巨量資料於惡意程式行為分析應用

蔡一郎(Yi-Lang Tsai)
國家高速網路與計算中心
National Center for High-performance Computing
yilang@narlabs.org.tw

摘要

巨量資料收集、儲存、分析與視覺化呈現技術，並且應用於資料科學的領域，由資料的收集、分析模型的建立、創造有價值的資訊，是目前資料科學應用中最典型的處理方式，以資訊安全領域而言，隨著資訊系統的複雜化以及網路頻寬的增長，加上多變化的資安威脅，成了目前雲端服務時代，所需要面對的重要議題，在有限的時間內如何利用有效的分析方式，解決快速成長的巨量資料，同時亦需解決不同的資料來源整合的問題，大尺度的誘捕網路在網際網路建立了預警的機制，提供資訊安全分析人員可以掌握網路上的危安因素，並且因應攻擊者持續改變的攻擊手法，透過資料的分析，找出隱藏其中的攻擊特徵，以建立資安風險評估機制，並降低資訊安全事件帶來的威脅。

關鍵詞：資訊安全(Information Security)、巨量資料(Big Data)、惡意程式(Malware)、雲端服務(Cloud Services)、誘捕網路(Honeynet)

壹、前言

資訊安全事件近年來層出不窮，而攻擊者所使用的手法也日益更新，除了傳統針對特定目標進行的攻擊之外，也運用了殭屍電腦(Bot)所組成的殭屍網路(Botnet)，大量的在網路上進行滲透與發動攻擊的動作，在資訊安全的事件分析上，除了需要掌握來自資訊安全設備、誘捕系統或是誘捕網路所造成的日誌資料之外，在事件的追蹤上，亦可以需要配合網路流量的資訊，或是到現場進行數位鑑識，以收集更多的資料，對於資安事件發生的原因，提供事件原由判斷上的依據。

由資訊安全偵測與防禦設備所產生的資料，依所在網路環境的使用人數、網路的流量以及網路應用服務的複雜性，除了不同的組織與單位在資料的成長上略有不同之外，相同的是這些資料都符合巨量化、多樣化、異質化的特性，對於資料科學上的研究而言，如何處理巨量資料，以及在有效的時間內找到有用的資訊，就成為處理資訊安全事件的重要關鍵因素。

雲端服務的架構提供使用上更多的彈性，除了資源的彈性調配之外，對於資料的處理上有別於以往的處理方式，對於分散在各地的資料，需要建立一有效的分析機制，以解決巨量資料的儲存、分析以及傳輸上所面臨的問題。

貳、巨量資料

從傳統的資訊安全防禦機制，建置了防火牆(Firewall)、入侵偵測/防禦系統(IDS/IPS)、垃圾郵件閘道等相關的設備，來自一個使用者的通訊行為，當到達目的地的過程中，依不同設備之間的特性，經過的設備都將可能留下相關的日誌紀錄，以資訊安全相關的設備而言，日誌的產生與其中所使用的網路安全政策與網路安全規則有關，因此不同的組織或是單位，都會因地制宜的設定個別的網路安全政策或是網路安全規則，以符合自己的需求。

因應使用者網路行為的多樣性，運用不同的誘捕系統提供特定項目的資訊，例如：部署 Dionaea[1]或 Nepenthes[2]誘捕系統，以應用於惡意程式的捕獲與掌握惡意程式的相關行為，部署 Kippo[3]誘捕系統，則應用於偵測來自網路上的遠端密碼探測，透過日誌的分析，可以掌握攻擊者對於目標系統進行的帳號與密碼猜測，攻擊者在進入系統中所進行的各項操作，包括對於系統的操作、遠端的檔案下載以及攻擊者在建立系統後門的階段，所採用的工具與執行的指令，運用深層資料的收集與分析，以及大尺度的資料收集，對於廣域的網路安全趨勢掌握上，可以透過不同點的資料收集，瞭解國內整體的資安風險。

對於不同的誘捕系統而言，所產一的資料格式可以能不同，因此當建置多種不同的誘捕系統的時候，必須解決不同誘捕系統對於相同性質的資料欄位命名的問題，以便對於所收集到的資料進行分析，本文將以 Dionaea、Kippo 以及 Amun 三種誘捕系統的資料進行分析。

表一：Dionaea 誘捕系統資料格式

```
[2014/02/13T23:48:03] 85.174.11.210:2082 -> 163.24.231.17:445 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=smbd
[2014/02/13T23:48:03] 1.162.204.224:19469 -> 163.24.230.23:445 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=smbd
[2014/02/13T23:48:03] 202.120.80.48:2454 -> 163.24.231.219:1433 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=mssqld
[2014/02/13T23:48:03] 186.92.150.33:2543 -> 163.24.230.19:139 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=pcap os_name="Windows 2000 SP4, XP SP1+"
[2014/02/13T23:48:03] 5.138.111.115:0 -> 163.24.231.22:0 severity=Malware_offered honeypot=dionaea
Download_url=http://5.138.111.115:4218/ramkh os_name="Windows 2000 SP4, XP SP1+"
[2014/02/13T23:48:03] 202.120.80.48:2450 -> 163.24.231.219:1433 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=mssqld
[2014/02/13T23:48:03] 177.189.42.197:0 -> 163.24.230.16:0 severity=Malware_downloaded honeypot=dionaea
Download_hash=5f501e59097554e90c2d649acc843984 Download_url=http://177.189.42.197:3617/rxlde
os_name="Windows 2000 SP4, XP SP1+"
[2014/02/13T23:48:03] 111.253.61.132:29795 -> 163.24.231.19:445 severity=Malicious_attack honeypot=dionaea
type=MS08-67 type=MS08-67 Dialogue=smbd
[2014/02/13T23:48:03] 169.254.91.67:4140 -> 163.24.230.68:139 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=pcap os_name="Windows 2000 SP4, XP SP1+"
[2014/02/13T23:48:03] 95.180.119.214:3167 -> 163.24.230.214:445 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=smbd os_name="Windows 2000 SP4, XP SP1+"
```

表二：Kippo 誘捕系統資料格式

```
[2014/06/01T00:00:01] 5.228.253.183:54564 -> 140.113.1.10:139 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=pcap os_name="Windows 2000 SP4, XP SP1+"
[2014/06/01T00:00:01] 5.228.253.183:46649 -> 140.113.1.10:445 severity=Possible_malicious_attack
honeypot=dionaea Dialogue=smbd os_name="Windows 2000 SP4, XP SP1+"
[2014/05/31T23:59:57] 180.153.194.154:40150 -> 140.113.6.22:22 severity=Malicious_attack honeypot=kippo
sshuser=stephanie sshpass=stephanie os_name="Linux 2.6 (newer, 3)"
[2014/05/31T23:59:57] 180.153.194.154:40150 -> 140.113.6.22:22 severity=Malicious_attack honeypot=kippo
sshuser=stephanie sshpass=stephanie ssh_tool=SSH-2.0-libssh-0.11 os_name="Linux 2.6 (newer, 3)"
[2014/05/31T23:59:58] 180.153.194.154:40302 -> 140.113.6.22:22 severity=Malicious_attack honeypot=kippo
sshuser=root sshpass=hamster ssh_tool=SSH-2.0-libssh-0.11 os_name="Linux 2.6 (newer, 3)"
[2014/05/31T23:59:58] 180.153.194.154:40302 -> 140.113.6.22:22 severity=Malicious_attack honeypot=kippo
sshuser=root sshpass=hamster os_name="Linux 2.6 (newer, 3)"
[2014/06/01T00:00:00] 180.153.194.154:40496 -> 140.113.6.22:22 severity=Malicious_attack honeypot=kippo
sshuser=root sshpass=welcome1 ssh_tool=SSH-2.0-libssh-0.11 os_name="Linux 2.6 (newer, 3)"
[2014/06/01T00:00:00] 180.153.194.154:40496 -> 140.113.6.22:22 severity=Malicious_attack honeypot=kippo
sshuser=root sshpass=welcome1 os_name="Linux 2.6 (newer, 3)"
[2014/06/01T00:00:01] 180.153.194.154:40729 -> 140.113.6.22:22 severity=Malicious_attack honeypot=kippo
sshuser=root sshpass=welcome ssh_tool=SSH-2.0-libssh-0.11 os_name="Linux 2.6 (newer, 3)"
[2014/06/01T00:00:02] 128.73.120.56:0 -> 140.113.6.6:0 severity=Malware_offered honeypot=amun
Download_url=http://128.73.120.56:9246/wzdocdjx os_name="Windows 2000 SP4, XP SP1+"
```

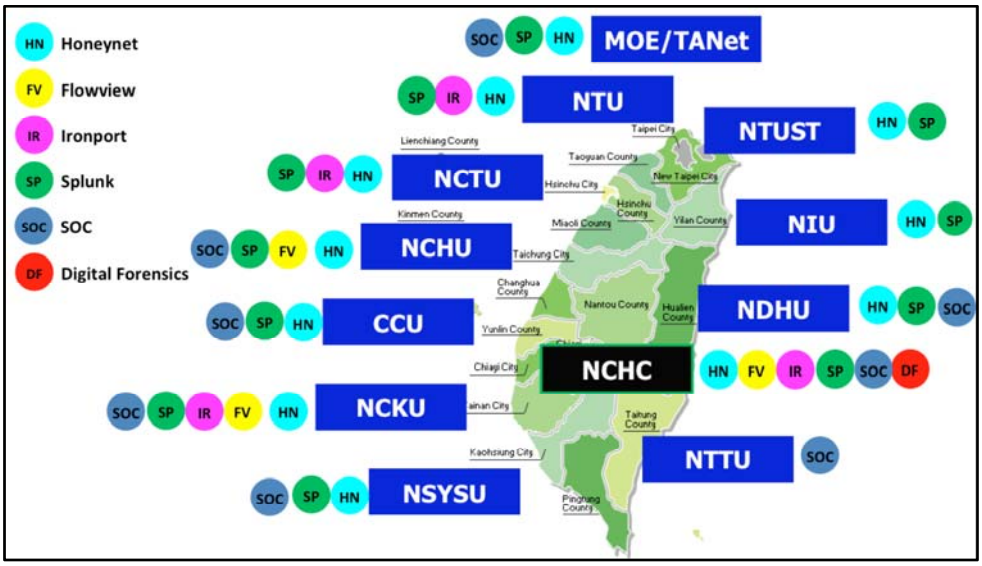
表三：Amun 誘捕系統資料格式

```
[2014/05/31T23:59:59] 114.42.236.180:1759 -> 140.120.21.19:445 severity=Malicious_attack honeypot=amun
[2014/05/31T23:59:59] 114.42.236.180:0 -> 140.120.21.19:0 severity=Malware_offered honeypot=amun
Download_url=http://114.42.236.180:7144/yxrlqf
[2014/05/31T23:59:59] 114.42.236.180:1761 -> 140.120.21.19:445 severity=Possible_malicious_attack
honeypot=amun
[2014/05/31T23:59:59] 114.42.236.180:0 -> 140.120.21.19:0 severity=Malware_offered honeypot=amun
Download_url=http://114.42.236.180:7144/yxrlqf
[2014/05/31T23:59:59] 114.42.236.180:1764 -> 140.120.21.19:445 severity=Possible_malicious_attack
honeypot=amun
[2014/05/31T23:59:59] 114.42.236.180:1762 -> 140.120.21.19:445 severity=Possible_malicious_attack
honeypot=amun
[2014/06/01T00:00:00] 114.42.236.180:1768 -> 140.120.21.19:445 severity=Malicious_attack honeypot=amun
[2014/06/01T00:00:00] 114.42.236.180:1765 -> 140.120.21.19:445 severity=Malicious_attack honeypot=amun
[2014/06/01T00:00:00] 114.42.236.180:1768 -> 140.120.21.19:445 severity=Possible_malicious_attack
honeypot=amun
[2014/06/01T00:00:00] 114.42.236.180:1767 -> 140.120.21.19:445 severity=Possible_malicious_attack
honeypot=amun
```

對於誘捕網路(Honeynet)的建置，其中依據偵測環境以及偵測目的上的需求，多數選擇建置一種以上的誘捕系統，除了增加資料收集上的多樣性之外，也可以掌握更多的資訊，例如：建置 Honeywall[4]的建置，可以建立連線行為的紀錄，透過統計與分析來自誘捕網路上的系統及網路應用服務日誌，就能夠掌握攻擊者所採用的攻擊手法，並且利用此攻擊手法的行為特徵，發展出可用以比對的特徵碼，以擴大偵測的範圍。

台灣學術網路與研究網路，範圍涵蓋了目前各級學校，範圍超過了四千所的學校以及五百萬的使用者，現階段誘捕網路部署在國內 11 個主要的網路節點，使用超過 6,000 個 IP 位址，透過誘捕系統進行資料的收集，在部署上採用的是混合不同型態誘捕系統的建置方式，在相同的偵測點部署不同用途的誘捕系統，以掌握多種常見的攻擊威脅，另配合資訊安全設備，包括了入侵偵測與防禦系統、垃圾郵件分析等設備，不同的偵測機制對於資訊安全的維運或網路威脅的預警機制，透過攻擊軌跡的追蹤，以掌握網路攻擊

事件的影響範圍，透過廣域的資料收集，以釐清事件的獨特性或廣泛性，大多數具備蠕蟲行為的惡意程式，將會碰觸到廣域的偵測設備，而如果是針對型的攻擊，則只會在特定的區域或網段被發現，接著可以配合所收集到的巨量資料，進行交叉的分析比對，將攻擊者的資訊當成比對的條件，例如：攻擊者的來源 IP 位址、所使用的通訊協定，以及通訊行為的特徵等，這些都是可以運用來與相關資訊進行比對時的參考。

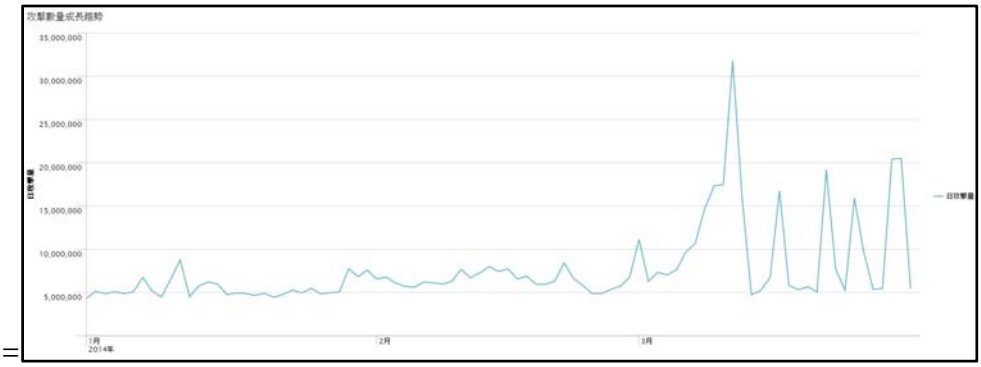


圖一：台灣學研網路之誘捕系統部署

誘捕系統的部署包括了低互動式誘捕系統以及高互動式的誘捕系統，從被動式的偵測到主動式的偵測，低互動式的誘捕系統，採用系統與應用服務模擬的方式，讓攻擊者以為所探測的目標存在可以利用的系統或應用服務漏洞，進而針對這個模擬出來的漏洞進行攻擊，而誘捕系統則藉此收集攻擊者所留下的資料，包括了下達的指令以及使用的工具程式，而低互動式的誘捕系統存在最大的問題就是無法模擬出所有可能的情況，如果是一個有經驗的攻擊者，當利用所模擬出來的漏洞完成入侵的動作之後，稍做簡單的測試，例如：在現有的目錄中新增檔案，再重新顯示檔案的清單，就能夠快速的辨識出誘捕系統的可能性大增，因此低互動式的誘捕系統，主要以偵測自動化的攻擊行為，而對於人為的攻擊行為，大多因為容易被識別無法收到較好的效益；而高互動式的誘捕系統大多以真實的作業系統與應用服務進行建置，並且故意留下可以被攻擊者運用的系統或應用服務漏洞，可以達到將攻擊者的行為真實紀錄下來的目的，不過因為是真正存在漏洞的系統，因此被攻陷之後成為跳板的機會大增，可能衍生更嚴重的資訊安全問題，因此在選擇建置高互動式誘捕系統時，需要特別的謹慎。

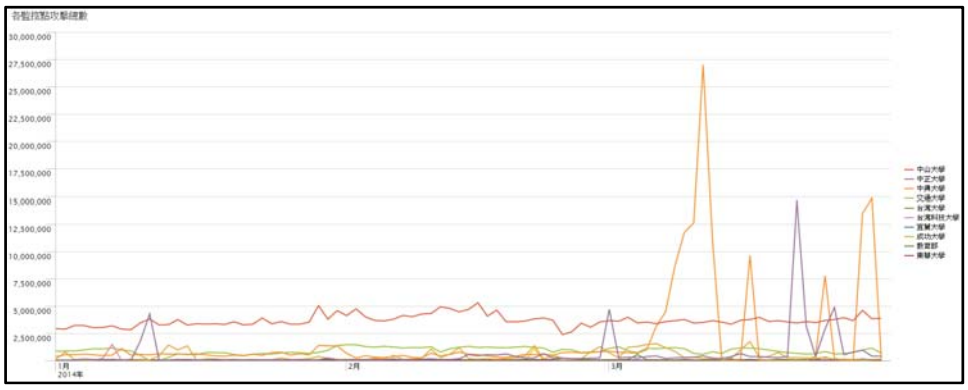
被動式的誘捕系統部署後並不會主動針對網路上的可疑網路或是主機進行分析，而是在網路上等著讓攻擊者或是惡意程式進行探測與發動攻擊的行動，對於資料的收集與資訊的掌握上處於被動的角色，如果未被探測到則無法引發後續的攻擊行動；而主動式的誘捕系統，透過資安分析人員所提供的疑似清單，進行後續的檢測程序，對於這些疑似威脅來源進行一一的掃瞄與清查，以確定疑似威脅來源的真實性，如果未能夠偵測出異常的活動，例如：不存在惡意程式的下載位址，或是系統資源的使用無法檢測出異常，這些看似系統或是應用服務可能使用的方式，將會是重要的檢查項目，目前主動探測的方式可以加快惡意程式或是資安威脅情況的確認，並且透過所檢測出來的結果，擬定因應的處置方式。[5]

將長時間所累積下來的巨量資料進行分析，以 2014 年第一季的資料統計為例，從偵測點所搜集到的資料量，可以知道攻擊數量的成長趨勢，透過長時間的資料分析，可以掌握攻擊趨勢是否具備週期性，其中的巨量攻擊，以 2014 年 3 月 11 日為例，當天偵測到的攻擊總量為 31,770,589 次，當天的攻擊量是以往的三倍之多，而在這天之後，仍發現有持續大量攻擊的行為存在。



圖二：攻擊數量成長趨勢圖

以這段期間所發生的巨量攻擊行為，由各偵測點的資料量進行分析，其中以中興大學偵測到高達 27,006,410 次的攻擊量為最多，因此可以斷定造成當天巨量成長的攻擊量主要以中興大學為主，不過再以後續四次的大量攻擊事件，中興大學的偵測點亦佔了三次，僅一次大量的攻擊事件發生在中正大學，此次發生在 2014 年 3 月 21 日，當天偵測到 14,633,234 次。目前在網路上的攻擊次數以及攻擊手法的多樣性，都朝逐漸成長的趨勢，尤其利用一些關鍵網路服務進行的放大攻擊，廣泛的利用現有的殭屍網路(Botnet)所進行的分散式阻斷服務(DDoS, Distributed Deny of Servie)攻擊造成的影響最大。



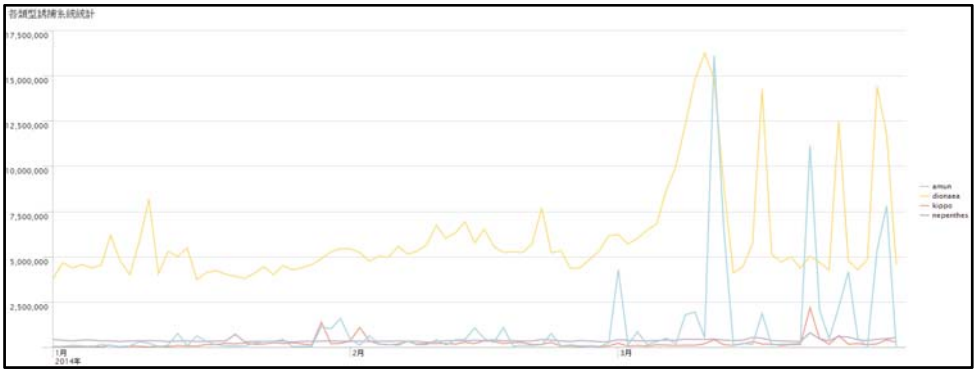
圖三：各監控點攻擊總數

不同的誘捕系統被設計來偵測不同的攻擊行為，或是針對不同的平台進行誘捕環境的建立，本論文選擇 Amun[6]、Dionaea 以及 Kippo 等三種誘捕系統進行資料的分析，資料量可參考表四所列，由不同的誘捕系統進行攻擊總數的調查，能夠取得攻擊行為的特性，例如：針對網站服務的攻擊行為或是與遠端進行帳號密碼猜測的探測行為, 採用低互動式誘捕系躲，以被動的方式偵測來自網路上的攻擊行為，依據惡意攻擊的行為給予回應，並藉此取得更多的資料供以後續的分析，此部份的資料包括了在完成攻擊的行動之後，在受駭主機上進行的操作或是下達的指令，對於攻擊者運用系統弱點的方式以及受駭系統在遭到攻擊後可能的後果。

表四：誘捕系統資料量(資料區間：2010/4/1 至 2014/3/31)

誘捕系統種類	遭受攻擊資料筆數
Amun	118,272,843
Dionaea	1,460,056,702
Kippo	22,643,450

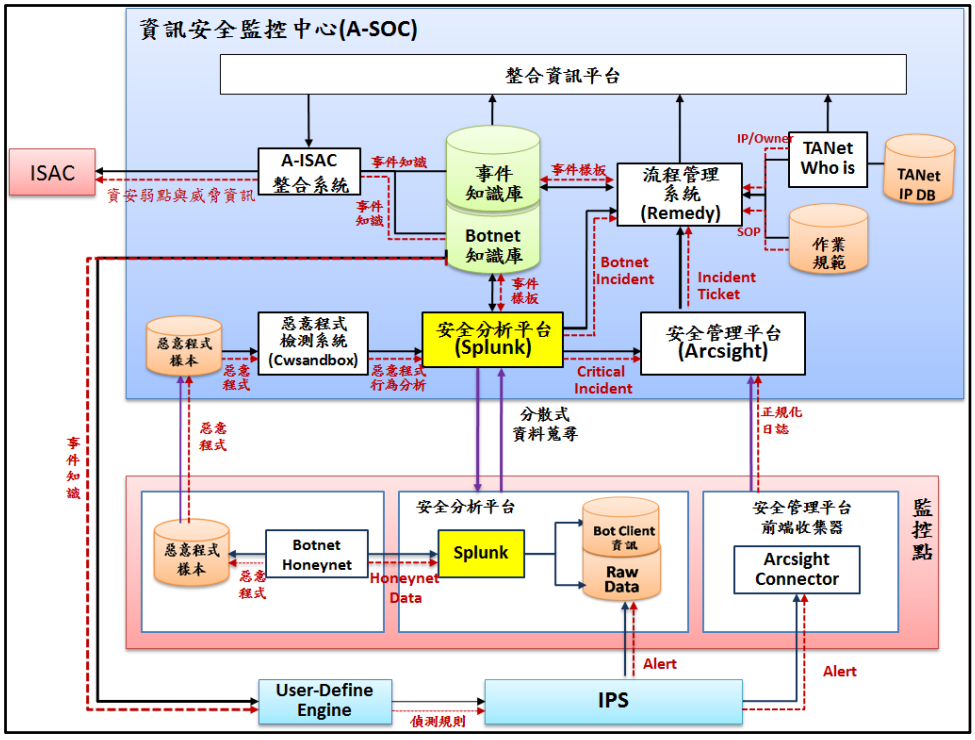
四年期間誘捕系統總共偵測到 1,600,972,995 筆攻擊紀錄，平均每年達 400,243,249 筆，每月則有 33,353,604 筆，因此從誘捕系統所偵測到的結果而言，每天存在網路上的異常攻擊仍然存在，且由長期的趨勢可以得知攻擊行為的次數有越來越多的情況。



圖四：各誘捕系統的統計

資訊安全維運中心(SOC, Security Operation Center)的巨量資料來自偵測範圍內的所有資訊安全設備，這些部署於網路上的偵測設備，隨時都在產生資料，因此針對不同型態的資料，必須設計出不同的處理流程，將不同的型態的資料以符合資料屬性的流程進行處理，其中主要的資料型態有惡意程式樣本檔、系統與資安設備的日誌紀錄以及網路攻擊的流量，惡意程式樣本檔為可執行或是需要配合應用軟體開啟的檔案，因此在處理上則採用沙箱測試(Sandbox)的方式，讓惡意程式樣本檔在所準備的環境執行或開啟，以掌握該樣本對於系統或網路環境所造成的影響；日誌紀錄為前端的偵測設備、網路設備以及提供服務的網路主機所產生的資料，此類型的資料能夠提供詳細的資訊，包括了通訊連線過程中的來源、目的地位址、通訊協定以及通訊的發生與結束的時間；而網路流量主要應用於擴充事件追蹤上的應用，利用已知的攻擊者、受害者與所採用的攻擊手法，配合網路流量的紀錄進行交叉的比對，可以擴大確認資安事件的影響範圍。

在資訊安全維運中心之內，除了即時的針對所收集到的資訊進行分析之外，對於每個事件所造成的影響進行追蹤也是相當重要的，事件本身所造成的影響程度與範圍須要能夠掌握與追蹤該事件後續的發展，從前端資料的收集、資料的傳輸、資料的分析以及事件單的成立，整個處理的流程必須在有限的時間內完成，所發佈的資訊安全事件單或是預警的情資，才能夠發揮預期的效益，其中需要許多不同的系統彼此結合，例如：查詢 IP 位址的 Whois 系統、流程管理系統以及資安事件的知識庫等，資訊的累積有助於提高事件單的正確度。



圖五：資訊安全維運中心架構圖

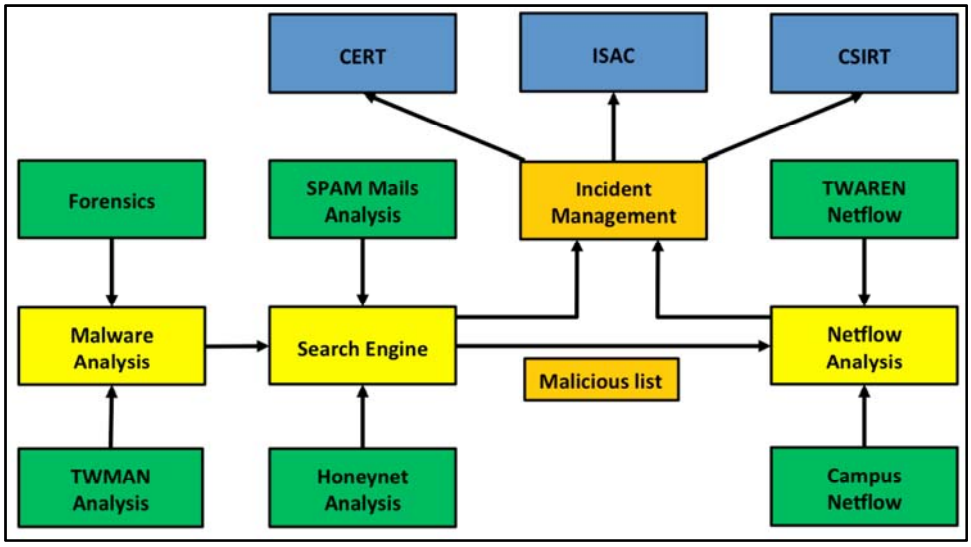
參、分散式的資料處理

對於巨量資料的儲存，除了儲存空間的考量之外，亦需要進行架構上的妥善規劃，對於傳統採用集中式保存資料的架構，分散式的資料儲存需要解決處理上的問題，包括了資料的儲存方式以及如何使用分散環境下的資料，

資訊安全威脅的應該，對於資訊安全維運中心而言是相當重要的，如何在短時間內完成事件威脅等級的分析，必須透過建立標準化的資料處理流程，以現階段的資料來源，涵蓋了誘捕系統、誘捕網路、資安設備以及網路流量的紀錄，這三種類型的資料屬性並不相同，誘捕系統與誘捕網路主要是針對特定弱點所提供的資訊，資安設備主要以特徵資料庫進行的比對結果，其中亦與部署據點所使用的網路安全政策有關，不同的網路安全政策將會影響能夠收集到的資訊；以點對點通訊協定(P2P, Peer to Peer Protocol)的行為偵測為例，部份的網路應用程式被設計成 P2P 的通訊方式，透過使用者對使用者的通訊方式，用以提昇網路通訊的效能，可以加速資料的傳輸以及增加資料的安全性，但是以目前網路上仍然相當活躍的 ZeroAccess 殭屍網路(Botnet)為例，所採用的就是 P2P 的通訊方式，而分散於各偵測點的誘捕系統、誘捕網路或是資安設備對於相同行為所收集到的資料相當龐大，尤其正當攻擊行為發生當下，每秒產生的資料筆數總數可達萬筆以上，因此在進行資料的分析時，將必須建立一致性的資料收集、分析與處理流程，以現階段台灣學術網路上的資安事件分析流程，可以由圖六來表示，其中包括了惡意程式的行為分析(Malware Analysis)、資料搜尋引擎平台、網路流量分析平台(Netflow Analysis)三個主要的部份，其中惡意程式的行為分析，主要的資料來自數位鑑識與沙箱行為測試的報告，目前除了採用 The Honeynet Project 所發展的 Cuckoo Sandbox[7]之外，亦採用了國家高速網路與計算中心[8]所自行發展的「台灣惡意程式分析網(TWMAN, TaiWan Malware Analysis Net)[9]」的分析環境，分別針對收集到的惡意程式進行為模式的分析，並綜合不同的分析報告，建立該樣本的行為模式報告，產出的報告與誘捕系統的日誌進行比對，確認攻擊的來源以及對於受駭系統造成的影響，其中網路行為的交叉比對可以找出疑似的殭屍網路中繼站(C&C, Command and Control)。

使用資料探勘與搜尋引擎的技術，針對分散於各資安偵測點的巨量資料進行分析，以降低資料必須利用網路回傳至資訊安全維運中心所需要耗費的網路頻寬，經過前端初步的針對所收集到的資料進行過濾，減少需要分析的資料量，其中需要利用分級與風險評估的方式進行資料的處理，由收集到的攻擊資訊再佐以網路流量的比對，有助於將建立點、線、面的關係，讓資訊安全事件的影響與涵蓋範圍完整的呈現，但是對於一個使用 10Gbps 以上的網路頻寬的高速網路環境而言，巨量的網路流量資料是交叉分析上的關鍵點，需要解決在有限的應變時間之內，如何進行巨量資料分析上的技術問題。

資安全事件的應變主要配合資訊安全維運中心進行處理，將已定義的資訊安全事件或是預警的情資進行發佈，提供佐證的資料伴隨的資訊安全事件單發佈給攻擊來源或是受害者網路的服務供應商或是管理單位，透過資訊安全事件的傳遞，加速資訊安全事件的處理，以及造成的持續影響。

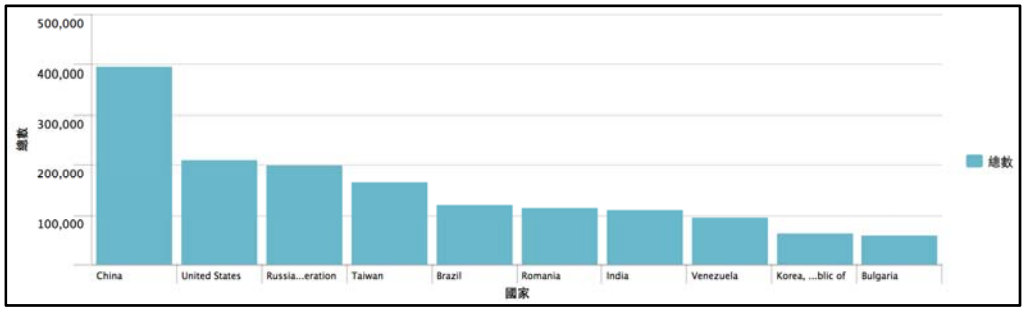


圖六：資料收集、分析與處理流程

肆、攻擊軌跡的追蹤

由惡意程式所造成的攻擊方式，依 Dionaea 誘捕系統所偵測到的攻擊階段，可以分成「 Possible_malicious_attack 」、「 Malicious_attack 」、「 Malware_offered 」 以 及「Malware_downloaded」，分別針對不同程度與層級的網路攻擊進行定義，不同階段的資料代表惡意程式或網路攻擊的不同的程度，整個網路攻擊具備不同的歷程，以一個成功的惡意攻擊行為而言，大多需要經過掃瞄探測、找尋目標主機弱點以及發動攻擊，最後再植入後門或是利用攻陷的主機進行其它的攻擊行為，而這些不同程度的行為，有發生的先後關係，但是並不見備連貫性，意即可能在整個攻擊行動中，會有不同程度的先後關係，但是每次出現的順序並不一定具有一致性。

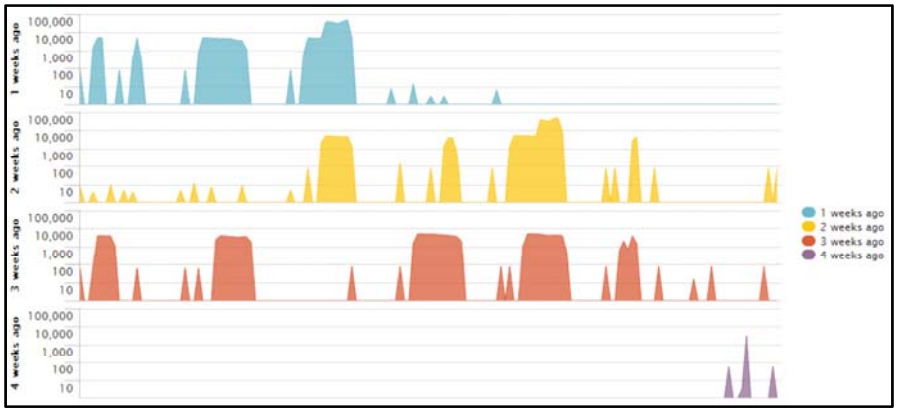
巨量資料的分析，可以先由統計的方式進行，針對攻擊來源的 IP 位址進行轉換，以攻擊來源國家的方式進行統計，可以先掌握發動攻擊的主要來源國家，以 2014 年 1 月 1 日到 2014 年 5 月 31 日的資料進行統計，我們可以發現攻擊來源的主要國家，已由往年的主要的攻擊來源美國、俄羅斯轉變成了中國大陸，這個是攻擊趨勢的上的轉變，不同的攻擊來源使用的攻擊的工具或是手法並不一定相同，再針對其中的攻擊手法進行深入的研究，就不難找出其中所使用的攻擊方式。



圖七：誘捕系統偵測各國攻擊來源的統計(資料日期：2014/1/1 至 2014/5/31)

惡意程式活動大多具備自我複製、散播以及感染的特性，因此由誘捕系統所掌握的資料進行分析，我們可以清楚的知道目前惡意程式針對攻擊目標所進行的動作以及可能影響的程度，如果只是在掃瞄目標主機的階段以及嘗試進行惡意的攻擊行為，可以視為預警追蹤的階段，在這個階段我們可以針對攻擊來源進行監測，持續的收集其所進行的各項攻擊的活動，以及所進行的攻擊方式；若惡意程式的活動已造成影響，在受攻擊的目標主機上，將會收集到惡意程式對於該系統或應用服務弱點進行攻擊時造成的影響，例如：使用者權限的提昇、後門程式的植入等，再進而對已攻陷的主機當成發動下一波攻擊的跳板，此時可以觀察到惡意的網路活動將會出現在原本遭受攻擊的目標主機上，這對於攻擊軌跡的追蹤上，可以區分成攻擊者以及受攻擊者兩個不同的部份，尤其以網路的通訊行為追蹤是屬於較重要的項目，如果攻擊的來源位址在將來亦會出現在遭入侵後的目標主機對外建立通訊連線的對象，就有極大可能的機會找到殭屍網路的中繼站。

由長時間的針對所收集的資料進行分析，可以由不同的面向切入，例如：指定曾經四個星期都出現過的攻擊來源位址，或是一周內攻擊次數超過 10,000 次的攻擊行為，這些都是能夠應用於找到可能存在的安全問題，長時間的針對誘捕系統進行的攻擊，以攻擊的特性而言大多屬於惡意程式所進行的自動化攻擊，偵測到的攻擊紀錄則會以針對相同弱點進行的攻擊行為，後續再配合網路流量的比對，以攻擊來源位址與通訊協定為比對條件，就能夠得到可能遭受相同攻擊來源與行為的受害主機範圍。



圖八：特定目標的長期資料追蹤

不同的攻擊活動在長期資料追蹤上所呈現的曲線圖並不一定相同，因此由不同的攻擊曲線，可以概略的區分出不同的網路的攻擊活動，或是屬於不同的殭屍網路所造成的網路行為，利用其所具備的特性我們可以針對目前收集到的資訊進行分析。

攻擊軌跡的追蹤，除了誘捕系統的廣泛部署之外，從資訊安全管理的角度而言，不同的資安設備亦能夠提供不同的屬性的資訊，例如：入侵偵測與防禦系統或是垃圾郵件的分析平台,都能夠提供不同類型的資訊供後續的分析;誘捕系統可以針對特定的系統、應用服務或是弱點進行模擬,因此可以掌握攻擊者對於這些目標進行的各種活動,例如：針對系統或應用服務弱點的攻擊活動，而入侵偵測與防禦系統，主要能夠透過網路行為特徵比對或行為分析的方式，針對現正進行的網路通訊行為進行分析，若有符合的特性將會發出預警資訊，對於高風險的網路行為亦可以配合主動的防禦機制進行阻擋，以目前殭屍網路的活動而言，仍然有許多的入侵管道為電子郵件，使用者在點選存在惡連結或是隱藏後門程式的附件檔案時,使用者的主機將會感染到惡意程式並且加入殭屍網路,而受到攻擊的遠端遙控，因此在電子郵件的防護上，針對郵件內容的網路連結必須進行分析，將所收集的連結位址，提供給主動分析的平台，模擬開啟該網路連結的環境，並且分析開啟連線後所產生的行為，接著對於有存在附加檔案的郵件，必須將附件檔案利用沙箱測試的技術，進行後續的分析，以瞭解附件檔案是否在存在安全上的顧慮。

網路攻擊的軌跡會呈現在所攻擊的目標系統以及途經的資訊安全設備與網路設備，因此對於完整攻擊軌跡的追蹤與事件發生原因的分析上，必須能夠收集到這些相關系統的資料，以便應用於資訊安全事件發生原因的分析，當收集的資料越多樣化時，除了處理時間上的要求之外，也需要解決在不同的資料屬性之間進行分析比對的機制，早期大多採用關聯式資料庫的作法，不過伴隨的資料的巨量成長趨勢，資料庫的效能已是處理巨量資料時所考量的重要因素。

伍、結論

巨量資料處理在資訊安全的領域，主要需要解決在有限時間內必須處理巨量異質資料的問題，包括誘捕系統的日誌紀錄、資訊安全設備的偵測紀錄以及網路的流量資料，每秒鐘數以萬筆計算的資料量，對於傳統的資訊安全分析機制造成了衝夠，本文以惡意程式分析以及資訊安全維運中心的實務運用的角度切入，提供目前可行方案中針對巨量資料的處理機制，對於分散於不同機制所造生的異質資料進行處理，並且仍然保有資訊安全事件在應變上的重要的時效問題，整合不同的分析機制以呈現事件的影響範圍，並做為後續持續追蹤上的重要資訊。

參考文獻

[1] Dionaea Honeypot, http://dionaea.carnivore.it/
[2] Nepenthes Honeypot, http://sourceforge.net/projects/nepenthes/
[3] Kippo Honeypot, https://code.google.com/p/kippo/
[4] Honeywall Honeynet, https://projects.honeynet.org/honeywall/
[5] Ying-Dar Lin, Chia-Yin Lee, Yu-Sung Wu, Pei-Hsiu Ho, Fu-Yu Wang and Yi-Lang Tsai, "Active versus Passive Malware Collection" in Proceeding of Computer Aware Computing,Volume 47 Number 4, Apr. 2014.
[6] Amun Honeypot, http://amunhoney.sourceforge.net/
[7] Cuckoo Sandbox, http://www.cuckoosandbox.org/
[8] 國家高速網路與計算中心, http://www.nchc.org.tw/
[9] 台灣惡意程式分析網 TWMAN, http://twman.nchc.org.tw/

[作者簡介]

蔡一郎，現職為財團法人國家實驗研究院國家高速網路與計算中心副研究員，亦為成功大學電腦與通訊研究所博士生，並同時擔任 The Honeynet Project Taiwan Chapter Leader 以及 Cloud Security Alliance Taiwan Chapter Founder and Director of Research，參與全球前瞻資訊安全技術之研究，從事資訊專業圖書著作達三十四本、雜誌期刊專欄達八十餘篇，長期投入資訊安全教育之推廣，並執行台灣學術網路(TANet, Taiwan Academic Network)資訊安全監控中心(A-SOC, Academic Security Operation Center)與殭屍網路(Botnet)偵測計畫與國內外多項資訊安全相關研究計畫負責人，亦獲選為台灣雲端安全聯盟理事長、中華民國資料保護協會監事、台灣聯合資訊安全發展協會監事、台灣科技化服務協會理事以及中華民國南部科學園區產學協會理事與監事，擁有專業技術證照達十餘張，深入學術研究、資訊安全實務以及產業趨勢分析。(Website: blog.yilang.org)