When to use the Ingestion API

Introduction

Google Chronicle is built on top of REST APIs, which allow customers to interact with the platform for search, detection, ingestion and other administrative functions. The Chronicle Ingestion API is used to push data into Chronicle, and forms an integral part of the data ingestion and processing pipeline. Customers have multiple ways of interacting with the Ingestion API:

- **Data Feeds** - managed through the Feed Management UI (or API), data feeds provide pre-built methods of fetching data from a cloud-based API, or cloud-based storage, and pushing the data into Chronicle via the Ingestion API. Customers are able to create these feeds quickly and easily to get common events or context data into Chronicle and reduce time-to-value. Examples of using data feeds to get data into Chronicle is pulling data directly from a REST API in a SaaS service, or writing event logs into an S3 bucket and pulling the data from there.
- **Forwarder** - the Chronicle forwarder is a Docker-based software component which is typically deployed within a customer's infrastructure, collects data from behind a firewall in a customer's environment, pushing the data to the Chronicle Ingestion API. Customers deploy this and, through configuring the forwarder, are able to push logs from traditional on-premises applications, devices, and systems into Chronicle.
- **Ingestion API** - while data feeds and the forwarder provide flexibility for customers to bring their data to Chronicle, there may be scenarios where these are not quite flexible enough to fetch or push data from applications, particularly where customers or partners have complex, bespoke applications which are either not exposed to the internet, or do not surface event or context data via a standard interface such as REST API or syslog. In such scenarios it may be simpler for the customer or partner to write custom code to send logs directly to Chronicle's Ingestion API.

Chronicle Ingestion API

Chronicle's Ingestion API is documented on the support site here, but it offers a couple of ways to format data before pushing it into the API:

- **Pre-formatted UDM events or entities** - Chronicle's Unified Data Model (UDM) provides a structured data format which can be used to describe events and entities, and forms the basis of searching and running detections against security data within Chronicle. Customers or partners may choose to pre-format the data they are sending to Chronicle into UDM in order to have ultimate control over the data structure once it is in Chronicle, and remove the reliance on the parsing process.
- **Unstructured log data** - posting unstructured log data to the Chronicle Ingestion API is how most data gets into Chronicle, whether this be posted directly, via the forwarder, or via a data feed. This API endpoint can be called directly for customers who wish to batch and send the data themselves without using the forwarder to do this.

As shown in the documentation, for both pre-formatted UDM and unstructured data, the events or entities should be batched up in a JSON body, and forwarded to the relevant endpoint. In the case of unstructured data, this JSON data should also include the log type associated with the data to correctly inform the data parsing process.

Parsing data in Chronicle

The parsing process will take a raw unstructured log received via the Ingestion API, and will then run the log through a Config Based Normalization (CBN) parser to normalize the data into the UDM format. The parser used for this could be one of the default parsers available out of the box with Chronicle, or could be through a custom parser written by the customer or partner, and submitted through the Chronicle Parser API. Advantages and disadvantages of the parsing process are listed below, note that if you do not wish to use this process, and require ultimate control of the format of the data, then the pre-formatted event or entity APIs should be used, either directly or via a partner solution.

Parsing Advantages

- Many out of the box default parsers are already available
- Does not require any pre-processing of event or entity data
- Minimal intervention needed by customers or partners

Parsing Disadvantages

- Where custom parsers are required (default parser is not currently available, or data is not available in a supported format), customers or partners need to build and manage parsers themselves
- Where using default parsers, these may change over time, leading to detection rules requiring maintenance

When to use the Ingestion API directly?

As we have seen, the parsing process has advantages and disadvantages, but when you want to have the ultimate control over how your application security data is mapped to UDM, and are able to write the data extraction and normalization into your application, or data pipeline, sending data to the Ingestion API directly as UDM events or entities is a great option.

Additionally, where you have an environment where it is not possible to deploy a forwarder, write data to an S3 bucket, or data is not available in the desired format via a REST API, then gathering the data yourself and then batching and forwarding it to the Ingestion API as unstructured logs can help.

Getting started with the Ingestion API

There are a number of tools and resources to help you get started with the Ingestion API:

- Google Chronicle official documentation - this documents the
- Context Aware Detection official documentation - this documents the Entity Ingestion API specifically
- Chronicle API Samples GitHub repository - includes sample Python scripts to get started with testing the Ingestion API

Conclusion

While sending data directly to the Chronicle Ingestion API is unlikely to be the default option for most data types, and in most environments, it does offer an alternative way to get data in, be that pre-formatted as UDM, or as unstructured data.

By pre-formatting the data, customers and partners can take control of how the resultant data will be presented in Chronicle, and provide an incredibly predictable and repeatable result without needing to think about how the parsing process may affect their data.

Half the battle in gathering and analyzing security data is getting data into the aggregation system, and the Chronicle Ingestion API provides another choice that customers can take advantage of to make this process as easy and flexible as possible.