Instead of searching across the connection log and multiple protocol logs, security teams can drill down into that context quickly and efficiently. After a quick entity search, other entities worth exploring may show up, and the small size of logs allows quick pivoting without breaking investigation flow.

Asset inventory

The logs also provide data analysts can use to create an accurate inventory of devices on the network.

By succinctly summarizing the activity for every host on the network in 15-minute windows, it's easy to answer questions such as:

- *How many hosts were active on subnet B today?*
- *Which devices have MAC addresses associated with Canon printers?*

By drawing information from network connections, DNS, DHCP, and more, teams can generate an inventory and check for changes for any time window that matters. It's also a quick way to take stock of all the potentially unmanaged devices on a network. For example, a few quick searches and a list of managed hosts merged with everything seen over the past 24 hours provides an idea of what may be unmanaged, along with some properties about those unmanaged assets.

Indexing

The known entity logs provide a faster way for an analyst to search for many indicators or across a wide time horizon and quickly pivot to a subset of the full logs for deeper investigation.

Consider it the TL;DR for Corelight logs. Searching for multiple IOCs across large spans of time can take too long in many SIEMs. Indexing helps quickly answer questions such as:

- *Did this IP address interact with any other hosts on Tuesday?*
- *Was this internal hostname looked up on the network in the past 30 days?*

Known logs help quickly identify the time window and relevant summaries for the traffic. With known logs as the starting point, it's easy to zoom in on other important data in the full logs, providing a shortcut to coffee break searches and saving valuable time in the analysis process. These summary logs are nearly two orders of magnitude smaller than the full logs, making it more affordable to store or keep them in a faster index for much longer.

Application Identification Package

Easily identifying the software applications running on the network provides helpful context to security analysts. While Corelight data identifies underlying ports and protocols, sometimes more detail is needed to understand (especially in encrypted traffic) which applications are using those protocols.

To this end, the Entity Collection includes our new Application Identification Package. Using various techniques, from DNS queries to certificate SNIs and protocol metadata, Corelight categorizes over 150+ types of applications and writes a new field directly to the connection log for easy correlation. These identified applications also feed into the Known Entity logs described above.

The Entity Collection can help you identify 150+ applications communicating on your network.

Zero Trust policy violation detection

Known entity logs can enable Zero Trust policy violation detection. For example, Corelight can deliver evidence to prove without repudiation that the user (known_users.log) connected from an authorized system (known_hosts.log) using an authorized protocol/application (known_services.log) at an authorized time and that the connection terminated after the authorized duration expired (conn.log).

Local Subnets Package

Many of Corelight's detections and data generation capabilities are driven by understanding the local network subnets. Without knowing what's "inside" and what's "outside" the particular vantage point of your Corelight Sensor, it can be hard to trigger the appropriate alert or provide the context for an investigation. While customers can manually configure local Corelight Sensor network settings, we find that customers don't always actually know all of their local networks (and sometimes just forget to configure them).

The Entity Collection's new Local Subnets Package uses sophisticated algorithms to determine what local subnets have been seen and then generates a summary (as a "notice"). The list can be used to configure the local network settings on a Corelight Sensor or to identify any potentially overlooked networks in the data.

Conclusion

The Entity Collection can help quickly identify new applications and subnets in use on local networks and accelerate incident response and threat hunting using the powerful data summaries in the Known Entities logs. In addition, many summary indexes, custom dashboards, and inventory scripts our customers are using can be replaced or simplified due to the power of the summarized data in the known logs.

Learn more about Corelight Collections

Corelight Collections are detection sets included with your Corelight subscription and can be activated depending on your needs. In addition to the Entity Collection, Corelight also offers the following collections:

Encrypted Traffic Collection
C2 Collection
Core Collection

Corelight provides security teams with network evidence so they can protect the world's most critical organizations and companies. On-prem and in the cloud, our open Network Detection and Response platform enhances visibility and analytics, leading to faster investigations and expanded threat hunting. Corelight's global customers include Fortune 500 companies, major government agencies, and large research universities. Based in San Francisco, Corelight is an open-core security company founded by the creators of Zeek®, the widely-used network security technology.

info@corelight.com | 888-547-9497