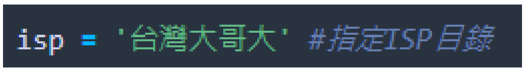
針對 ssdeep_paylodClean_py.py & combine_analysis_py1/2.py

執行參數說明

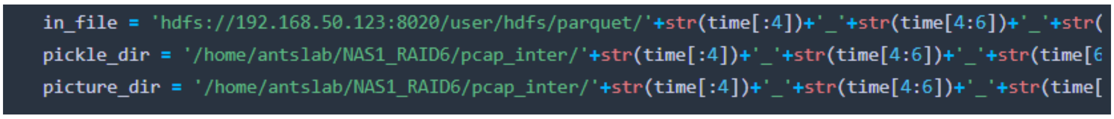
ssdeep_paylodClean

1. 需於程式內指定欲執行的 isp 名稱(與放置目錄一致)

   甲、

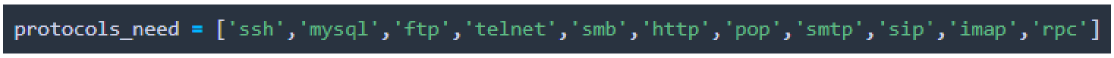
```
isp = '台灣大哥大' #指定ISP目錄
```

2. 需指定 session parquet 的路徑，及欲儲存 pickle 及 picture 的目錄位置

   甲、

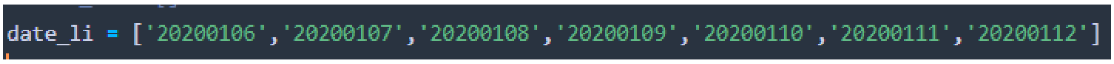
```
in_file = 'hdfs://192.168.50.123:8020/user/hdfs/parquet/'+str(time[:4])+'_'+str(time[4:6])+'_'+str(
pickle_dir = '/home/antslab/NAS1_RAID6/pcap_inter/'+str(time[:4])+'_'+str(time[4:6])+'_'+str(time[6
picture_dir = '/home/antslab/NAS1_RAID6/pcap_inter/'+str(time[:4])+'_'+str(time[4:6])+'_'+str(time[
```

3. 需指定欲選擇執行的 protocols 名稱

   甲、

```
protocols_need = ['ssh','mysql','ftp','telnet','smb','http','pop','smtp','sip','imap','rpc']
```

4. 需指定欲執行的日期，並以下述日期格式

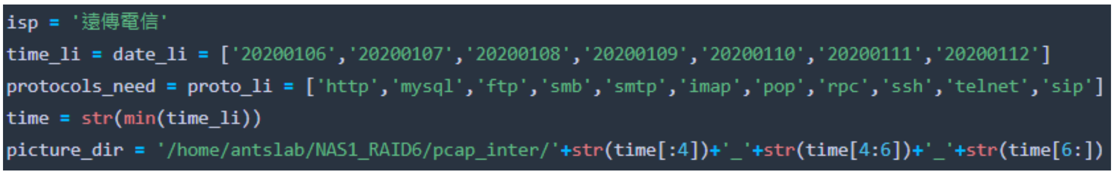
   甲、

```
date_li = ['20200106','20200107','20200108','20200109','20200110','20200111','20200112']
```

5. 得到之 ssdeep clustering 結果(依照 fuzzy hash 相似度比較法分群出手法)可供 combine_analysis_py1 使用

combine_analysis_py1

1. 須於程式內指定 isp 名稱、執行的日期、protocols 名稱如下

   甲、

```
isp = '遠傳電信'
time_li = date_li = ['20200106','20200107','20200108','20200109','20200110','20200111','20200112']
protocols_need = proto_li = ['http','mysql','ftp','smb','smtp','imap','pop','rpc','ssh','telnet','sip']
time = str(min(time_li))
picture_dir = '/home/antslab/NAS1_RAID6/pcap_inter/'+str(time[:4])+'_'+str(time[4:6])+'_'+str(time[6:])
```

2. 可輸出:

   甲、用以繪製 cluster pattern 的圖之 pickle 檔(跨天同 protocol 大小圓圈圖)

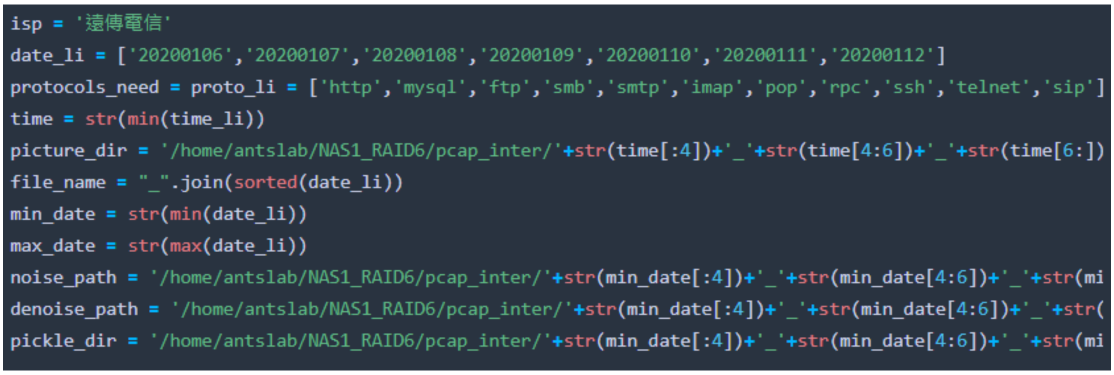
   乙、cluster 之 key session 的 time list，交由專家人工檢視 payload 內容以得知各 cluster 在做的事情

   丙、noise cluster pickle，交由專家人工檢視 http protocol 中不重要的雜訊 cluster 以濾除

combine_analysis_py2

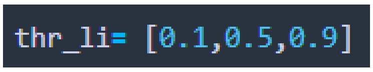
1. 需於程式中指定 isp 名稱、日期、protocols、picture 目錄路徑、經專家人工檢視後的 noise cluster dict 路徑、無 noise cluster dict 路徑、pickle 目錄路

徑，範例如以下

甲、

```
isp = '遠傳電信'
date_li = ['20200106','20200107','20200108','20200109','20200110','20200111','20200112']
protocols_need = proto_li = ['http','mysql','ftp','smb','smtp','imap','pop','rpc','ssh','telnet','sip']
time = str(min(time_li))
picture_dir = '/home/antslab/NAS1_RAID6/pcap_inter/'+str(time[:4])+'_'+str(time[4:6])+'_'+str(time[6:])
file_name = "_".join(sorted(date_li))
min_date = str(min(date_li))
max_date = str(max(date_li))
noise_path = '/home/antslab/NAS1_RAID6/pcap_inter/'+str(min_date[:4])+'_'+str(min_date[4:6])+'_'+str(mi
denoise_path = '/home/antslab/NAS1_RAID6/pcap_inter/'+str(min_date[:4])+'_'+str(min_date[4:6])+'_'+str(
pickle_dir = '/home/antslab/NAS1_RAID6/pcap_inter/'+str(min_date[:4])+'_'+str(min_date[4:6])+'_'+str(mi
```

2. 需於程式中指定 attack pattern 分群的嚴格程度，用以製作敏感度分析使用，默認 threshold 為挑選 0.1、0.5、0.9 (list 最後所指定的 thr 將做為主要分析之 threshold)

甲、

```
thr_li= [0.1,0.5,0.9]
```

3. 可顯示:

甲、期間中各 protocols 的 session 數量

乙、Multi-hot 統計 df pickle (col: 手法名稱；row: 攻擊者 IP)

丙、將 col 和=1,row 和=1 以下的濾掉之上述統計 df pickle

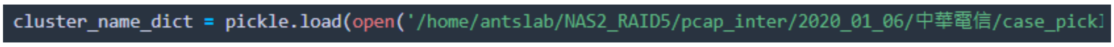
丁、不同 threshold 所對應得到之 attack pattern 數量及 Loner IP 數量

戊、未來可用以繪製世界地圖視覺化的 Attack Pattern pickle

己、統計指定期間中此 ISP 共有多少 sessions、且每個 IP 又執行了多少個 session (輸出 dictionary 有 m 個 session(key)的 IP 有幾個(value))

庚、Loner IP 所涵蓋的 clusters、loner ip 所對應的 cluster name df pickle

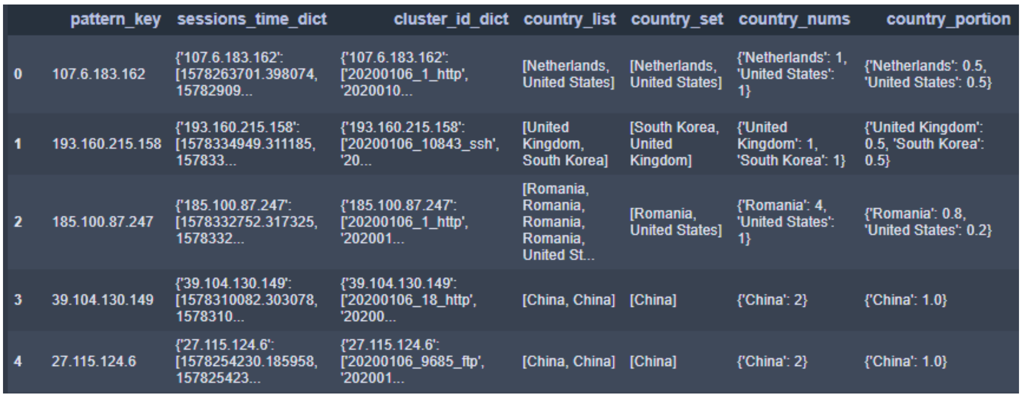
4. 將經專家人工審視各 cluster 之 intention dictionary 讀入，需指定該 dict 路徑

甲、

```
cluster_name_dict = pickle.load(open('/home/antslab/NAS2_RAID5/pcap_inter/2020_01_06/中華電信/case_pickl
```

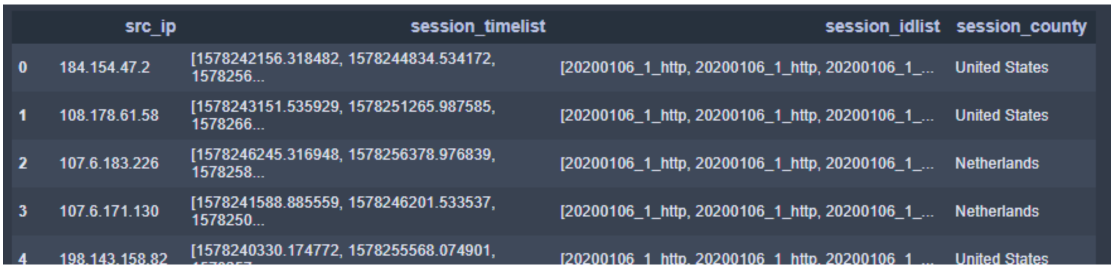
5. 可再進一步輸出:

甲、IP 群統計做圖用 df (包含 mitre、cluster、地域等統計資訊)pickle

i. 

	pattern_key	sessions_time_dict	cluster_id_dict	country_list	country_set	country_nums	country_portion
0	107.6.183.162	{'107.6.183.162': [1578263701.398074, 15782909...	{'107.6.183.162': ['20200106_1_http', '2020010...	[Netherlands, United States]	[Netherlands, United States]	{'Netherlands': 1, 'United States': 1}	{'Netherlands': 0.5, 'United States': 0.5}
1	193.160.215.158	{'193.160.215.158': [1578334949.311185, 157833...	{'193.160.215.158': ['20200106_10843_ssh', '20...	[United Kingdom, South Korea]	[South Korea, United Kingdom]	{'United Kingdom': 1, 'South Korea': 1}	{'United Kingdom': 0.5, 'South Korea': 0.5}
2	185.100.87.247	{'185.100.87.247': [1578332752.317325, 1578332...	{'185.100.87.247': ['20200106_1_http', '202001...	[Romania, Romania, Romania, Romania, United St...	[Romania, United States]	{'Romania': 4, 'United States': 1}	{'Romania': 0.8, 'United States': 0.2}
3	39.104.130.149	{'39.104.130.149': [1578310082.303078, 1578310...	{'39.104.130.149': ['20200106_18_http', '20200...	[China, China]	[China]	{'China': 2}	{'China': 1.0}
4	27.115.124.6	{'27.115.124.6': [1578254230.185958, 157825423...	{'27.115.124.6': ['20200106_9685_ftp', '202001...	[China, China]	[China]	{'China': 2}	{'China': 1.0}

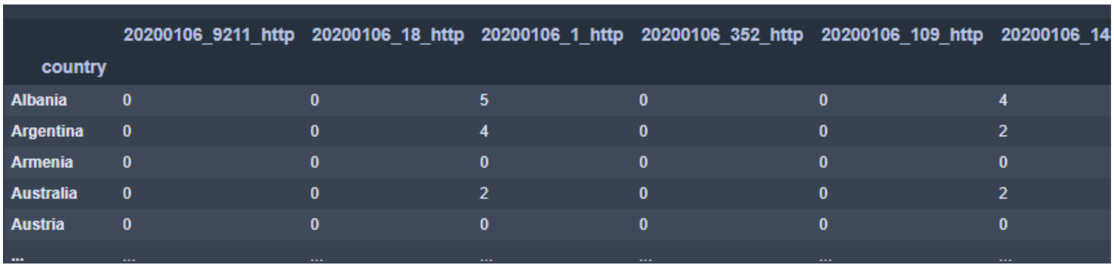
乙、loner df 資訊 pickle(可畫圖包含 time list、IP、國家、cluster 手法名稱)

i. 

	src_ip	session_timelist	session_idlist	session_county
0	184.154.47.2	[1578242156.318482, 1578244834.534172, 1578256...	[20200106_1_http, 20200106_1_http, 20200106_1_...	United States
1	108.178.61.58	[1578243151.535929, 1578251265.987585, 1578266...	[20200106_1_http, 20200106_1_http, 20200106_1_...	United States
2	107.6.183.226	[1578246245.316948, 1578256378.976839, 1578258...	[20200106_1_http, 20200106_1_http, 20200106_1_...	Netherlands
3	107.6.171.130	[1578241588.885559, 1578246201.533537, 1578250...	[20200106_1_http, 20200106_1_http, 20200106_1_...	Netherlands
4	198.143.158.82	[1578240330.174772, 1578255568.074901,	[20200106_1_http, 20200106_1_http, 20200106_1_...	United States

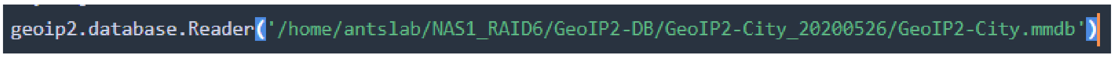
丙、手法(clusters)出現次數頻率統計 pickle(該手法於該國家有幾個 IP 執行過)

丁、不同國家會做那些守法的 cluster

i. 

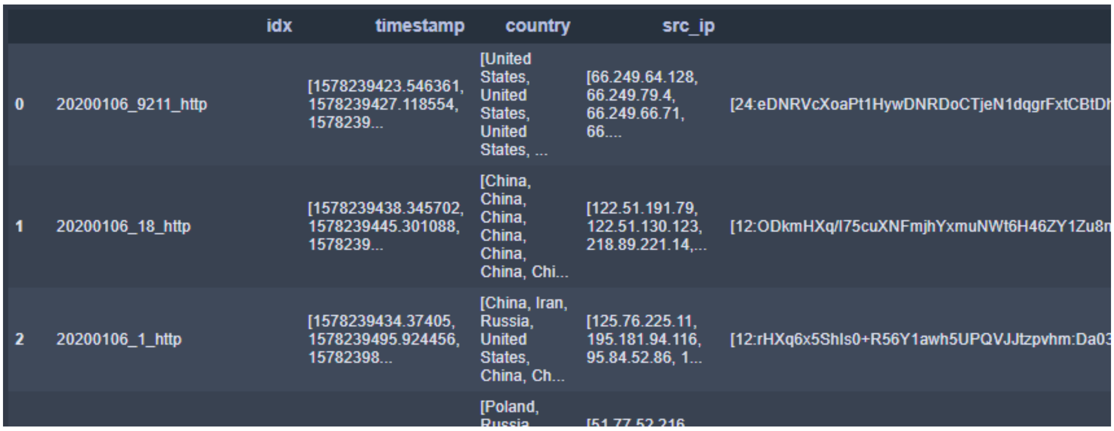
	20200106_9211_http	20200106_18_http	20200106_1_http	20200106_352_http	20200106_109_http	20200106_14
country						
Albania	0	0	5	0	0	4
Argentina	0	0	4	0	0	2
Armenia	0	0	0	0	0	0
Australia	0	0	2	0	0	2
Austria	0	0	0	0	0	0
...	...	...	...	...	...	...

6. 需於程式中指定 geoIP 的 DB 路徑

甲、

```
geoip2.database.Reader('/home/antslab/NAS1_RAID6/GeoIP2-DB/GeoIP2-City_20200526/GeoIP2-City.mmdb')
```

7. 便可再更進一步輸出包含有手法經緯度的繪圖用資訊 df pickle (未來用以繪製全球地圖視覺化)

甲、

	idx	timestamp	country	src_ip	
0	20200106_9211_http	[1578239423.546361, 1578239427.118554, 1578239...	[United States, United States, United States, ...	[66.249.64.128, 66.249.79.4, 66.249.66.71, 66....	[24:eDNRVcXoaPt1HywDNRDoCTjeN1dqgrFxtCBtDł
1	20200106_18_http	[1578239438.345702, 1578239445.301088, 1578239...	[China, China, China, China, China, China, Chi...	[122.51.191.79, 122.51.130.123, 218.89.221.14,...	[12:ODkmHXq/I75cuXNFmjhYxmuNWt6H46ZY1Zu8n
2	20200106_1_http	[1578239434.37405, 1578239495.924456, 15782398...	[China, Iran, Russia, United States, China, Ch...	[125.76.225.11, 195.181.94.116, 95.84.52.86, 1...	[12:rHXq6x5Shls0+R56Y1awh5UPQVJJtzpvhm:Da03
			[Poland, Russia	[51.77.52.216	