Alert Analysis

- Mukul Raj Sharma

Dataset:

- time - Anomalous flow observed time,
- device_mac - MAC address of the device which involved in the anomalous flow,
- device_type - Type of the device (1 - Axis camera, 2 - Cisco camera, 4- Telemecanique sensor)
- remote_endpoint - IP address of remote server involved in the anomalous flow,
- eth_src - Source MAC address,
- eth_dst - Destination MAC address,
- eth_type - Ethernet type,
- ip_src - Source IP address,
- ip_dst - Destination IP address,
- ip_proto - IP protocol,
- src_port - Source port number,
- dst_port - Destination port number,
- packet_count - number of packets observed,
- byte_count - byte count observed,
- threat_name - Name of alert,
- threat_description - Description of alert.

Correlations:

A brief overview of the correlations between various fields giving us an idea which areas might have interesting patterns. The text and IP fields have been converted to unique numerical data points before plotting this chart.

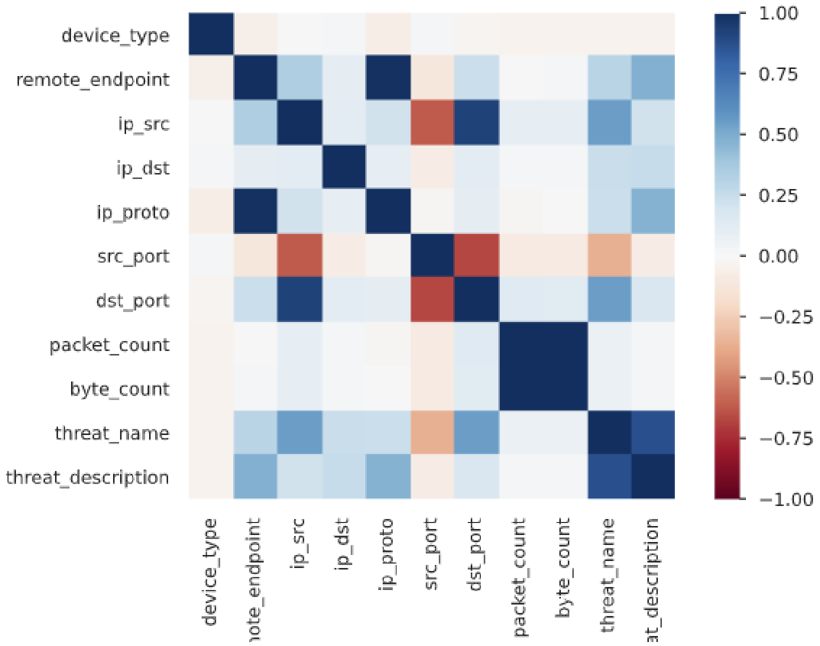
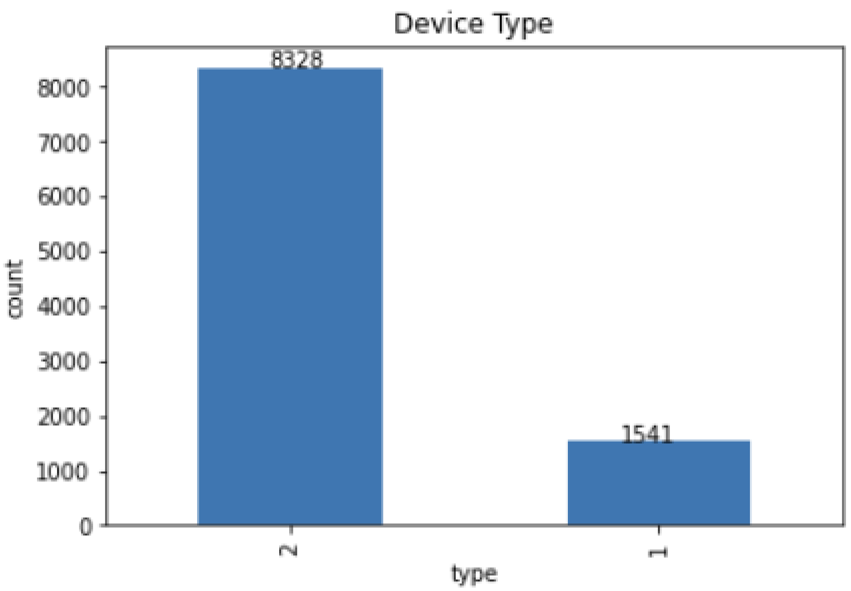


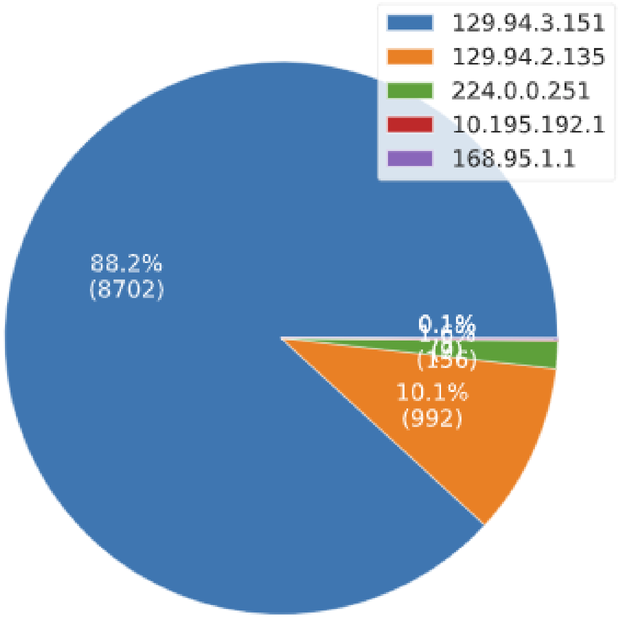
Fig 1. correlation matrix

- Strong relation between threat names and descriptions (expected).
- Packet count and byte count are directly related (b.c. multiplte of p.c.).
- Source port and destination ports have strong negative correlation.
- Remote endpoint and IP protocol are strongly related.
- Port and threat type/name are also related which is also expected.

Data counts and spread:

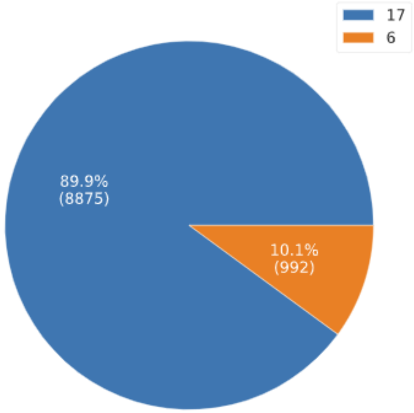
- Cisco camera is significantly overpowering the dataset.
- No datapoint for Telemecanique sensor(category 4).



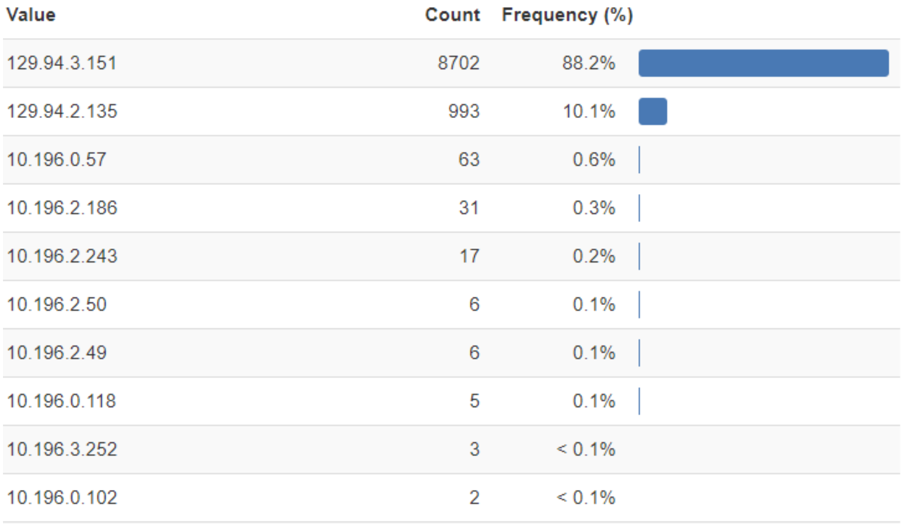


Remote_endpoint

- Remote endpoint for most (88%) threats belongs to a particular IP.
- The top 2 endpoints and the 89% IP protocols belonging to category 6 justify their strong correlation in the matrix *(fig 1.)*



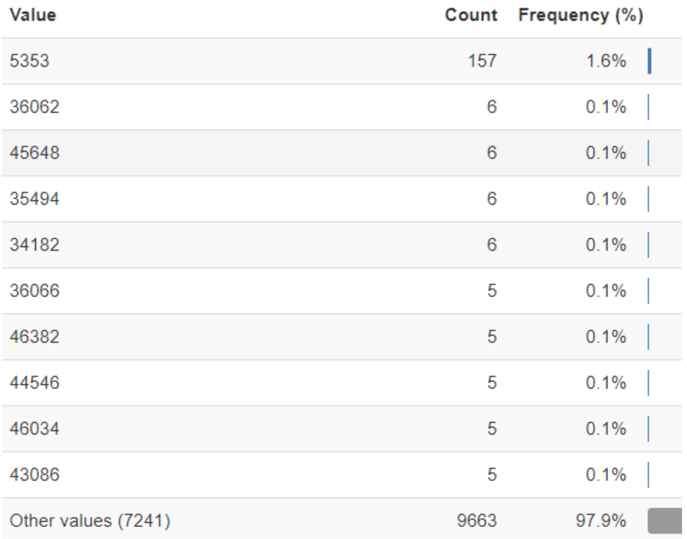
Ip_proto



Value	Count	Frequency (%)
129.94.3.151	8702	88.2%
129.94.2.135	993	10.1%
10.196.0.57	63	0.6%
10.196.2.186	31	0.3%
10.196.2.243	17	0.2%
10.196.2.50	6	0.1%
10.196.2.49	6	0.1%
10.196.0.118	5	0.1%
10.196.3.252	3	< 0.1%
10.196.0.102	2	< 0.1%

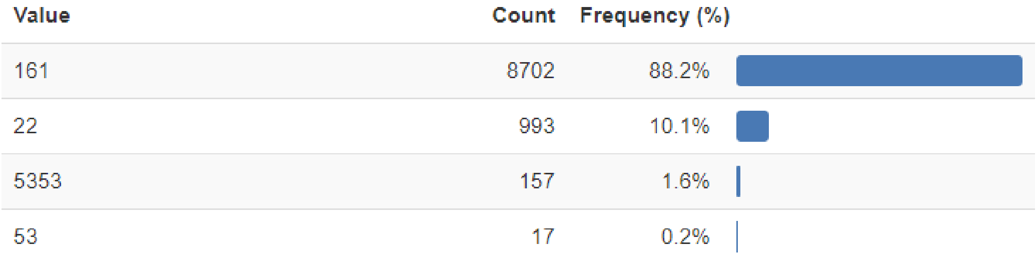
ip_src

- Pattern for ip_src and remote_endpoint are almost identical with the same sources making up 90% of the datapoints.
- IP destination is much more spread out and no such pattern can be seen.



Value	Count	Frequency (%)
5353	157	1.6%
36062	6	0.1%
45648	6	0.1%
35494	6	0.1%
34182	6	0.1%
36066	5	0.1%
46382	5	0.1%
44546	5	0.1%
46034	5	0.1%
43086	5	0.1%
Other values (7241)	9663	97.9%

src_port

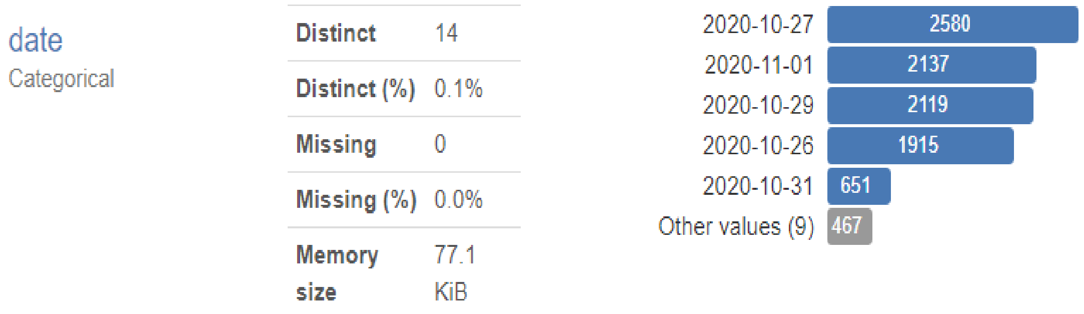


Value	Count	Frequency (%)
161	8702	88.2%
22	993	10.1%
5353	157	1.6%
53	17	0.2%

dst_port

- As it can be seen that all the threats on port 161(SNMP) are raised from IP 129.94.3.151 and on 22(SSH) from 129.94.2.135.
- This also justifies their strong corelation in *fig 1.*

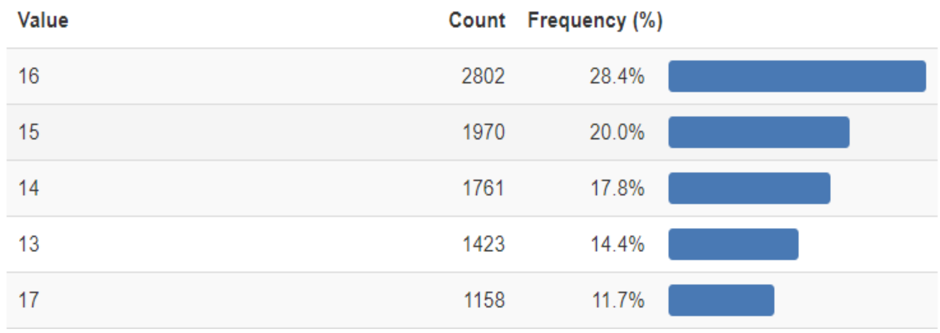
Date time (studied after splitting):



date
Categorical

Distinct	14
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	77.1 KiB

2020-10-27	2580
2020-11-01	2137
2020-10-29	2119
2020-10-26	1915
2020-10-31	651
Other values (9)	467

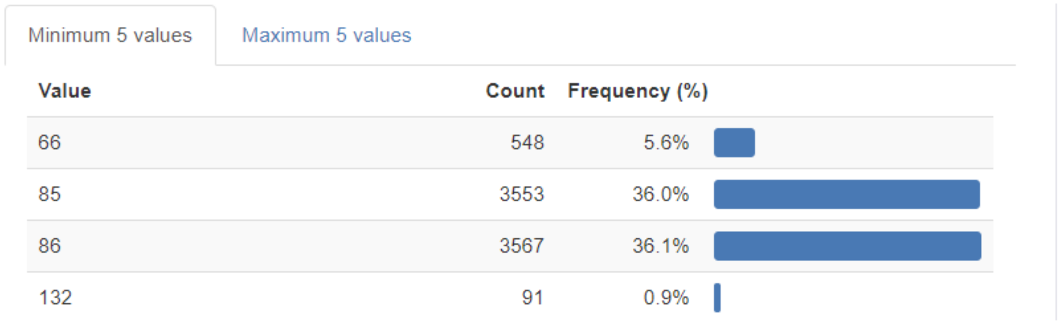


Value	Count	Frequency (%)
16	2802	28.4%
15	1970	20.0%
14	1761	17.8%
13	1423	14.4%
17	1158	11.7%

hours

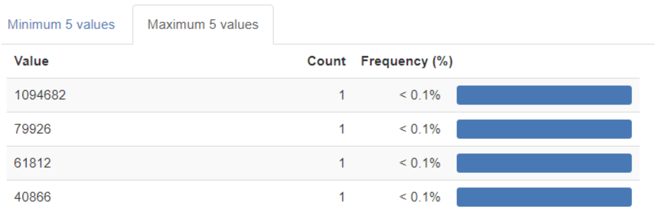
- Not a lot of information can be gained from the date field.
- The hours field however shows that most alerts are raised during the daytime between 13:00 and 17:00.

Byte Counts:



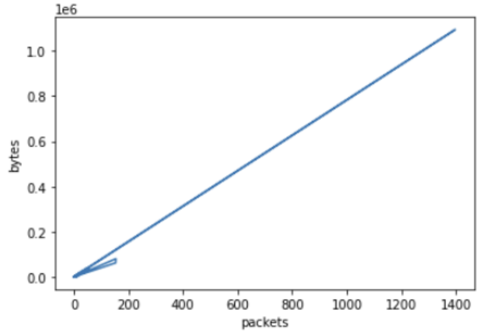
Minimum 5 values | Maximum 5 values

Value	Count	Frequency (%)
66	548	5.6%
85	3553	36.0%
86	3567	36.1%
132	91	0.9%



Minimum 5 values | Maximum 5 values

Value	Count	Frequency (%)
1094682	1	< 0.1%
79926	1	< 0.1%
61812	1	< 0.1%
40866	1	< 0.1%



- Most alerts lie between 66-150 byte count.
- Some extreme values can however be observed going as high as 1094682.
- Byte counts are linearly related to packet counts when packet count exceeds 200.