視覺化分析資安設備日誌與異常偵測

國立臺灣大學 計資中心網路組

童鵬哲

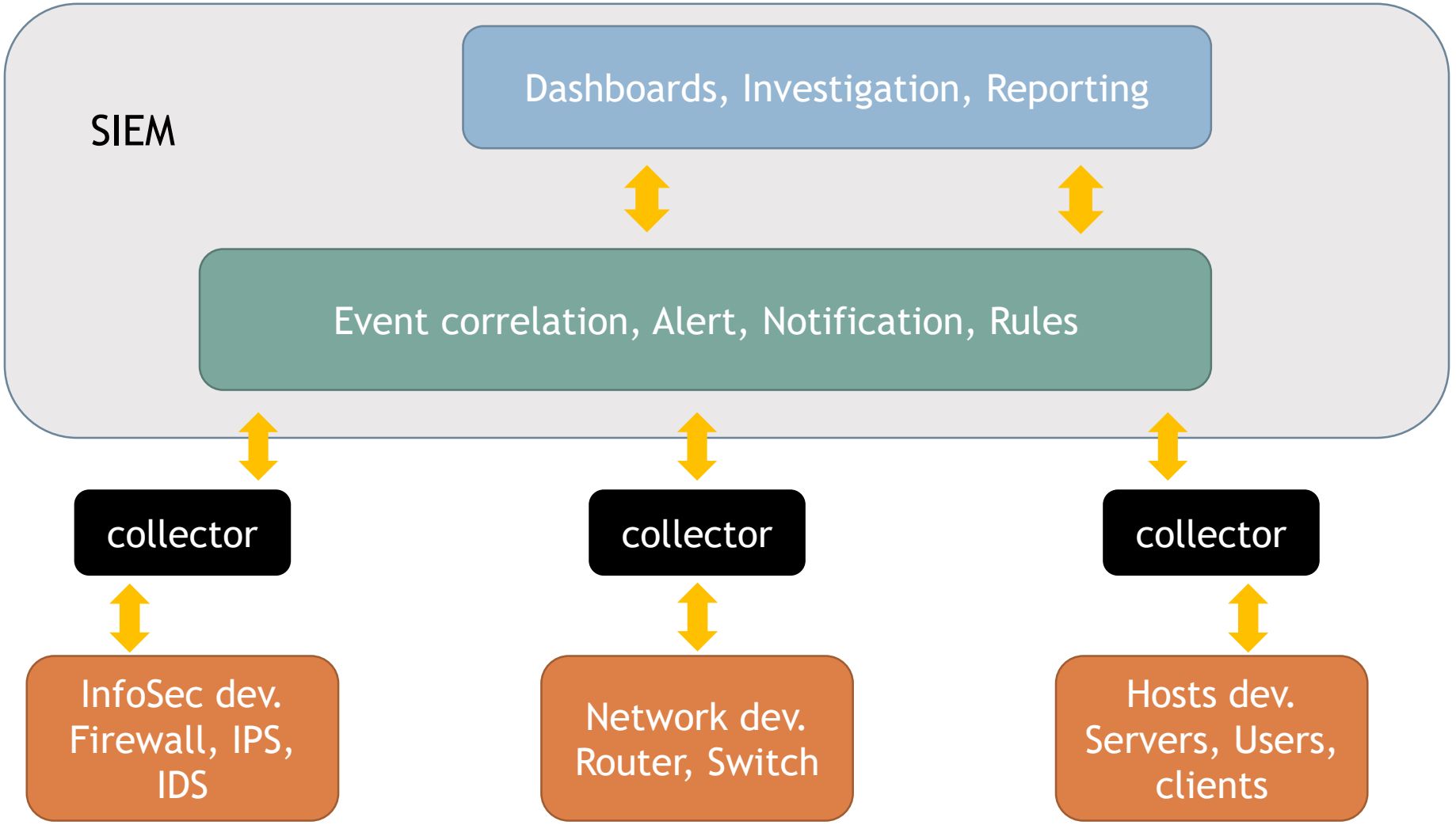
Outline

- 資安事件監看
- Elastic Stack/IPS syslog to ELK
- 視覺化分析/關聯分析/事件分析
- 異常偵測
- 結論
- Colab

3

資安事件監看

SIEM (Security Information Event Management)



CEF (Common Event Format)

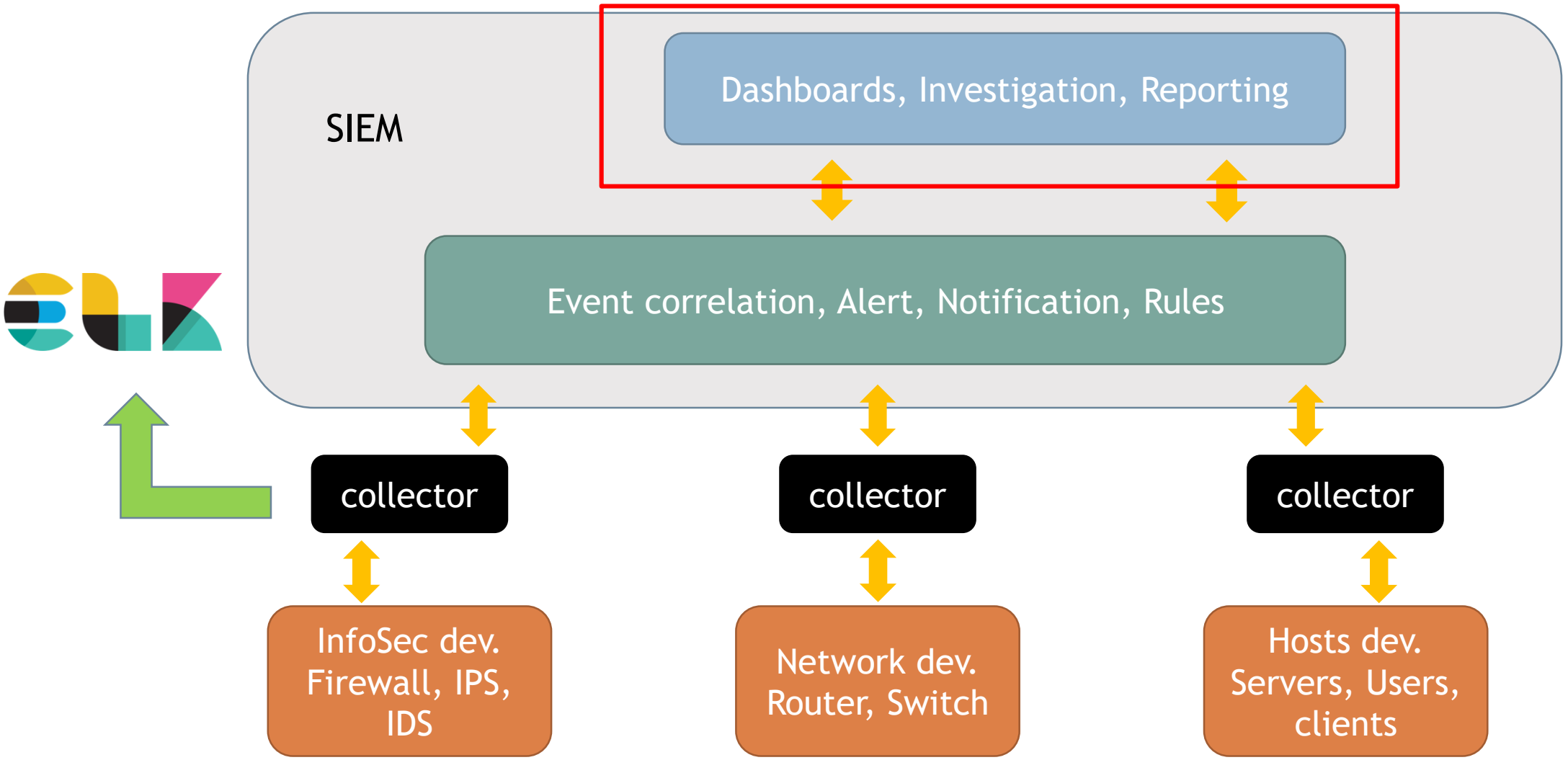
統一日誌格式

- 眾多設備各種不同的日誌格式。
- 方便日誌的蒐集，必須整合對應為一致的欄位。

CEF

CEF標準對於設備日誌常用的欄位，如：時間、IP位址、IP Port、訊息名稱、威脅等級...等，皆有詳細的定義。

SIEM (Security Information Event Management)



7

1.Elastic Stack
2.syslog to ELK

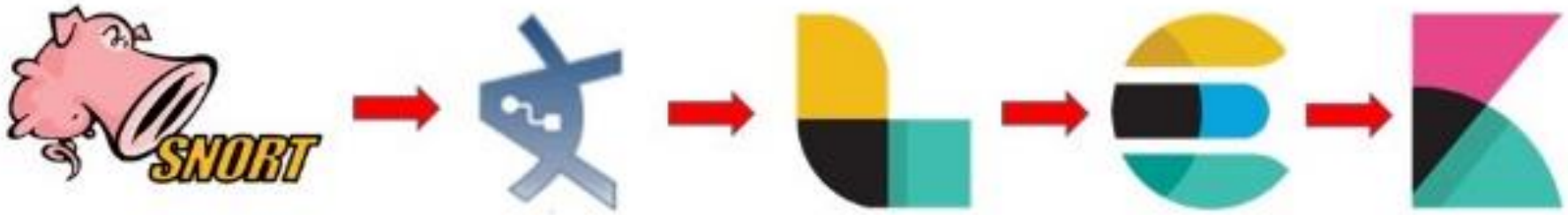
Elastic stack

- **Elasticsearch**
  - 資料儲存、搜尋使用
  - RESTful API
- **Logstash**
  - inputs → filters → outputs
  - Parse
- **Kibana**
  - 視覺化
  - 圖表

Stack

IPS syslog to ELK

- IPS 轉拋 syslog 至 SIEM 平台前，會先送至中介設備 (collector) 進行事件之正規化。
- 透過 collector 將事件再轉拋至 Logstash。
- 透過 Logstash 之 CEF codec 套件與 ArcSight 套件進行數據解析。
- 最後，送進 Elasticsearch 資料庫儲存。



CEF plugin

Cef codec plugin

- Plugin version: v6.1.1
- Released on: 2020-04-09
- Changelog

For other versions, see the Versioned plugin docs.

If your input puts a delimiter between each CEF event, you'll want to set this to be that delimiter.

For example, with the TCP input, you probably want to put this:

```
input {
  tcp {
    codec => cef { delimiter => "\r\n" }
    # ...
  }
}
```

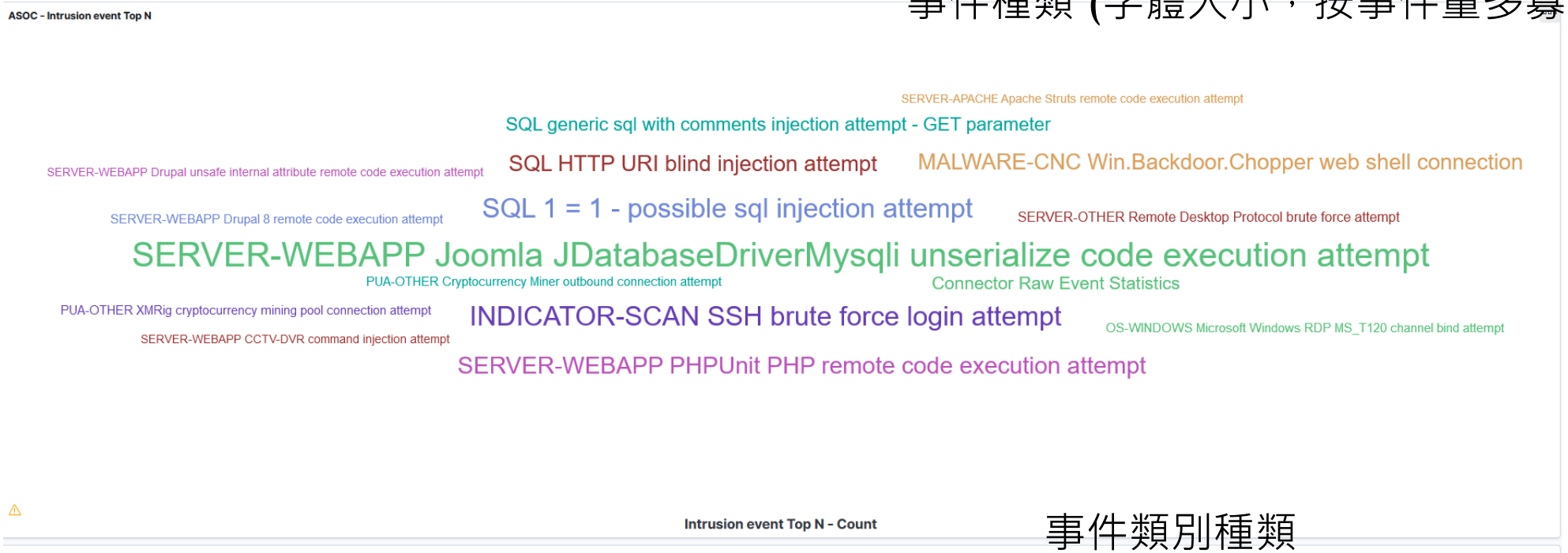
Ref : https://www.elastic.co/guide/en/logstash/current/plugins-codecs-cef.html

11

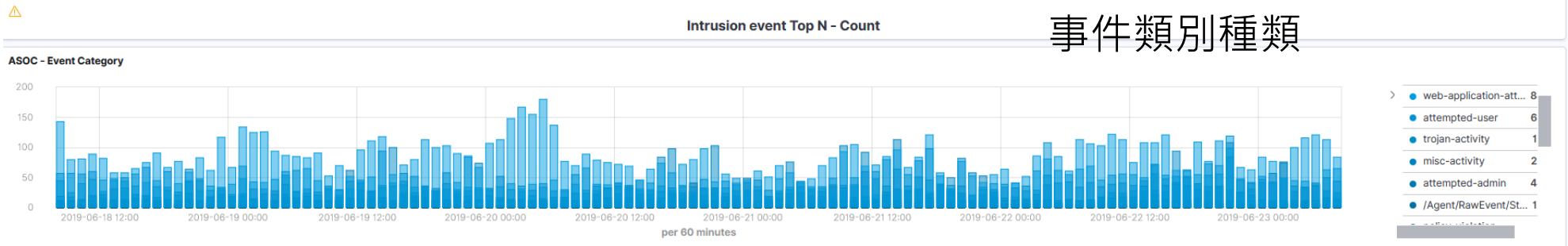
視覚化分析

Kibana儀表板

事件種類 (字體大小，按事件量多寡調整)



事件類別種類

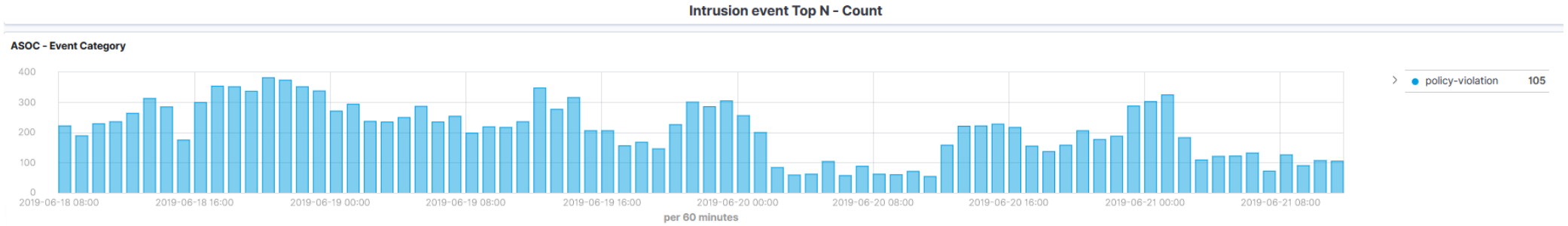


Kibana儀表板

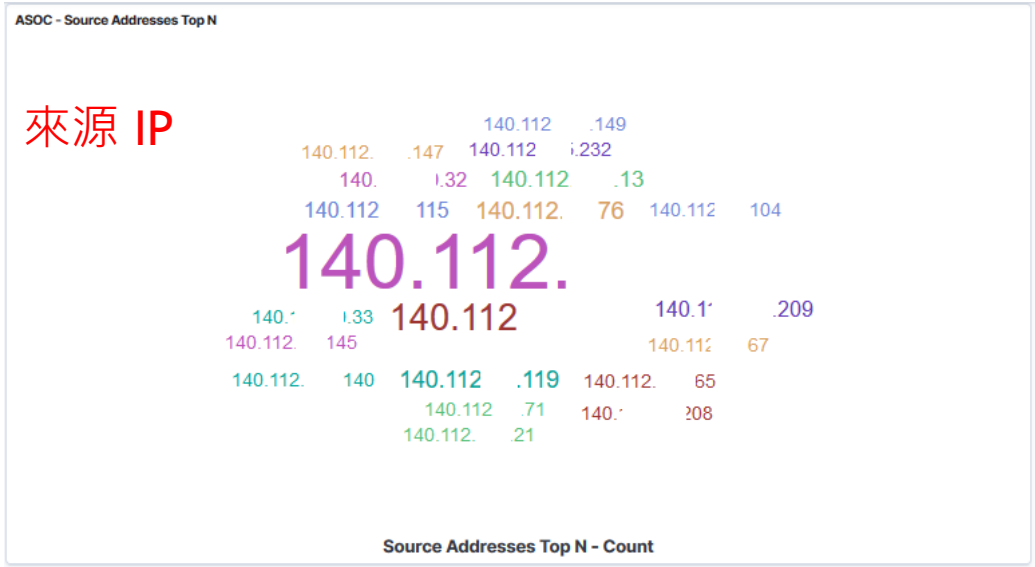
ASOC - Intrusion event Top N

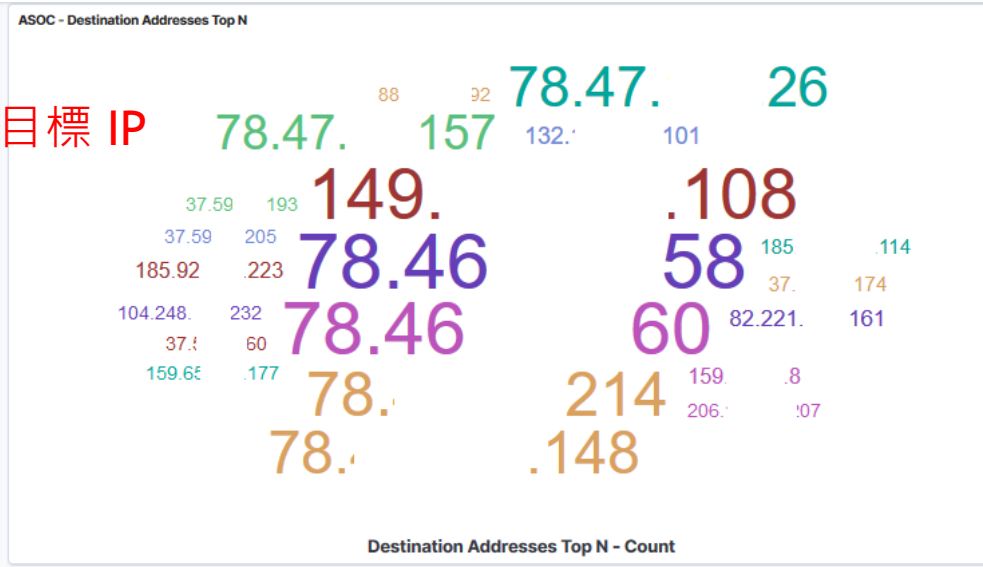
點擊愈調查之事件名稱

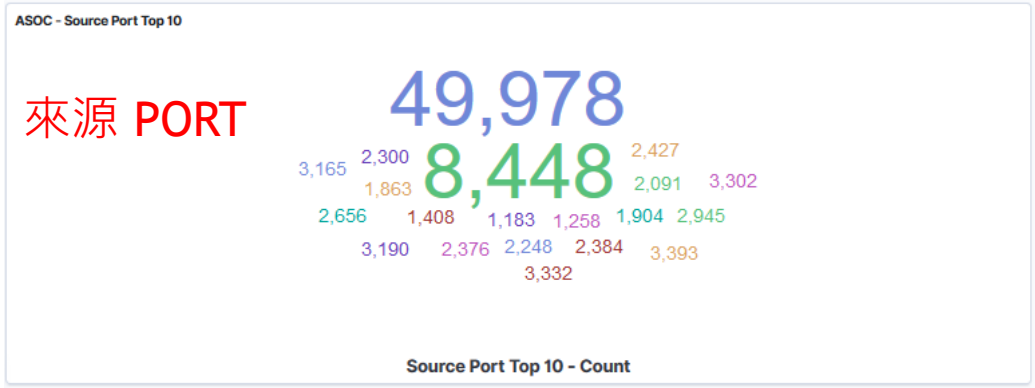
PUA-OTHER Cryptocurrency Miner outbound connection attempt



Kibana儀表板

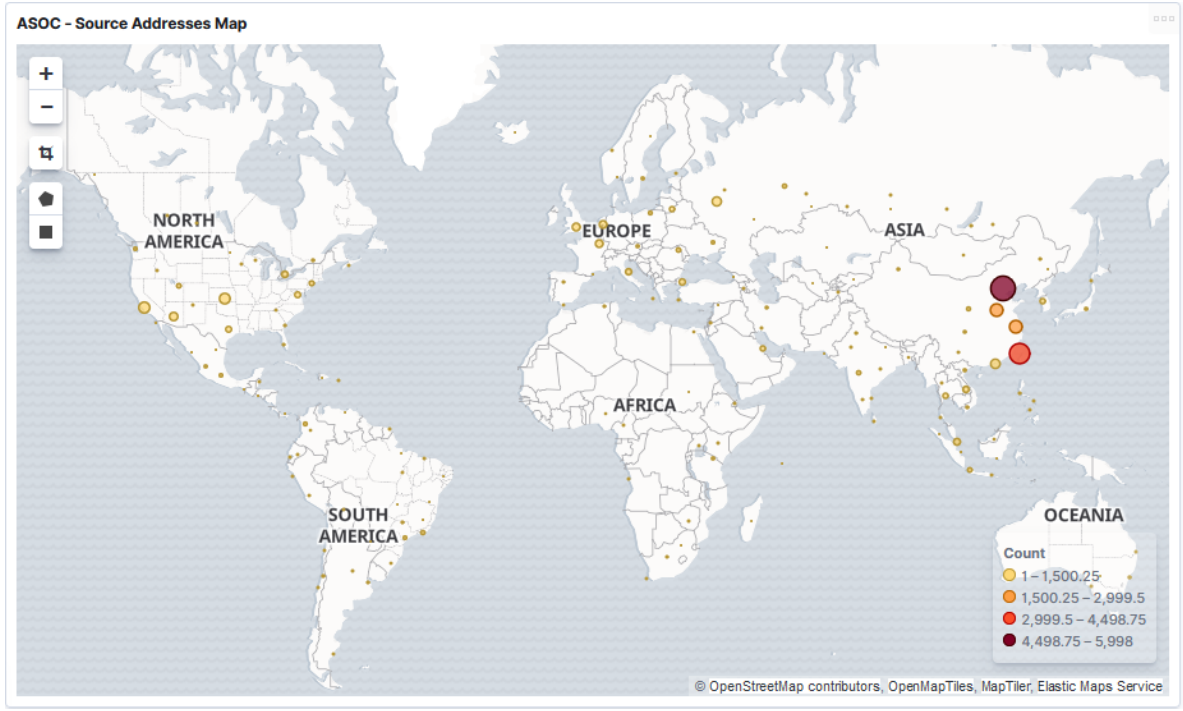

ASOC - Source Addresses Top N
來源 IP
140.112.
Source Addresses Top N - Count

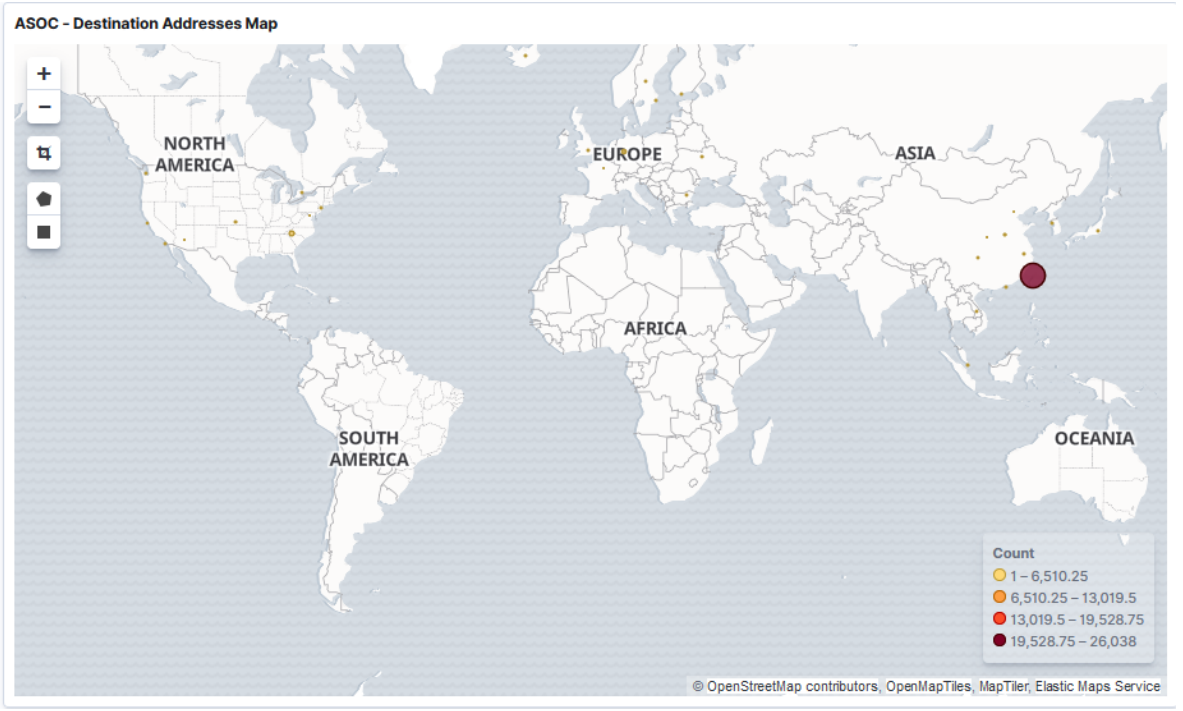

ASOC - Destination Addresses Top N
目標 IP
78.47.
149.
.108
78.46
58
60
214
.148
Destination Addresses Top N - Count


ASOC - Source Port Top 10
來源 PORT
49,978
8,448
Source Port Top 10 - Count

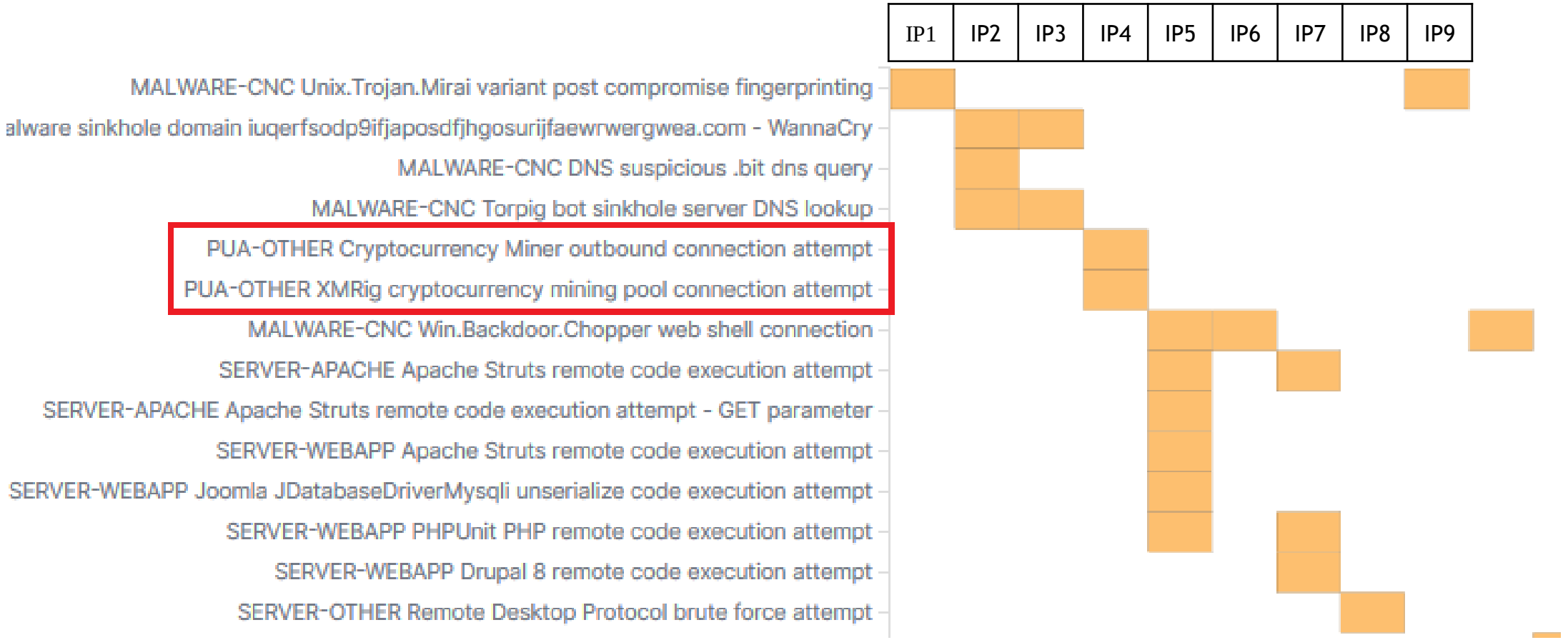

ASOC - Destination Port Top 10
目標 PORT
443
456
454
458
450
459
457
455
452
451
453
Destination Port Top 10 - Count

Kibana儀表板


ASOC - Source Addresses Map
NORTH AMERICA
EUROPE
ASIA
AFRICA
SOUTH AMERICA
OCEANIA
Count
1 – 1,500.25
1,500.25 – 2,999.5
2,999.5 – 4,498.75
4,498.75 – 5,998
© OpenStreetMap contributors, OpenMapTiles, MapTiler, Elastic Maps Service


ASOC - Destination Addresses Map
NORTH AMERICA
EUROPE
ASIA
AFRICA
SOUTH AMERICA
OCEANIA
Count
1 – 6,510.25
6,510.25 – 13,019.5
13,019.5 – 19,528.75
19,528.75 – 26,038
© OpenStreetMap contributors, OpenMapTiles, MapTiler, Elastic Maps Service

Kibana儀表板


IP1
IP2
IP3
IP4
IP5
IP6
IP7
IP8
IP9
MALWARE-CNC Unix.Trojan.Mirai variant post compromise fingerprinting
alware sinkhole domain iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com - WannaCry
MALWARE-CNC DNS suspicious .bit dns query
MALWARE-CNC Torpig bot sinkhole server DNS lookup
PUA-OTHER Cryptocurrency Miner outbound connection attempt
PUA-OTHER XMRig cryptocurrency mining pool connection attempt
MALWARE-CNC Win.Backdoor.Chopper web shell connection
SERVER-APACHE Apache Struts remote code execution attempt
SERVER-APACHE Apache Struts remote code execution attempt - GET parameter
SERVER-WEBAPP Apache Struts remote code execution attempt
SERVER-WEBAPP Joomla JDatabaseDriverMysqli unserialize code execution attempt
SERVER-WEBAPP PHPUnit PHP remote code execution attempt
SERVER-WEBAPP Drupal 8 remote code execution attempt
SERVER-OTHER Remote Desktop Protocol brute force attempt

17

關聯分析

利用API結合外部程式分析

- 利用 ELK方便的API搜索功能。
- 快速從 Elasticsearch資料庫中抓取出欲分析的資料。
- 透過外部程式或自行撰寫之程式，進行更多樣化的分析。
- 使用 Python程式語言撰寫了自動化腳本。

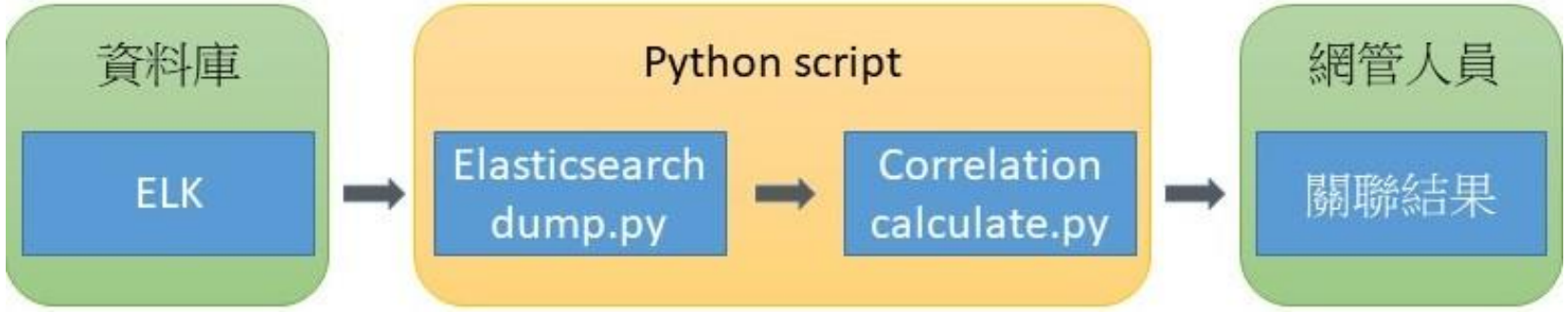
利用API結合外部程式分析

- 腳本程式1 (Elasticsearch-dump.py)

利用 Elasticseach API，從每筆CEF資料中過濾出需要的欄位：來源IP、目標IP以及事件名稱。

- 腳本程式2 (Correlation-calculate.py)

透過程式1所得到結果，自動整理成事件名稱與目標IP之樞紐分析表後，再進行事件與事件之間相關係數(Pearson correlation coefficient)計算，並得出相關係數排序表及相關圖表。



Dataset

Dataset

- IPS 平時儲存於 Elasticsearch 之 CEF 日誌。
- 連續五日約有7萬筆。
- 來源IP：4212 筆。
- 目標IP：13180 筆。
- 規則種類：112 種。
- 整理Dataset，以目標IP與規則種類對應之樞紐分析表格呈現。

name	0	1	2	3	4	5	6	7	8	9
destinationAddress										
0	22.0	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN
1	129.0	56.0	NaN	2.0	34.0	NaN	NaN	7.0	239.0	NaN
2	NaN	NaN	379.0	4.0	NaN	NaN	NaN	1.0	NaN	NaN
3	NaN	NaN	NaN	3.0	NaN	NaN	NaN	1.0	NaN	NaN
4	1197.0	164.0	NaN	5.0	144.0	NaN	4.0	30.0	5.0	NaN
5	NaN	NaN	149.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN
6	NaN	NaN	353.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	NaN	NaN	1240.0	2.0	NaN	NaN	NaN	1.0	NaN	NaN
8	NaN	2.0	NaN	3.0	3.0	NaN	NaN	3.0	NaN	NaN
9	NaN	NaN	NaN	NaN	NaN	152.0	NaN	NaN	NaN	NaN
10	1.0	NaN	NaN	2.0	NaN	NaN	123.0	1.0	1.0	NaN
11	NaN	NaN	66.0	5.0	NaN	NaN	NaN	1.0	NaN	NaN

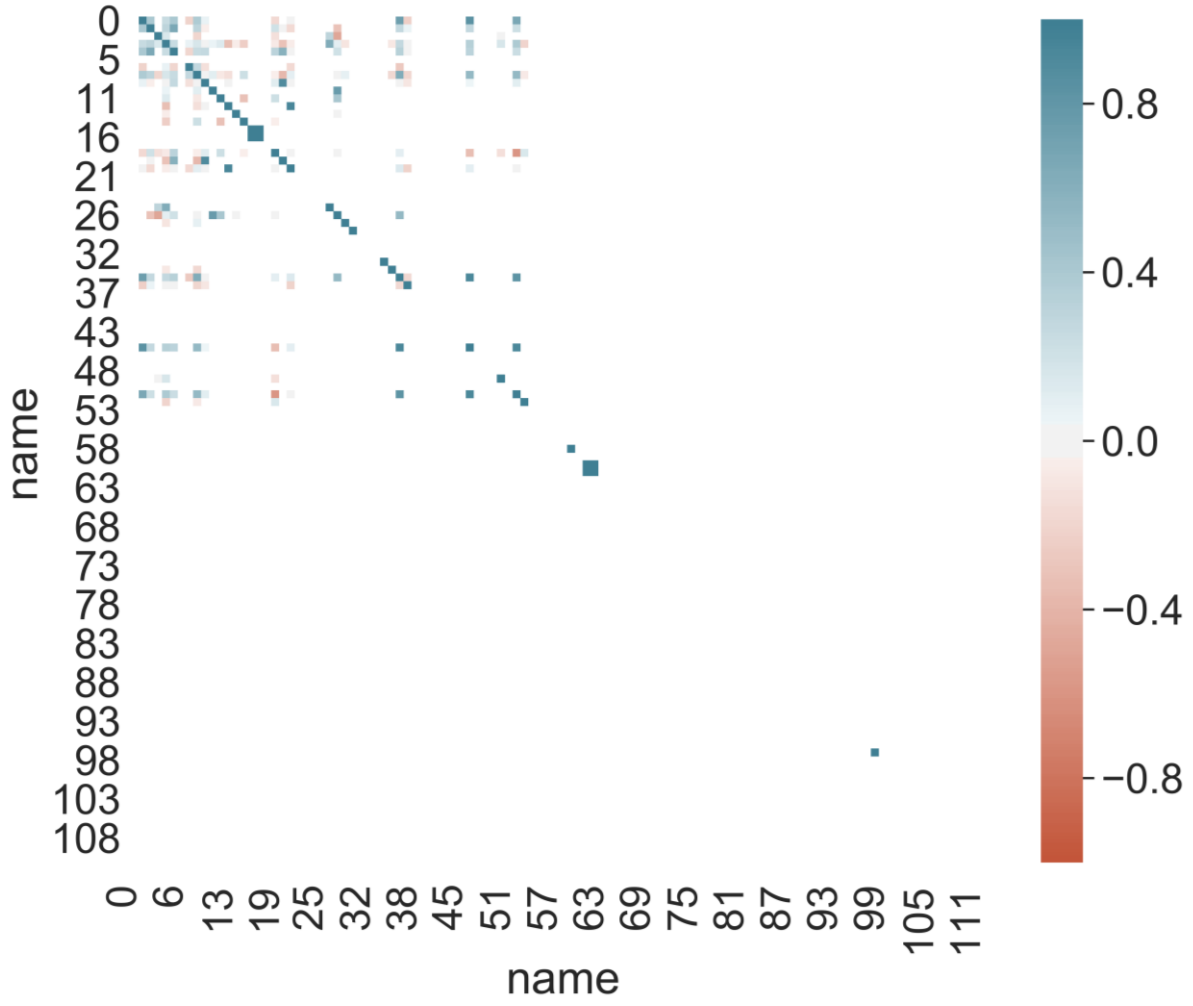
Evaluate

- Pearson correlation coefficient

$$r_{xy} \overset{\text{def}}{=} \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- x, y : two columns (rules)
- x_i, y_i : the event numbers at detinationAddress i , $\forall\, i = 1, \cdots, n$
- $\overline{x}, \overline{y}$: sample means of x and y
- $-1 \le r_{xy} \le 1$

Result


name
0
5
11
16
21
26
32
37
43
48
53
58
63
68
73
78
83
88
93
98
103
108
0
6
13
19
25
32
38
45
51
57
63
69
75
81
87
93
99
105
111
name
0.8
0.4
0.0
−0.4
−0.8

Result

Name1	Name2	R
SERVER-WEBAPP Netgear DGN1000 series routers authentication bypass attempt	**SERVER-WEBAPP Netgear DGN1000 series routers arbitrary command execution attempt**	1
PUA-OTHER Cryptocurrency Miner outbound connection attempt	**PUA-OTHER XMRig cryptocurrency mining pool connection attempt**	0.99
SERVER-WEBAPP Drupal 8 remote code execution attempt	SERVER-WEBAPP Drupal unsafe internal attribute remote code execution attempt	0.93
SERVER-APACHE Apache Struts remote code execution attempt - GET parameter	SERVER-WEBAPP Apache Struts remote code execution attempt	0.92
SERVER-WEBAPP Apache Struts remote code execution attempt	SERVER-APACHE Apache Struts remote code execution attempt	0.92

24 事件分析

事件分析 1

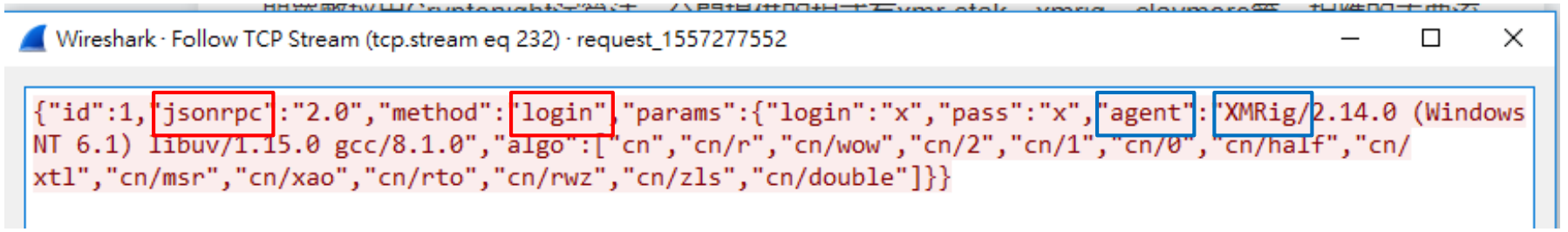
- 挖礦規則

- 規則1 (PUA-OTHER Cryptocurrency Miner outbound connection attempt)

  Patterns ： jsonrpc、login…等。

- 規則2 (PUA-OTHER XMRig cryptocurrency mining pool connection attempt)

  Patterns ： XMRig、agent…等。



事件分析 2.1

Netgear DGN1000 漏洞

- 規則1：(Netgear DGN免驗證提權漏洞規則)
  SERVER-WEBAPP Netgear DGN1000 series routers authentication bypass attempt

- 規則2：(Netgear DGN執行任意指令漏洞規則)
  SERVER-WEBAPP Netgear DGN1000 series routers arbitrary command execution attempt

- Exploit 方式：
  執行setup.cgi程式時，使用特定函數以及特殊字串，便能繞過驗證並進行提權，甚至執行任意命令。

事件分析 2.2

- 規則1 - 免驗證提權的特殊字串 ： currentsetting.htm

- 規則2 - 執行任意指令字串 ：
  利用syscmd函數試圖執行busybox命令

```
GET /setup.cgi?next_file=netgear.cfg&todo=syscmd&cmd=busybox&curpath=/&currentsetting.htm=1 HTTP/1.1
Connection: keep-alive
Accept-Encoding: gzip, deflate
Accept: /
User-Agent: Mozilla/5.0
```

28

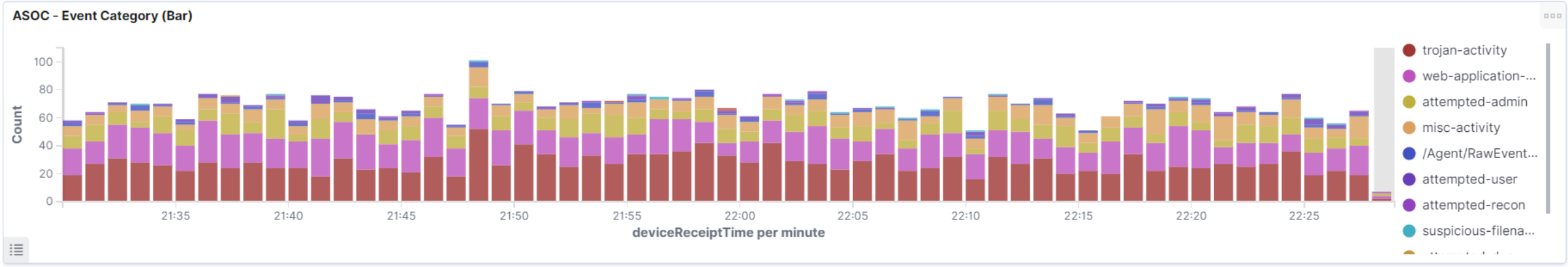
Anomaly detection

異常偵測

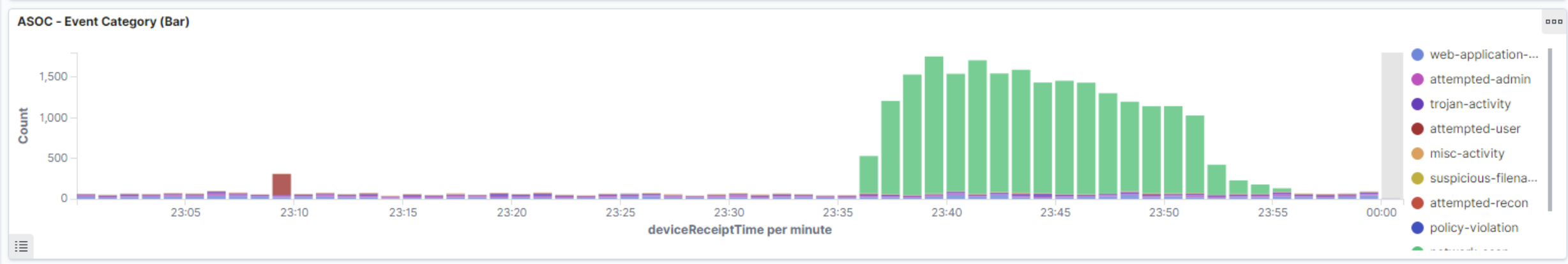
- 正常狀況下，不管是事件量或是事件量比例，應該會呈現一個穩定狀態或穩定比例。
- 假設有比例改變的情形，表示事件生態可能發生改變。

- 方法：
- 利用 RNN(LSTM)搭配 Autoencoder 搭建異常分析架構。
- 目前利用 IPS 的intrusion event 資料進行訓練及預測，嘗試找出異常之狀況(如：事件量驟增、大量掃描、事件量驟降、異常比例.....等狀況)。

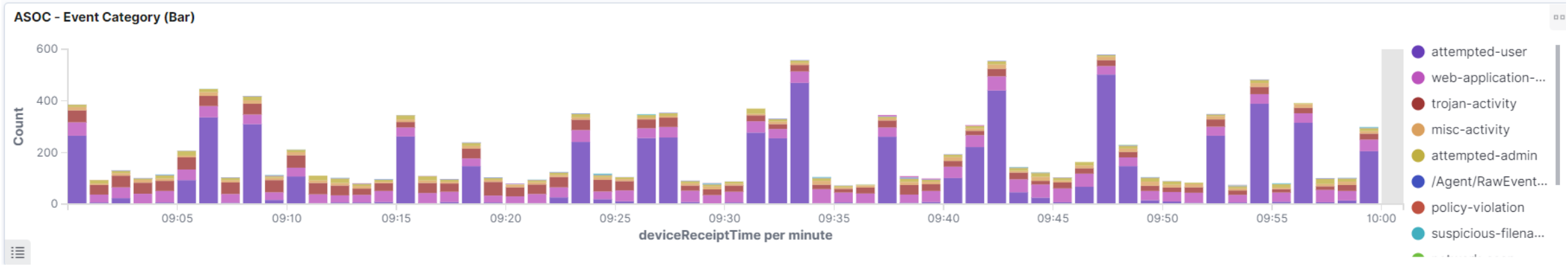
Normal state



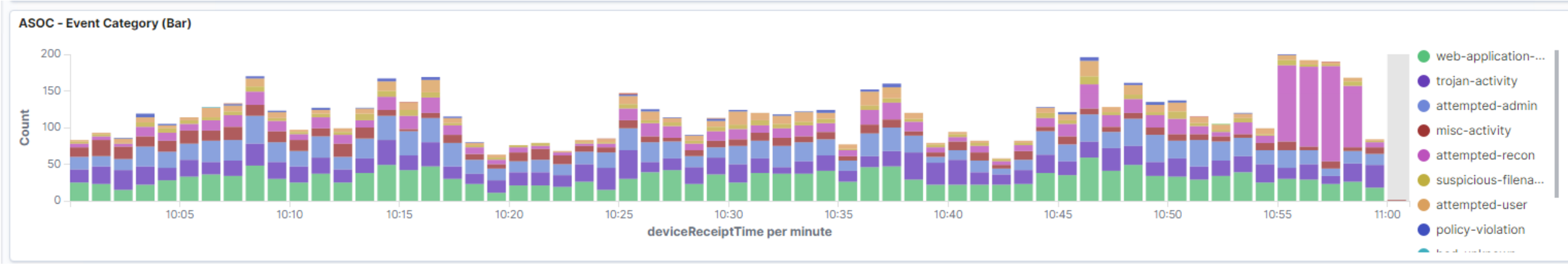
Anomaly state (事件量驟增)



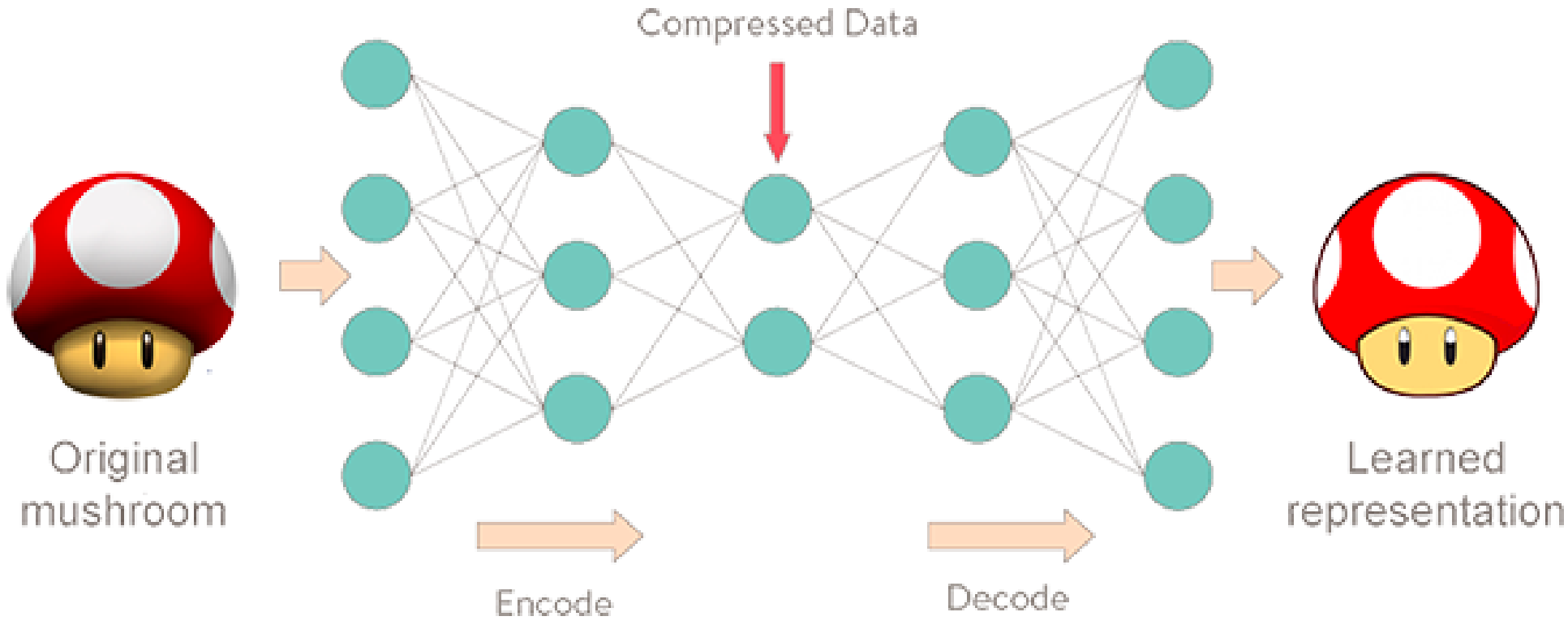
Anomaly state (異常事件突起)



Anomaly state (比例異常)

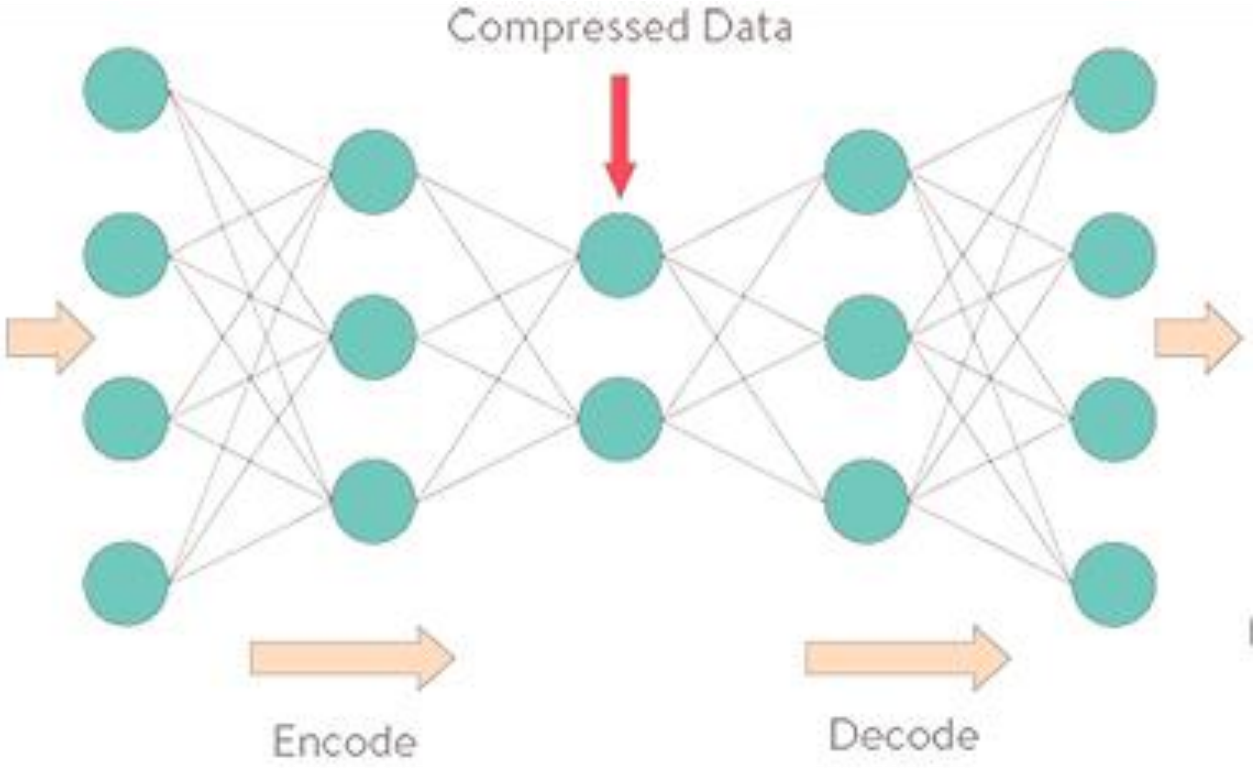


Autoencoder

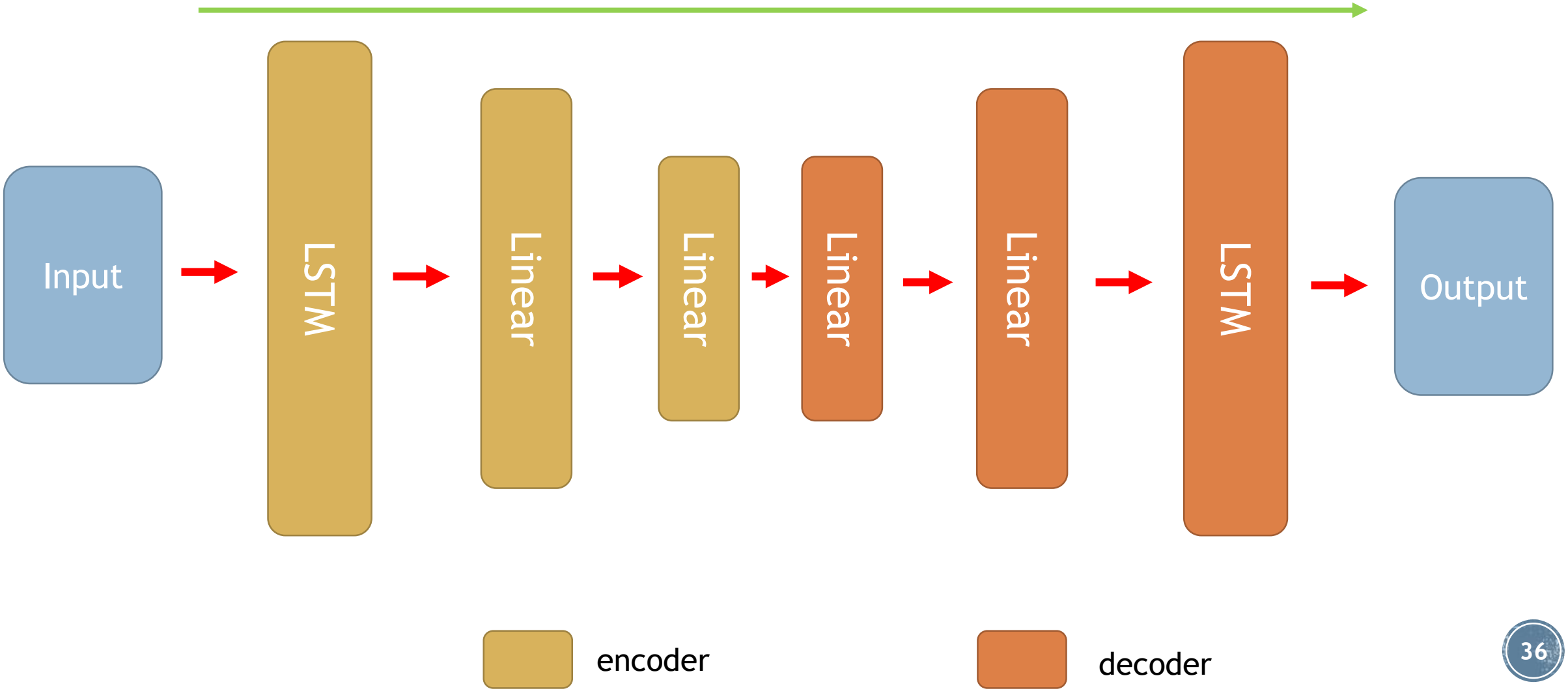


Ref : https://towardsdatascience.com/deep-autoencoders-using-tensorflow-c68f075fd1a3

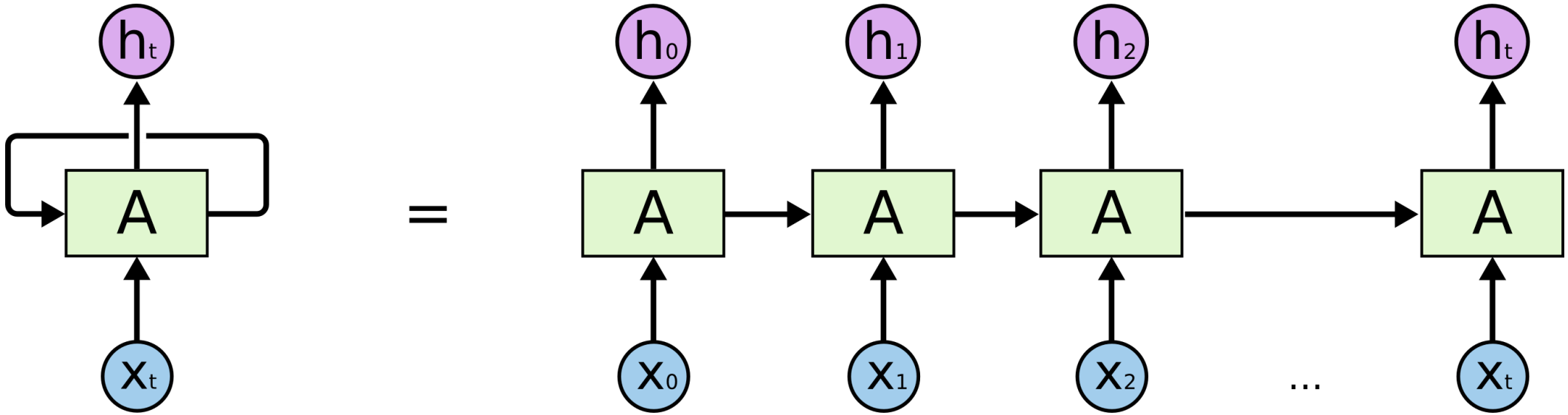
Autoencoder



Network graph _ autoencoder



RNN



RNN works better for time seris prediction.

Dataset

- Data (category種類)

time	not-suspicious	unknown	bad-unknown	attempted-recon	cessful-recon-limit	successful-recon-largescale	attempted-dc
2019/1/1 01:00	0	0.00106564	0.000213129	0.0436914	0.0217391	0	0
2019/1/1 02:00	0	0.000120226	0	0.00745401	0.0034064	0	0
2019/1/1 03:00	0	0.000538648	0.000807972	0.0535955	0.0280097	0	0
2019/1/1 04:00	0	0.000482859	0.000724288	0.049493	0.0251086	0	0
2019/1/1 05:00	0	0.000525486	0	0.0517604	0.0239096	0	0
2019/1/1 06:00	0	0.000401338	0	0.0266221	0.0120401	0	0
2019/1/1 07:00	0	0.000776297	0	0.0258766	0.0109975	0	0

- Features (event category) , total : 38 .

['not-suspicious', 'unknown', 'bad-unknown', 'attempted-recon', 'successful-recon-limited', 'successful-recon-largescale', 'attempted-dos', 'successful-dos', 'attempted-user', 'unsuccessful-user', 'successful-user', 'attempted-admin', 'successful-admin', 'rpc-portmap-decode', 'shellcode-detect', 'string-detect', 'suspicious-filename-detect', 'suspicious-login', 'system-call-detect', 'tcp-connection', 'trojan-activity', 'unusual-client-port-connection', 'network-scan', 'denial-of-service', 'non-standard-protocol', 'protocol-command-decode', 'web-application-activity', 'web-application-attack', 'misc-activity', 'misc-attack', 'icmp-event', 'inappropriate-content', 'policy-violation', 'default-login-attempt', 'sdf', 'malware-cnc', 'client-side-exploit', 'file-format']

Training

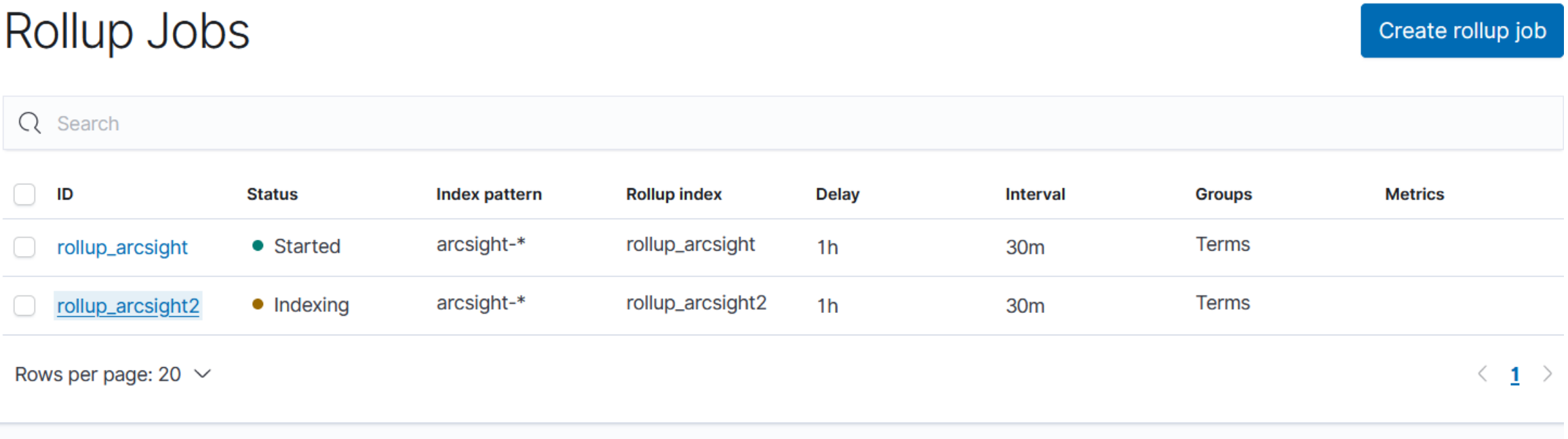
- Epoch : 100
- Train loss average : 0.013…
- Training Data datetime range:

2019-01-01 0:00
to
2020-06-04 16:00

```
Epoch:  99 | train loss: 0.0120
Epoch:  99 | train loss: 0.0130
Epoch:  99 | train loss: 0.0150
Epoch:  99 | train loss: 0.0132
Epoch:  99 | train loss: 0.0115
Epoch:  99 | train loss: 0.0136
Epoch:  99 | train loss: 0.0148
Epoch:  99 | train loss: 0.0115
Epoch:  99 | train loss: 0.0142
Epoch:  99 | train loss: 0.0126
Epoch:  99 | train loss: 0.0129
Epoch:  99 | train loss: 0.0140
Epoch:  99 | train loss: 0.0148
Epoch:  99 | train loss: 0.0132
Epoch:  99 | train loss: 0.0143
Epoch:  99 | train loss: 0.0132
執行時間：360.153073 秒
```

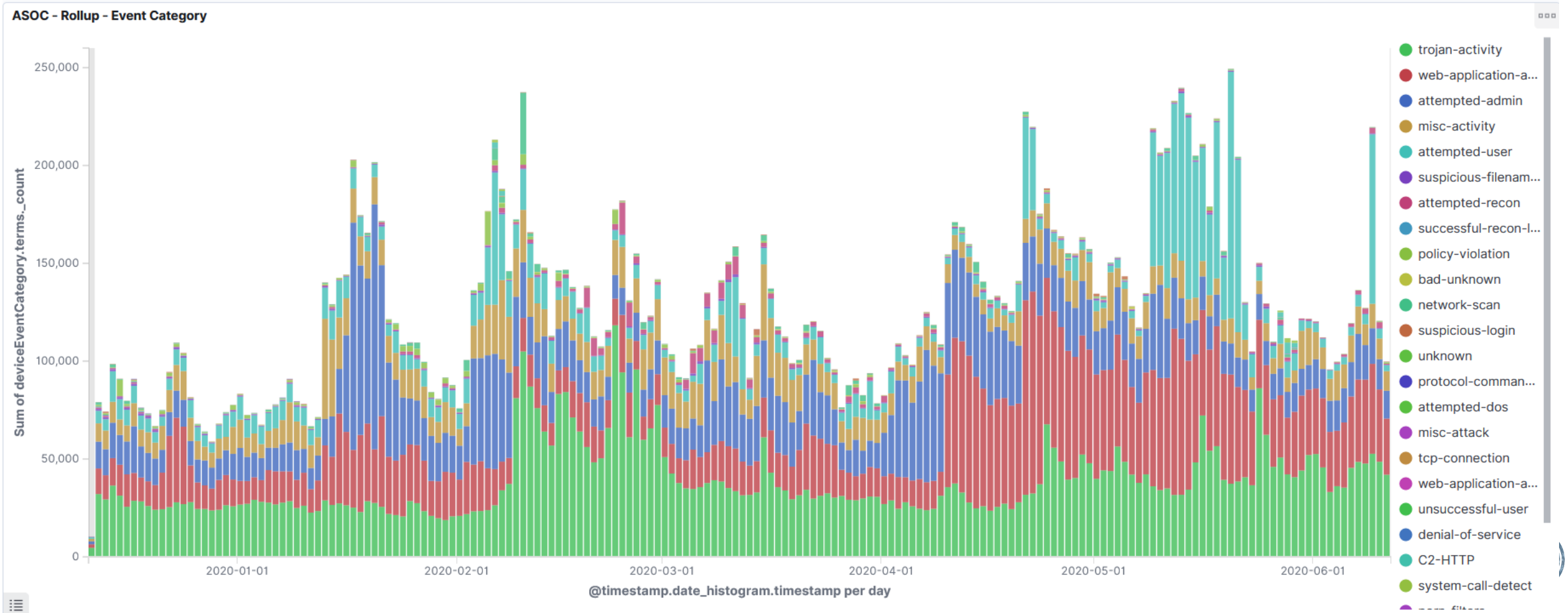
Elasticsearch 調整

- 新增rollup index，可快速aggregate 所需欄位，自成一個新index (table)，供預測使用。



Elasticsearch Rollup

- rollup index，因為僅有特定欄位，儲存節省空間，可快速搜尋長期之資料。

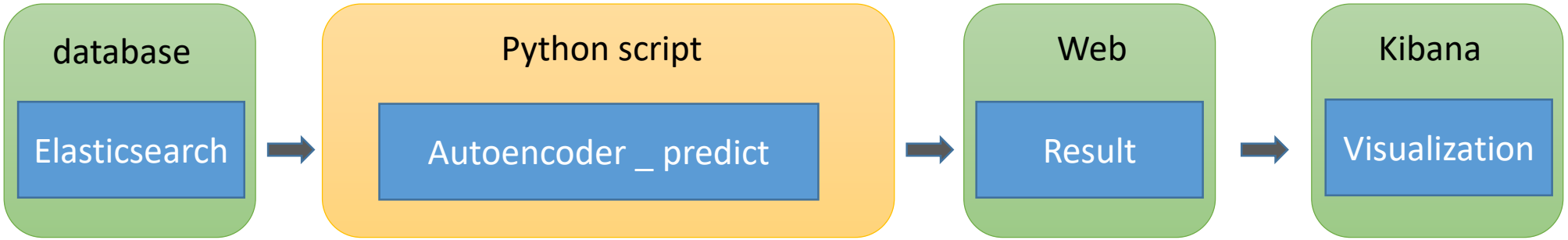


Predict Workflow

Testing Data datetime range:

直接讀取online elasticsearch 資料庫

2019/07/01 - Now



Demo-2

異常偵測 for IPS event

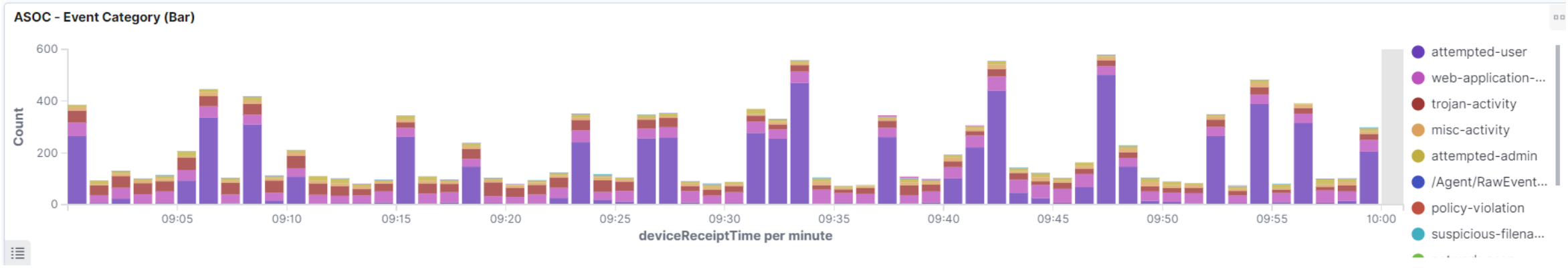
2020-05-20 08:00 ~ 2020-05-26 09:00

異常時段(小時)：10

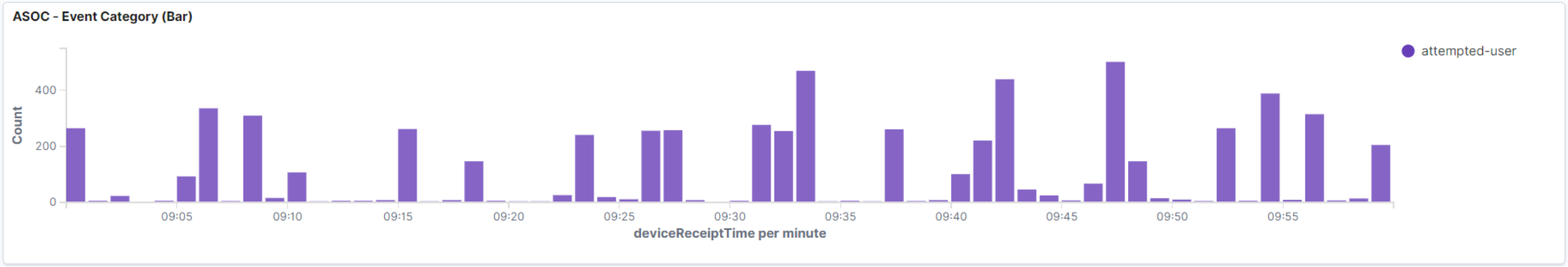
Loss threshold : 0.05

Time	Link	Loss
2020-05-20 09:00:00	Link to kibana	0.0512
2020-05-20 10:00:00	Link to kibana	0.0530
2020-05-20 11:00:00	Link to kibana	0.0556
2020-05-20 12:00:00	Link to kibana	0.0525
2020-05-20 13:00:00	Link to kibana	0.0545
2020-05-20 14:00:00	Link to kibana	0.0533
2020-05-20 15:00:00	Link to kibana	0.0545
2020-05-20 16:00:00	Link to kibana	0.0584
2020-05-20 18:00:00	Link to kibana	0.0564
2020-05-20 19:00:00	Link to kibana	0.0533

Demo-4


ASOC - Event Category (Bar)
600
400
200
0
Count
09:05
09:10
09:15
09:20
09:25
09:30
09:35
09:40
09:45
09:50
09:55
10:00
deviceReceiptTime per minute
attempted-user
web-application-...
trojan-activity
misc-activity
attempted-admin
/Agent/RawEvent...
policy-violation
suspicious-filena...

Demo-5


ASOC - Event Category (Bar)
attempted-user
Count
400
200
0
09:05
09:10
09:15
09:20
09:25
09:30
09:35
09:40
09:45
09:50
09:55
deviceReceiptTime per minute

Demo-6


ASOC - Intrusion event Top N
SERVER-ORACLE Oracle WebLogic Server remote command execution attempt
FILE-OTHER Adobe Acrobat Pro TIFF embedded XPS file out of bounds read attempt
NETBIOS SMB NTLMSSP authentication brute force attempt
BROWSER-IE Microsoft Edge App-v vbs command attempt
EXPLOIT-KIT Terror EK exe download attempt
SERVER-WEBAPP Joomla JDatabaseDriverMysqli unserialize code execution attempt
SERVER-OTHER Mikrotik RouterOS directory traversal attempt

加入一般統計與RMSE

預期目標與做法

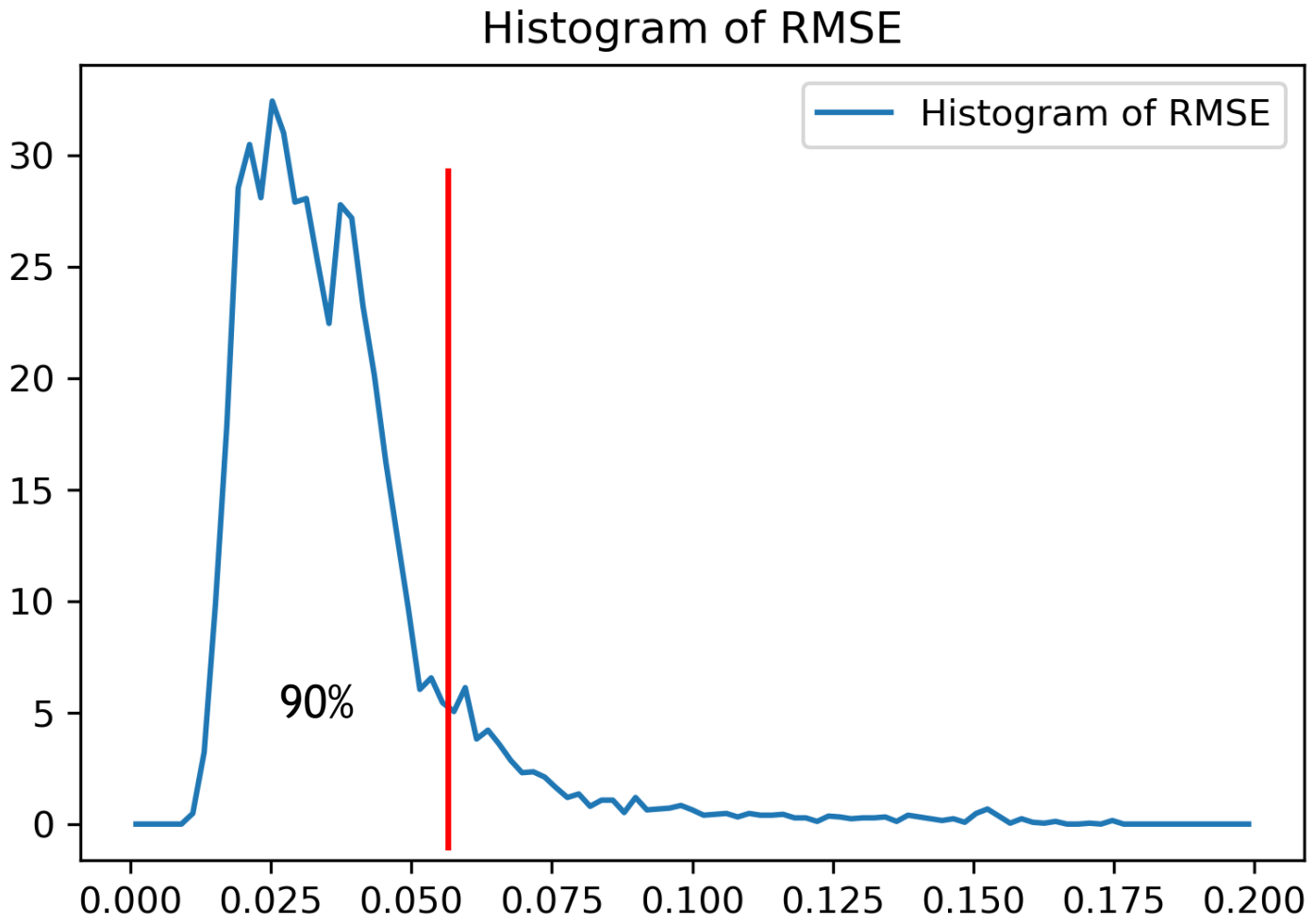
- 評估一般統計與 Deep learning 之預測比較

計算方法

- 計算 Trainging Dataset 每個 fields 之平均
- 再計算每個 entries 的 RMSE。
- RMSE：

$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}$$

RMSE



RMSE	Percentage
0.04545	80%
0.04949	85%
0.05757	90%
0.07171	95%
0.12626	99%

預測結果 新增RMSE欄位

異常偵測 for IPS event

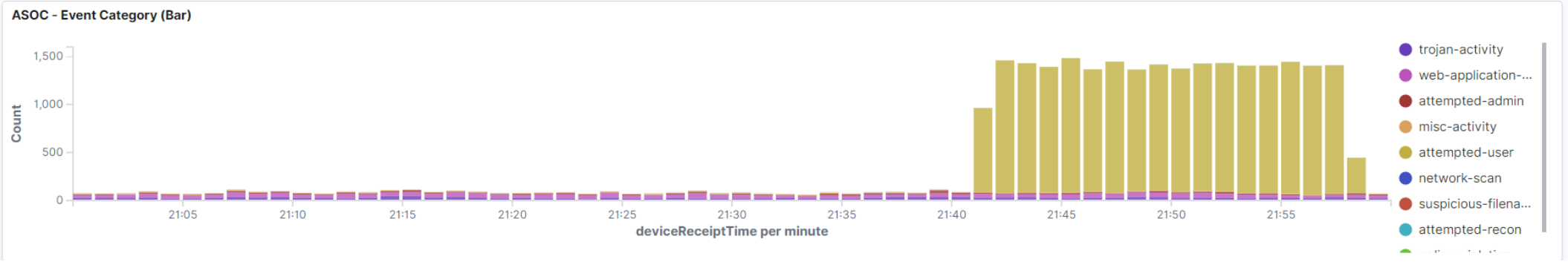
2020-05-30 08:00 ~ 2020-06-11 20:00

異常時段(小時)：6

Loss threshold : 0.018

Time	Link	Loss	RMSE
2020-06-09 21:00:00	Link to kibana	0.0349	0.1319
2020-06-09 19:00:00	Link to kibana	0.0326	0.1229
2020-06-09 10:00:00	Link to kibana	0.0261	0.0968
2020-06-09 07:00:00	Link to kibana	0.0255	0.0970
2020-06-09 04:00:00	Link to kibana	0.0244	0.0827
2020-06-09 08:00:00	Link to kibana	0.0186	0.0657

DEMO 2


ASOC - Event Category (Bar)
1,500
1,000
500
0
Count
21:05
21:10
21:15
21:20
21:25
21:30
21:35
21:40
21:45
21:50
21:55
deviceReceiptTime per minute
trojan-activity
web-application-...
attempted-admin
misc-activity
attempted-user
network-scan
suspicious-filena...
attempted-recon

Conclusions

- 將設備日誌資料導入 ELK。
- 視覺化分析資料。
- 利用 Elasticsearch API，使外部程式配合。
- 計算事件之間關聯性。
- 找尋高關聯性之事件。
- 分析事件。
- 透過更多計算方法，找尋最異常，需被觀察之事件，
- 利用機器學習，找尋異常事件。
- 可避免使用 Threshold based的偵測方式。

Feature Works

➢ 預計可將此異常比例分析，應用至其他log分析上，如： netflow...等。

Colab

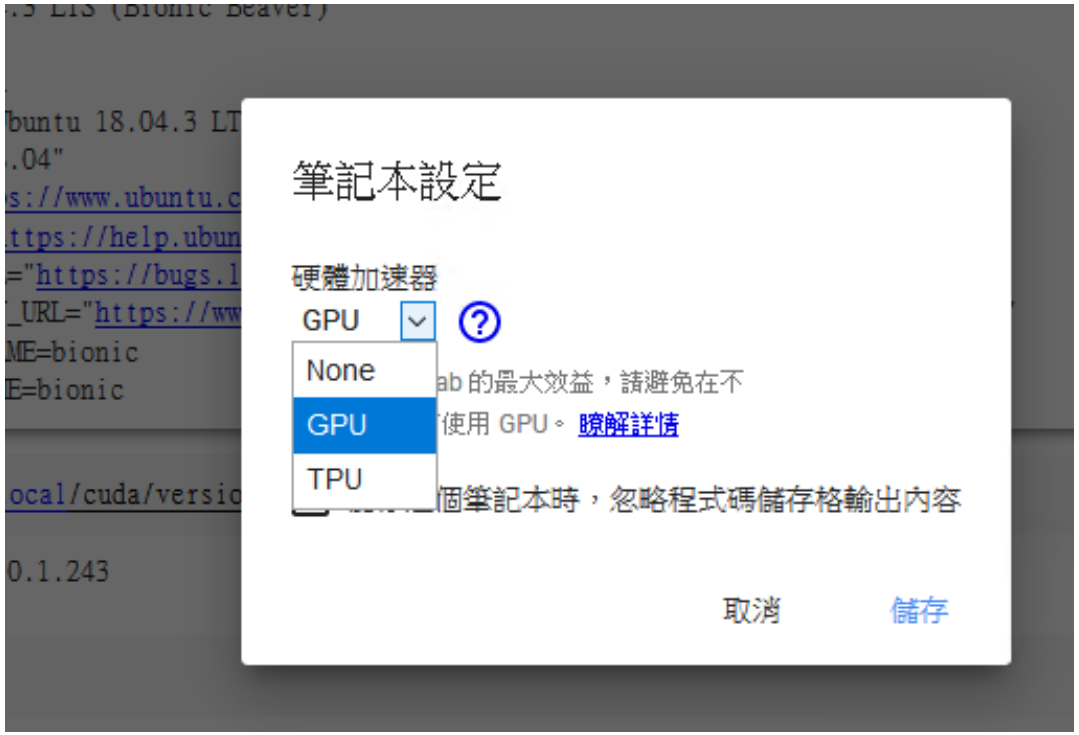
What's Colab

- Google 的免費雲端 Jupyter Notebook (https://colab.research.google.com)

深度學習利器 (窮人版)

- Colab 有提供GPU、TPU的運算資源。
- 對於沒有錢買顯示卡也想做深度學習運算的同學來說是一大福音。

- 但也不是可以一直使用：
- Colab 運算資源會優先提供給付費會員
- Colab Pro：



Colab CPU VS. GPU

CPU

```
Epoch:  99 | train loss: 0.0147
Epoch:  99 | train loss: 0.0112
Epoch:  99 | train loss: 0.0121
Epoch:  99 | train loss: 0.0126
Epoch:  99 | train loss: 0.0147
Epoch:  99 | train loss: 0.0137
Epoch:  99 | train loss: 0.0137
Epoch:  99 | train loss: 0.0122
Epoch:  99 | train loss: 0.0131
Epoch:  99 | train loss: 0.0142
Epoch:  99 | train loss: 0.0140
Epoch:  99 | train loss: 0.0142
Epoch:  99 | train loss: 0.0115
Epoch:  99 | train loss: 0.0129
Epoch:  99 | train loss: 0.0151
Epoch:  99 | train loss: 0.0140
執行時間：5050.658701 秒
```

18倍

GPU

```
Epoch:  99 | train loss: 0.0136
Epoch:  99 | train loss: 0.0119
Epoch:  99 | train loss: 0.0131
Epoch:  99 | train loss: 0.0122
Epoch:  99 | train loss: 0.0145
Epoch:  99 | train loss: 0.0137
Epoch:  99 | train loss: 0.0138
Epoch:  99 | train loss: 0.0146
Epoch:  99 | train loss: 0.0126
Epoch:  99 | train loss: 0.0127
Epoch:  99 | train loss: 0.0136
Epoch:  99 | train loss: 0.0126
Epoch:  99 | train loss: 0.0131
Epoch:  99 | train loss: 0.0137
Epoch:  99 | train loss: 0.0132
Epoch:  99 | train loss: 0.0136
/usr/local/lib/python3.6/dist-packa
  "type " + obj.__name__ + ". It wo
執行時間：277.567550 秒
```