以關聯式資料庫分析及預測惡意程式系統

周政緯 劉恒華 鍾曜年 劉奕賢 李忠憲
國立成功大學電機工程學系 / 電腦與通信工程研究所
{johnny_800503, kennyliu, joey54780, dannyliu}@hsnet.ee.ncku.edu.tw,
jsli@mail.ncku.edu.tw

摘要

近年來殭屍網路的猖獗，其惡意程式通常會隨著電子郵件、通訊軟體、電腦系統漏洞等多種方式來入侵電腦，使駭客能夠遠端控制多部受害主機來進行如分散式阻斷服務、金融詐騙等等攻擊行為。傳統上對殭屍網路的研究主要多由本體論等方式著手。本研究則是以建立一個關聯式資料庫系統，將不同惡意程式分析平台的分析報告整合起來，再以資料探勘的方式去找出其對內主機註冊機碼的行為模式，或對外通訊行為的擴展模式，進而去預測出該惡意程式的未來擴展區域。

關鍵詞：殭屍網路、資料探勘、關聯式資料庫

Abstract

Bots usually reside in the e-mails, communication software, computer system vulnerabilities, etc., and they allow hackers to remotely control lots of victims' computer to perform attacks like DDOS, financial fraud, etc. Generally, researches of botnet mainly with ontology, our approach is to build a relational database, which integrate analysis reports from different analysis platforms, and utilize data mining tools to find the feature and pattern of the bots. Then, further predict where the bots will spread.

Keywords: Botnet;Data Mining;Relation Database.

1. 前言

在 21 世紀網路極速成長的社會中，不論是民間企業、政府組織或一般小老百姓，整體營運、生活都被網路所包覆，對於資訊科技的依賴，已成為100%不可缺少的資源，也因此，對於資訊科技的探討也成了重要的課題。在建置方便的資訊基礎設施時，當然以功能性、方便性為重，但卻在另一方面，各種資訊安全在傳輸及使用時的的問題，也慢慢衍生出來，成為反向的另一些問題。如果有安全上認知或背景知識的人都了解，這是一體兩面的，如何在發展便利、快速的情形下，始終保有高度隱匿資料的辦法，亦成為日益發展備受探討的問題。微軟曾在 RSA 歐洲大會上發表的一篇研究報告，顯示美國地區受到殭屍攻擊的數量已高達 220 萬[1]；另一方面，日本也在近期報告提出[2]，日本目前受到名為「Gameover 宙斯」的殭屍病毒攻擊，數量約為 15 萬，但別小看這支殭屍病毒，日本警察廳曾表示，此種病毒在入侵電腦之後，會直接洩漏個人電腦相關密碼，亦同時會有原始殭屍病毒攻擊，也就是受到犯罪著的直接操縱，進行各種網路犯罪活動，這也非僅日本受到此病毒攻擊，相關報到提出，全世界約有一百萬台的電腦也在不知覺中受到這類攻擊，以上兩例皆可看出殭屍網路目前對全球的威脅，已非同小可，日後也必將成為資訊安全重大課題之一。

在眾多網路安全的議題下，我們著重於殭屍網路的研究，是由於近年惡意軟體的成長相當快速，對於許多機密資訊，已構成嚴重威脅，這也引起國內外各方面資訊安全專家更重視這個議題。這也非僅發生在資訊業界、企業組織，甚至是政府單位高度重視，而我們著重於殭屍網路的原因，也是我們認為在這麼多資安問題中，這算是對於安全威脅最重的一環，因為現今網際網路普及、發達之下，從總角之年至耄耋之年都有機會使用網際網路，但大多數使用者對於資訊安全意識相當薄弱，故眾多資訊科技中，又以個人電腦最易於受到病毒攻擊，而攻擊者也抓住這點，在將病毒注入於個人電腦後，利用受到感染的電腦當作跳板，以指數倍散播病毒，猶如細胞分裂般擴散至其他電腦，這樣的攻擊，不僅原先宿主電腦資料會不斷外流，攻擊者也可直接控制其電腦做任何行為，甚至擴散出的病毒，在感染其他電腦後，都可直接或間接受到攻擊者的監控，以及各種盜竊資料，再者，所謂大規模殭屍病毒攻擊，也是攻擊者同時利用各感染電腦，小至幾十台，多至幾百幾千甚至上萬台電腦同時下指定做出不可逆的攻擊反應，或進行分散式阻絕服務攻擊(DDoS)[3][4][5]，讓資訊安全單位也無法回溯抓到攻擊者。這也造成目前網路犯罪率提高，駭客利用不知情的受害者電腦，做出許多盜取帳號密碼、軟體序號，小至個人隱私照片，大至銀行帳戶，各式跨國間的金融犯罪[6]也因應而生，而政府方面則是受到監聽、竊取國家經濟、國防、安全等，各式機密資料。

以企業來說，若駭客利用殭屍網路對企業網站做特定攻擊[7][8]，可能進行分散式阻絕服務攻擊行為[3][4][5]，導致外部使用者無法順利連至網站，短暫或長時間的客戶無法藉由網站與企業公司溝通，甚至，假冒特定公司發出偽造釣魚信件，藉由發信地於其公司內部，騙取客戶高度信任，導致一

般民眾誠實交出個人資料、密碼等，這對企業與客戶都是相當大的傷害，對於商譽以及淺在的商業利益損失都是無法估計的，由此可見殭屍網路攻擊對於網路的服務供應商、各大企業都具有高度的威脅；以個人電腦來說，若個人電腦受到殭屍網路病毒入侵，或許個人電腦資料量不大，內部儲存的資料也不是那麼機密，但是攻擊者卻可以利用受感染的電腦，做為 Agent，猶如控制大批殭屍一般，做出統一性行為，像是同時向目地 server 要取資料，當然幾十個 command 不構成威脅,但是幾十萬上百萬的 command，或許是一般主機成受不了的，當然這些行為對於宿主都不會產生過多傷害，只是使自己電腦淪為犯罪著的工具，這應該是所有人都不願意見到的吧。

上述已經提到許多殭屍網路對於資訊安全的影響已極嚴重威脅，對於惡意程式問題，已經有許多業界開始著重於這部分，政府部分也針對這方面有持續研究的進展。不過目前公布的研究多半於將重點放在阻斷以及抓出殭屍網路病毒，也就是針對已經受到攻擊或者有被攻擊跡象的電腦做出防禦的方式，當然也有少部分的研究在方向瞄準以回溯方式，導出攻擊者電腦的位置，但直白的說，以目前多數研究來看，是對於直接性防禦的內容是相當不足的。有鑑於此，本研究的目的為建立一個結構化惡意程式分析平台的完整資料庫，運用第三方惡意程式分析系統，將其分析平台報告開發出相關轉換套件，並將取得資料內容做關聯式資料庫[9]數據，以供給後續相關研究可順利運用關聯式資料探勘工具，順利將此研究內容順利延續，進而獲得更大效益。

2. 文獻探討

此研究利用惡意程式分析平台對殭屍網路惡意程式做蒐集以進行分析，再進而萃取各分析報告中重要的資料，加以彙整成此次分析的報告，最後利用結構化的關聯式資料庫存取分析後的資料，故本研究的架構建立在惡意程式平台、惡意程式資料庫。

2.1 惡意程式類型

一般惡意程式及電腦病毒雖有許多不通之處，但統一視為傷害電腦會產生的惡意行為。主要又可以區分為病毒、特洛伊木馬、蠕蟲[10]。以病毒來說，主要攻擊為竊取資料、毀損檔案或更嚴重的格式化受害者電腦,或消耗記憶體佔據 CPU 使用率，使其電腦效能低落或無法進行任何行為，大大影響受害者電腦。特洛伊木馬式的攻擊，則是受害者電腦被攻擊之後，像在產生後門之類的方式於宿主電腦，以洩出受害電腦內部資料。蠕蟲，最大威脅部份就是自我複製，受害主機會自動複製垃圾檔案，可能塞爆原始受害主機或自動發出至其他電腦。更進一步來說，依照趨勢科技[11]將惡一程式分類為:惡意軟體（Malware）、病毒（Virus）、蠕蟲（Worm）、木馬（Trojan Horse）、垃圾郵件（Spam）、網路釣魚（Phishing）、網路釣魚（Phishing）、網址嫁接（Pharming）、間諜軟體（Spyware）、廣告軟體（Adware）、殭屍電腦和殭屍網路（Bot 和 Botnet）、勒索軟體（Ransomware）。

2.2 殭屍網路

殭屍網路算是傳統惡意程式進化組合後的惡意攻擊方式，他利用特洛伊木馬後門方式進入電腦，又可透過網路蠕蟲方式散播至其他宿主電腦，以近年來說，算是極為強大的攻擊方式，加在上目前網路普及率極高，以及各式使用者對於資訊安全意識薄弱，也大大增加駭客攻擊成功率。

以下可稍為介紹殭屍網路在攻擊時，整體的分析分式，大略可分為攻擊者(Botmaster)、C&C Server、受害者電腦(bot)。其中攻擊者會利用 C&C Server 做為跳板入侵受害者電腦，又因受害者不知自己電腦早已被入侵，在不知情情況下也成為犯罪工具，故稱受害電腦受到殭屍網路攻擊[12]。

殭屍網路的組成可由下列幾項說明:

1. Botmaster: 指攻擊者，主要為下令惡意程式，利用各 C&C Server 入侵受害者電腦。
2. Command and Control (C&C) Server: 大多為跳板 SERVER，攻擊者利用這些 SERVER，入侵受害者電腦，傳遞指令制 C&C 後，藉由 C&C 執行，也因此造成難以抓出攻擊者的原因。不過殭屍網路的拓樸非常多元，C&C Server 也少部份由 Botmaster 扮演。
3. Bot: 受到殭屍網路攻擊的電腦，一般為了維持宿主的不知情,通常會盡量不在宿主電腦上做破壞，也因此大多數受害電腦皆不知情。

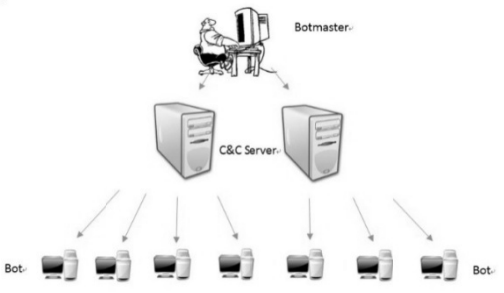

圖1. 殭屍網路組成

圖 1 說明一般殭屍網路組成型式。此研究會先將惡意程式定義清楚後，將網羅到的惡意程式送至惡意程式分析平台，並將結果做關聯式分析，再存入關聯式資料庫，在進行最後一項 Data Mining[13][14]的動作，並且將預測功能實踐出來。

2.3 惡意程式分析平台

此研究為達到預測惡意程式，藉由第三方惡意程式分析平台[15][16][17][18][19]做分析結果，並整合成為殭屍網路結構化資料庫。故此研究透過國家實驗研究院高速網路與計算中心(NCHC)及國立成功大學資通安全研究與教學中心(TWISC@NCKU)及 International Secure System Lab 的 Anubis，分析平台分別為:TWMAN、Cuckoo、Malbed、CWsandbox、Anubis，其中 TWMAN、Cuckoo 及 CWSandbox 為 NCHC 所提供，Malbed、CWSandbox 為 TWISC-NCKU 所提供。

- TWMAN: 臺灣惡意程式分析網，英文名稱為 Taiwan Malware Analysis Net。是一套為安全而生的分析平台，能夠檢測很多惡意程式以及系統網路，並利用自動收集方式分析收到 data，已提供資安人員研究參考。
- CWSandbox:不需建立環境極可直接自動分析惡意程式，可依需求決定留下哪些惡意程式執行分析檔案，存放於資料庫，以供查詢。
- Malbed:Malbed 會將惡意程式做分解，並分析記錄更動的檔案、註冊表單等，記錄 IP、PORT、等資料後自行分析做觀察，並了解其惡意程式行為和發展的趨勢。

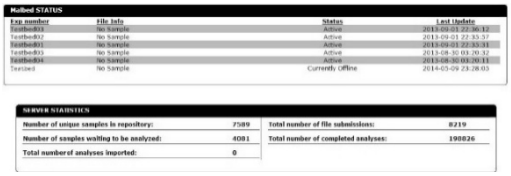

圖2. Malbed 分析平台畫面

- Anubis: 在虛擬環境下做二進位文件分析，並即時觀察分析結果，此分析評在最重要的是著重於安全相關方面程式變動，利用線上沙盒分析惡意程式，並且提供使用者直接上傳方式，使更多人直接使用。

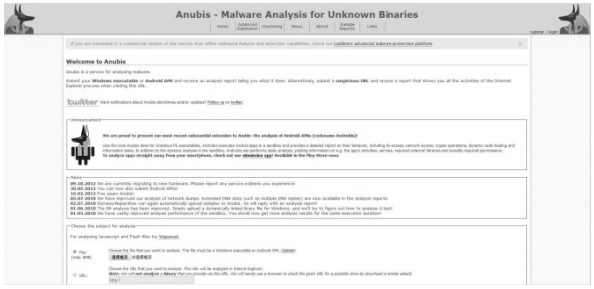

圖3. Anubis 分析平台內容顯示

3. 系統架構

本研究所提出之結構化殭屍網路資料庫系統(SBDS)，主要分為殭屍網路資料分析模組(BDAM)、殭屍網路資料展示模組(BDPM)、殭屍網路資料轉換模組(BDTM)、殭屍網路報告擷取模組(BRGM)和殭屍網路資料整合模組(BDIM)等五大元件。其中殭屍網路資料轉換模組(BDTM)將殭屍網路報告擷取模組(BRGM)和殭屍網路資料整合模組(BDIM)的功能整合並取代。本研究系統架構圖如下所示：

圖4. 結構化殭屍網路資料庫系統架構

以下針對本研究所提出之結構化殭屍網路資料庫系統(SBDS)之五大模組分別進行說明：

1. **殭屍網路資料展示模組(BDPM)**

該模組主要功能在於將結構化殭屍網路資料庫系統(SBDS)中，依使用者要求條件去取得相關資料，再透過網頁應用程式的方式加以顯示，可以提供使用者查詢相關的統計資訊或各惡意程式的分析結果及下載原始報告等功能。

圖5. BDPM 以網頁呈現使用者要求的資料

2. **殭屍網路資料批次轉換模組(BDTM)**

該模組主要功能提供批次轉換第三方資料內容，以避免大量傳送樣本請求分析報告時，造成第三方分析平台效能瓶頸，因應提供第三方分析平台的 TWISC@NCKU、NCHC 等合作單位要求，本年度增加此一殭屍網路資料庫批次轉換模組(BDTM)，至第三方資料庫擷取出本系統的結構化殭屍網路資料庫系統(SBDS)需要進行分析的欄位，並加以排序整理，以供其他模組進行資料統計與呈現等等。在今年度中，殭屍網路報告擷取模組(BRGM)和殭屍網

路資料整合模組(BDIM)的功能整合於此模組中。

3. **殭屍網路報告擷取模組(BRGM)**
該模組主要功能在於負責將惡意程式樣本上傳至其他的惡意程式分析軟體進行相關分析，再取回其分析報告，以作為本研究後續作業的主要資料來源。透過惡意程式分析軟體的相關分析的結果，可令使用者更加清楚特定程式的意圖與行為。在此一模組中，將以批次及事件驅動兩種管道來觸發該模組，以執行其負責之分析報告擷取的工作。

4. **殭屍網路資料整合模組(BDIM)**
該模組主要功能在於負責整合來自於殭屍網路報告擷取模組(BRGM)所擷取的惡意程式分析報告，並加以處理轉換，最終輸入本研究所建置的結構化殭屍網路資料庫系統(SBDS)中。本研究將依惡意程式分析報告的資料項目及相關文獻回顧，訂定結構化殭屍網路資料庫的資料綱要，利用實體關係模型、資料型別和條件約束將其行為描述進行一個有系統化的整理架構，以提供相關分析數據整合的依據，同時保留原有資料，以供後續進一步比對的需求。

5. **殭屍網路資料分析模組(BDAM)**
該模組主要功能在於將 BDTM 所輸入的資料庫系統(SBDS)進行資料探勘(Data Mining) 並以 FP-Growth 演算法[20]分析關聯性資料庫的海量資料，隨後取得殭屍網路重要的關鍵性欄位，並做惡意程式未來的擴展情形與風險評估(Attack Map &Predictive list)。

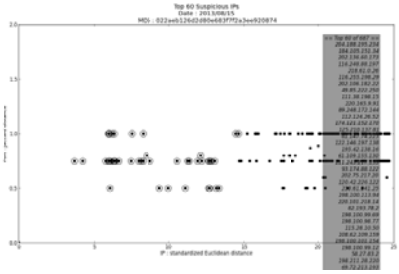

圖6. 預測清單(Predictive List)

我們以惡意程式的行為特徵去預測未來可能感染此惡意程式的 IP，根據我們的預測方法，發現取行為相似度排行前 60 名的 IP 準確率為最高。預測清單(Predictive List)的預測被感染 IP，將其地理位置以地圖的方式標記出來呈現，以清楚看出該惡意程式的未來擴展區域與方向為何。

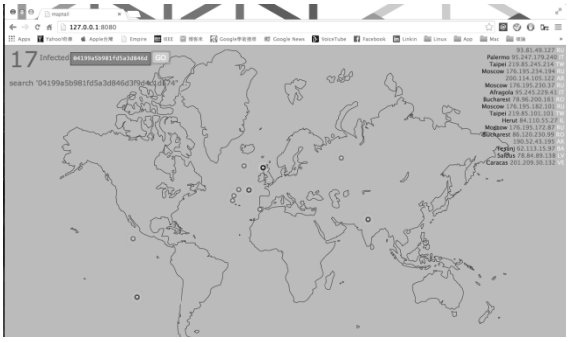

圖7. 攻擊預測地圖(Attack Map)

本研究將經由前兩年所建置的四大模組為基礎，於今年度中配合前二年執行現況，增加殭屍網路資料庫批次轉換模組(BDTM)，同時強化殭屍網路報告分析模組(BDAM)，並擴充惡意程式的樣本數量，以使本計畫所建置的結構化殭屍網路資料庫系統(SBDS)更加完整。進而在結構化殭屍網路資料庫系統(SBDS)的基礎上，利用此一資料樣本進行分析與分類等相關作業和評估，以進一步提升殭屍網路相關資訊的交換及新型殭屍網路識別率，並提供系統安全的風險評估機制，以提供企業組織做為安全防護的依據。結構化殭屍網路資料庫系統(SBDS)系統環境圖如下所示：

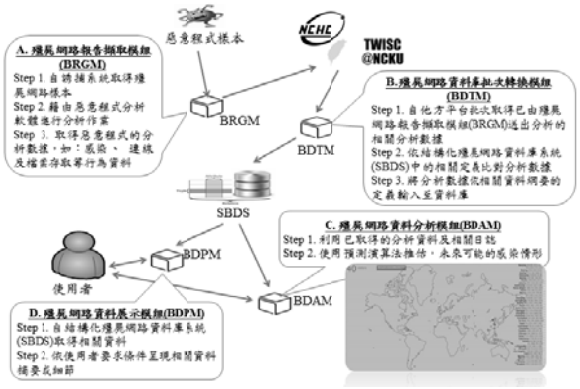

圖8. 結構化殭屍網路資料庫系統環境

本研究以關聯式資料庫系統為存儲所有分析報告的載具。在資料綱要及資料庫架構的設計上，本研究以正規化模型為依據，依序進行第一正規化、第二正規化及第三正規化三項程序，確認資料中沒有重複值組、單一筆資料中沒有多重屬性、資料庫中可避免資料重覆儲存及更新可能造成資料不一致等問題。本研究設計之結構化惡意程式資料庫[21]的實體關係圖(ERD)如下所示：

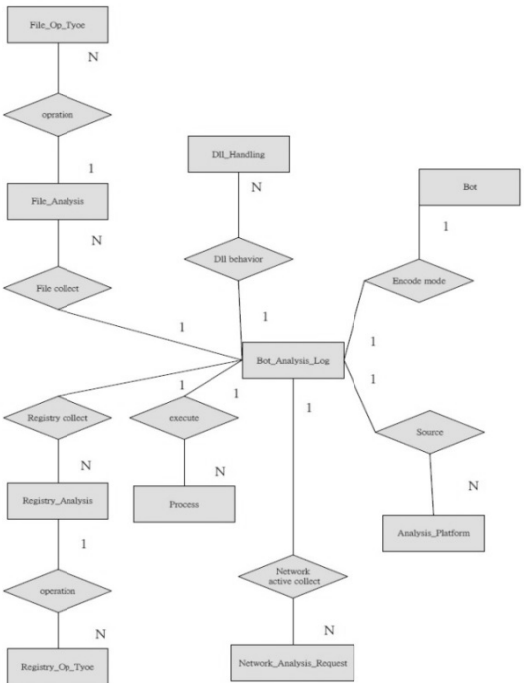

圖9. 結構化殭屍網路資料庫系統環境

4. 結論

本研究有鑑於關聯式資料庫系統中資料探勘工具的成熟，唯當前相關研究主要是運用語意網、本體論等方式，而鮮少運用在商業資料分析中較為成熟的關聯式資料探勘工具。探究其原因，產業界中主要的惡意程式分析的相關平台，其分析報告主要原始輸出格式即為延伸標記語言(XML)或網頁格式(HTML)，甚至為純文字格式(txt)，研究者因此受限於資料來源及轉換處理不易等限制所致。有鑑於此，本研究主要目的為建立一結構化惡意程式資料庫系統，運用第三方惡意程式分析平台的分析結果，開發相關轉換套件，將相關分析報告，轉換成為關聯式資料庫數據內容，再以資料探勘的方式去找出惡意程式的特徵以及行為模式，進而去預測出該惡意程式的未來擴展區域。

誌謝

感謝科技部計畫 MOST 100-2218-E-006 -029 -MY3、MOST 103-2221-E-006 -146 -MY3 及經濟部計畫 102-EC-17-A-02-S1-222 提供經費支持本研究的進行。感謝國家實驗研究院高速網路與計算中心與國立成功大學資通安全教學研究中心(TWISC@NCKU)提供樣本資料。

參考文獻

[1] Adrienne Hall, “In Pursuit of Cyber Crime,” RSA 2010, London, UK, Oct. 12-14, 2010.

[2] 資安人, “借鏡日本JPCERT經驗打擊殭屍網路,” Retrieved 2014/07/20 fromhttp://www.informationsecurity.com.tw/article/article_detail.aspx?aid=6081#ixzz38Ip3xck7

[3] Felix C. Freiling, Thorsten Holz & Georg Wicherski, “Botnet Tracking: Exploring a Root-Cause Methodology to Prevent Distributed Denial-of-Service Attacks,” ESORICS '05, Milan, Italy, Sep. 12-14, 2005.

[4] Guofei Gu, Junjie Zhang & Wenke Lee,2008, “BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic,” NDSS'08, San Diego, USA, Feb. 8-11, 2008.

[5] Alomari, Esraa, Gupta B. B. & Karuppayah Shankar,“Botnet-based distributed denial of service (DDoS) attacks on web servers: classification and art,”International Journal of Computer Applications, Vol. 49, No. 7, pp.24-32,2012.

[6] Chandrasekaran, Balakrishnan, John R. Josephson& V. Richard Benjamins, “What are ontologies, and why do we need them?,” IEEE Intelligent Systems and their Applications, Vol. 14, No. 1, pp.20-26, 1999.

[7] Richardson Robert,2008 CSI computer crime and security survey, Computer Security Institute, 2008.

[8] Guofei Gu, Junjie Zhang & Wenke Lee, “Bot Sniffer: Detecting Botnet Command and Control Channels in Network Traffic,” NDSS'08, San Diego, USA, Feb. 8-11, 2008.

[9] DavidMaier,Theory of Relational Databases, Computer science press, 1983.

[10] Hoffman Lance J,Rogue programs: viruses, worms and Trojan horses, Van Nostrand Reinhold Co., 1990.

[11] Trend Labs, “認識惡意軟體,” Retrieved 2014/07/20 fromhttp://blog.trendmicro.com.tw/?p=5795

[12] Lei Zhang, Shui Yu, Di Wu & Watters P., “A Survey on Latest Botnet Attack and Defense,” 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Changsha, China, Nov. 16-18, 2011.

[13] Bin, Shen, Liu Yuan& Wang Xiaoyi, “Research on data mining models for the internet of things,” 2010 IEEE International Conference on Image Analysis and Signal Processing (IASP), Zhejiang, China, Apr. 9-11, 2010.

[14] Han, Jiawei& Micheline Kamber, Data Mining, Southeast Asia Edition: Concepts and Techniques, Morgan kaufmann, 2006.

[15] Cuckoo, “Cuckoo Sandbox,” Retrieved 2014/07/20 from http://www.cuckoobox.org.

[16] NCHC, “Clonezilla 再生龍,” Retrieved 2014/07/20 from http://clonezilla.nchc.org.tw.

[17] TWISC@NCKU, “Malbed,” Retrieved 2014/07/20 from http://malbed.twisc.ncku.edu.tw.

[18] NCHC, “臺灣惡意程式分析網,” Retrieved 2014/07/20 from http://twman.nchc.org.tw.

[19] University of Erlangen-Nuremberg, “Malware Analysis System, CWSandbox: Behavior-based Malware Analysis,” Retrieved 2014/07/20 from http://mwanalysis.org/.

[20] Borgelt Christian, Siegfried Nijssen & Mohamed J. Zaki,“An Implementation of the FP-growth Algorithm,”KDD 2005,Chicago, USA, Aug. 21, 2005.

[21] 劉奕賢, 蔡舜智,江啟賓,張家瑋,安家駒,李忠憲, “結構化惡意程式資料庫系統,” TANET2013臺灣網際網路研討會, 台中, 臺灣, Oct. 23-25, 2013.