Communications of the CCISA
Vol. 25 No. 2 May. 2019

網路勒索病毒的特徵分析與知識本體模型建構

林孝忠[1]、王平[2*]、洪維謙[3]
[1,2,3]崑山科技大學 資訊管理學系
[1] fordlin@mail.ksu.edu.tw、[2]pingwang@mail.ksu.edu.tw、[3]yamn18345@gmail.com

摘要

近年隨著連網行動裝置與無線感測技術普遍化，帶動物聯網(Internet of Thing, IoT)與雲端服務的整合契機，通常物聯網設備使用精簡型的作業系統，無法安裝掃毒引擎，加上管理者缺乏持續的作業系統更新，容易產生資安漏洞並遭受攻擊，可以成為駭客攻擊的網路跳板，並造成企業或個人隱私資訊外洩！故本研究以樹莓派實作一個物聯網為基礎之智慧家庭的資網路防護系統，針對近期發生的勒索病毒(Ransomware)威脅，透過 Cuckoo 沙盒分析攻擊特徵(attack vectors)，再運用正規化概念分析 (Formal Concept Analysis, FCA)建構勒索病毒之知識本體模型(ontological model)，其目的是希望能建立電腦病毒知識本體為一概念化的正規抽象模型，明確定義病毒與攻擊行為間之關聯，作為病毒類別與變種鑑定的參考依據，以強化析網路病毒防護與資安管理。

關鍵詞：勒索病毒、知識本體、正規化概念分析、Cuckoo、攻擊特徵

* 通訊作者 (王平, pingwang@mail.ksu.edu.tw)

Regular Paper

Communications of the CCISA
Vol. 25 No. 2 May. 2019

Using Signature Analyses to Construct an Ontological Model of Ransomware

Hsiao-Chung Lin[1], Ping Wang[2*], Wei-Qian, Hong[3]

[1, 2, 3]Department of Information Management, Kun Shan University, Tainan, Taiwan

[1]fordlin@mail.ksu.edu.tw, [2] pingwang@mail.ksu.edu.tw, [3]yamn18345@gmail.com

Abstract

The growing popularity of employing of the mobile device enables the development of the Internet of Thing (IoT). Generally, IoT devices use an embedded operating system, cannot completely install anti-virus engines, and uers have not continuously updated the operating system. Consequently, system vulnerabilities prone to attacks and may lead to the privacy of business or personal information leakage. Accordingly, the present study proposes an IoT-based security defence system with Raspberry Pi to analyse the attack vectors of Ransomware using Cuckoo malware dynamic analysis platform. Importantly, an ontology-based method for developing domain ontologies using Formal Concept Analysis (FCA) technique is proposed. Experimental data show that our model is capable of performing the missions including of i) explicitly identifying the relations between Ransomware and their malicious behavior , ii) categorizing the Ransomware and the variations, and (iii) assist manager analyse the security controls for virus protection from cyber threats.

Keywords: Ransomware, Ontology, FCA, Cuckoo, Attack vector

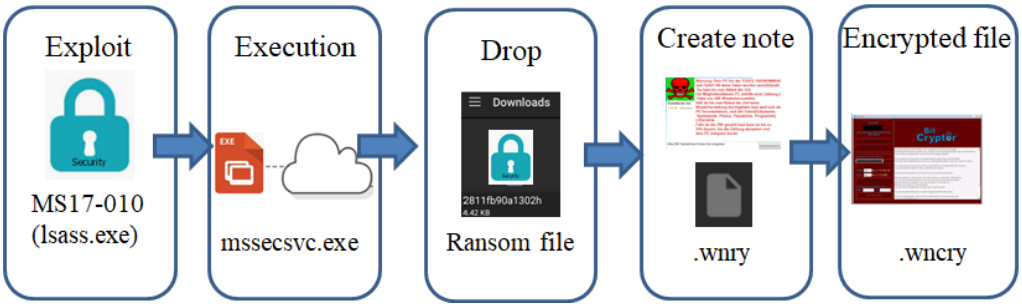
壹、前言

因智慧型手機與雲端運算應用的普及，加上物聯網的興起，物聯網應用從行動裝置延伸到各領域，物聯網裝置的資安防護方案是否完備，或是使用者是否意識到物聯網裝置安全設定或系統更新的重要性，皆是物聯網安全(IoT security)值得重視的議題。常見之物聯網安全威脅多為硬體限制與運行較精簡之作業系統，缺乏持續系統更新，容易產生資安漏洞並遭受攻擊，可能造成企業或個人隱私資訊外洩！外加，使用者可用應用程式(APP)遠端監控，被控裝置的即時狀態回饋及再下達控制命令，若物聯網被駭客入侵，隱私可能被外洩。因近期物聯網裝置之安全漏洞(weak points)陸續被發現，來自雲端威脅(cyber threats)可能入侵企業伺服器與連線後台資料庫，造成工廠及供應鏈之資訊包括物聯網裝置識別、設定參數、安裝地點、使用電力、供應商及時間等，可以窺視企業生產製程、製造參數與產品資訊，將可能造成企業的鉅額損失。

在物聯網中可連網之 IP 設備如偵測器內含精簡型電腦，如街頭攝影機含樹莓派等皆須面對惡意程式之入侵、攻擊、竊取隱私資料等威脅，做為中繼跳板，若再誘使下載網路勒索病毒(Ransomware)，引發後果不堪設想。以著名的 WannaCry 勒索病毒為例，WannaCry 被認為利用了美國國家安全局的「永恆之藍」（EternalBlue，亦被編號為 共同系同弱點 CVE-2017-0144 和 MS17-10）工具以攻擊 Microsoft Windows 作業系統的電腦，被針對的副檔名共有 176 種，包括 Microsoft Office、資料庫、壓縮檔、多媒體檔案和各種程式語言常用的副檔名。「永恆之藍」利用了某些版本的微軟伺服器訊息區塊（Windows Server Message Bloc，SMB）協定中的數個漏洞，而當中最嚴重的漏洞是允許遠端電腦執行惡意程式碼，接著再將真正的勒索病毒檔案送入受害系統，它會用 .WNCRY 副檔名來對相關檔案進行加密，勒索病毒之感染途徑如圖一。[1]

趨勢科技於 2017 年 4 月中首次監測到勒索病毒（RANSOM_WCRY.C），最初它透過網路釣魚攻擊誘使使用者從 Dropbox 網址下載惡意程式；後來，趨勢科技發現這個肆虐全球的勒索病毒「WannaCry/Wcry」已進化為結合了 SMB 伺服器漏洞 EternalBlue 與新勒索病毒家族（RANSOM_WCRY.I / RANSOM_WCRY.A）的新變種。[13]

由於網路勒索病毒 WannaCry 肆虐，台電也受害，全公司有 779 台電腦中毒，受害單位遍及台電上下，包括總管理處、水火力電廠及業務單位等，中毒電腦皆已網路斷線並隔離搶修。教育領域有 10 所學校（包含台灣大學、台灣師範大學等）的 59 台電腦中毒，但因為都不屬於核心系統，電腦重灌以後就沒問題。[14]



圖一: 勒索病毒之感染途徑

修改自: [25]

1.1 研究動機及目的

故本研究針對近期發生的網路勒索病毒(Ransomware)威脅，以樹莓派實作一個物聯網為基礎之智慧家庭的資安網路防護系統，為瞭解勒索病毒 先架設密罐蒐集連線之網路威脅資訊，再透過 Cuckoo 沙盒[2] 分析病毒攻擊特徵(attack vectors)，並運用正規化概念分析 (Formal Concept Analysis, FCA)[10,15,17] 客觀建構勒索病毒之知識本體(ontology)雛型，其目的是希望能建立電腦病毒知識本體為一概念化的正規抽象模型，明確定義病毒與攻擊行為間之關聯，系統化了解每一網路勒索病毒攻擊特徵與運作順序，以做為病毒類別與變種鑑定的參考依據，作為評估網路病毒威脅之資安風險等級，研析資安管理做法。

1.2 研究特色

本研究透過 GitHub theZoo 專案(https://github.com/ytisf/theZoo)[3] 蒐集到的真實(live)病毒樣本，病毒知識本體發展是參考 Standard university 與 University of Manchester (2000)發展本體論(ontology)為基礎，透過 Cuckoo 沙盒分析萃取出相關聯的特徵，搭配 Protégé 知識庫發展環境[4]，運用知識本體理論與 FCA 確認病毒種類與特徵間關聯，將特徵與病毒種類之關聯轉換成為機器可讀的知識架構，降低建置病毒知識分類架構時的困難和降低建置時間，搭配機率理論估算病毒種類與特徵間關聯之支持度，以利使用者正確識別已知及變種病毒種類。

1.3 論文章節架構

本篇論文共分四節：第一節前言，說明論文的研究背景、動機及目的。第二節文獻探討，首先，針對勒索病毒歷史文獻作探討；次之，對知識本體做相關介紹；再者，針對惡意程式檢測方法作相關比較。第三節介紹 Cuckoo 沙盒仿真環境，以動態分析手機病毒特徵，第四節建置勒索病毒之知識本體，說明勒索病毒特徵分析實作的過程，最後作出研究結論與建議。

貳、文獻探討

本節介紹勒索病毒、沙盒動態分析及 Cuckoo 沙盒分析技術。

2.1 勒索病毒

勒索病毒於近幾年內開始流行，如相當知名的 WannaCry 便是其中一隻勒索病毒，這些勒索病毒通常皆透過如 email 內夾帶偽裝檔案讓受害者執行後便寫入機碼內，開始進行加密行為，加密成功後便出現提示，使受害者發現系統內檔案已被加密，並要求受害者付款解鎖，各病毒間差異處主要為，加密系統利用之漏洞不同，傳遞手法也有所不同，目前大多為針對 Windows 系統，也有部分針對 Linux、Mac OS，雖勒索病毒看似近期內才開始大肆猖獗的盛行，但早在 1989 年時便有此種概念的病毒出現，該病毒為 AIDS Trojan 會替代開機執行檔，一旦執行達 90 次以上，便會隱藏並加密 C 槽內所有文件的名稱，要求使用者聯繫 PC Cyborg 付款，將 189 美金寄到在巴拿馬的郵政信箱，這是比較古老的病毒勒索方式，在電腦技術發達的現在，勒索病毒的傳遞與贖金取得皆更加無影無蹤[13]，新的勒索病毒可能會透過殭屍網路進行傳遞，要求支付數位貨幣如比特幣等，以下為本研究整理勒索病毒如表一。

表一: 勒索病毒威脅列表

作業系統	病毒名稱	描述
MS-DOS	AIDS Trojan	● 宣稱受害者的某個軟體已經結束了授權，並加密磁碟上的檔案，要求繳 189 美元的費用給 PC Cyborg 以解除鎖定
Win32	Gpcode.AG	● 會對所有可讀取的目錄內特定格式的檔案進行加密，並在每個被加密的目錄下建立為 readme.txt 提供付費解鎖的途徑
Win32	Gpcode.AK	● 為 Gpcode.AG 之變種，與前代不同之處在於使用 1024 位元的 RSA 加密
Windows	WinLock	● 顯示色情圖片遮擋使用者的電腦螢幕，並提示受害者利用大約 10 美元的簡訊付費以接收解鎖的密碼
Windows	CryptoLocker	● 以 email 傳送偽裝 pdf 格式的.exe 檔，啟動後會以隨機名稱寫入我的文件，並於登陸檔內新增機碼開機啟動，連接勒索者所控制的伺服器成功連接，產生 2048 位元 RSA 加密金鑰，加密
Windows	Petya	● 感染 MBR，受感染系統下次啟動時便加密 NTFS 檔案系統檔案表，完全阻止系統啟動 Windows，直至支付贖金
Windows	WannaCry	● 是一種利用 NSA 的(EternalBlue)漏洞透過網路 Windows 電腦進行加密利用 AES-128 和 RSA 演算法惡意加密用戶檔案以勒索比特幣
Windows	Bad Rabbit	● 誘使受害者下載偽裝成 Adobe Flash 的更新程式，一旦安裝後，電腦內的檔案文件便會被加密

資料來源：本研究整理

2.2 病毒動態分析方法

分析電腦病毒行為特徵常運用沙盒(sandboxing)以「病毒動態行為特徵分析(behavior analysis)」，沙箱是一個抽象的概念，是指「在一個特定的環境中，根據所需的安全性限制程式的行為」，讓程式放在模擬真實環境的沙箱內執行，對病毒做動態的追蹤，觀察是否有異常行為，透過此方式可不影響真實的系統環境檢測出病毒。知名沙盒包括CWSanbox、Cuckoo、INetSim、TRUMAN，病毒動態行為特徵分析法流程如圖二，分析方法特徵整理如表二。

圖二: 電腦病毒動態分析之流程

表二:電腦病毒動態分析技術特性

	電腦病毒動態分析特性
特色	・ 惡意程式須可以執行 ・ 在控制的環境下(沙盒)以虛擬主機執行並監控觀察惡意程式之感染行為 ・ 常使用模式與核心模式以分析不同病毒行為特徵，說明如下: ・ User space Sandbox 會在系統函式庫中插入程式碼，以得到運作時應用程式 API 的控制權或是透過 Debugger 執行軟體的 samples。但 User space sandbox 會被 Malware 軟體偵測到，進而故意模擬正常的行為，以規避 Sandbox，此時管理者就需要 Kernel sandbox。 ・ Kernel sandbox 解決此問題方法是在 Kernel 內監控 System calls，減少 Sandbox 被 Malware 察覺的機會
優勢	・ 可察覺未知惡意程式感染行為樣態 ・ 錯誤回報(false positives)減少
限制	・ 約 30%病毒具有反偵測執行環境之虛擬主機能力 ・ 動態所需分析時間長

黃獻德(2010) [9]整合 CWSandbox 沙箱與開發模擬的虛擬網路環境，建構自動化惡意程式分析環境 TWMAN (Taiwan Malware Analysis Net)，順利分析本地惡意程式，取得惡意行為特徵。亞俊、孫宏民(2012) [20]介紹了一種在作業系統內核心層內檢測 Android

應用程序中惡意程式的專案，通過即時分析設備內生成的日誌並搭配惡意程式檢測演算法，來確定是否存在應用程序的行為應被視為惡意行為。謝維揚、曾文貴(2013) [21]提出了一個新的有效動態分析(dynamic analysis)Android 惡意程式行為的方法，實作出了一個名為 MalCatcher 的系統套用了上述的方法，其主要是修改了 Android 系統原始碼並重新編譯，然後運行在模擬器上建立一獨立且受控制的環境供 APP 導入執行，並且將網路封包監控軟體 snort 重新交叉編譯為 Android 系統可執行之版本運行在模擬器上，用以監控網路封包是否有洩漏使用者隱私資料，其餘國內相關研究請參考[22]~[24]。

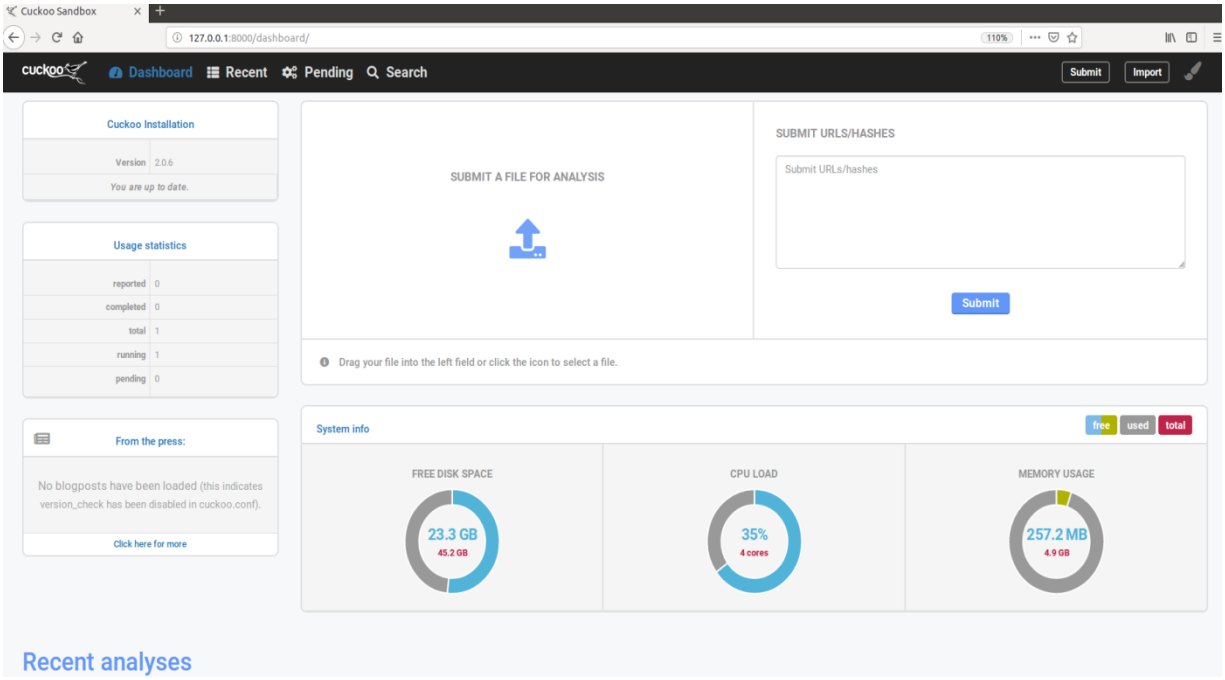
動態分析讓程式在虛擬的環境下執行並監控觀察其行為，由於程式必須持續執行，因此耗費的運算時間明顯相對比靜態分析多。沙箱可依目的不同分為多種技術，像是要進行檢查而停在某個程式呼叫執行點、防止潛在的危險而中止程式、在程式運行時監控及紀錄程式活動。舉例目前資安實驗室常使用廣泛之 CWSandbox 沙箱[5]是針對 Windows 或 Server 為主，運作可分成兩個部份:

1.靜態分析，利用程式特定字串來進行文字探勘(text mining)分析，並且把分析結果寫入紀錄檔 logfile。
2.若靜態分析沒有偵測到符合的特定字串，則進行動態行為分析，使程式獨立在虛擬環境沙箱下執行，並紀錄其異常行為特徵，並輸出至檔案。

本研究不採用病毒原始碼(Code Analysis)靜態分析法，因為 Code Analysis 大致運用反組譯(decompilation)、解密(decryption)、樣式比對(pattern matching)以及系統呼叫分析(system call analysis)等方法。靜態分析進行分析時，會把原始碼所有狀態轉換成模型，同時把原始碼轉換成抽象的語意，並會把程式執行的所有可能發生的因素都考慮進來，因此靜態分析是具保守且可靠的。[6]以上的技術有一個共通點是不會直接執行應用程式，因此靜態分析較快速且簡單，反之缺點是當病毒以多形(polymorphism)或變形(metamorphism)的方式呈現時，靜態分析會失效，故防衛者發展出程式語意分析(semantic analyzed)來對抗變種病毒。若沒有外部情報或手動檢查，一般靜態分析函數還是可規避且無法了解未知惡意程式感染行為樣態，對新的變種病毒將無法偵測。

2.3 Cuckoo 沙盒動態分析

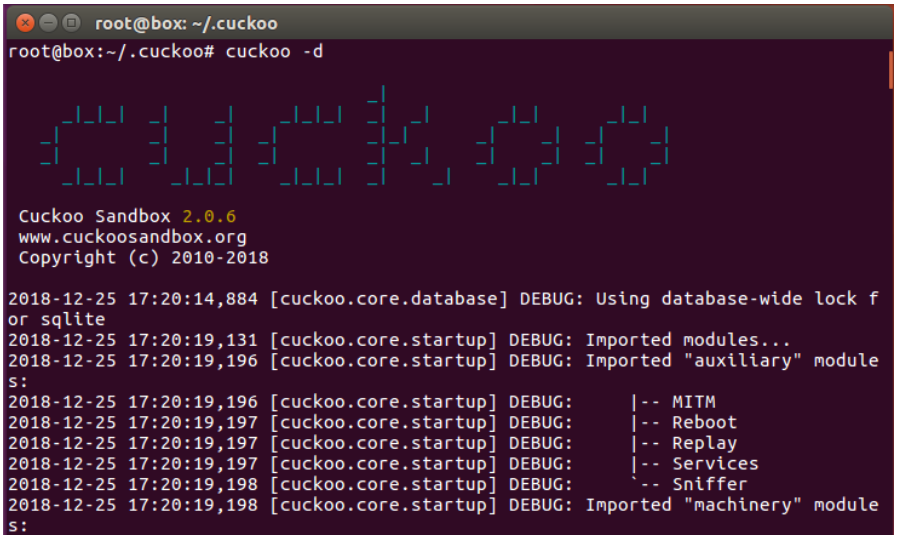
Cuckoo 沙盒是在參加 2010 年 Google 程式專案夏令營的作品，最近在 GPL 許可下開源發佈。任何人可以將其加到自行研發的專案中，打造自己的惡意軟件行為分析工具。基本上，Cuckoo 是一個輕量級的 windows 二進制文件行為自動動態分析工具，如圖三。



圖三: cuckoo web 分析介面

它能夠給出程序運行過程中詳細的關鍵 API 調用和網絡活動。 由於 Cuckoo 的開源特性和廣泛的模塊化設計，可以客製化分析環境，可將各階段的分析結果處理以和產出報告。為使用者可透過指令要求 Cuckoo 提供服務，可以按照您希望的方式輕鬆地將沙箱架構再到現有的計算機組態和後台中，只要滿足開放源碼 GPLv3 許可要求。建置一個可以執行程式的環境紀錄可疑行為及資訊洩漏(information leakage)，透過 Cuckoo 沙盒可以觀察程式所造成的影響的行為，如圖四，分析完成可提供以下資訊[2]:

1. 由惡意程式生成的所有進程執行的 win32 API 呼叫的追蹤。
2. 惡意程式在執行期間創建，刪除和下載的文件。
3. 惡意程式進程的內存轉儲。(Memory dumps of the malware processes)
4. 網路流 PCAP 格式紀錄。(Network traffic trace in PCAP format)
5. 執行惡意軟件期間拍攝的 Windows 桌面的螢幕截圖。
6. 主機的完全內存轉儲，包括自動運行 Volatility。(Full memory dumps of the machines, including automatic running of Volatility).
7. 工具更多資訊及下載：http://www.cuckoobox.org/download.php



圖四： Cuckoo 沙盒開啟畫面

參、勒索病毒知識本體模型之建置

本節介紹知識本體發展環境 Protégé應用於勒索病毒資料庫之發展與概念格之建置過程。

3.1 知識本體理論與 Protégé

Protégé知識本體發展環境是由史丹佛大學發展，已應用於知識探索之各項應用，常做為知識體系建立與概念認知的表達。知識本體應有的基本要素為：類別(Class)、資料槽(Slot)、實例(Instance)及基本元素(Axiom)[7]，以電腦病毒為例說明病毒知識本體的基本要素定義與案例如表三.。

表三: 電腦病毒知識本體的基本要素

	基本定義	案例
類別 (Class)	・類別:具有相同屬性的物件群體	・類別:電腦病毒一般區分主機型或網路型病毒 ・子類別:僵屍病毒、蠕蟲、特洛伊木馬可視為網路型病毒的子類別，巨集型病毒或開機病毒可視為主機型病毒的子類別
資料槽 (Slot)	・在知識本體論中用來描述概念的屬性或概念之間的關聯 ・其中父類別與子類別間的關聯也是一種 Slot	・蠕蟲子類別包括 Code Red、Code Red II、Klez、Sadmind 、 Sircam 和 Nimda 等 ・僵屍病毒包括 IRC、Mirai、Viro 等
實例 (Instance)	・知識本體中為一個概念或類別的案例，實例會繼承類別的所有屬性或關聯	・電腦中毒案例的某一隻蠕蟲都是感染電腦病毒的實例，但病毒名稱、受感染平台及感染途徑都可能不同。
基本元素	・於知識本體中是原則或限制，其	・電腦病毒一般是指惡意程式，用以破壞

(Axiom)	功能在於制定概念間關聯或限制 ・與 Slot 不同之處在於，Slot 清楚定義兩個類別之間的關聯	或竊取他人資料 ・僵屍病毒的限制是受感染主機(Zomie)之感染途徑是經過僵屍電腦(跳板)傳輸命令，僵屍電腦再接受遠端主機控制。

資料來源: 整理自[15,17]

本研究應用知識本體模型於電腦病毒特徵儲存、病毒識別；Protégé 資料庫定義三個主要的知識模組: (1)說明特定的類別，(2)描述每一個概念的性質與屬性稱為資料槽(slot)，(3)類別所產生的知識實例(instance)。類別與類別之間可存在繼承關係，子類別可繼承父類別所定義之資料槽，資料槽可用以描述類別與類別之間的關係，包含類別與類別的實例是一組完整的知識概念，亦即知識基礎。

基本上，Protégé工具繪製知識本體之類別關係圖類似資源規劃系統(ERP)之產品用料清單(bill of material, BOM)，其可表達出每一個類別所階層中包含的資料槽、類別和類別之間的關聯(縱向與橫向)、描述屬性限制面向、屬性的型態達到共同的認知，可釐清類別與類別之上、下位及相鄰關係，減少概念之衝突，架構出概念清楚的知識本體。Protégé 工具可外掛 FcaView Tab 功能模組，將知識本體轉換成非同形式於知識本體的本文(context)，並藉由物件中相同屬性的關聯產生隱含的資料槽。[18]本文中的物件代表知識本體的類別、屬性則代表知識本體的特徵，兩者被作為本文交叉關聯表的表頭；本文交叉關聯表中若同一列的物件跟同一欄的屬性有關聯則成為一個概念，先前研究整理行動病毒之本文交叉關聯表，如表四。

表四: 行動病毒之本文交叉關聯表[18]

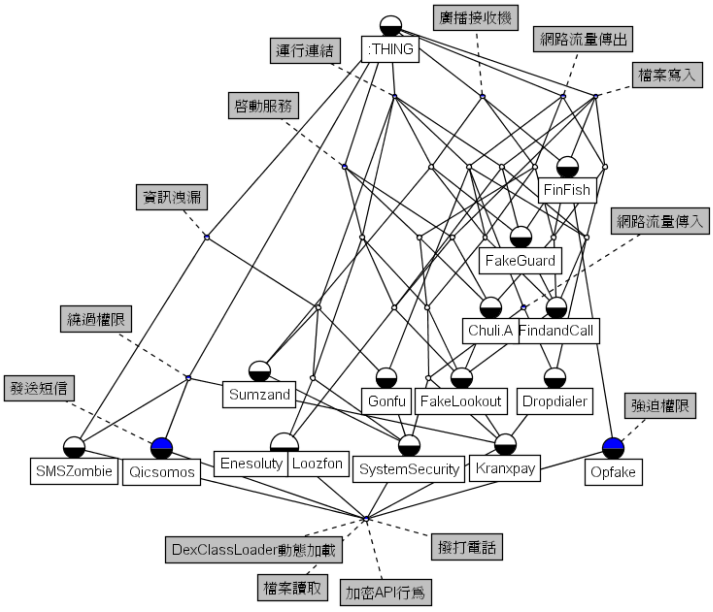
物件 / 屬性	SMS Zombie	Chuli.A	Enesoluty	System Security	FakeGuard	Sumzand	FakeLookout
撥打電話							
發送短信							
資訊洩漏	X		X	X		X	
繞過權限	X						
強迫權限							
啟動服務		X	X	X			X
廣播接收機		X		X	X	X	X
DexClassLoader 動態加載							
運行連結			X	X	X	X	X
網路流量傳出		X		X			X
網路流量傳入				X			

3.2 正規化概念分析(Formal Concept Analysis, FCA)

Gruber(1993)提出知識本體的概念化(conceptualization)、共享(Shared)、正式的(Formal)及明確 (Explicit)的特性。正規化概念分析是在 1982 年由 Rudolf Wille 提出一種以格理論(lattice theory)為基礎用於資料分析的數學方法，FCA 強調以人的認知為中心，可以從資料集(Data set)中發現概念結構(conceptual structures) ，是建構知識結構的重要

方法。[8] 這些結構可以透過概念格(conceptual lattices)作圖形化的呈現，常被應用在知識呈現及知識挖掘的研究，概念格能夠分析複雜的結構及釐清集合中相關屬性的關聯如圖五。

國內學者林建宏、方定國 (2001) [15]運用知識本體與 FCA 正規化進行電腦病毒(malware)種類與病毒感染特徵樹的分析，建立以知識本體為基礎建構病毒分類知識庫系統，改善變種病毒識別不易問題。黃盈豪、許通安(2008)[19]提出一套自動化建置知識分類架構的方法，結合混合式機器學習法則，應用於電腦病毒領域的查詢與推論，經由既有的病毒特徵產生電腦病毒概念與屬性之間的關聯階層架構，使用 SWRL 規則推論出電腦病毒間隱含的關係，並且提供相關的解決方案給予使用者。王平、王宇治(2013)[18]針對過手機病毒過 DroidBox 仿真環境(instrumented emulator)，以動態分析並紀錄 Android Apps 可疑行為與分析病毒特徵，再運用知識本體與 FCA 正規化分析電腦病毒種類與病毒感染特徵的關聯，建立電腦病毒知識本體為一概念化的正規抽象模型(如圖五)，明確定義病毒與攻擊行為間之關聯，系統化了解每一行動病毒屬性與屬性相互間之關聯。



圖五: 行動裝置病毒之概念格[18]

在圖五概念格中「根」(thing) 是所有子概念的最高層概念，由於底層概念是沒有任何關聯，因此稱為空概念，每個圓圈代表是一包含物件及屬性集合的正規化概念對(pair)。其中，黑與藍圓圈是概念全文中有共同屬性的概念物件，它是明確地被分類系統所指出；空白圓圈則表示概念全文內所隱含的物件，表示可能有其它屬性的抽象化概念。

Ganter & Wille(1999) 定義一個正式本文(formal context) K 是由(G，M，I)三元素所構成(triple)，本文包含一組概念 G、一組屬性 M 與一組關係 I，I 用以表示概念 G 與屬性 M 間之關聯，表示為二元關係圖(bipartite graph) $I \subset G \times M$。 [10]

一個本文的正式概念(formal context)定義為一組數對 (G_i, M_i) 受限於

1. $G_i \subseteq G$
2. $M_i \subseteq M$
3. 每一概念 G_i 擁有屬性 M_i
4. 概念 G 不在 G_i 範圍，則概念 G 不會擁有屬性 M_i
5. 概念 G 擁有一組屬性 M 不包含於 M_i, 則存在一個概念 G_i 不會擁有屬性 M_i

一個正式概念為一組 G_i 和 M_i 的配對(pair)，其中滿足上述限制條件，則 G_i 稱為概念的範圍(extent)，而 M_i 稱為概念的涵義(intent)。一個概念本文(context)可運用交叉關聯表(cross reference table)加以表示，其中概念置於列(rows)，屬性置於行(columns), 關聯以一布林值 (Boolean value)表示。一個概念可透過一個相關屬性之最大化之子集合。

針對概念的子集 $A \subseteq G$ 和屬性的子集 $B \subseteq M$，可以定義兩個衍生運算(derivation operators)如下所示：[10]

$A' = \{m \in M \mid gIm \text{ for all } g \in A\}$, and dually

$B' = \{g \in G \mid gIm \text{ for all } m \in B\}$.

應用衍生運算可構成兩個閉合運算符號：

$A \mapsto A'' = (A')'$ *for* $A \subseteq G$ 範圍閉合(extent closure),及

$B \mapsto B'' = (B')'$ *for* $B \subseteq M$ 涵義閉合(intent closure)

衍生運算符定義了概念集合與屬性集合之間連接為伽羅瓦格(Galois)，在法語中概念格稱為 treillis de Galois（伽羅瓦格）。

正式概念的全等(Equivalent)

通過上述衍生運算，Rudolf Wille 給出了一個正式概念的定義：本文（G，M，I）中一個形式概念可由對集合（A，B）來加以描述，限制條件為：

$A \subseteq G, B \subseteq M, A' = B, \text{ and } B' = A$.

直觀地說，正式概念由對集合（A，B）滿足以下任一情況即為一個正式概念：

- A 的每一個概念都有 B 中每一個屬性，
- 對於 G 中每一個概念不存在於 A 子集合，屬性子集合 B 中有一些屬性無法對應到概念(for every object in G that is not in A, there is some attribute in B that the object does not have)
- 對於屬於 M 集合屬性但不屬於子集合 B 的每個屬性，A 中一些概念沒有對應至該屬性。(for every attribute in M that is not in B, there is some object in A that does not have that attribute)

正式本文的概念格(Concept lattice of a formal context)

正式概念的排序

一個本文 K 的概念（A_i，B_i）可透過包含範圍來進行比大小(排序)，或通過等效地涵義的雙重包含來排序。 更精確地說，概念上的排序 ≤ 可定義如下：對於 K 的任何兩個概念（A_1，B_1）和（A_2，B_2），($(A_1, B_1) \leq (A_2, B_2)$，當 $A_1 \subseteq A_2$)等同於($(A_1, B_1) \leq (A_2, B_2)$，當 $B_1 \supseteq B_2$.)

按此正式概念的排序，每組正式概念都有一個最大的共同子集合。(上限) 若概念格滿足 meet 和 join 操作的公理，就定義為一個完整概念格。

涵義運算(Implications)

隱喻運算 A→B 涉及兩組屬性 A 和 B，並表示每一個擁有 A 屬性的概念也具有來自 B 的每一個屬性。當(G, M, I)是形式概念本文，而 A，B 是子集合時，屬性集合 M（即 $A,B \subseteq M$），如果 $A' \subseteq B'$，則暗示 $A \rightarrow B$ 成立。 對於每個有限形式概念本文，所有有效含義的集合都具有規範基礎(canonical basis)，一系列具有意義的隱喻，所有有效隱喻都可以通過自然推理得出 (Armstrong rules)這是一種基於含義的知識探索方法。

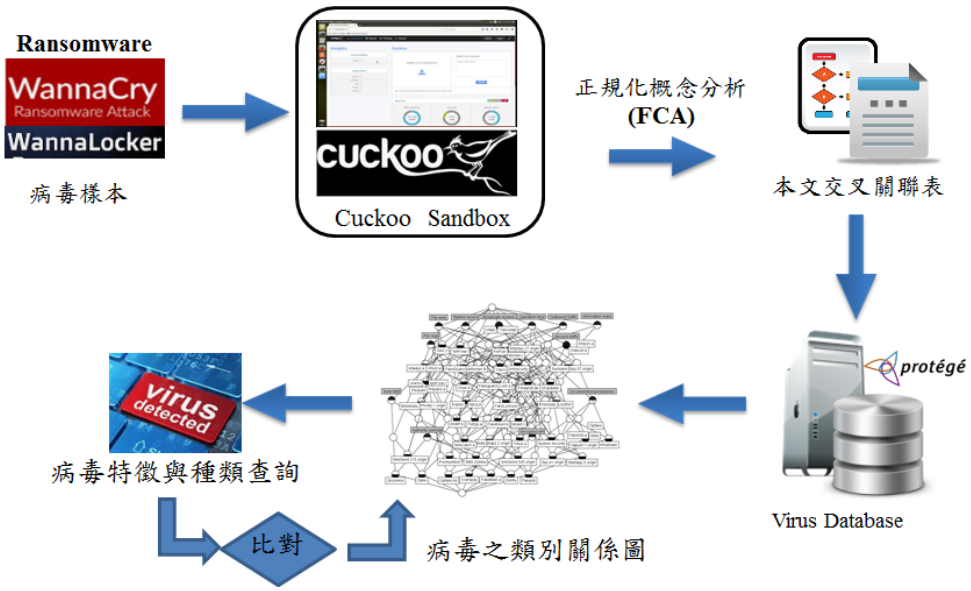
由上述說明正規化概念分析是建立在數學基礎，對組成本體的概念、屬性及關係等用形式化的語意表述，然後根據使用情境，建構出知識本體之概念格(concept lattice)，清楚地以階層式表達出知識本體的結構，概念格可以圖形方式顯示為“線圖”，這可能有助於理解數據間的上、下從屬關係。

Saquer and Degun(2001)提出 FCA 可用於處理知識獲得、知識組織及知識表達，經正規劃概念分析之交叉關聯表，讓知識本體的階層式關聯的由上而下呈現[17]，藉由概念格來建置知識庫，提供一種與傳統的、統計的資料分析和知識的表現方法，FCA 已成為人工智慧學科的重要研究方法，在機器學習、資料採擷、資訊檢索等領域得到廣泛的應用。Priss (2003)指出在資訊科學的領域中，FCA 應用包括在數學方格可用於分類，分類系統可根據關係之分類獲得一致性的分析結果[11]。

形式概念分析的數學理論對於之知識分析是有幫助的，例如在物種分類常應用分群，須將概念格分解成沒有信息損失的較小群組，或者小群組嵌入到另一大的家族分析，例如類似物種之概念格的相似性分析(similarity analysis)等，有興趣同仁可參閱[11]。

肆、驗證與確認

本節透過 Cuckoo 動態實驗分析病毒特徵，運用知識本體的資料庫之系統化建立病毒種類與種類、種類與屬性間之關聯，其實驗規劃流程為圖六，藉由機率理論估算電腦病毒與各特徵間關聯之支持度。



圖六: 實驗流程圖

4.1 勒索病毒分析案例

實驗之網路勒索病毒樣本由 GitHub 的 theZoo 專案下載,實驗日期: 2018.9~2019.3。由於發現幾乎所有版本的真實(live)惡意程式都很難通獲得,theZoo 專案決定以透過查詢和安全的通道以提供給資安研究者下載惡意程式。 theZoo 由 Yuval tisf Nativ 創建,現在由 Shahak Shalev 負責後續維護。theZoo 專案是公開化蒐集真實惡意程式的儲存庫(respority),其初期惡意程式來源主要由 https://thezoo.morirt.com 來提供,透過眾人羣力共同來上傳樣本,使惡意程式分析成為可能。[3]本實驗第一階段共蒐集 29 家族網路勒索病毒,但部分病毒具有反偵測沙盒能力,造成無法順利執行並萃取出行為特證做後續分類使用,至 109.4.4 經過多次嘗試與測試共取出 19 勒索病毒家族(Cerber、Cryptowall、Locky、Petrwrap、Petya、Radamant、Satana、Vipasana、WannaCry、TeslaCrypt、BadRabbit、Ryun、CryptoNar、Hermes、LockerGoga、Termite、KeyPass、CMB Dharma、GandCrab 等);以著名的 WannaCry 勒索病毒為例,本研究利用 Cuckoo 動態分析病毒樣本,監測到木馬程式會以的加密型勒索軟體兼蠕蟲病毒(Encrypting Ransomware Worm)攻擊主機。該病毒利用 AES-128 和 RSA 演算法惡意加密用戶檔案,以使用 Tor 加密通訊,勒索受害者提供比特幣。

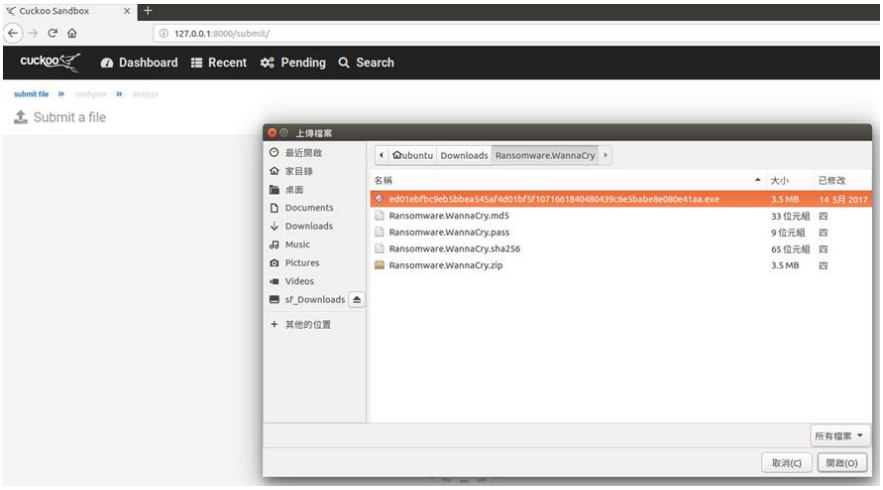
當受害者被安裝惡意程式時,會進行以下活動,包括(1) 該病毒進入目標主機之後,就會對主機硬碟和儲存裝置中許多格式的檔案進行加密 (2) 利用網路檔案分享系統的漏洞,傳播到任意的其他聯網的主機 (3) 處於同一區域網路的相鄰主機也會被感染。該惡意程式安裝完後會連線至後台,此時受駭主機已遭受到程序監控,傳送出行動裝置使用的基本訊息至特定的遠端監控主機,如圖七。

圖七: WannaCry 勒索病毒受駭的畫面 [26]

此案例利用動態紀錄模擬真實行動裝置系統來監看勒索病毒樣本執行過程，利用 windows api 生成密鑰、進行反偵測虛擬機環境、竊取本地瀏覽器個資資料、與遠方且未進行本地 DNS 查詢的主機通訊、在 Windows 啟動時自動執行、從作業系統安除大量下載惡意文件，作為建置 protégé病毒特徵庫之動態分析屬性，其實驗步驟如下：

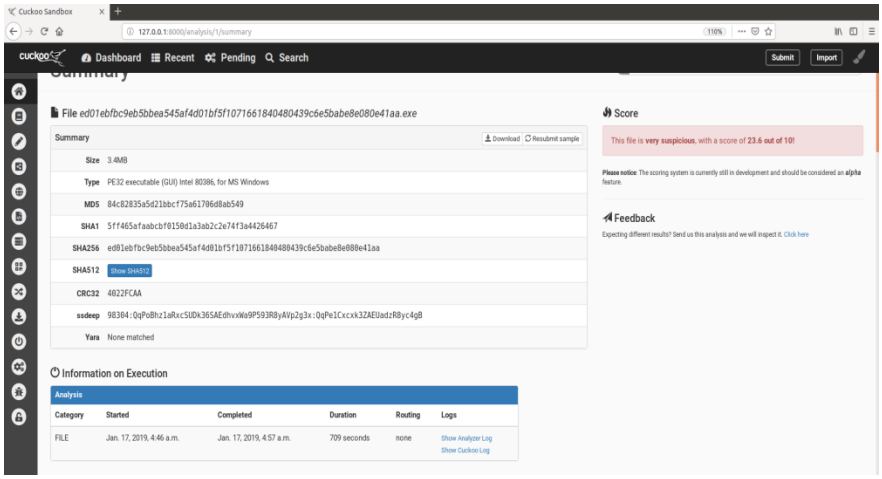
步驟 1.以除錯模式開啟 cuckoo。

步驟 2. Cuckoo Web 執行勒索病毒特徵分析如圖八。



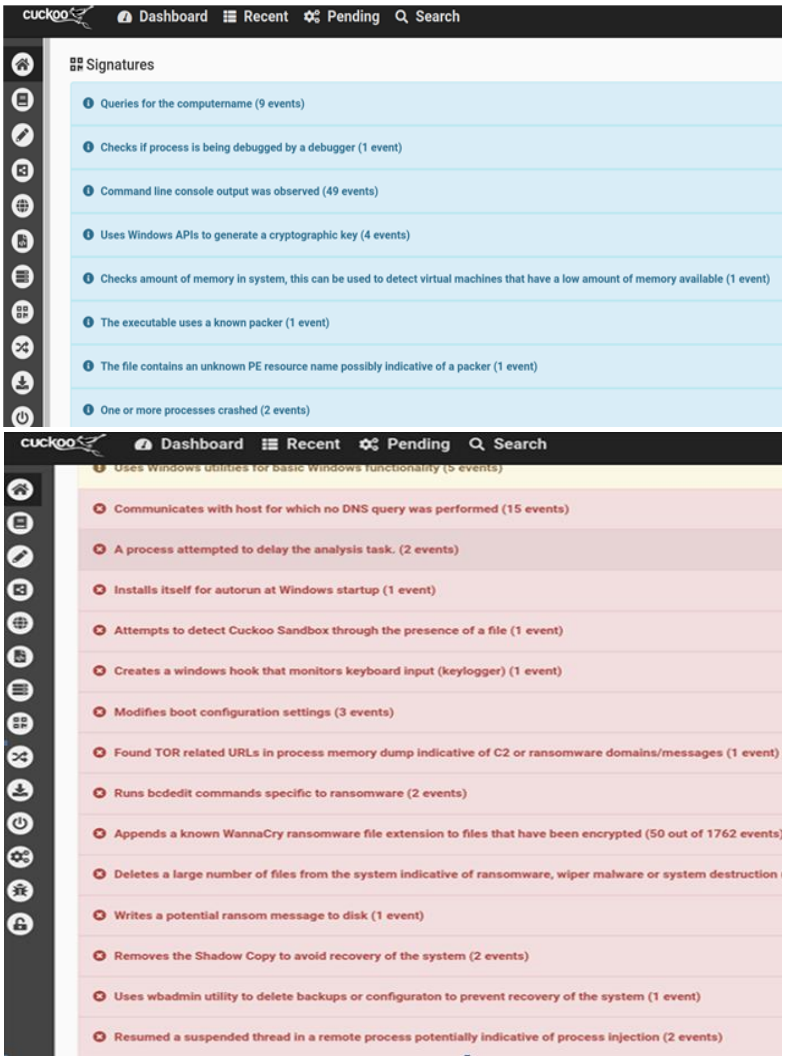
圖八: 開啟 cuckoo web 分析 WannaCry 勒索病毒

步驟 3. 勒索病毒特徵分析綜合結果之顯示如圖九。



圖九: 顯示分析之綜整結果

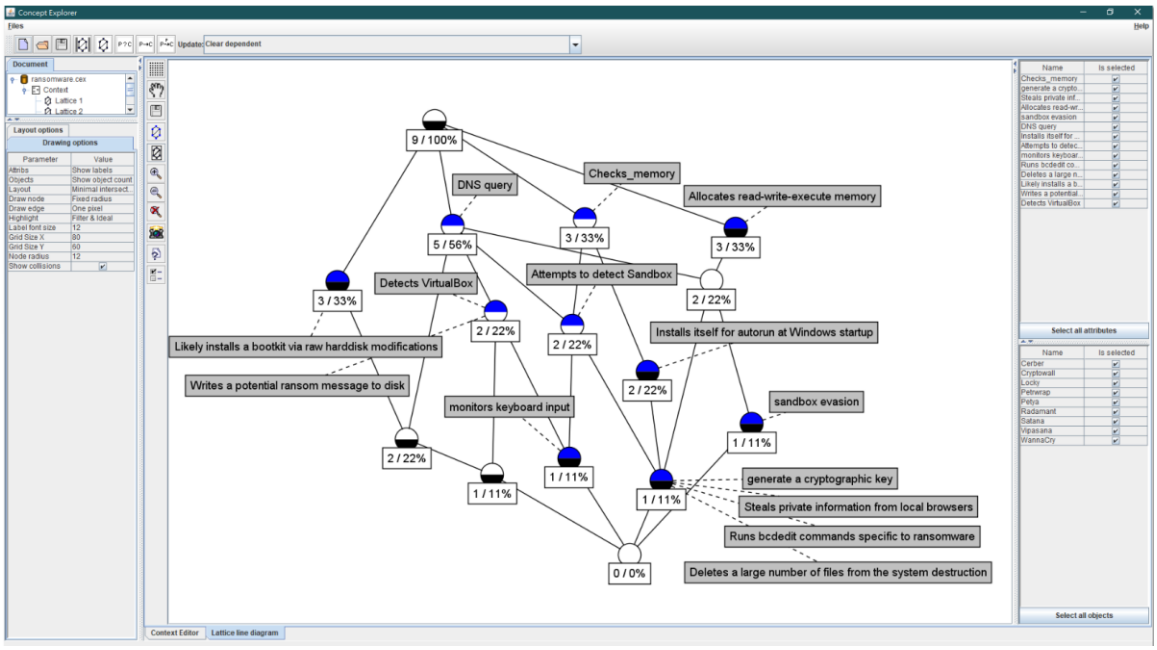
步驟 4. 網路勒索病毒從中毒至發作詳細感染過程之清單如圖十。



圖十: cuckoo 顯示勒索病毒執行流程之異常行為屬性

步驟 5. 以 Protégé編輯病毒資料庫

透過 cuckoo 行為分析，可歸納出共同特徵作為 FCA 的輸入，合併使用知識本體概念工具 Concept Explorer(http://conexp.sourceforge.net/)與外掛 FcaView Tab 功能模組編輯作為分析的輸出，以描述惡意軟體和行為屬性之間的關係。依據前面步驟 1~4 的分析結果，彙總出勒索病毒之行為屬性欄位，如表五。依據表五內的欄位輸入至 Protégé 特徵庫內，並將其轉換為 xml 格式提供給 Concept Explorer 如圖十一，圖十一為病毒特徵的一般化(generalization)階層結構，可協助管理者識別未知惡意軟體的輪廓。



圖十一: Concept Explorer 工具 [12]

表五: Cuckoo 動態分析勒索病毒行為特徵清單

編號	行為屬性描述
1	檢查系統的記憶體容量(Checks amount of memory in system)
2	使用 Windows API 生成加密金鑰 (Uses Windows APIs to generate a cryptographic key)
3	從本地瀏覽器竊取資料(Steals private information from local Internet browsers 0
4	分配讀寫記憶體(通常用於解壓縮) (Allocates read-write-execute memory, usually to unpack itself)
5	搜尋正在執行的惡意程序，行為包括進行沙盒反偵測、程式碼注入、記憶體傾印等惡意行為(Searches running processes potentially to identify processes for sandbox evasion, code injection or memory dumping)
6	與未執行 DNS 查詢的主機通訊 (Communicates with host for which no DNS query was performed)
7	在 Windows 開機時自動執行(Installs itself for auto-run at Windows startup)
8	試圖通過現存檔案來偵測 cuckoo

	(Attempts to detect Cuckoo Sandbox through the presence of a file)
9	運行特定於勒索惡意程式的 bcdedit 命令 (Runs bcdedit commands specific to ransomware)
10	從系統刪除大量文件，顯示勒索軟體、wiper 病毒、破壞系統(Deletes a large number of files from the system indicative of ransomware, wiper malware or system destruction)
11	透過硬碟修改安裝 bootkit (Likely installs a bootkit via raw harddisk modifications)

步驟 6. 勒索病毒家族概念全文之建立

為清楚說明所蒐集勒索病毒家族概念全文，十九個家族病毒經由 Cuckoo 動態分析出來的特徵以建立概念全文，病毒命名方式皆使用防毒軟體公司賽門鐵克(Symantec)病毒數據庫為依據，每一行代表一個病毒物件，每一列代表病毒的屬性特徵，交叉關聯矩陣註記「X」表示同一列的物件關聯同一行的物件，如表六，繪製其概念格如圖十二。

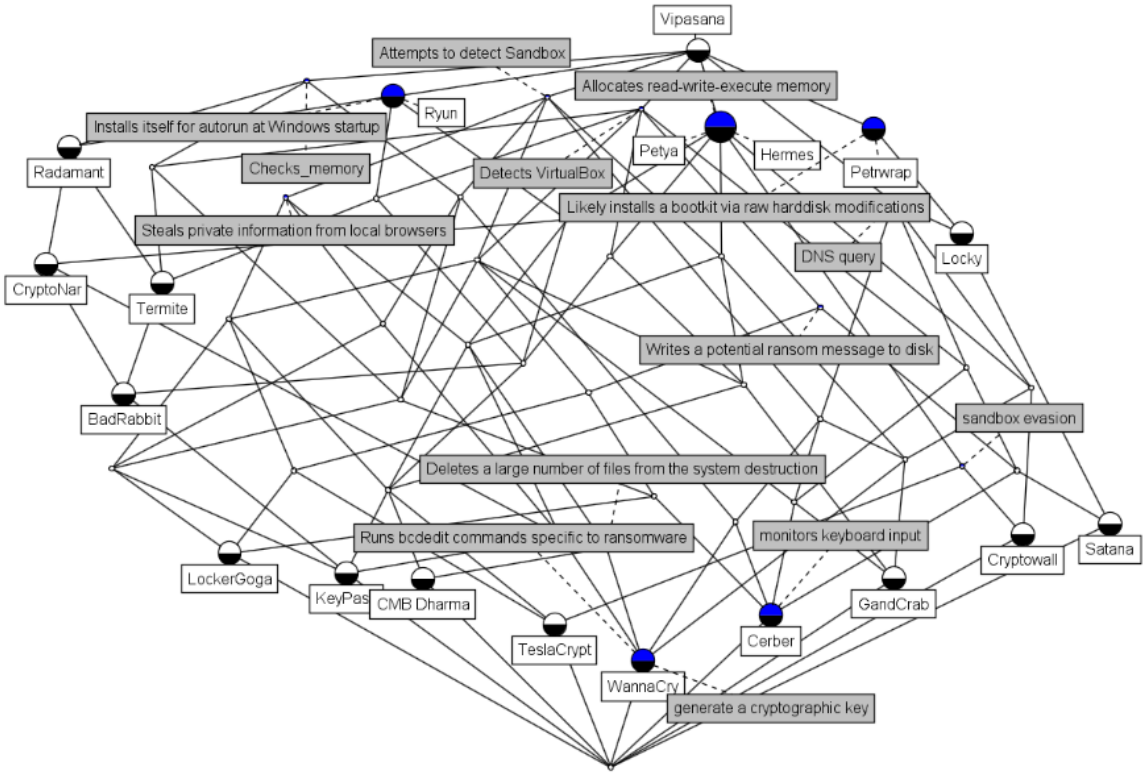
表六: 勒索病毒家族概念全文

	Checks_m...	generate a...	Steals priv...	Allocates r...	sandbox e...	DNS query	Installs its...	Attempts to...	monitors k...	Runs bcde...	Deletes a l...	Likely insta...	Writes a p...	Detects Vir...
Cerber	X					X		X	X				X	X
Cryptowall				X	X	X								
Locky						X						X		
Petrwrap												X		
Petya				X										
Radamant	X						X							
Satana						X						X	X	X
Vipasana														
WannaCry	X	X	X	X		X	X	X		X	X			
TeslaCrypt			X	X	X		X	X					X	X
BadRabbit	X			X			X							X
Ryun							X							
CryptoNar	X			X			X							
Hermes				X										
LockerGoga	X		X					X					X	X
Termite	X						X							X
KeyPass	X		X	X			X	X						X
CMB Dharma			X	X			X	X			X			X
GandCrab				X		X		X						X

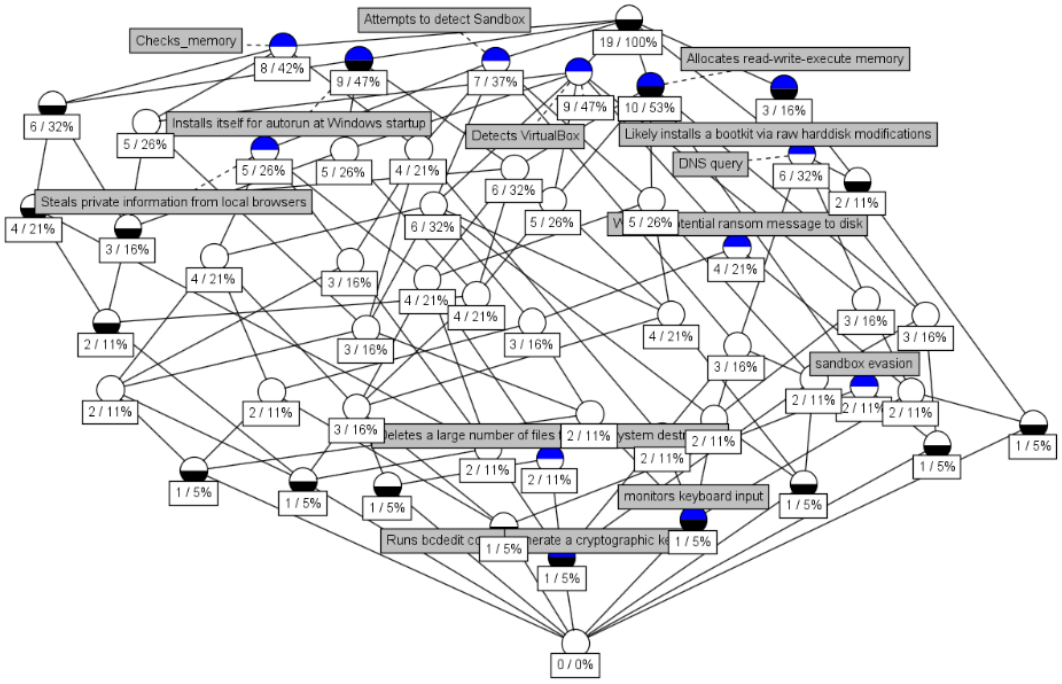
步驟 7. 勒索病毒家族概念格之建立

為了解勒索病毒家族與特徵之完整關聯，本研究將九個家族進行特徵動態分析，實驗結果先概念全文，再以 Protégé 資料庫提供之 FCA 進行病毒間相同屬性之合併與一般化程序，結果如圖十二。在圖十二中，WannaCry 病毒物件其關聯 Radamant 物件及 Cerber 物件，代表 WannaCry 物件與 Radamant、Cerber 物件有共同屬性，全文的最底層則是物件跟屬性間沒有關聯所產生底層概念。將實驗一病毒特徵之動態分析重新分析本體樹每一特徵間之關聯支持度的計算是採用公式 (1)，計算的結果如圖十三。病毒感染行為透過一系列行為事件 *s* 來表達，透過滑動視窗 *win* 來觀察單一事件(α) 發生機率之支持度 *sup* 為：

$$p_i(\alpha, s, win) = \sup(\alpha) = \frac{|\{\alpha \quad occur \quad in \quad win)\}|}{|\{w(s, win)\}|} \tag{1}$$



圖十二: 勒索病毒家族概念格(19 家族)

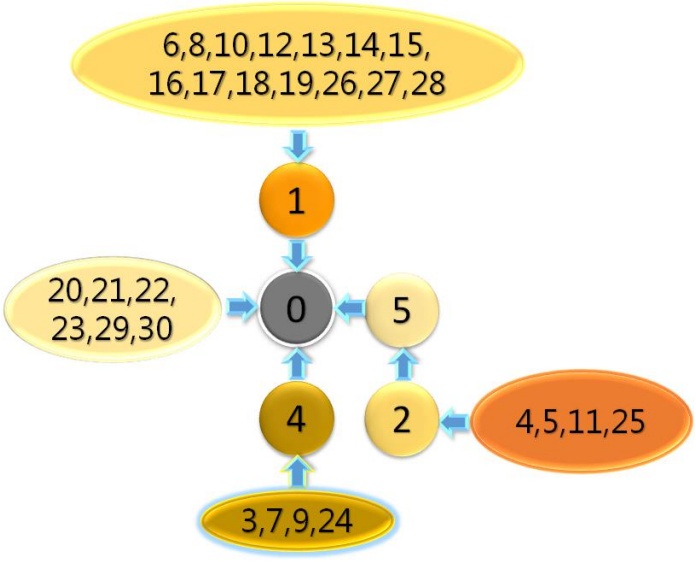


圖十三: 以機率顯示勒索病毒樣本概念格之關聯(19 家族)

4.2 病毒感染行為間之因果關係

在圖十二使用者不易看出感染病毒發作的行為(屬性) 間之因果關係，Protégé 資料庫提供病毒屬性間關係如圖十四，首先將 Protégé 系統將病毒行為屬性加以編號以利繪圖了解相互因果關係，整理行為屬性因感染勒索病毒，將引發行為屬性順序繪製成因果關係圖(cause and effect diagram)，如圖十五所示，由圖十五中可知屬性 1(generate a cryptographic key)其受其他 14 個相關屬性所影響(屬性 6,8,9,10,11,12,13,14,15,16,17,18,19,26,27,28)，結果將造成感染勒索病毒(屬性 0) ；另外屬性 2(Steals private information from local browsers)受其他 4 個屬性所影響(屬性 4,5,11,25)；屬性 4(Sandbox evasion)受其他 4 個屬性所影響(屬性 3,7,9,24)；此外，屬性 2,5,4 將間接造成感染勒索病毒(屬性 0)。

圖十四: 屬性間之上下因果關係(Protégé)



圖十五: 勒索病毒之屬性的因果關係圖

伍、結論

本研究依據知識本體架構,搭配Cuckoo病毒動態分析與模型正規化概念分析方法，透過病毒之行為特徵分析，運用 Protégé知識本體發展環境建構一個勒索病毒之知識本體雛型，建立以電腦病毒行為為基礎的正規抽象模型，用以評估勒索病毒之危害及研析防護的措施。因部分勒索病毒會主動偵測沙盒，進而故意模擬正常的行為，以規避 Sandbox 被察覺的機會，故捕獲並歸納出一勒索病毒家族實務上是不容易，目前僅完成部分樣本加以特徵分析，此唯一持續性的研究工作，未來將導入深度學習神經網路(Deep Neural Networks, DNN)，以機器學習方式來複雜多類變種病毒行為特徵，增加病毒抽象模型的應用領域與偵測正確性。

[誌謝]
本研究承蒙行政院國家科學委員會計畫（MOST 107-2218-E-001-002 TWISC2.0, and MOST 107-2410–H -168-002）經費補助，謹此致謝。

參考文獻

[1] https://zh.wikipedia.org/wiki/WannaCry (2018/7/28)
[2] HoneyNet, “Cuckoo Sandbox”, https://github.com/ cuckoosandbox/cuckoo (2018/8/13)
[3] Y. Nativ, “GitHub- ytisf/theZoo: A repository of LIVE malwares for your own joy and pleasure” https://github.com/ytisf/theZoo (2018/8/01)
[4] Standard University, Protégé, https://protege.stanford.edu/(2018/7/11)
[5] C. Willems, T. Holz, and F. Freiling, “Toward automated dynamic malware analysis using CWSandbox, ” IEEE Security & Privacy, vol. 5, no. 2, pp. 32-39, 2007.
[6] Z. Tzermias, G. Sykiotakis, M. Polychronakis, and E. P. Markatos, “Combining static and dynamic analysis for the detection of malicious documents, in *Proceedings of the Fourth European Workshop on System Security*. 2011, ACM: Salzburg, Austria. pp. 1-6.
[7] N.F. Noy, D.L. McGuinness. “Ontology Development 101: a guide to creating your first ontology”, *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05*, 2001.
[8] M. Uschold, M. Grueninger, “Ontologies: principles, methods and applications”, *Knowledge Engineering Review*, vol. 11, no.2, pp. 93-155, 1996.
[9] H. D. Huang, “Ontology-based intelligent system for malware behavioral analysis”, *WCCI 2010 IEEE World Congress on Computational Intelligence*, July, 18-23, Barcelona, Spain, 2010.
[10] http://en.wikipedia.org/wiki/ Formal_concept_ analysis. (2018/8/7)
[11] T. Priss, “Formal Concept Analysis in Information Science”. 2005, http://www.upriss.org.uk/fca/fca.html. (2018/8/7)
[12] S. Yevtushenko, “The Concept Explorer” , http://conexp.sourceforge.net/ (2018/8/13)

[13] TechNews，“史上第一勒索蠕蟲 WannaCry / Wcry 大舉入侵，趨勢科技教你週一拒當資安人質”， 2017 年 05 月 15 日，http://technews.tw/2017/05/15/ wannacry-wcry-keep-watch/ (2018/7/28)

[14] 陳鷺人，“勒索病毒襲台 台電也中鏢”，中時電子報, 2017 年 05 月 15 日, https://www.chinatimes.com/realtimenews/20170515003208-260405?chdtv (2018/7/29)

[15] 林建宏、方國定，“正規化概念分析建構電腦病毒特徵之知識本體”，*國立雲林科技大學資訊管理系碩士論文*，2009。

[16] Trend Labs 趨勢科技全球技術支援與研發中心, “勒索病毒 ransomware /勒索軟體”，2015 年 02 月 01 日, https://blog. trendmicro.com.tw /?cat=2267 &paged=8 (2018/8/7)

[17] 戚玉樑，“以本體技術為基礎的知識庫建置程序及其應用”，*資訊科技與社會*，第五卷第二期”. 2005，頁 1~18。

[18] 王平、王宇治，“行動裝置病毒知識本體雛型之建構-以 Android 系統為例”，*第 24 屆國際資訊管理學術研討會* (ICIM2013),真理大學, 2013。

[19] 黃盈豪, 許通安，“以知識本體為基礎建構病毒分類知識庫系統”，*中原大學資訊管理研究所碩士論文*, 2007。

[20] 亞俊、孫宏民，“核心層即時 Android 惡意軟體偵測之研究”，國立清華大學資訊系統與應用研究所碩士論文，2012。

[21] 謝維揚、曾文貴，“ MalCatcher:以存取以及網路洩漏隱私資料行為為基礎的 Android 惡意程式行為偵測”，國立交通大學網路工程研究所碩士論文，2013。

[22] 鄧全良、徐雄健，“以虛擬程式載入器及病毒行為模式分析法的防毒系統”，銘傳大學 資訊工程學系碩士論文， 2004。

[23] 張世杰、洪士灝，“Ape: Android 系統惡意程式之自動化測試環境”，國立臺灣大學資訊工程學研究所碩士論文，2013。

[24] 吳廸、林大為，“惡意程式碼的檢測研究”，健行科技大學資訊工程所碩士論文，2013。

[25] https://buzzorange.com/techorange/2017/05/15/trendmicro-wannacry/. (2019/3/9)

[26] https://applealmond.com/posts/5171

[作者簡介]

王平，交通大學資訊管理研究所博士，現為崑山科技大學資訊管理系教授，同時兼任副研發長，研究方向為深度學習神經網路、資訊安全、網路服務及技術創新與專利佈局之研究。

洪維謙，現為崑山科技大學資訊管理系研究生，研究專長為電腦病毒特徵分析、網路入侵偵測與系統弱點掃描之研究。

林孝忠，中山大學資訊管理研究所博士，現為崑山科技大學資訊管理系助理教授，研究專長為物聯網安全、網路入侵偵測與 SDN 網路之研究。