Setting up the Use Case

Intro

One of the key design principles of Apache Metron is that it should be **easily extensible.** We envision many users using Metron as a platform and building custom capabilities on top of it; one of which will be to add new telemetry data sources. For the purpose of this documentation, we will walk you through how to add a new data telemetry data source: Squid proxy logs.

Setting up the Use Case Scenario

Customer Foo has installed Metron and they are using the out-of-the-box data sources (PCAP, YAF/Netflow, Snort, and Bro). They love Metron! **But now they want to add a new data source to the platform: Squid proxy logs.**

Customer Foo's Requirements

The following are the customer's requirements for Metron with respect to this new data source:

1. The proxy events from Squid logs need to be ingested in real-time.
2. The proxy logs must be parsed into a standardized JSON structure that Metron can understand.
3. In real-time, the Squid proxy event needs to be enriched so that the domain names are enriched with the IP information.
4. In real-time, the IP within the proxy event must be checked for threat intel feeds.
5. If there is a threat intel hit, an alert needs to be raised.
6. The system should provide the ability to configure rules via a custom DSL to prioritize/score different types of alerts.
7. The end user must be able to see the new telemetry events completely enriched from the new data source. But most importantly, the user should be able to see the alerts prioritized by the high priority with the corresponding contextual data.
8. All of these requirements will need to be implemented easily without writing any new Java code.

But wait....Customer Foo still wants more. Customer Foo wants to extract intelligence from the Squid telemetry stream and apply this intelligence in real-time to the threat triaging function. The additional requirements are:

1. Be able to profile my data and incorporate statistical features of the profile into my threat triage rules
2. Be able to deploy a machine learning model that derives additional insights from the stream
3. Incorporate my machine learning model into my threat triage rule along with threat intel, static rules, and statistical rules

What is Squid?

Squid is a caching proxy for the Web supporting HTTP, HTTPS, FTP, and more. It reduces bandwidth and improves response times by caching and reusing frequently-requested web pages. For more information on Squid see Squid-cache.org.

How Metron Enriches a Squid Telemetry Event

When you make an outbound http connection to https://www.cnn.com from a given host, the following entry is added to a Squid file called access.log.

The following represents the magic that Metron will do to this telemetry event as it is streamed through the platform in real-time:

Key Points

Some key points to highlight:

- We will be adding a net new data source without writing any code. Metron strives for easy extensibility and this is a good example of it.

- This is a repeatable pattern for a majority of telemetry data sources.