Lab 15

Threat Intelligence

Lab Objectives

- Manually parse through Firehol data via Opensearch
- Query the LLM to request IPs from the Firehol Index

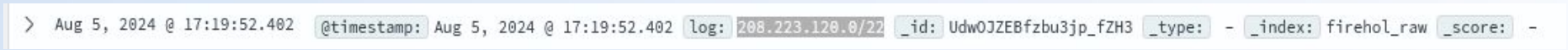
Last but not least...

In our “firehol_knn” index, we have the firehol level 1 IPList ingested. This is a very barebones list consisting of just IPs, with no other information. We will go through both querying it, as well as the effect of having not enough detail within the index for RAG to function properly.

Let’s identify a source of interest

Open up the Discover panel again, and open up the “firehol_raw” index. From here, pick any IP/subnet that matches your interest. For our case, we will grab this one:
`208.223.120.0/22`

Note: Please try to grab a different IP than what is listed on this slide



LLM Query

HTTP Method and Endpoint:

- `POST /_plugins/_ml/agents/RtwEIJEBfzbu3jp_wwAi/_execute`
- Executes a specific agent identified by the ID `RtwEIJEBfzbu3jp_wwAi`.

Request Body:

- **`parameters`:**
  - **`question`:** "Does the IP 208.223.120.0 exist in this dataset?"
    - The question being submitted for processing by the agent, inquiring about the existence of a specific IP address in a dataset.

```
POST /_plugins/_ml/agents/SPtHoJgBBHdPVe5Coawy/_execute
{
  "parameters": {
    "question": "Does the IP 208.223.120.0 exist in this
dataset?"
  }
}
```

About that IP Before

From the previous lab where we looked at the web logs- you should have saved a few IPs from that dataset. Try to ask the LLM about them here. We will be using the firehol dataset with this query. In theory, if you do have any matches, the LLM should pull them up.

Limitations Limitations Limitations

Depending on the IP you chose, you may run into the limitation mentioned previously regarding the embedding model. In short, we have too much data for the embedding model to process properly. Below is what happens when it is unable to process all of the data - only two documents are returned by the embedding model.

```
        "name": "MLModelTool",
        "result": """To answer this question, we need to analyze the given context which consists of two documents from an Elasticsearch index named
          "firehol_knn". Each document contains information about a specific IP range and its score calculated by K-Nearest Neighbors (KNN) algorithm.

Looking at the provided JSON documents, we can see that the first document has an IP range of "208.238.44.0/22" and the second document has an IP range of "203
  .22.187.0/24".

Comparing these IP ranges to the one we want to check, which is "208.223.120.0", we can see that it doesn't match either of the existing ones. Therefore, based
  on the available data, we cannot conclude that the IP range "208.223.120.0" exists in this dataset.

So, the answer is no, the IP 208.223.120.0 does not exist in this dataset."""
```

Additional Threat Intel

In order to have some better information about the IPs we have recorded- we also have an index that has additional information about some of the IPs from AbuseIPDB. Not every IP in our original dataset is in AbuseIPDB. As such, you may have to round robin through some in order to get a full result. Below is the AgentID(query) to use with the AbuseIPDB dataset.

```
POST /_plugins/_ml/agents/D4JSeJgBlaNTCsEIkPr2/_execute
{
  "parameters": {
    "question": "Give me more information about
114.119.142.197"
  }
}
```

Lab End