Day 2: Log Analysis
“O Data, All Ye Faithful”

12/02/2023 | Link | TryHackMe Advent of Cyber 2023

Table of Contents

Executive Summary

Indicators of compromise (or IOC's) are sought after by blue team cybersecurity professionals in order to determine if a cyber attack has taken place. Similarly, Indicators of Attack (or IOA's) are also sought after, in order to determine if a cyber attack is presently ongoing. IOC's and IOA's are often found through analysis of system logs or through publicly shared IOC's.

Relevance

Although incredibly powerful elements of cybersecurity defense, when every single application, on every single host, on every single network generates thousands of log events every single day, it's impossible for cybersecurity professionals to analyze every single log event. Time spent looking though past data prevents cybersecurity professionals from fortifying networks for potential future dangers.

Data science tools and processes are essential for being able to quickly use data to protect systems. SIEMs are essential for being able to:

1. Collect data from multiple sources;
2. Process and normalize data into a single index;
3. Data mine and identify the relationships within the data;
4. Analyze data to identify past, present, and predictive trends;
5. Communicate data findings through visualizations and reports.

Demonstration

This lab demonstrates the usage of Pandas, a Python library, to manipulate, process, and structure data in order to analyze a packet capture to identify how many packets were captured in total, which IP address generated the most amount of traffic, and identify which protocol was used most frequently in the capture, ICMP.

Table of Contents

Lab Narrative

Given data in the file network_traffic.csv, we are going to use Pandas to sort the data to analyze it.

Count

Import Pandas library, and assign the aliases to make code easier to write.

```
#import Pandas under alias pd
import pandas as pd

#assign df alias to dataframe
dataframe = df

#assign df to call and read 'network_traffic.csv'
df = pd.read_csv('network_traffic.csv')

#display top 5
df.head(5)
```

Running the last line of code above, returns the following lines:

	PacketNumber	Timestamp	Source	Destination	Protocol
0	1	05:49.5	10.10.1.7	10.10.1.9	HTTP
1	2	05:50.3	10.10.1.10	10.10.1.3	TCP
2	3	06:10.3	10.10.1.1	10.10.1.2	HTTP
3	4	06:10.4	10.10.1.9	10.10.1.3	ICMP
4	5	06:10.4	10.10.1.1	10.10.1.7	ICMP

```
#call count function to df
df.count()
```

We learn that the number of packets captured is 100 (see PacketNumber 100).

```
df.count()
```

```
PacketNumber    100
Timestamp       100
Source          100
Destination     100
Protocol        100
dtype: int64
```

Table of Contents

Groupby

```
#group and display the Source column and the size column
df.groupby(['Source']).size()
```

We can see that 10.10.1.6 sent the most traffic during this packet capture.

```
df.groupby(['Source']).size()
```

```
Source
10.10.1.1      8
10.10.1.10     8
10.10.1.2     12
10.10.1.3     13
10.10.1.4     15
10.10.1.5      5
10.10.1.6     14
10.10.1.7      5
10.10.1.8      9
10.10.1.9     11
dtype: int64
```

```
#group and display the Protocol column and the size column
df.groupby(['Source']).size()
```

We can see that the protocol most used during this capture was ICMP.

```
df.groupby(['Protocol']).size()
```

```
Protocol
DNS     25
HTTP    24
ICMP    27
TCP     24
dtype: int64
```

Table of Contents

Conclusion

This lab demonstrated the effectiveness of Python as a data science tool in the context of cybersecurity. Alternatively, we could have opened the csv file in Excel and sorted and regrouped the 100 packets by hand. However, in a real world situation, the packet capture would likely have hundreds or thousands of entries, making sorting by hand extremely time consuming.

While there aren't any immediately unusual findings from this packet capture alone, it is worth noting that ICMP is typically associated with ping requests. This could possibly be a precursor to ICMP flooding targeted at the host 10.10.1.4, but with not enough data, it would be hasty judgment to immediately begin executing incident response processes. Blue team should continue to further monitor traffic to/from 10.10.1.4 and ICMP protocol usage.

Table of Contents