運用大數據，偵測 SSH 字典攻擊告警分析系統

陳品瑄 梁明章 陳俊傑
財團法人國家實驗研究院國家高速網路與計算中心
{hsuan, liangmc, jjchen }@narlabs.org.tw

摘要 —管理者在抵禦入侵者時依然是透過密碼系統做第一道防線防護。字典攻擊手法為利用字典中經常出現的字串加以組合藉以猜測伺服器管理者可能使用的密碼，這類攻擊手法的惡意行為，不僅消耗網路頻寬及系統資源，同時也將造成資安威脅。SSH(Secure Shell)是 Linux 常用的加密式的通訊協定。伺服器一般都會開啟 SSH 服務方便管理者遠端管理，但是容易被用各種方法來取得進入的權限。為了能有效解決 SSH 字典攻擊的問題，本研究針對 SSH攻擊的威脅，透過日誌收集與分析的方式，掌握網路使用行為以及攻擊相關事件紀錄。以傳統 Log 分析方式很難快速得知異常狀況；加上 Log 並非結構化資料。因此，本研究以 ELK Stack[1]為基礎，在台灣高品質學術研究網路（TaiWan Advanced Research and Education Network，TWAREN）[2]網路環境下建置大數據整合資料平台，整合不同的網管技術(Syslog、Netflow、SNMP)，不僅分析系統日誌文件，並同時關聯 SSH 通信協定的 Netflow 網路流量資料，整合分析在該攻擊時點 SSH 通信協定的流量紀錄，進一步分析 SSH攻擊事件網路流量，同時快速定位攻擊者所在地(國家、城市)，利用攻擊來源 IP 經緯度以視覺化的呈現在攻擊地圖上，讓管理者可以快速追蹤攻擊來源。利用 ELK 快速查詢的能力來加速判斷是否遭受有 SSH 攻擊的事件發生，及時予以告警，以避免 SSH攻擊造成網路上的威脅。同時透過流經 TWAREN 骨幹網路上的 SSH 通信協定的 Netflow 流量資料，快速分析其它受害者，共同協防以避免資訊安全攻擊事件發生。將 TWAREN 海量資料背後所隱藏的資訊，藉由即時分析，快速掌握威脅來源，有效應用巨量資料分析技術偵測 SSH 攻擊，最後，並實際 TWAREN 網路環境資安事件管理機制，驗證偵測 SSH 攻擊結果準確度。

關鍵詞：SSH、Big Data、TWAREN、ELK、ElasticSearch、LogStash。

一、 簡介

隨著資訊科技的蓬勃發展，資安意識的逐漸普及與提升，SSH 是幾乎所以的網路設備或是伺服器都會開啟的服務，現今，利用 SSH 登入攻擊的案例也有增多的趨勢。有越來越多的組織佈署了安全防禦設備或是上網行為管控機制，針對封包裡第七層(Application)的內容進行檢查與分析。第七層的訊息除了能幫助 IT 人員得知組織中所發生的資安威脅事件之外，也可以有更完整的了解人員的網路使用行為。一般維運工程師在網路異常或遭受攻擊時的做法，登入設備常利用 grep/sed/awk 等 Linux 工具查看日誌檔資料，查看原因。目前大多數的網路與資安設備、電腦作業系統都已支援透過 syslog 協議的方式將日誌輸出並結合日誌儲存系統，因此，可透過報表工具、告警達到更好的管理效率。IT 管理人員則透過日誌收集與分析的方式，快速掌握網路使用行為以及資安相關的事件。不同的網管技術所提供的分析資料意義也有所不同，整合不同的網管技術(syslog、SNMP、Netflow)，也將是新一代 IT 維運方式。Flow 記錄兩個 IP 之間的傳輸用量統計，資料內容帶有 Source IP、Source Port、Destination IP、Destination Port、Protocol、Packet、Byte 等。而 Syslog 日誌主要是記錄系統中任何時間發生的大小事件。因此，本研究在 TWAREN 網路環境下，整合 Syslog 與 Flow 的訊息，以 ELK Stack 為基礎，利用 Elastsearch 即時查詢和 Logstash，完成 Syslog 和 Netflow 的接收、儲存、分析與查詢的功能。透過分析系統日誌記錄檔並關聯 Netflow 網路流量來偵測 SSH 字典攻擊及自動告警系統。

二、 字典攻擊

對於 SSH 服務常見就是字典攻擊或是暴力破解攻擊，攻擊者先以探測性的方式去掃描各個機器的 SSH 服務(預設通訊埠 22)是否有開啟，再以字典攻擊方式，攻擊開啟 SSH 服務的伺服器，試圖取得伺服器的登入權限。遠端攻擊者通過不同的密碼無限次地進行嘗試登入。也就是攻擊者運用字典檔內的單字詞組合，進行網路系統帳號密碼的猜測行為，攻擊期間透過不斷的錯誤嘗試，直到取得正確的帳號密碼，或無法破解為止，才會放棄攻擊。一般字典攻擊可分為線上(Online)與離線(Offline)兩種攻擊模式[5]：

A. 線上字典攻擊(On-line)
- 攻擊者在網路上直接對受駭者做帳號密碼登入驗證，以便檢查猜測的密碼是否可正確登入
- 攻擊者在嘗試帳號密碼登入時，會留下大量錯誤紀錄於受駭者日誌紀錄檔中

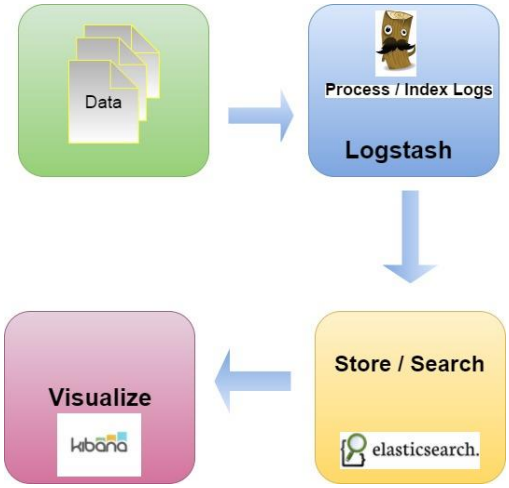
B. 離線字典攻擊(Off-line)
- 不需與伺服器做互動帳號密碼驗證
- 攻擊者透過受駭者中其他程式漏洞盜取受駭者的密碼檔
- 可離線使用離線字典攻擊工具，於攻擊者電腦猜測使用者密碼

在本研究中採用線上字典攻擊模式進行偵測，ELK 所收集到在 TWAREN 網路中的設備日誌紀錄檔。

三、 系統建置與實作

3.1 資料收集

好的海量資料處理平台可以事半功倍的提升開發人員和維運人員的效率。以傳統 log 分析方式很難快速得知異常狀況；加上 log 並非結構化資料，因此，本系統透過 ELK Stack 來確保高效率且穩健的運作，整合不同資料來源。ELK 是 Big Data 的重要解決方案之一，架構是基於 ElasticSearch(E)、LogStash(L)[1]、Kibana(K)所組成如圖一所示。



圖一： ELK Stack

A. Syslog

在本研究中利用 Logstash 蒐集多種不同來源的日誌及網路流量，透過 Log 檔的時間戳記(TimeStamp)的資料，來分析事件發生的時間和內容。在 ELK 中 input plugin 有 syslog plugin 可以使用，但是每台網路設備都會有自定義自己的記錄格式，導致收到的記錄訊息格式不盡相同，因此，必須透過 Logstash 去做解析。在 Elasticsearch 集群中可以包含多個索引(indices)(資料庫)，每一個索引可以包含多個類型(types)(表)，每一個類型包含多個文件(documents)(行)，然後每個文件包含多個欄位(Fields)(列)。

表 I
傳統 SQL 資料庫與 Elastsearch 對照表

傳統 SQL 資料庫	Elastsearch
DB	Index
Table	Type
Primary key	Id
Row	Document
Column(Schema)	Field

因此，將原本非結構化的 syslog 透過 Logstash 轉成結構化的資訊，資料格式如下：

表 II
Syslog 資料格式

Field	說明
syslog_facility	定義產生事件的子系統(subsystem)。
syslog_facility_code	定義產生訊息之系統的一部分代碼
syslog_host	傳送訊息之系統的 IP 位址
syslog_hostname	傳送訊息之系統的名稱。
syslog_message	訊息的文字
syslog_pri	產生訊息的處理序識別碼
syslog_severity	事件的嚴重性層級
syslog_severity_code	事件的嚴重性層級代碼
syslog_program	定義產生事件的程式
syslog_timestamp	產生事件的日期和時間。
type	產生事件 Table

B. NetFlow

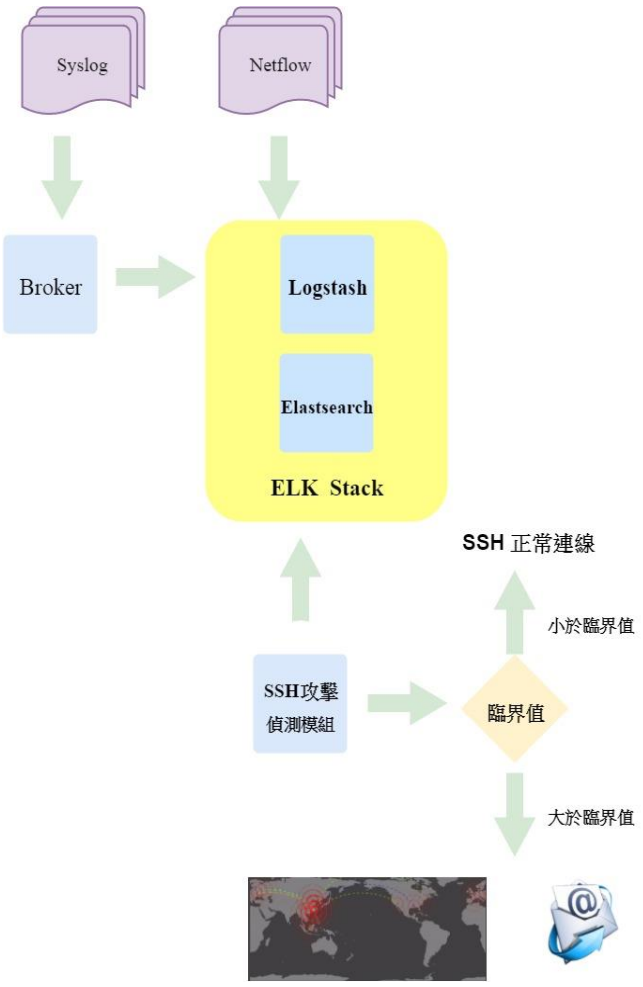
在本研究 ELK 中同時整合 NetFlow 流量資料，來了解當下網路用量。在本研究中，ELK 利用 NetFlow v9 格式的資料屬性進行分析 SSH 攻擊。每一筆 flow 是依相同的來源 IP 位址(source IP address)、來源埠號(source port number)、目的 IP 位址(destination IP address)、目的埠號(destination port number)、協定種類(protocol type)、服務種類(type of service)、及路由器輸入介面(router input interface)的封包資訊，透過以上七個欄位的封包，來判斷這個封包是否屬於任何已記錄的 Flow，有的話則將新收集到的封包的相關流量資訊整合到對應的 Flow 記錄中，其 bytes 數和 packets 數都會累計記入該 flow 中，如果找不到封包對應的 Flow 記錄，便產生一個新的 Flow 記錄來儲存相關的流量資訊。

在本系統中，透過網路流量來觀察 SSH 攻擊網路流量的變化，透過 NetFlow，分析關於攻擊相關時點 SSH 網路用量的統計及封包數及封包大小統計分析，以此來提供管理者更進一步判斷是否為 SSH 正常網路流量。並藉由 Netflow 網路流量，追查出其它受害者。

3.2 系統流程

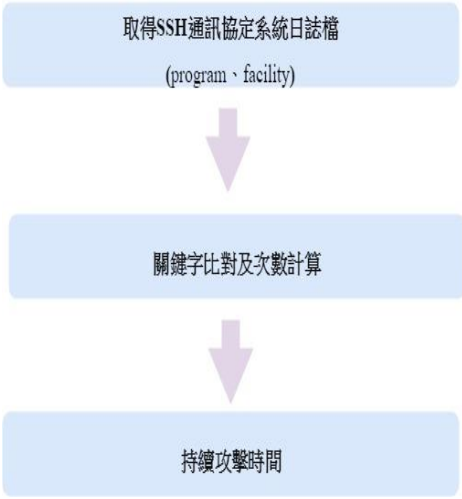
SSH 攻擊分析是當攻擊者發動 SSH 掃描探測攻擊時，而產生的錯誤記錄日誌檔。並透過分析 SSH 連線記錄，以取得進行 SSH 攻擊的網路來源。但記錄檔分析缺點在於每天累積的海量日誌資料、事件紀錄。若頻繁的進行日誌檔分析，將可能造成系統效能低落。因此，本研究利用 ELK Stack 來解決海量的記錄檔資料處理與分析。透過收集 TWAREN 所有的路由器及主機的日誌檔資料及 TWAREN 網路流量資料分析 SSH 攻擊。

本研究提出的 SSH 攻擊分析，運作流程如圖二所示：



圖二： 系統流程

研究分析日誌資料資料來源為 TWAREN 網路環境中的 Router 設備及主機，透過 Logstash 解讀設備日誌資料並儲存在 Elastsearch 中。



圖三： SSH 攻擊偵測流程

根據 RFC3164 的定義 Syslog facility 有下列 23 種

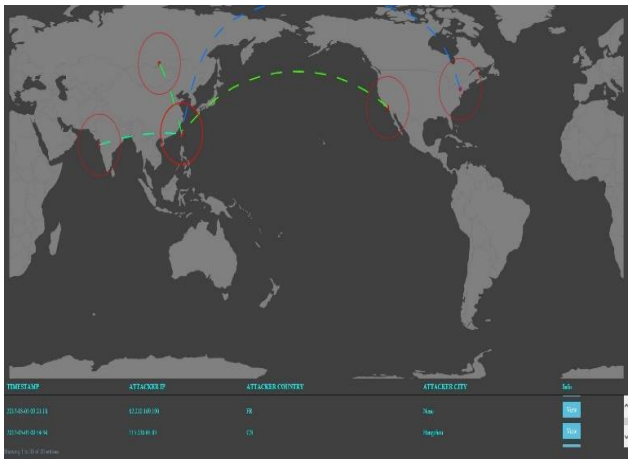
1. kernel messages：系統核心所產生的訊息
2. user-level messages ：使用者自訂的訊息
3. mail system：系統產生的訊息
4. system daemons：deamon 所產生的訊息
5. security/authorization messages (note 1)：登入或認證相關訊息
6. messages generated internally by syslogd：syslogd 產生的訊息
7. line printer subsystem：印表機訊息
8. network news subsystem：新聞服務的相關訊息
9. UUCP subsystem
10. clock daemon (note 2)
11. security/authorization messages (note 1)
12. FTP daemon
13. NTP subsystem
14. log audit
15. log alert
16. clock 2
17. local 0：local x 都是由本機設定
18. local 1
19. local 2
20. local 3
21. local 4
22. local 5
23. local 6

訊息紀錄檔案都會透過 syslog() 系統呼叫將訊息丟給其他程式來進行接收處理，當攻擊者嘗試以 SSH 服務登入時，失敗的訊息紀錄在訊息配置紀錄 SyslogFacility 中的 security/authorization messages，因此，在本系統偵測 SSH 攻擊者探測多台機器是否有提供 SSH 登入服務，進而使用帳號密碼字典檔案方式執行暴力破解登入時，登入失敗訊息會配置在 security/authorization，因此，本系統分析 SSH 連線系統日誌及 SyslogFacility 為 security/authorization messages 配置紀錄進行關鍵字比對及次數計算。

SSH 攻擊偵測模組會定期查詢 Elk 中收到的 SSH 通訊協定的所有設備日誌檔資料，在海量分布日誌中解析，判斷日誌資料中是否有入侵特徵的關鍵信息並更進一步觀察攻擊持續時間，以判斷是否需要告警通知。臨界值設定根據歷史攻擊資料及 TWAREN 安全政策來進行調整設定臨界值(Threshold)與其比較過濾，並透過臨界值的變數，來決定是否要將此行為視為 SSH 字典攻擊行為。當達到攻擊次數的臨界值時，則會發送告警通知信，並同時視覺化的呈現在網站上，透過攻擊地圖追蹤攻擊來源。在本研究中也利用 NetFlow 網路流量資料做為維運者後續觀察是否為正常 SSH 網路流量。因為，本研究以 SSH 攻擊為方向，因此以網路流量中連線目的埠為 SSH 預設通訊埠 22 網路流量資料為分析資料。透過視覺化的呈現，可以得到攻擊的來源、受駭的目的地、攻擊的網路流量等資訊。

四、 SSH 攻擊實例驗證

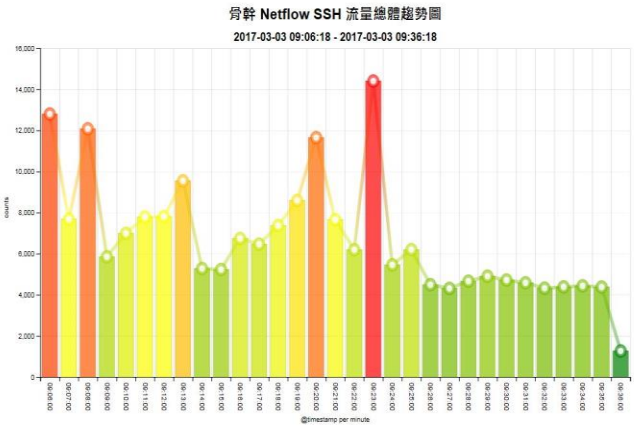
在本研究中將偵測分析到的攻擊者和所本中心所引入的 ArcSight 資安事件控管平臺針對 SSH 攻擊監控結果做驗證，以當 SSH 攻擊偵測模組偵測到有 SSH 攻擊事件發生時，即時定位攻擊者所在地(國家、城市)，透過所記錄的經緯度，以網路地圖方式呈現攻擊來源。以 2017 年 03 月 03 日攻擊情形為例，呈現方式如圖四所示。



圖四：SSH 攻擊網路地圖

以 2017 年 3 月 3 日在 TWAREN 資安事件中紀錄發生 SSH 攻擊且本研究中也偵測到相同攻擊來源，發現在 09:21 位於法國(62.210.169.190) ，有人使用暴力登入的方式，不斷的以大量的 SSH 攻擊 TWAREN 骨幹網路中的主機。透過即時統計該攻擊時點在 TWAREN 骨幹上 SSH 網路流量，觀察 SSH 網路流量變化，如圖五所示。

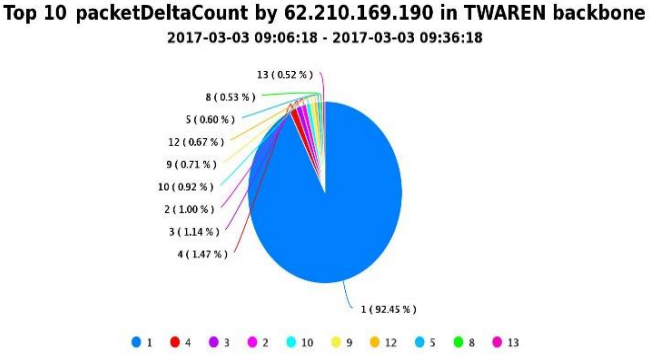
偵測攻擊時間	疑似攻擊者	疑似攻擊者國家	疑似攻擊者城市
2017-03-03 09:21:18	62.210.169.190	FR	None



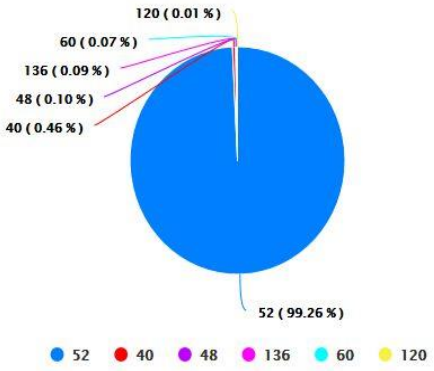
圖五： 攻擊時點骨幹 Netflow 趨勢圖

圖五為即時提供該攻擊時點前後 15 分鐘，TWAREN 網路上 SSH 通信協定 Netflow 流量趨勢圖，x：時間，y：次數。

我們藉由流經 TWAREN 骨幹網路的流量特性，蒐集 SSH 字典攻擊的流量特性資料，利用 TWAREN 網路 Netflow 流量資料，對 Netflow 資料中連線目的通訊埠為 22 的網路流量進行分析，統計分析攻擊者 SSH 連線網路封包資料，由圖六、圖七可以得知，偵測到來自法國的攻擊者在攻擊當下時間區間內，嘗試對 TWAREN 網路中的 IP 進行 SSH 連接，SSH 網路流量暴增，且九成以上在 TWAREN 所管轄網路流量中，流向多個 IP 的網路流量中 bytes 和 package 數較小且固定，和一般正常 SSH 網路流量的 bytes 和 package 數數值有所差異。藉此整合 Netflow 網路流量提供綜合性分析，透過即時分析網路流量 bytes 和 package 數，提供維運者進行後續網路狀況稽核，更進一步判斷是否為 SSH 正常流量。

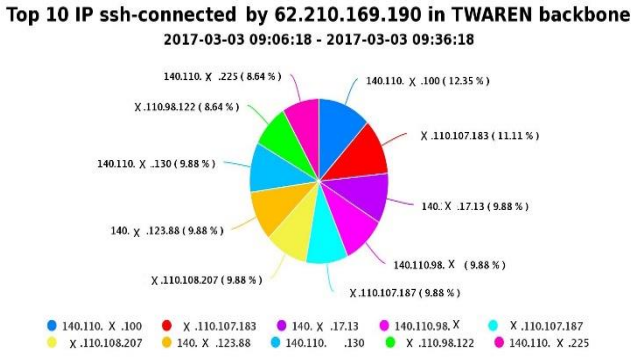


圖六 Netflow 網路流量封包統計



圖七 Netflow 網路流量封包大小統計

同時，在 TWAREN 骨幹網路上，透過 Netflow 網路流量分析，追蹤到的 SSH 攻擊者(如圖八)，在攻擊的時間區間當下，還有那些設備也遭到攻擊，透過分析出的受害清單，共同協防以避免遭受到重大攻擊。



圖八 其它疑似受駭名單

結論與未來研究工作

安全威脅千變萬化、網路入侵事件頻傳，單獨的網路監控已不足以應付日常維運所遭遇到問題。當系統發生不當或遭受攻擊時，管理者必須藉助日誌記錄檔，查明真相，以解決或預防此類事件影響系統正常運作，以確保系統的安全。在本研究中，利用 SSH 協定與 SSH 字典攻擊的特性，透過 Elk stack，整合 syslog 、Netflow 不同的網管技術來做分析，利用系統日誌檔記錄系統中所發生事件的人、事、時、地等相關網路使用行為訊息，並搭配 Elasticsearch[4]巨量資料搜尋分析，檢視 SSH 網路攻擊可能事件，了解系統已經遭遇到或潛藏的問題所在，並據此做出決策，自動告警。同時關聯分析 Netflow 網路流量，來了解關於網路用量的訊息，並同時搜尋出在 TWAREN 骨幹網路中其它正在遭受攻擊受駭清單，在發生事件時儘快通報，或是能透過資安事件通報機制避免其他單位受駭。資訊安全攻擊手法層出不窮，每種技術各有專長也都有不足之處。由於本系統透過關鍵字比對、及攻擊次數設定等臨界值，因此，移植性高，未來，可以透過 ELK Stack 整合不同的網路技術，利用大數據分析在海量的資料來源中，偵測其它網路攻擊，將已發生或是潛藏的問題挖掘出來，為網路與資訊安全提供新的思維，將有助於 IT 人員窺知網路活動的全貌，掌握網路使用細節，以滿足目前 IT 管理的網路維運需求。

參考文獻

[1] Elastic, https://www.elastic.co/
[2] TWAREN, TaiWan Advanced Research and Education Network. http://www.twaren.net/.
[3] Logstash, https://www.elastic.co/products/logstash
[4] Elasticsearch, https://www.elastic.co/products/elasticsearch
[5] Pinkas, B. and Sander, T.(2002) “Securing passwords against dictionary attacks,” Proceedings of the 9th ACM Computer and Security Conference
[6] 陳品瑄,梁明章,陳俊傑, “TWAREN 大數據整合資料平台” TANet2016.