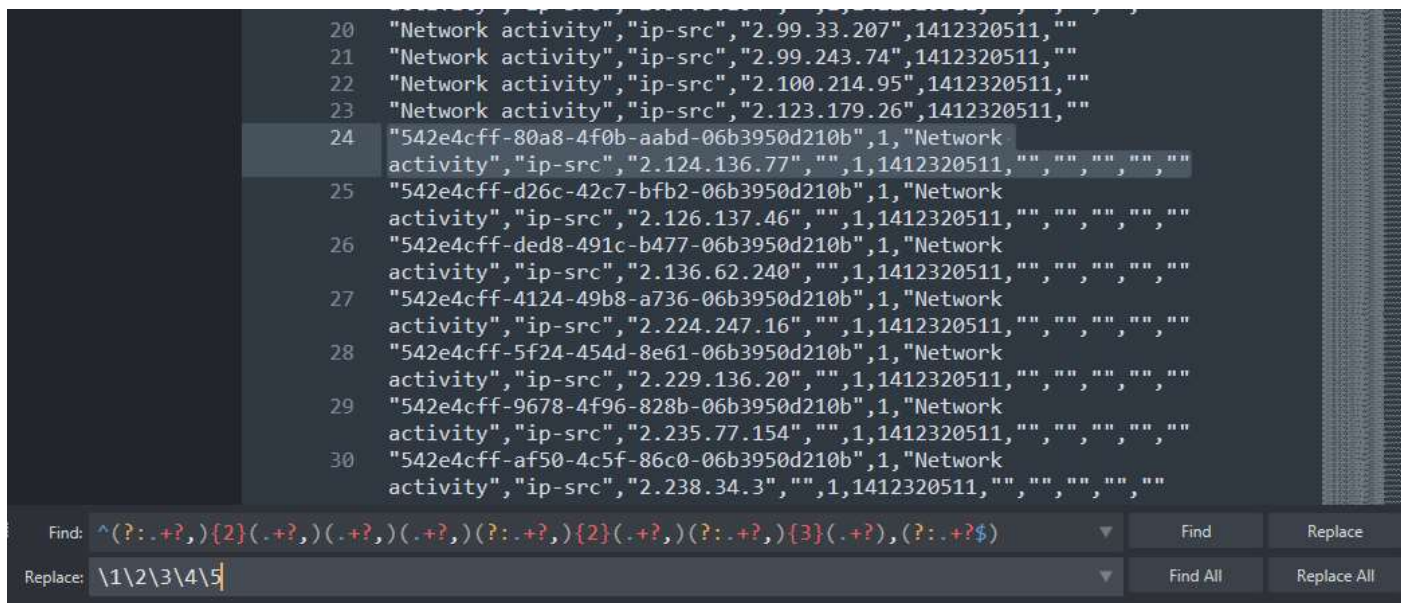


SANS DFIR MISP Data Cleanup Example

Wednesday, August 10, 2022 1:39 PM

Get Just a Subset of Data

1. First row indicates data types per field. Decide what we want to keep.
 - a. Here we have:
uuid,event_id,category,type,value,comment,to_ids,date,object_relation,attribute_tag,object_uuid,object_name,object_meta_category
 - b. There are 13 fields
 - c. I only want category,type,value,date,object_name (fields 3,4,5,8,12)
2. Build a regex for CSV fields that will get each field associated with a particular capture group:
 - a. `^(?:.+?),{2}(?:.+?)(?:.+?)(?:.+?){2}(?:.+?)(?:.+?){3}(?:.+?)(?:.+?){2}$`
 - b. Explanation
 - i. `^` - Start at the beginning of the line
 - ii. `(?:)` - Create a non-capturing group (find matches, but don't capture them for further use)
 - iii. `.+?` - Match all characters up to (and including) the next comma.
 - iv. `{2}` - Repeat the previous pattern (in this case, the non-capture group) 2 times.
 - v. `()` - Create a capturing group where matches are available for reuse.
 - vi. Repeat capturing and non-capturing groups as needed
 - vii. `(.+?)` - The last capturing group is `_not_` bounded by a comma (so that when we insert it at the end of the line, we have clean line endings)
 - viii. `$` - End of the line
3. Find and replace each line with the capture groups, in the order that you want to see the data
 - a. Test your expression:

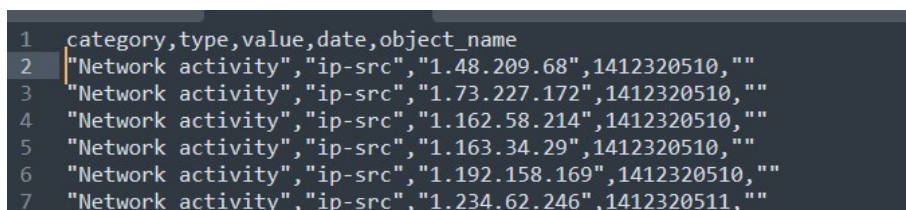


```
20 "Network activity","ip-src","2.99.33.207",1412320511,""
21 "Network activity","ip-src","2.99.243.74",1412320511,""
22 "Network activity","ip-src","2.100.214.95",1412320511,""
23 "Network activity","ip-src","2.123.179.26",1412320511,""
24 "542e4cff-80a8-4f0b-aabd-06b3950d210b",1,"Network
activity","ip-src","2.124.136.77",1412320511,"", "", "", "", ""
25 "542e4cff-d26c-42c7-bfb2-06b3950d210b",1,"Network
activity","ip-src","2.126.137.46",1412320511,"", "", "", "", ""
26 "542e4cff-ded8-491c-b477-06b3950d210b",1,"Network
activity","ip-src","2.136.62.240",1412320511,"", "", "", "", ""
27 "542e4cff-4124-49b8-a736-06b3950d210b",1,"Network
activity","ip-src","2.224.247.16",1412320511,"", "", "", "", ""
28 "542e4cff-5f24-454d-8e61-06b3950d210b",1,"Network
activity","ip-src","2.229.136.20",1412320511,"", "", "", "", ""
29 "542e4cff-9678-4f96-828b-06b3950d210b",1,"Network
activity","ip-src","2.235.77.154",1412320511,"", "", "", "", ""
30 "542e4cff-af50-4c5f-86c0-06b3950d210b",1,"Network
activity","ip-src","2.238.34.3",1412320511,"", "", "", "", ""
```

Find: `^(?:.+?),{2}(?:.+?)(?:.+?)(?:.+?){2}(?:.+?)(?:.+?){3}(?:.+?)(?:.+?){2}$`

Replace: `\1\2\3\4\5`

- b. Run it against the whole doc:



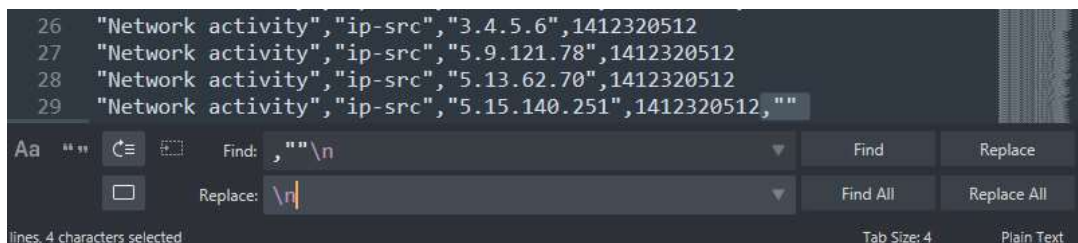
```
1 category,type,value,date,object_name
2 "Network activity","ip-src","1.48.209.68",1412320510,""
3 "Network activity","ip-src","1.73.227.172",1412320510,""
4 "Network activity","ip-src","1.162.58.214",1412320510,""
5 "Network activity","ip-src","1.163.34.29",1412320510,""
6 "Network activity","ip-src","1.192.158.169",1412320510,""
7 "Network activity","ip-src","1.234.62.246",1412320511,""
```

```

1 category,type,value,date,object_name
2 "Network activity","ip-src","1.48.209.68",1412320510,""
3 "Network activity","ip-src","1.73.227.172",1412320510,""
4 "Network activity","ip-src","1.162.58.214",1412320510,""
5 "Network activity","ip-src","1.163.34.29",1412320510,""
6 "Network activity","ip-src","1.192.158.169",1412320510,""
7 "Network activity","ip-src","1.234.62.246",1412320511,""
8 "Network activity","ip-src","2.25.130.80",1412320511,""
9 "Network activity","ip-src","2.25.141.149",1412320511,""
10 "Network activity","ip-src","2.28.163.125",1412320511,""
11 "Network activity","ip-src","2.28.232.183",1412320511,""
12 "Network activity","ip-src","2.33.214.24",1412320511,""
13 "Network activity","ip-src","2.38.41.213",1412320511,""
14 "Network activity","ip-src","2.97.3.104",1412320511,""
15 "Network activity","ip-src","2.99.33.207",1412320511,""
16 "Network activity","ip-src","2.99.243.74",1412320511,""
17 "Network activity","ip-src","2.100.214.95",1412320511,""
18 "Network activity","ip-src","2.123.179.26",1412320511,""
19 "Network activity","ip-src","2.124.136.77",1412320511,""
20 "Network activity","ip-src","2.126.137.46",1412320511,""
21 "Network activity","ip-src","2.136.62.240",1412320511,""
22 "Network activity","ip-src","2.224.247.16",1412320511,""
23 "Network activity","ip-src","2.229.136.20",1412320511,""
24 "Network activity","ip-src","2.235.77.154",1412320511,""
25 "Network activity","ip-src","2.238.34.3",1412320511,""
26 "Network activity","ip-src","3.4.5.6",1412320512,""
27 "Network activity","ip-src","5.9.121.78",1412320512,""

```

- c. Note: the empty quotes at the row ends are there because there is no "object_name" data for those objects. If you'd like to clean that up, you can use this regex:



```

26 "Network activity","ip-src","3.4.5.6",1412320512
27 "Network activity","ip-src","5.9.121.78",1412320512
28 "Network activity","ip-src","5.13.62.70",1412320512
29 "Network activity","ip-src","5.15.140.251",1412320512,""

```

Find: ,""\n
Replace: \n

lines, 4 characters selected Tab Size: 4 Plain Text

4. There'll be a lot of lines from Yara / Sigma / ect. that still need addressing, too. You can find / remove those with a regular expression similar to this:
- `^([\\|*\\$\\(\\)\\s\\{\\}\\}â€¢ð\\-\\.\\.\\<\\#\\>\\^\\?\\[\\]\\@]|meta:|strings:).*`
 - If you want to get aggressive: `^[^\\"]`