結構化惡意程式資料庫系統

劉奕賢　蔡舜智　江啟賓　張家瑋　安家駒　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
{dannyliu, sctsai, takumi, william, jjan}@hsnet.ee.ncku.edu.tw, jsli@mail.ncku.edu.tw

摘要

近年來惡意軟體急遽增加，特別是殭屍網路，對於資訊安全的危害甚鉅。由於駭客工具快速的發展與大量散播，加上經濟的誘因及網路犯罪防治的困難等因素。殭屍網路成為駭客用來作為分散式阻絕服務攻擊、發送垃圾郵件等犯罪工具。傳統上對殭屍網路的研究主要由本體論等方式著手，本研究有鑑於關聯式資料庫系統中資料探勘工具的成熟，故著手建置以關聯式資料庫系統為基礎的結構化惡意程式資料庫系統，以做為後續研究運用該類型工具的基礎。

關鍵詞：殭屍網路、資料探勘、關聯式資料庫、分散式阻絕服務攻擊、網路安全

Abstract

Recently, number of malwares surges dramatically, especially the botnet. It becomes a vast challenge to network security. Due to rapidly developed and massive spreading hacking tools and financial inducement, network crimes increase tremendously. Botnet is a powerful criminal tool for DDOS or SPAM. In contrast to the research of botnet mainly with ontology, this research is based on data mining technology with relational database.

Keywords: Botnet; Data Mining; Relation Database; DDOS; Network Security

1. 前言

不論是民間企業組織或政府單位，日常營運中對於資訊科技依賴程度與日俱增，資訊科技已是不可或缺的重要基礎建設之一。在建置資訊基礎設施的同時，主要以功能性及便利性為主要考慮，然而在個資法施行後，對機敏性資料及個人識別資訊的防護及管理，已成為另一個考量的重點。功能性、便利性的訴求，與安全性考量相對來說，是不同的角度，如何保護其資料的安全，並達到機密性、完整性與可用性的要求，已成為企業組織與政府單位的重要問題[17][18]。微軟[1]曾在 RSA 歐洲大會上發表的一篇研究報告，更顯示美國地區殭屍網路數量以 220 萬部居冠，由此可見其猖獗的程度。

近年惡意軟體的快速增長，對資訊安全的機密性、完整性與可用性已產生嚴重的威脅。這種趨勢已引起國外資訊界及企業組織、資訊業界與政府單位的重視。在眾多的惡意程式中，尤以殭屍網路(botnet)帶來的安全威脅最為嚴重。當前因網路連線型態的改變，撥接上網造成連線持續不一定的問題，在當前 ADSL、FTTB 等網路接取服務普及下，持續連網已成為網路上的常態。大多數的使用者資訊安全意識薄弱，故個人電腦容易淪為感染殭屍網路的受害者。新一代的惡意程式與傳統的電腦病毒最大的差異在於對宿主的行為。傳統上電腦病毒以破壞、干擾電腦運作、造成資料損壞為主要目的，使用者極易察覺其電腦運作不正常，而懷疑自身中毒；相較之下，殭屍網路以運用宿主做為跳板、竊取宿主的機敏資料為主要目的，在日常操作中，使用者極難察覺異常，進而讓殭屍網路這種新一代的惡意程式潛伏其中。加上經濟上的誘因及網路犯罪防治的困難等多種助因，促使駭客工具快速的發展與散播。殭屍網路慢慢成為駭客用來作為竊取帳號密碼、軟體序號或其他機敏資料的工具，甚至成為進行分散式阻絕服務攻擊(DDoS)[5][6]、發送垃圾郵件(SPAM)的跳板主機。

假若駭客運用殭屍網路對某特定企業網站進行分散式阻絕服務的攻擊行為，導致一般正常的使用者無法順利瀏覽該公司網站，或假冒某公司發出釣魚郵件進行詐騙犯罪，這種行為對商譽造成的影響，其潛在商業損失是難以估計的。由此可見殭屍網路對網路的服務供應商、企業形象與廣告商等產生的威脅及影響。殭屍網路主要是由個人電腦所組成的網路大軍，許多人由於電腦防護不足或缺少基本的資安觀念，往往在不知情的情況下成為殭屍網路的一員。在經濟利誘或網路軍備競賽的引導下，駭客們可以利用大量的殭屍電腦進行垃圾郵件發送或分散式阻絕服務攻擊等的惡意行為[5]，而這些非法行為透過殭屍電腦的隱蔽性和自我摧毀等功能，是很不容易被發現的[6]。從微軟的安全研究報告提到目前的殭屍網路大多用來進行網路金融犯罪，手段包含發送垃圾郵件，釣魚網頁，竊取個人機密資料等等。且自從網域快速轉換對應(fast-flux)技術出現之後，殭屍網路更加難以偵測，因為殭屍網路透過網域名稱系統(DNS)快速變換對應的網路位址(IP)，進而導致傳統以鎖定特定網路位址的防治措施無法奏效，甚致會對無辜使用者造成傷害。

有鑑於殭屍網路對資訊安全的嚴重影響及威脅，且產業界中不乏針對惡意程式分析的相關平台。唯當前相關研究主要是運用語意網、本體論等方式，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，主要是惡意程式的分析報告，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。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，以供後續相關研究得以運用關聯式資料探勘工具，進而試圖從中獲得更大的效益。

2. 文獻探討

本研究將運用惡意程式分析平台對殭屍網路惡意程式的範本進行分析，進而萃取其中重要的資料加以收集彙整，再以結構化的關聯式資料庫方式存取相關分析資料，故本研究將就惡意程式類型、殭屍網路原理及架構與惡意程式分析平台等進行相關探討，以做為本研究結構化惡意程式資料庫系統設計的基礎。

2.1 惡意程式類型

傳統上，惡意程式又常被稱為電腦病毒。早期依惡意程式行為的不同，主要可區別為病毒、特洛依木馬及網路蠕蟲三種[13]。其中病毒這種程式，具有自我複製能力。可能會對損毀檔案、格式化磁碟而造成損害。或消耗記憶體造成電腦效能低落，影響使用者正常作業等。特洛依木馬則不具備複製能力，以竊取受害者的資料為主要目的。網路蠕蟲則以其可透過網路連線、電子郵件、檔案分析或即時通訊等各種通訊管道，快速進行自我複製來傳播而得名。從趨勢科技的定義[14]來說，目前常見的電腦安全威脅主要分為垃圾郵件、間諜程式和廣告軟體、網路釣魚、病毒與其他惡意程式等六大類型。新一代的惡意程式，又稱為進階持續性滲透攻擊(APT)[12]其結合傳統的網路蠕蟲、特洛伊木馬及其他類型的相關技術，如 Rootkit 等，於後門工具的基礎上強化，並透過相互融合而發展成為目前最為複雜的攻擊模式之一。

當前常見的惡意程式傳播方式，有下列常見幾項[9]:

1. 主動式漏洞攻擊：
   攻擊者會先搜尋一般電腦作業系統或應用程式的缺陷，利用該漏洞，植入惡意程式。此一手法，又以零日攻擊(Zero-day Attack)[3]最為精典，即運用缺陷被發現後，相關廠商完成修補的時間差，對相關系統進行攻擊。
2. 跨網站腳本攻擊(XSS)：
   跨網站腳本攻擊(XSS)[8]即攻擊者在提供網頁服務的站台中，於特定超文件標記網頁中(HTML)嵌入惡意的程式腳本，當訪問者訪問這些網站時，使用者會被重新導向到惡意網站而執行惡意腳本，使得惡意程式被下載到主機上，並被自動執行。
3. 郵件病毒：
   許多常見的電腦病毒及其他惡意軟體係以電子郵件附件 (即與郵件訊息一同傳遞的檔案) 的形式散佈。如果附加至郵件訊息的檔案內含病毒。輔以社交工程的技巧，誘使不知情的使用者執行附件，而感染該惡意程式。
4. 網路蠕蟲：
   是一種電腦程式碼，會將自身附加到程式或檔案，在電腦之間散佈，同時感染途經的電腦。病毒會把自身附於母體程式，然後嘗試感染其他電腦。
5. 即時通訊軟體與社群網路：
   此一管道即運用 Facebook 等好友清單及個人訊息的功能，配合社交工程手法，誘使不知情的受害者好友，點選連結或執行程式而感染之。

2.2 殭屍網路

殭屍網路是從傳統惡意代碼形態包括網路蠕蟲、特洛伊木馬和後門工具的基礎上進化，並透過相互融合發展而成的目前最為複雜的攻擊模式之一。近年來隨著網路連線型態的改變，持續連網的連線服務如 ADSL、FTTB 等服務普及與連線速率的提昇，加上使用者缺乏危機意識，大量的個人電腦容易淪為受到殭屍網路感染的受害者，進行成為殭屍網路危害資訊安全的代罪羔羊。

殭屍網路是駭客出於惡意，大量傳播殭屍程式來感染並控制受害電腦，以形成一個連線的網路群體，駭客可以透過一對多的控制命令來指示受害電腦從事特定的行為。因受害電腦並不知情，故稱殭屍網路。殭屍網路主要由下列三大元件所組成：

- Master： 為駭客本身，主要目的為下達惡意程式指令、更新惡意程式碼給殭屍網路成員(bot)。
- Command and Control (C&C) Server： 有時會由 Master 本身來扮演這一個角色，其主要負責管理控制整個殭屍網路的伺服器，並將 Master 的指令傳遞給 bot。殭屍網路的通訊拓樸十分多元，包括星狀、階層式等多種不同的架構，且 C&C Server 也會經常變動以逃避追蹤及反制。
- Bot： 一群無辜受到殭屍網路程式感染而被遙控的受害者電腦。殭屍網路程式有別於傳統以破壞電腦正常運作為目的的電腦病毒不同，其反而不希望讓受害者察覺本身遭受感染，而成為殭屍網路的

成員，在依照 Master 的指令於特定時間進行特定的工作。

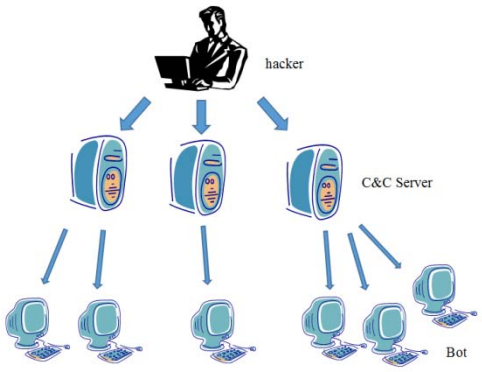

圖1. 殭屍網路的組成元件

故本研究亦會結合多種的惡意程式分析平台的分析結果，以取得殭屍網路在不同分析平台硬體上執行時的相關數據。並記錄如網路通訊行為、主機檔案存取行為等相關的系統操作情況。

2.3 惡意程式分析平台

本研究主要目的，即整合第三方惡意程式分析平台的分析結果，並整合成為一結構化的殭屍網路資料庫。故本研究主要將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 平台[7]等第三方專業資訊安全研究單位所提供的分析平台進行相關惡意程式範本的分析作業，以下將介其所擁有的惡意程式分析平台特色。

- 高速網路與計算中心(NCHC)

國家實驗研究院下的高速網路與計算中心是國內重要的資訊安全研究單位之一，其主要具備了 TWMAN(抬丸郎)、Cuckoo 及 CWSandbox 等平台，並結合其自組開發提供的自由軟體 Clonezilla(再生龍)系統還原套件來強化用用實機作為分析平台硬體的能力。以下分別以這三大平台的特色進行介紹：

1. TWMAN(抬丸郎)[11]：由於現今的一些殭屍病毒已具備偵測當下環境是否為虛擬化，且具備偵測是否有沙盒在監視，為了解決此一問題 TWMAN 利用實際的電腦進行感染，執行惡意程式進行分析，並利用 Clonezilla[10]快速恢復的能力將受感染的電腦還原。此外更可進一步針對惡意程式樣本進行切片比對，與資料庫中已有的惡意程式進行相關的比對，找出變種或加殼的新型惡意程式。

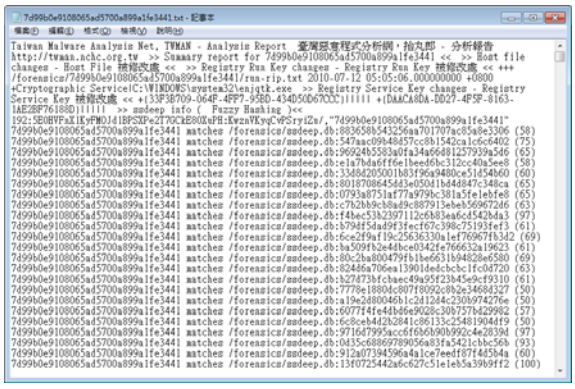

圖2. TWMAN 的惡意程式分析報告

2. Cuckoo[4]： 此一平台為二進制行為自動動態分析的工具，能夠呈現程序運行中詳細的關鍵應用程式介面和網路活動，也可以處理分析 PDF、office 和其他微軟的文件，在執行惡意程式時產生網路流量的儲存，並在分析的同時拍攝畫面和呈現程序運行中詳細的相關應用程式介面。此外該平台因有發現部分不同的惡意程式具有相同的 MD5 值，而無法區分，進一步加入 SHA1 及 SHA256 兩種的識別方式。

圖3. Cuckoo 的惡意程式分析報告

3. CWSandbox[16]：CWSandbox 以自動分析惡意程式為基礎，在虛擬的環境且有限的時間下，密切監控其所有的系統呼叫，並能夠自動生成一份詳細的報告，大大的簡化了惡意程式分析師的工作。其強大的能力可以讓使用者分析所有的微軟的應用程式介面其中包括感染的：office 文件、PDF 檔、惡意的 URL 和 Flash 廣告。

圖4. CWSandbox 的惡意程式分析報告

- 國立成功大學資通安全研究與教學中心(TWISC@NCKU)

  國立成功大學資通安全研究與教學中心主要具備 Malbed 及 CWSandbox 兩大分析平台，其特色在於分析作業並非是於虛擬的環境下進行，而是運用美國 Utah 大學所授權的 Emulab 軟體配合實體的系統來架構分析所需的環境。：

  1. Malbed[15]：以自動化惡意程式分析為基準，並利用自行開發、整合之 Malbed 進行分析記錄更動的檔案、註冊表以及在惡意程式在執行時所做的系統呼叫，在網路方面將來源端和目的端的 IP 位置、區域名及連接埠記錄下來，以觀察分析中的惡意程式，是否有對外連接的行為，並藉由上述的分析了解惡意程式發展的趨勢。

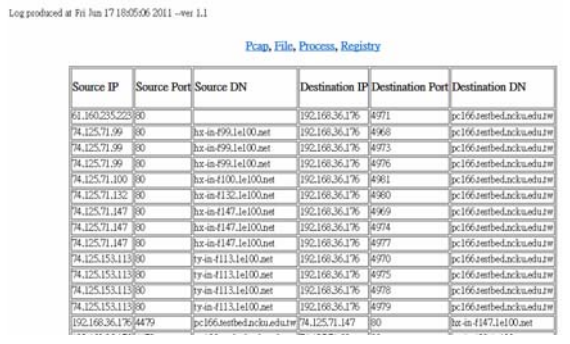

圖5. Malbed 的惡意程式分析報告

  2. CWSandbox：因 CWSandbox 與國網使用平台相同，唯分析的硬體系統改以 Testbed 方式進行分析而有所差異，平台本身特色就不再加以贅述。

- Anubis 惡意程式分析分台

Anubis[7]主要是在虛擬的環境下進行二進位文件的分析，並可及時觀察其分析和執行，分析的重點主要著重在安全相關方面程式的變動，也因為將範圍縮小所以可以得到更精確的結果，也讓分析程序更加簡單，藉由線上的沙盒執行與測試惡意程式，並記錄其執行時的系統呼叫與檔案和註冊表的更動藉以判斷此一惡意程式的目的。此外 Anubis 亦有提供介面供第三方上傳惡意程式樣本以進行分析，並於分析後提供相關數據以供研究參考。

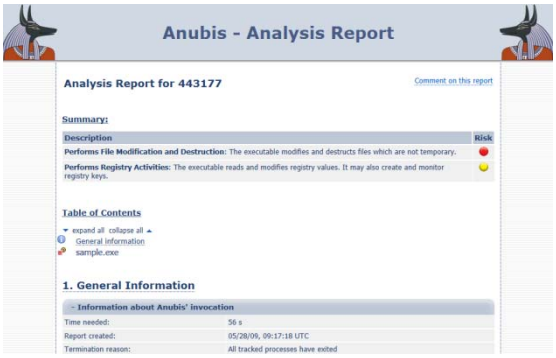

圖6. Anubis 的惡意程式分析報告

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)，主要分為殭屍網路報告擷取模組(BRGM)、殭屍網路資料整合模組(BDIM)、殭屍網路資料展示模組(BDPM)及殭屍網路資料庫(BotDB)等四大元件。本研究系統架構圖如下所：

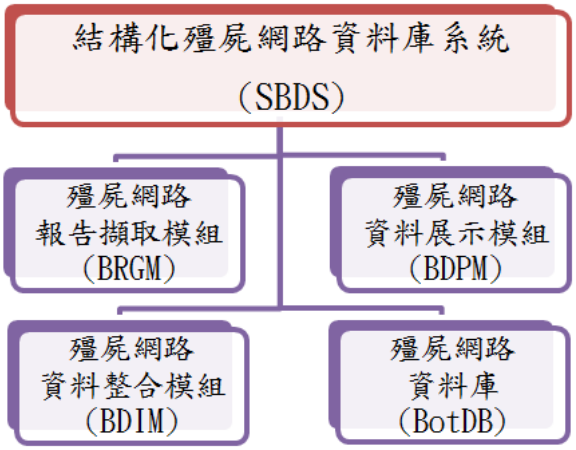

圖7. 結構化殭屍網路資料庫系統架構

以下針對本研究所提出之結構化殭屍網路資料庫系統(SBDS)之四大模組分別進行說明：

1. 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。

2. 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。

3. 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)中的相關資料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。

4. 殭屍網路資料庫(BotDB)
該資料庫即本系統的主要核心，負責儲存所有惡意程式的分析報告內容。運用關聯式資料庫的設計規範加以實作。並輔以正規化的相關流程，確保資料儲存的一致性及避免重覆儲存的問題。

結構化殭屍網路資料庫系統(SBDS)系統環境圖如下所示：

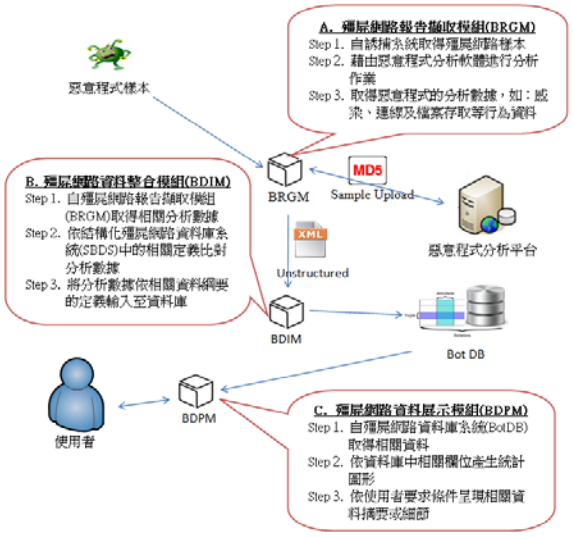

圖8. 結構化殭屍網路資料庫系統環境

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[2]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：

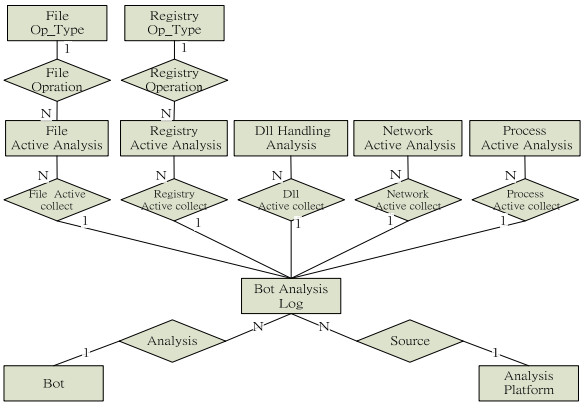

圖9. 結構化殭屍網路資料庫系統環境

本研究的設計中，以 Bot Analysis Log 這個資料表為紀錄每份惡意程式分析報告的核心，再由其與 Bot 資料表的關聯，來識別該份報告屬於每一隻惡意程式；並透過其與 Analysis Platform 的關聯說明該分析報告是由那一個惡意程式分析平台所產生的。此外惡意程式分析報告中會有關於執行緒(Process)、網路活動(Network Active)、動態連結函式庫(Dll)、註冊機碼(Registry)及檔案操作(File)等相關行為，再分別以關聯的型式，分別存放至對應的資料表中。同時在註冊機碼(Registry)及檔案操作(File)這兩種活動中，再以其操作行為類型，各別加以關聯。以此設計為藍圖，再結合關聯式資料庫管理系統加以實作這結構化惡意程式資料庫。

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，以供後續相關研究得以運用關聯式資料探勘工具，進而試圖從中獲得更大的效益。

誌謝

感謝國科會計畫 NSC100-2218-E-006-029-MY3 提供經費支持本研究的進行。

參考文獻

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

[2] Catriel Beeri, Philip A. Bernstein & Nathan Goodman, 1978, “A sophisticate's introduction to database normalization theory,” Proceeding VLDB '78 Proceedings of the fourth international conference on Very Large Data Bases, Vol 4, pp. 113-124.

[3] Constantin Musca, Emma Mirica adn Razvan Deaconescu, 2013, “Detecting and Analyzing Zero-Day Attacks Using Honeypots,” 2013 19th International Conference on Control Systems and Computer Science (CSCS 2013), May 29-31, Bucharest, Romania

[4] Cuckoo, “Cuckoo Sandbox,” Retrieved 2012/03/15 from http://www.cuckoobox.org.

[5] 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.

[6] 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.

[7] International Secure Systems Lab, “Anubis: Analyzing Unknown Binaries,” Retrieved 2011/10/15 from http://anubis.iseclab.org.

[8] Lwin Khin Shar & Hee Beng Kuan Tan, 2013, “Predicting SQL injection and cross site scripting vulnerabilities through mining input sanitization patterns,” Information and Software Technology, Vol. 55, Iss. 10, pp. 1767-1780.

[9] Mentor Wang, 2008, “Botnet 介紹,” , Retrieved 2010/11/26 from http://mentorwang.blogspot.com/2008/07/botnet.html.

[10] NCHC, “Clonezilla 再生龍,” Retrieved 2011/12/15 from http://clonezilla.nchc.org.tw.

[11] NCHC, “臺灣惡意程式分析網,” Retrieved 2011/12/10 from http://twman.nchc.org.tw.

[12] Sood A.K. & Enbody R.J., 2013, “Targeted Cyberattacks: A Superset of Advanced Persistent Threats,” IEEE Security & Privacy, Vol. 11, Iss. 1, pp. 54-61.

[13] Trend Micro, “病毒/特洛依木馬程式/網路蠕蟲網頁,” Retrieved 2013/08/13 from https://imperia.trendmicro-europe.com/tw/threats/home-user/common-threats/viruses/index.html.

[14] Trend Micro, “常見的電腦安全威脅,” Retrieved 2013/08/13 from https://imperia.trendmicro-europe.com/tw/threats/home-user/common-threats/index.html.

[15] TWISC@NCKU, “Malbed,” Retrieved 2012/03/30 from http://malbed.twisc.ncku.edu.tw.

[16] University of Erlangen-Nuremberg, “Malware Analysis System, CWSandbox: Behavior-based Malware Analysis,” Retrieved 2012/01/05 from http://mwanalysis.org/.

[17] Von Solms B., 2006, “Information security - The fourth wave,” Computers & Security, Vol. 25, No. 3, pp.165-168.

[18] Yeh, Q. J. & Chang, A. J. T., 2007, “Threats and countermeasures for information system security: A cross-industry study,” Information & Management, Vol. 44, pp.480–491.