以異常為基礎之即時通訊惡意網址偵測

官大智
國立中山大學資訊工程學系

陳嘉玫
國立中山大學資訊管理學系

林家賓
國立中山大學資訊工程學系

王則堯*
國立中山大學資訊管理學系

摘要

由於即時通訊（Instant Messaging, IM）的普遍性及立即性，現今已成為駭客散佈惡意軟體（malware）的平台。並且為了躲避防毒軟體的偵測，已較少使用傳送惡意檔案的方式，而是以傳送惡意網址（malicious URL）為目前普遍的擴散途徑。這些惡意網址可能會下載病毒檔案或是連到釣魚網站（phishing website）。一旦使用者被 IM 惡意程式攻陷，惡意網址就會透過受害者的連絡人清單繼續擴散出去，而且有時候還會搭配社交工程的手法，使得收訊者很難判斷此連結是否為惡意。而目前缺乏有效的解決方案，能夠即時地偵測 IM 惡意網址。本研究提出一個即時偵測 IM 惡意網址的方法。此方法基於網址的異常特徵及傳訊者的異常行為，定義了一組行為模式來描述可能的惡意行為，並且利用計分演算法來評估異常特徵的重要性，藉此預測網址是否為惡意。為了增加偵測速度，惡意行為模式可以有效地用來識別已知的惡意網址，另外計分演算法產生的分數模型，可以被用來偵測未知的惡意網址。實驗結果顯示，本研究提出的方法能夠達到低誤警率（false positive rate）和低誤判率（false negative rate）。

關鍵詞：即時通訊、惡意網址、即時通訊蠕蟲

* 本文通訊作者。電子郵件信箱：harry0716@gmail.com
2011/02/18 投稿；2011/08/12 修訂；2011/10/17 接受

Anomaly Based Malicious URL Detection in Instant Messaging

D. J. Guan
Department of Computer Science and Engineering, National Sun Yat-sen University

Chia-Mei Chen
Department of Information Management, National Sun Yat-sen University

Jia-Bin Lin
Department of Computer Science and Engineering, National Sun Yat-sen University

Tse-Yao Wang*
Department of Information Management, National Sun Yat-sen University

Abstract

Instant messaging (IM) has been a platform of spreading malware for hackers due to its popularity and immediacy. To evade anti-virus detection, hacker might send malicious URL message, instead of malicious binary file. A malicious URL is a link pointing to a malware file or a phishing site, and it may then propagate through the victim's contact list. Moreover, hacker sometimes might use social engineering tricks making malicious URLs hard to be identified. The previous solutions are improper to detect IM malicious URL in real-time. Therefore, we propose a novel approach for detecting IM malicious URL in a timely manner based on the anomalies of URL messages and sender's behavior. Malicious behaviors are profiled as a set of behavior patterns and a scoring model is developed to evaluate the significance of each anomaly. To speed up the detection, the malicious behavior patterns can identify known malicious URLs efficiently, while the scoring model is used to detect unknown malicious URLs. Our experimental results show that the proposed approach achieves low false positive rate and low false negative rate.

Keywords: Instant Messaging, Malicious URL, IM Worms

* Corresponding author. Email: harry0716@gmail.com
2011/02/18 received; 2011/08/12 revised; 2011/10/17 accepted

壹、緒論

隨著網路的普及，帶動各種網路應用程式的發展，造成網路的使用為生活不可或缺的一部份。而即時通訊（Instant Messaging; IM）就是已被普遍使用的網路應用程式之一。IM 蠕蟲主要的擴散途徑有兩個，一是直接以檔案傳送的方式，二是傳送網址（Universal Resource Locator; URL）訊息的方式；所謂網址訊息是網際網路上使用的標準資源位址標示，也稱做網頁位址。目前普遍的傳播途徑是以傳送惡意網址的方式為主。這類惡意網址有可能直接就是下載病毒檔案，或者連到藏有病毒檔案的網頁或偷取個人資料的釣魚網站，再利用社交工程的手法，誘騙收訊者點選該網址。

先前的研究（Williamson et al. 2004; Mannan & Oorschot 2005; Xie et al. 2007）都試著解決 IM 蠕蟲擴散及偵測的問題，不過都無法即時地偵測出 IM 惡意網址的存在，本研究主要目的在於，當 IM 使用者收到網址訊息時，希望能夠即時分辨該網址是否屬於惡意。而惡意的定義為，該網址不是由本人親自傳送的，可能是電腦被感染病毒蠕蟲，由惡意程式傳送的，或者可能是 IM 帳號密碼被竊取，而從其他電腦傳送的。同時希望能夠達到即時偵測的效果，使得在使用者收到網址訊息後能立即偵測出是否有問題，讓使用者沒有機會點擊惡意網址連結，也就能減少受害的可能性。另外希望能夠不依賴黑、白名單的幫助，發展較智慧的偵測方法，可以有效辨識已知攻擊外，還能偵測未知的攻擊，並達到較高的偵測率及較低的誤判率。

貳、文獻探討

一、IM 架構

IM 是架構在網路上用來達到即時交流的方法，屬於客戶端－伺服器（client－server）結構。伺服器負責訊息的路徑選擇，而客戶端負責傳送及接收訊息。基於伺服器扮演的角色不同，IM 系統的資料傳送主要有以下兩種方式（Williamson et al. 2004; Hindocha & Chien 2003）：

1. 伺服器代理人（Server Proxy）：伺服器主要掌控所有的通訊內容，也就是任何訊息的傳送都經由伺服器做轉傳。兩個使用者間不需要做直接的連線，可以避免防火牆的阻擋。但是由於都是透過伺服器轉送，可能會造成伺服器負擔太大，而且使用者的隱私也無法得到保護。
2. 伺服器經紀人（Server Broker）：伺服器只負責做雙方連線的建立，減少了伺服器端的負擔及洩漏隱私的危險。現在大部份的 IM 系統都是採用

Server Proxy 的方式，有少數會使用到 Server Broker 的架構，像是 Yahoo! Instant Messenger 就是先以 Server Broker 的方式建立連線，如果建立連線失敗再改以 Server Proxy，透過伺服器來轉傳，而 IM 軟體所提供的檔案傳送功能，伺服器都是扮演 Broker 的角色，建立好雙方的連線後，兩個客戶端就以點對點（peer-to-peer）的方式傳送檔案。

二、IM 網路拓撲

IM 系統形成的網路拓撲（topology）可將之視為圖（graph），其中圖的節點為各個 IM 使用者，有邊相連表示兩個使用者互相出現在彼此的連絡人清單內。而已有研究指出，IM 網路具有無尺度網路（scale-free network）的性質（Williamson et al. 2004; Smith 2002; Morse & Wang 2005）。所謂無尺度網路，其節點連結的分佈遵循冪次定律（power law）：任何節點與其他 k 個節點相連的機率，與 $1/k^n$ 成正比，其中 n 通常為接近於 2 的常數（Barabasi & Bonabeau 2003）。也就是說，連絡人清單大小只有某使用者一半的人，在 IM 網路中數量卻有該使用者的 4 倍之多。簡單地說，擁有較多連絡人的使用者佔少數，大多數的使用者都具有較少的連絡人。其連結分佈的情形類似圖 1 所示，圖 1 左可見，大部份的節點只有少數連結，而少數節點則擁有大量連結，就這種意義來說，此網路是「無尺度」的，這種網路的定義特性是，若將連結的分佈取對數畫在雙對數坐標軸上，結果會成一直線，如圖 1 右（Barabasi & Bonabeau 2003）。

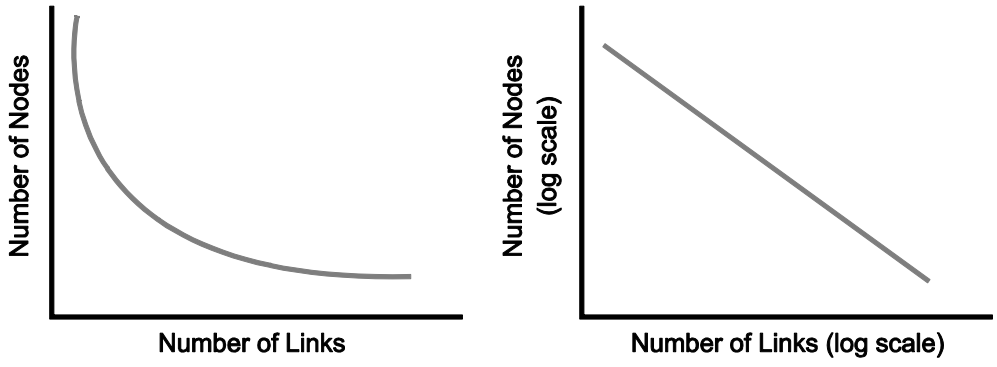


圖 1：節點連結數的冪次分佈圖（e.g., Barabasi & Bonabeau 2003）

有學者發現，在無尺度網路裡並不存在傳染病的臨界值（Pastor-Satorras & Vespignani 2001）。也就是說，電腦病毒在 IM 網路中，可以擴散得非常快速，而且存活時間也可以相當得久。因為 IM 網路的集散節點（hub）會連到許多其他節點，也就是擁有大量的連絡人，當一個使用者被感染病毒，他至少會感染到一個

集散節點，一旦集散節點被感染後，病毒便會傳播給網路內大部份的使用者，當中也有可能包含其他的集散節點，由此可知電腦病毒在 IM 網路中擴散是相當迅速的。同時 Hindocha 與 Chien（2003）指出，感染 50 萬台機器平均只需花費 1 分鐘的時間。Williamson、Parry 與 Byde（2004）的實驗指出，大約 10 秒左右，即可感染整個包含 710 個節點的網路之大部份的機器。Morse 與 Wang（2005）針對包含 2309 個節點的 AIM 網路，也只需相當短的時間，即可感染整個網路。因此再一次說明了，透過 IM 網路來擴散病毒，可以在短時間內感染所有的機器。

三、IM 蠕蟲

IM 蠕蟲（IM worms）是網路蠕蟲的一種，網路蠕蟲是一種透過網路傳播的惡意程式碼，它與病毒的不同在於，網路蠕蟲可以不需使用者的幫助，而自行複製本身進行擴散。至於 IM 蠕蟲就是利用 IM 系統和 IM 通訊協定的特性和漏洞進行攻擊，並在 IM 網路中擴散的網路蠕蟲（eg., Mannan & Oorschot 2005）。IM 蠕蟲的傳播機制主要有以下四種（e.g.,卿斯漢等 2006; Mannan & Oorschot 2005）：

1. 惡意檔案傳送（需要使用者互動）：IM 蠕蟲直接將蠕蟲副本以檔案形式傳送給使用者，利用社交工程手法誘騙使用者接收並執行該檔案。惡意檔案通常會偽裝成不同的檔案格式，像.zip、.exe、.scr、.pif、.jpg 等，來試圖躲避防毒軟體的偵測。不過由於該機制需要使用者的參與，只要使用者有足夠的安全防範意識，而不執行接收到的惡意檔案，便不會遭受 IM 蠕蟲的感染。
2. 惡意網址訊息（需要使用者互動）：IM 蠕蟲傳送包含網址的文字訊息給使用者，此網址可能直接連結到一 IM 蠕蟲檔案，或指向 IM 蠕蟲在受害者主機上建立的網站服務位址，同時也利用社交工程手法，誘使接收者點選該惡意網址。此機制雖然需要使用者互動，但惡意網址通常能夠躲避防毒軟體的偵測，因此近年來，駭客大多以傳送惡意網址的方式來攻擊，這也是本研究主要探討的部份。
3. IM 客戶端程式的漏洞 （不需使用者互動）：IM 蠕蟲利用客戶端的漏洞，在受害者的主機上執行 shellcode，可達到不需使用者參與而順利在遠端執行惡意程式。例如，2004 年 ICQ 的 sound scheme file location 漏洞，可以允許駭客利用 IE 已知的漏洞，在使用者的電腦執行惡意腳本（e.g., secunia 2009）。2005 年 MSN 的 PNG 漏洞和 GIF 漏洞，均被證實可以讓 IM 蠕蟲在沒有使用者參與的情況下，傳輸並執行惡意檔案。
4. 作業系統的漏洞（不需使用者互動）：IM 蠕蟲基於目標使用者所在主機的作業系統漏洞，執行 shellcode 並建立通道傳送蠕蟲副本，同樣可以達

到不需使用者的互動，而在遠端執行惡意程式碼。例如，Bropia.F 蠕蟲就利用 Windows 作業系統一些已知的漏洞，像 SQL、IIS 或 WebDAV 等著名的緩衝區溢出漏洞，來發動攻擊或散播其他蠕蟲程式。

參、研究方法

本研究主要目標在於即時偵測經由 IM 軟體接收到的惡意網址，此類惡意網址通常是從合法的 IM 帳號送來，其內容可能誘使接收者下載病毒檔案或是竊取 IM 帳號密碼的釣魚網站。雖然駭客經常搭配社交工程技巧使用，不過對於警覺性高的使用者仍然可以察覺出此網址訊息有問題，因此透過觀察收集到的對話記錄樣本，可以找出惡意網址訊息與正常網址訊息差異之處，其中包括網址及傳訊者行為的異常。根據找到的特徵，發現有些特徵同時存在的話，其屬於惡意的可能性相當大，因此可以利用一些已知的惡意特徵組合，先判斷接收到的網址訊息是否具有這些特徵，以加快整體的偵測速度。同時使用一些啟發式的特徵，並根據特徵的顯著性，設定相對的分數值，用以預測一個網址訊息是否為惡意。接下來先介紹本研究的系統架構及流程。

一、系統架構與流程

本偵測系統由三個主要模組組成，如圖 2 所示：（一）特徵擷取模組（Feature Extraction Module），（二）惡意行為比對模組（Malicious Behavior Matching Module）和（三）異常特徵計分模組（Anomalous Feature Scoring Module）。

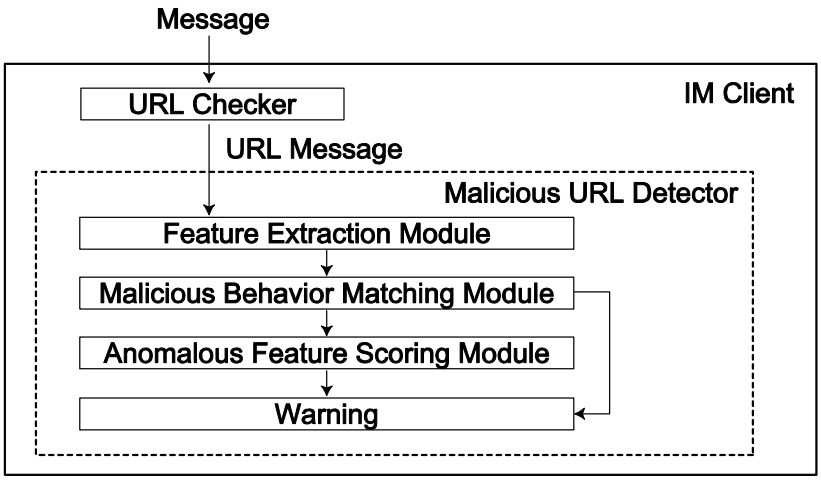


圖 2：IM 惡意網址偵測架構

本系統偵測流程如圖 3 所示，當 IM 使用者收到訊息時，會先判斷此訊息內是否包含網址，如果有，則表示可能為惡意攻擊的網址，接著取出相關的特徵來判斷。偵測的第一階段會先以事先定義好的惡意行為模式來做比對，只要此網址訊息滿足任何一個惡意行為，即被認定為惡意，便發出警告。否則，再經由第二階段的分數模型來做預測，此分數模型會先基於一組訓練資料做訓練，並給予門檻值，若此網址訊息得到的總分小於等於門檻值的話，系統將判定為惡意，同樣送出警告訊息。

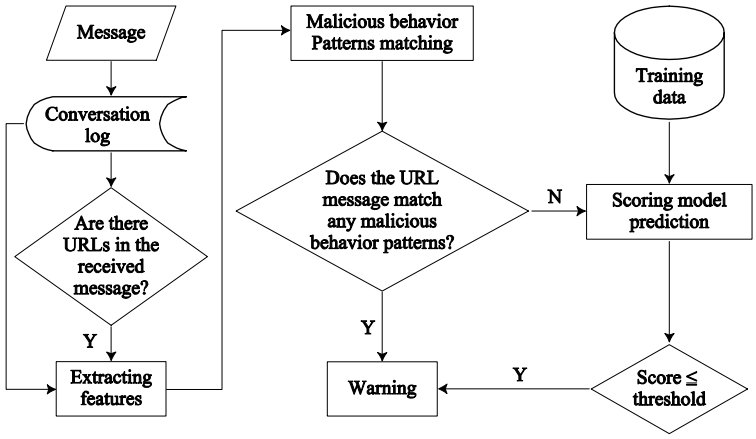


圖 3：IM 惡意網址偵測流程圖

（一）特徵擷取模組

藉由觀察包含惡意網址的對話記錄及正常的網址訊息記錄，找出可以用來判斷是否異常的特徵。觀察的對話記錄來自於區域網路內的閘道端監控設備收集，加上使用者的 IM 對話記錄，而這些對話記錄以一天為單位，記錄兩人所有的訊息內容，記錄的欄位包括日期、時間、傳訊者帳號、收訊者帳號和訊息內容。另外也從網路上搜尋以往曾經出現過的 IM 惡意網址，以便了解更多的惡意網址攻擊模式。所擷取的特徵可以分為行為相關特徵、網址相關特徵兩類，與，並分別詳細描述於下：

1. 訊息文字包含 IM 帳號（ID in Text），定義為特徵 F1：通常惡意網址訊息的文字部份會包含傳訊者或收訊者的帳號。例如：傳訊者（ciei@livemail.tw）送出的訊息為“ciei check out these awesome pics from the awesome party LOL http：//bingchilin.gone-wild-patry-pics.com”，其中文字部份包含傳訊者的使用者名稱“ciei”這是惡意網址訊息常見的情形，其意圖是想利用社交工程的手法，引誘收訊者點選該惡意網址。所以，此特徵會檢查網址訊息的文字部份是否包含傳訊者或收訊者的帳號

或使用者名稱。

2. 首次傳送的網址訊息（First URL Message），定義為特徵 F2：大部份惡意網址都是只有一句訊息，沒有其他前後文，加上通常惡意網址都會從久未連絡的好友帳號傳過來，因此，檢查收到的網址訊息是否為傳訊者一天當中第一次傳送的訊息，如果第一次傳送的訊息就有包含網址，則就有可能是惡意的網址，故將此當成判斷異常與否的特徵之一。
3. 網址包含 IM 帳號（ID in URL），定義為特徵 F3：為了讓惡意網址看起來像是合法的，駭客通常會將 IM 使用者名稱當成某個網域的子網域（sub domain），或出現在網址的路徑名稱（path name），或是詢問字串（query string）中。例如：http://username.pri-party-pic.com 有時候還有可能出現電子郵件形式的 IM 帳號，例如：http://username@hotmail.com.fddcol.com 這樣的目的都是為了讓收訊者誤以為該網址是合法的，所以該特徵記錄網址是否包含傳訊者或收訊者的使用者名稱或電子郵件位址。
4. IP 形式的網址（IP-based URL），定義為特徵 F4：不管正常或惡意，都有可能出現以 IP 表示的網址，但是若將 IP 經過編碼處理的話，就是屬於惡意行為。而常見的編碼方式有，以十六進位編碼，例如：http：//140.117.169.165 為一個正常形式的 IP，編碼後變成 http：//0x8C.0x75.0xA9.0xA5，其中每個十進位數字以 0x”後面接著十六進位表示。或者以 DWORD （double word）長整數表示，也就是將一個 IP 轉成 2 進位後再以長整數的形式表示，例如：http：//140.117.169.165，可以做這樣的計算（140*224+117*216+169*28+165*20）得到一 DWORD IP：http：//2356521381。因此，只要 IP 為非一般標準的表示式，都將視它為惡意的。
5. 令人混淆的網址（Confused URL），定義為特徵 F5：有些惡意網址會偽裝成看起來像轉址到另一個合法網站，或偽裝成與合法網站有關，但實際上仍然是連結到惡意的目的地，這也是一般釣魚網站常用的手法。例如：http://sdmeccanica.com/lndex.php?SignIn&ru=http://www.ebay.com.au/&trksid=m37 是一個 eBay 的釣魚網站。此項特徵主要檢查網址除了最前面的“http://”和主機名稱（hostname）之外，是否包含額外的“http（s）：”或“www.”字樣。
6. 網域的信譽（Reputation of Domain），定義為特徵 F6：此特徵將藉由 Google 搜尋來評估一個 domain 的信譽。以網址的域名（domain name）當做搜尋的關鍵字，檢查搜尋結果是否有與關鍵字相同 domain 的網站來做依據。因為大部份惡意網址的存活時間較短且缺少其他網站連向它們，所以利用 Google 搜尋惡意網址的網域名稱，通常不會出現在排序較

前的搜尋結果中。例如：圖 4 左是一個惡意網址，http：//host.piclooks.com/?wriv01 取其域名“piclooks.com”為關鍵字，而搜尋結果前幾筆的網址都沒有與關鍵字相同的網域。反之，一個合法的網址：http：//www.nsysu.edu.tw（如圖 4 右），在搜尋結果的前幾筆都有相同域名（nsysu.edu.tw）的網址。所以，此特徵僅比對搜尋結果的前兩筆，檢查是否有相同網域的網址，以此判斷一個網域的信譽。

圖 4：惡意域名“piclooks.com”的與合法域名“nsysu.edu.tw”的搜尋結果

7. 傳訊者延遲時間（Delay Time of Sender），定義為特徵 F7：學者 S.Gianvecchio et al.（2008）發現，依照傳訊的機制來看，聊天機器人可以分為，固定時間傳送（periodic bots）、隨機時間傳送（random bots）和回應特定內容的傳送（responder bots），其中 periodic bots 會以相同的時間間隔傳訊，也就是傳訊的延遲時間是固定的。

   為了識別出這種 periodic bots，對傳訊者延遲時間這個特徵，將檢查每個值是否一樣，但由於可能因為頻寬的關係，收訊會有延遲的情形，所以容許有一秒的誤差，也就是說每一次傳訊的時間差都跟前一次最多相差 1 秒的話，稱此傳訊者的延遲時間是規律的（regular）。對於正常使用者而言，延遲時間是較不規律的，依此可清楚辨識出惡意的傳訊行為。
8. 傳訊者回應時間（Response Time of Sender），定義為特徵 F8：類似特徵 F7，有一種機器人它傳訊的發生是基於收訊者的動作，它會先傳訊息給收訊者，並等待回應，如果收訊者沒回應，它就沒有後續動作，如果有回應了，它才會再傳送訊息，而且該機器人回應收訊者的時間間隔都是固定的。因此將檢查回應訊息的時間間隔是否相同，所以對於回應時間有規律性的傳訊者，其傳送的網址將會視它為惡意的。

9. 網域註冊天數（Age of Doamin），定義為特徵 F9：駭客們會註冊許多的網域來配置攻擊來源，以躲避黑名單阻擋，故此類網域的存活時間通常都很短。根據 APWG 於 2008 年上半年的統計報告指出，一個釣魚網站的存活時間平均為 49.5 小時，其中位數為 19.5 小時（e.g., Group 2009）。我們假設駭客們使用的惡意網域，其生命週期為 49.5 小時，約 2 天。我們透過 WHOIS 資料庫檢查某網址，若其網域註冊時間低於 2 天（Age of Domain ≤ 2），則視為惡意的。同樣地，這個特徵在釣魚網站偵測的相關研究中，也是常被使用到的一個特徵（e.g., Zhang et al. 2007; Fette et al. 2007; Basnet et al. 2008; Pan & Ding 2006）。
10. 主機名稱的“-”個數（Dashes in Hostname），定義為特徵 F10：惡意網址經常以“-”連接多個字，成為一個較長的字串當做網域，而且有時候這個較長的字串看起來像是一個句子，例如：http：//wriv01.real-cool-newyear-party-pics.com，其中“-”有 4 個。所以，此特徵將計算在主機名稱當中出現的“-”個數。
11. 最長的網域標籤（Longest Domain Label），定義為特徵 F11：網域名稱（domain name）由多個網域標籤（label）組成，每個標籤之間以句點（.）隔開。擷取此特徵是因為，有些惡意網址會使用較長的字串當網域，有別於一般正常的使用，而且過長的網域也較不容易記憶，例如：http：//31837.hzaseruijintunhfeugandeikisn.com/5/54878/，其最長的網域標籤為 28。因此，該特徵會記錄一個網址的域名中，標籤長度最長是多少。

表 1：特徵表

特徵	對應欄位	特徵值
F1	Username in Text	{0,1}
F2	First URL Message	{0,1}
F3	Username in URL	{0,1}
F4	IP-based URL	{0,1}
F5	Confused URL	{0,1}
F6	Domain in Google Search Result	{0,1}
F7	Entropy of Delay Time	{（P_i）,-1}
F8	Entropy of Response Time	{（P_i）,-1}
F9	Age of Domain	{D_i,-1}
F10	Dashes in Hostname	{x≥0}
F11	Longest Label in Hostname	{x>0}

（二）惡意行為比對模組

實際進行偵測實驗時，發現部分特徵的組合，對於惡意樣本有較高的偵測率，在進行樣本偵測時，先以部分特徵組合，作為初步的惡意行為比對，將可提高整體偵測效能。以下為本研究定義的 2 種惡意行為比對模組：

1. 訊息內含網址且文字訊息中包含使用者名稱（特徵 F1+特徵 F2）：經由觀察，含有惡意網址的通訊經常只有單獨一句訊息，並在訊息中包含使用者的名稱，使警覺心較低的收訊者，看到自己熟悉的名稱，就認為無害。雖然，在正常的通訊中，分享網址訊息的情況也不少，但特意含有使用者名稱的情況，仍屬異常。故我們認為，訊息內含網址且內容包含 IM 使用者名稱，並且還是一天之中首次傳送的訊息，該訊息網址為惡意連結的可能性極高。
2. 網址包含使用者名稱且網域信譽低（特徵 F3+特徵 F6）：部分提供個人服務的網站合法網站，如網路相簿、網頁空間、部落格服務等，會以會員的帳號建立一個唯一的網址供個人使用，有些將帳號當成該網域的子網域，或是把帳號當成網址的路徑名稱（例如，http://username.blogspot.com 是提供個人部落格服務的網站，http://www.wretch.cc/album/username 是提供個人相簿服務的）。

   在惡意網址中，也經常出現“網址包含使用者名稱”這種情況，是攻擊者企圖偽裝成合法網址，誘使受害者上當。在“網址包含使用者名稱”此一情況下，為了區別出合法與惡意網址，則加上檢查該網域的信譽，作為輔助。利用 Google 搜尋可疑網址的域名，比對前兩筆搜尋結果的域名是否與該域名相同，基本上，一個合法網址的域名都是會出現在前幾筆搜尋結果的。因此，當接收到含有使用者名稱的可疑網址，先以 Google 搜尋其域名，若搜尋結果沒有出現相同網域的網址，則視此網址為惡意的。

以上為本研究所定義的 2 個惡意行為比對模組，用以輔助系統加速偵測的效能，最後再將無法判定的樣本，再以本研究提出的異常特徵計分模組來加以判斷。

（三）異常特徵計分模組

本研究提出的異常特徵計分模組，針對未被惡意行為比對模組偵測出惡意的樣本，再次進行分析。此模組依系統訓練結果（系統先以一組包含良性及惡意樣本的 training dataset 施以訓練），分別給予 11 個特徵分數值，分數具有兩種指標意涵；當取分數的絕對值時，其值越高，該特徵的偵測效果越好，反之，則偵測效果較差；觀察分數的正負值表現時，當特徵為正值，則表示該特徵對樣本的判

斷，傾向為良性；若為負值，則表示該特徵對樣本的判斷，傾向為惡性。當進行實際樣本分析時，系統將分析得到的 11 個特徵的分數加總，與本研究所訂定的門檻值加以比對，若小於等於門檻值，則將該樣本視為惡意。

下列為我們提出的異常特徵計分演算法：

每個特徵都會得到一個特徵對應函數，

$$F_i(x_{ij}) = S_{ij}，$$

其中 x_{ij} 為特徵 F_i 可能的結果，其函數值 S_{ij} 為不同特徵值 x_{ij} 對應的分數。對於待測的樣本將檢查 11 個特徵的值，並給予相對應的分數，最後得到一個總分：

$$S = \sum_{i=1}^{11} S_{ij}$$

並設定門檻值為 0，若 $S \leq 0$ 則該網址就被視為是惡意的。

實際觀察樣本時發現，某些特徵值可能的結果太多，導致資料分佈過於分散，使得該項特徵值的表現無顯著性（例如，F_7：Entropy of Delay Time）。此時將先對所有的特徵值做分群的動作，改以不同的群組計算相對應的分數。或者有些特徵的結果，已知效果不會太明顯，也就是良性樣本或惡意樣本中，都有可能出現該特徵值（例如，當 F_1：Username in Text=0 時，良性或惡意樣本中，都會出現文字訊息不包含使用者名稱的情況，於是只對特徵值等於 1 時才算分，而特徵值 0 就相當於給 0 分）。因此計算分數前，需預先決定要算分的特徵值為何，以 $R_i=\{r_{i1}, ...\}$ 表示對特徵 F_i 欲算分的特徵值集合，集合中的元素 r_{ik} 可能是某特徵值群組或可能只有一個特徵值。

接著準備一組訓練資料，包含良性樣本 B 和惡意樣本 M，因為我們將評估一個特徵在兩類不同資料中的顯著程度，將選擇相同樣本數的兩類資料，分別以 N_b 代表良性樣本數，N_m 代表惡意樣本數，則 $N_b=N_m$。計分演算法 Score(B，M，R_i) 將以兩類的訓練樣本及欲算分的特徵值集合 R_i 當做演算法的輸入，其輸出 R_i 中的 r_{ik} 相對應的分數 s_{ik}，針對第 i 個特徵 F_i，演算法過程如圖 5 所示。

Score (B, M, R_i)

Input: B, M, $R_i = \{r_{ik}\}$

Output : s_{ik}

for all r_{ik} do

1. 計算 B 中符合特徵值 r_{ik} 的樣本數：n_b
2. 計算 M 中符合特徵值 r_{ik} 的樣本數：n_m
3. 計算 r_{ik} 的分數

$$s_{ik} = \frac{n_b - n_m}{N_b (= N_m)}$$

4. 設定特徵對應函數 $F_i(r_{ik}) = s_{ik}$

圖 5：計分演算法

演算法依序對R_i中的每個特徵值組 r_{ik} 計算，分別在良性資料及惡意資料中有多少樣本有 r_{ik} 這組特徵值，記為 n_b 和 n_m，而因為一項特徵效果最好的情形發生在，某一類資料全部都有此項特徵，而另一類資料全都沒有此項特徵，因此，接著計算該組特徵值的顯著性$n_b - n_m$，其絕對值越大，表示兩類資料分別具有該特徵值的樣本數相差越大，也就是說該特徵值效果越好，當$|n_b - n_m| = N_b(N_m)$時效果最好，而其值如果為正，表示有較多的良性樣本擁有此特徵值，也可以說此特徵值屬於良性的特徵；如果其值為負，表示有較多的惡意樣本有此特徵值，固屬於惡意特徵；若其值為零，表示有此特徵值的良性與惡意樣本數一樣，可以表示此項特徵值效果為零。最後計算該特徵值的顯著性對於最佳效果值的比例，也就是$s_{ik} = (n_b - n_m)/ N_b$，$s_{ik}$ 就是該組特徵值 r_{ik} 的分數，此分數值會介於+1 到−1 之間，一個分數為+1 或−1 表示該特徵值效果最好，接著設定特徵對應函數$F_i(r_{ik}) = s_{ik}$，所以，對待測網址訊息，檢查其 11 個特徵，並給予每個特徵值相對應的分數，計算總分：

$$S = \sum_{i=1}^{11} F_i(r_{ik}) = \sum_{i=1}^{11} s_{ik}$$

若總分 $S \leq 0$，則將此網址視為是惡意的。

表 2：計分演算法範例

j	x_{0j}	f_b	f_m	k	r_{0k}	n_b	n_m	n_b-n_m	s_{0k}
1	0	5	0						
2	1	14	1	1	0~2	28	5	23	23/30
3	2	9	4						
4	3	2	17	2	3~4	2	25	-23	-23/30
5	4	0	8						

以表 2 為例說明此演算法，假設特徵F_0可能的結果有$\{x_{01}, x_{02}, \ldots, x_{05}\}=\{0, 1, \ldots, 4\}$，訓練資料數$N_b=N_m=30$，每個可能的值分別計算在 B 和 M 中有多少樣本具有此特徵值，分別列於表中f_b和f_m欄位，又假設現在不想對每個不同的特徵值給予分數，想要先將原有的特徵值做分組處理，一個集合$R_0=\{r_{01}, r_{02}\}=\{0\sim2, 3\sim4\}$表示想要算分的特徵值組集合，現在將$x_{0j}$分成特徵值 0~2 一組，3~4 一組，再分別計算相對應的分數，針對特徵值範圍 0~2 和 3~4 分別計算在 B 和 M 中相符的樣本數，列於表中n_b和n_m欄位，在 B 中有 28 筆樣本其F_0這個特徵的值為 0、1 或 2；有 2 筆樣本其值為 4 或 5，在 M 中有 5 筆樣本其F_0的特徵值為 0~2；有 25 筆其值為 4~5。接下來計算n_b-n_m的值，分別為 23 和－23，最後計算r_{01}和r_{02}所對應的分數s_{01}和s_{02}，其$s_{01}=23/30=0.77$，$s_{02}=-23/30=-0.77$，並記錄該特徵對應函數：

$$F_0(x)=\begin{cases}0.77 & \text{if } 0\le x\le 2\\ -0.77 & \text{if } 3\le x\le 4\end{cases}$$

其中 x 為特徵F_0可能的值。

（四）計分特徵值選擇策略

在計分演算法中，算分前需決定每個特徵$\mathbf{F_i}$欲算分的特徵值集合$\mathbf{R_i}$，接下來將說明本研究所選擇的$\mathbf{R_i}$，而那些沒有在$\mathbf{R_i}$中的特徵值，相當於直接以 0 分來看，其每個特徵的$\mathbf{R_i}$如表 3 所示。

表 3：$\mathbf{R_i}$的選擇策略

Features	Selected R_i
F_1: Username in Text	R_1={1}
F_2: First URL Message	R_2={1}
F_3: Username in URL	R_3={1}
F_4: IP-based URL	R_4={1}
F_5: Confused URL	R_5={1}
F_6: Domain in Google Search Result	R_6={0}
F_7: Entropy of Delay Time	R_7={{0~A_{7m}},{A_{7m}~A_{7b}},{A_{7b}~}}
F_8: Entropy of Response Time	R_8={{0~A_{8m}},{A_{8m}~A_{8b}},{A_{8b}~}}
F_9: Age of Domain	R_9={{0~30},{30~60},...{330~360},{360~}}
F_{10}: Dashes in Hostname	R_{10}={All appeared values}
F_{11}: Longest Domain Label	R_{11}={All appeared values}

特徵F_1~F_5皆為二元特徵，而其值為 0 時，表示沒有出現該特徵，而沒有這些特徵的網址訊息是良性的或惡意的可能性都很大，因此將不考慮這些特徵值為 0 時的情形，而直接給予 0 分，反而在特徵值為 1 時，其顯著程度可能會比較大，所以針對F_1~F_5只計算特徵為 1 時的分數。

特徵F_6檢查待測的域名是否能與 Google 搜尋結果的前兩筆域名相符。經由觀察發現，有些惡意網址的域名會使用可免費申請的域名，而這些提供免費域名服務的網站，所註冊的域名可能都有一段時間了，當利用 Google 搜尋該域名，會符合前幾筆搜尋結果出現該域名的結果，將使得F_6=1 時，無法確切判定為惡意或是良性。因此，將只計算F_6=0 時的分數，F_6=0 相當於給 0 分。

F_7和F_8是計算傳訊者延遲時間和回應時間的 entropy，而這些特徵可能的值太多，所以會先將所有的特徵值分組，分組的依據為，所有良性訓練資料 entropy 的平均值A_b，和所有惡意訓練資料 entropy 的平均值A_m，若$A_m < A_b$，則將原本特徵值分成三組，{0~A_m，A_m~A_b，A_b~}，原本特徵值大於 0 且小於等於A_m的為一組，大於A_m且小於等於A_b為一組，剩下大於A_b的視為一組，反之亦然；若A_m=A_b時，則只分成兩組。

F_9是計算網域已註冊多少天，分組以 30 天為間隔單位，將所有特徵值分成若干組{{0~30}，{30~60}，...{330~360}，{360~}}，大於 360 天的則視為一組。最後特徵F_{10}和F_{11}也是屬於連續的特徵，但可能的值不算太多，所以這兩個特徵將對所以可能的特徵值算分，其R_{10}={x_{10j}}，R_{11}={x_{11j}}。

肆、研究結果

一、樣本收集

為了評估本研究的偵測方法，需要收集兩組分別包含良性網址及惡意網址的 IM 對話記錄。由於並無公開的測試資料可用，因此要收集夠多的樣本是有些許困難。透過某單位網管部門協助，在不違反道德原則的條件下，以學術研究的目的，於閘道端設置監控設備，收集區域網路內使用者的 IM 對話記錄。本研究只挑選出內容含有網址訊息的對話記錄，這些記錄是以兩個使用者一天的對話內容為單位，記錄著傳訊日期、時間、傳訊者帳號、收訊者帳號及訊息內容。

另外，為了增加樣本數量，我們自行建立幾個新的 IM 帳號，做為誘捕惡意樣本之用。以虛擬機器（VMware）執行這些誘捕帳號的客戶端程式，並將這些誘捕帳號加入正常使用的 IM 帳號之連絡人中，然後在虛擬機器內執行 IM 病毒程式或輸入帳號、密碼至 IM 釣魚網站，藉此來收集更多的惡意樣本。

將所收集到的樣本，以人工分類方式分為良性及惡意。分類的依據，以親自拜訪該網址網頁內容及觀察對話內容為主，有些已失效的 URL 則利用 Google 搜尋相關資訊，進一步確認該失效的 URL 是否屬於惡意或良性。總計收集了 202 個包含良性網址的對話記錄及 222 個包含惡意網址的對話記錄。我們的樣本使用的系統主要以 MSN 及 YIM 為主，其中又以 MSN 為大宗，相關樣本個數列於表 4。

表 4：對話記錄樣本數及實驗資料個數

Class	Chat logs			Experimental data
	MSN	YIM	Total	
Benign	193	9	202	356
Malicious	186	36	222	315

由於每個對話記錄中包含不只一句網址訊息，所以事先擷取每一句網址訊息及其相關的特徵，存到資料庫中當做後續實驗的資料。本研究使用 Perl 程式語言撰寫擷取特徵並存入資料庫的程式，使用 MySQL＋phpMyAdmin 來管理資料庫。於是資料庫中共有 356 筆良性網址及相關特徵的資料和 315 筆惡意網址及相關特徵的資料，而這些資料將是之後實驗的依據。

二、實驗評估

為了評估本研究偵測方法的效能，將以兩個實驗來做量測，第一個實驗將針對不同的訓練資料個數，分別執行多個回合取其結果平均值，以檢視本方法偵測效果的影響。第二個實驗將以相同的測試資料，測試一些聲稱能夠偵測惡意網頁和釣魚網站的瀏覽器工具列，並與本方法做比較。

而通常評估系統效能有以下幾個準則，如表 5 所示。本研究定義良性的網址為 positive，惡意的網址為 negative，當給定一良性網址，若系統錯誤判斷成惡意的，則稱為 false positive（FP）；若否，則稱為 true positive（TP）。而給定一個惡意網址，若系統錯誤判斷成良性的，則稱為 false negative（FN）；否則，稱為 true negative（TN）。

表 5：評估系統效能之準則

Before testing	As benign after testing	As malicious after tesing
Benign (positive)	true positive (TP)	false positive (FP)
Malicious (negative)	false negative (FN)	true negative (TN)

本研究將使用兩項指標來評量系統效能，False Positive Rate （FPR）和 True Negative Rate（TNR），分別定義如下：

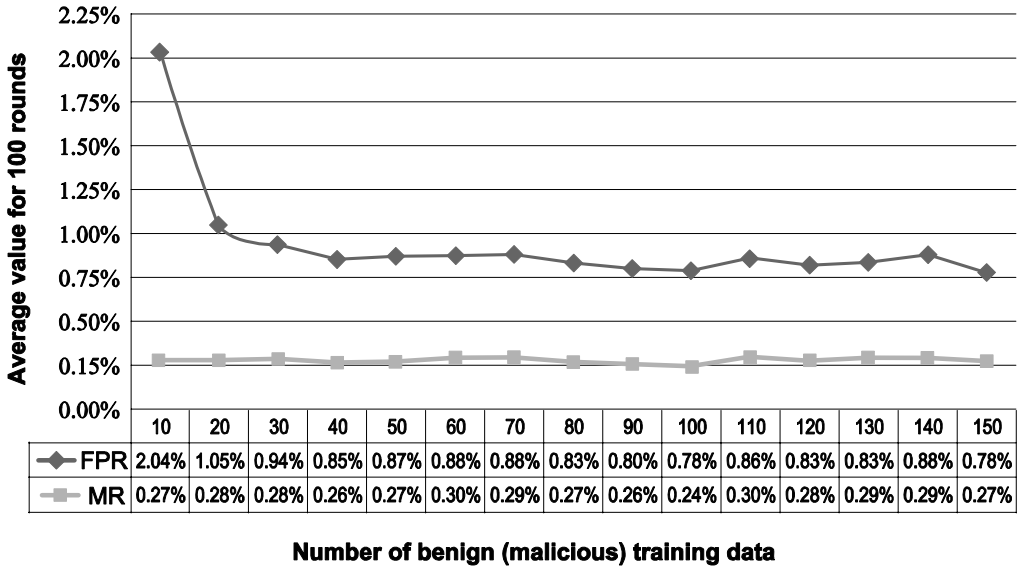
$$\text{FPR} = \frac{\text{FP 的個數}}{\text{所有良性的測試資料個數}}$$

$$\text{TNR} = \frac{\text{TN 的個數}}{\text{所有惡意的測試資料個數}}$$

三、實驗一：不同訓練資料個數之影響

此實驗將以不同個數的訓練資料集，來測試本方法的偵測效果，每組訓練資料包含相同個數的良性資料和惡意資料。將分別從 356 筆良性資料和 315 筆惡意資料，隨機各取 10，20，...，150 筆資料當訓練資料集，其餘資料當測試資料集，每次測試將執行 100 回合後取其 FPR 和 TNR 的平均值。另外同時記錄測試過程中，被第一階段的惡意行為比對到的資料個數，以檢視所定義的惡意行為模式，是否能夠識別出惡意網址，而不會錯將良性網址判斷成惡意。

圖 6 顯示不同的訓練資料集之平均 FPR 值，由結果可看出，當訓練資料集個數太少時，如：良性和惡意的資料各取 10 筆，其 FPR 值會偏高一點，因為偵測方法的第二階段，特徵值分數的設定代表該特徵值的效果，且與良性和惡意的資料個數有關，若訓練資料個數偏少，得到的分數可能不夠表示特徵值真正的效果，因而導致 FPR 較高，不過平均 FP 個數大約只有 7 個而已。

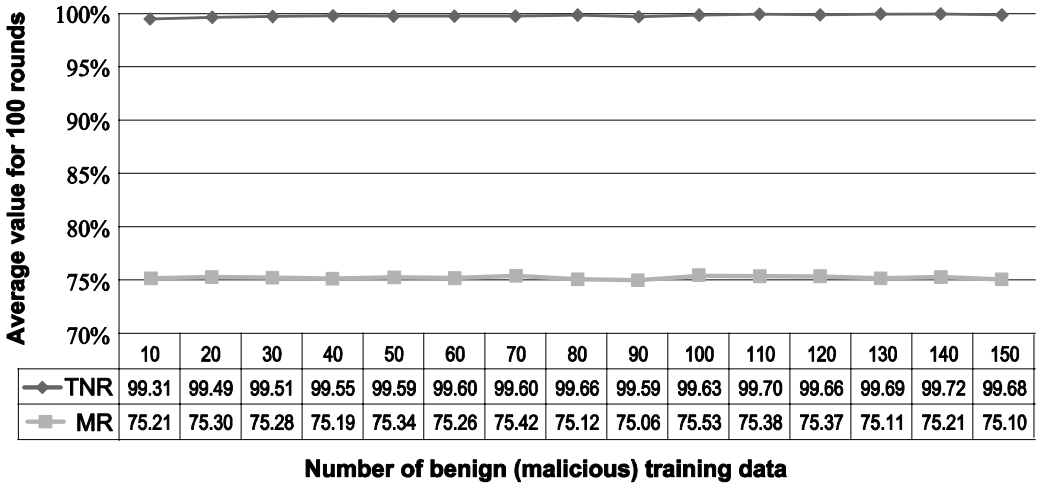


	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
FPR	2.04%	1.05%	0.94%	0.85%	0.87%	0.88%	0.88%	0.83%	0.80%	0.78%	0.86%	0.83%	0.83%	0.88%	0.78%
MR	0.27%	0.28%	0.28%	0.26%	0.27%	0.30%	0.29%	0.27%	0.26%	0.24%	0.30%	0.28%	0.29%	0.29%	0.27%

圖 6：不同訓練資料集的平均 FPR

本研究建議，訓練資料集的兩類資料至少各選 30 筆，即可達到平均 1%以內的 FPR，約從 0.78%到 0.94%。另外在圖 6 中的 matched 欄位，顯示被第一階段惡意行為比對到的部份，在每次測試中平均有 0.24%到 0.3%的良性資料符合某個惡意行為，不過實際上只包含一筆良性資料被誤判，該資料剛好符合惡意行為中的“Regular Delay Time”。就整體來說，FPR 已達到可接受的效果。

圖 7 則顯示平均 TNR 的結果，同樣最差的情形發生在訓練資料集只各取 10 筆資料的時候，不過平均也只有 2 個 FN。而當兩類資料各取 30 筆以上當訓練資料集，則 TNR 平均值約從 99.51%到 99.72%，可見對於惡意網址的判別有不錯的效果。另外被惡意行為比對到的資料個數，在每次測試中平均都有 75%的惡意網址被識別出來，再次說明偵測方法中的第一階段，可以有效判斷惡意網址的出現，而第二階段的計分演算法，也能正確的預測那些沒有直接符合惡意行為的網址，是否為惡意的。



	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
TNR	99.31	99.49	99.51	99.55	99.59	99.60	99.60	99.66	99.59	99.63	99.70	99.66	99.69	99.72	99.68
MR	75.21	75.30	75.28	75.19	75.34	75.26	75.42	75.12	75.06	75.53	75.38	75.37	75.11	75.21	75.10

圖 7：不同訓練資料集的平均 TNR

四、實驗二：與瀏覽器工具列之比較

第二個實驗將針對一些聲稱可以偵測惡意網站或釣魚網站的瀏覽器或瀏覽器工具列，檢驗其偵測的效能，為了公平起見，將使用相同的測試資料集做測試，並與本方法做比較。測試環境在 Windows Vista 作業系統平台。

實驗方式將重新對良性和惡意資料隨機各選 50 筆當作本方法的訓練資料集，剩下的資料則為測試資料集，包含 306 個良性資料和 265 個惡意資料，而測試資料集的網址即為用來測試瀏覽器工具列的來源。為了符合真實的情形，先將每個網址複製到 IM 軟體中，然後再點選該網址，此動作是為了模擬當 IM 使用者收到一網址時，直接點選此網址且自動開啟瀏覽器瀏覽。而在測試過程中發現，大部份的惡意網址其實都已失效，但基於黑名單機制的工具列，應該仍然可以檢查該網站曾經分析過的結果，唯有基於啟發式的 SpoofGuard 工具列可能會受到影響，不過由於是使用相同的測試資料與本方法做比較，所以仍可評估這些瀏覽器工具列相對的效能。

值得注意的是，SpoofGuard 工具列測試前會先檢查使用者的瀏覽歷史記錄，如果目前測試的網站是曾經瀏覽過，則它會假設該網站是安全的，而直接以綠色圖示表示。因此，本實驗測試將在 IE 8 提供的 InPrivate 瀏覽模式下執行，在此模式下可以阻止 IE 儲存任何與瀏覽工作有關的資訊，包括歷史記錄，所以這將使得 SpoofGuard 每次都會對受測網址做完整測試。

實驗結果顯示於表 6，分別記錄各個工具列不同警告程度的個數。當中可以看見 IE 8 和 Google 的測試結果，雖然良性網址都沒有被誤判，但惡意網址也只有少數被偵測到，可以說 IE 8 和 Google 的惡意網站清單中，只包含這些測試資

料的少數惡意網址。而 McAfee（此研究採用 McAfee 的產品 SiteAdvisor）和 ThreatExpert 其偵測效果也都沒有很好，McAfee 只有出現 7 次紅色警告和 1 次黃色警告，且灰色佔了 133 筆，表示有大部份的網址還沒有被 McAfee 評估過。ThreatExpert 只有出現黃色警告 25 次，反而對於良性的網址出現了 13 次紅色警告，使得其 FPR 高了一點。而 SpoofGuard 總共發出 185 次紅色警告，表示能夠有效偵測大部份的釣魚網站，但同時也有大部份的誤判，對於良性網址也有 78 次紅色警告。

表 6：瀏覽器工具列測試之實驗數據

	benign: 306				malicious: 265			
	red	yellow	green	gray	red	yellow	green	gray
IE 8	0	-	306	-	1	-	264	-
Google	0	-	306	-	8	-	257	-
McAfee	2	2	256	46	7	1	124	133
ThreatExpert	13	3	290	0	0	25	240	0
SpoofGuard	78	217	11	-	185	71	9	-

最後將計算各工具列的 FPR 和 TNR，並與本研究的結果比較。對於使用顏色來區別可疑程度的工具列，將會以紅色及黃色顯示的結果當作依據，若良性網址出現紅色或黃色的警示，則視為 FP；若惡意網址出現紅色或黃色警示，稱之為 TN，其 TNR 也可以稱做該工具列的偵測率。整體比較如圖 8 所示

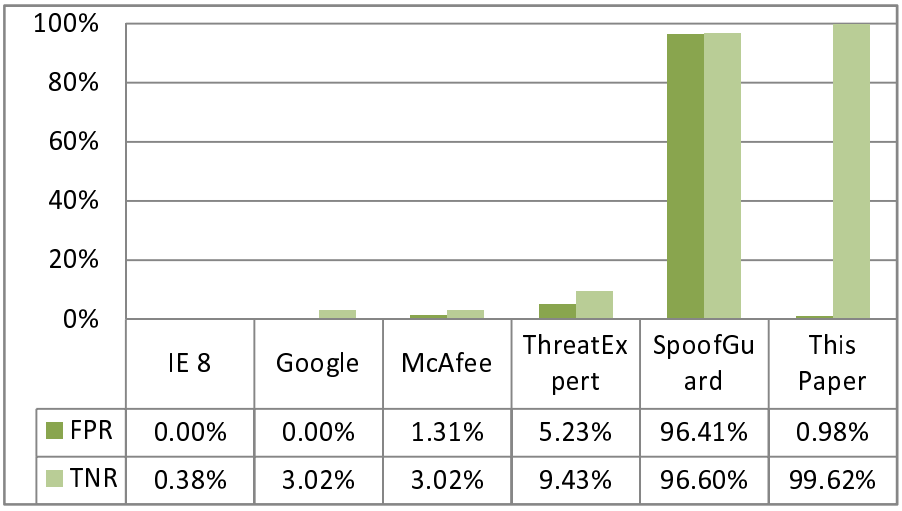


圖 8：本方法與瀏覽器工具列的 FPR 和 TNR 之比較

其中 FPR 方面，其值要越小越好，而只有 IE 8 和 Google Chrome 的結果比本方法還低，不過差距沒有超過 1%，而 TNR 方面，本研究的偵測方法得到最好的效果，表示針對相同的測試資料，本研究的方法能夠有效地偵測出惡意的網址並且維持相當低的 FPR。

五、實驗討論

觀察特徵值分數設定的結果，分數絕對值較大且為負值的特徵值，包含F_2=1（First URL Message）、F_3=1（Username in URL）、F_6=0（域名經由 Google 搜尋不到）、F_9=0~30（0 ≤ Age of Domain ≤ 30），也就是說這些特徵值是屬於效果較好且較惡意的特徵。另外，分數絕對值較大且為正值的特徵值，有$F_7 > A_{7b}$（Entropy of Dealy Time > 良性資料的平均 entropy）、$F_8 > A_{8b}$（Entropy of Response Time > 良性資料的平均 entropy）、F_9>360（Age of Domain > 360 天），這些特徵值是屬於效果較好且偏良性的特徵。

因此，對於良性網址訊息來說，若有兩個較惡意的特徵，或者有一個較惡意的特徵再加上一個稍微可疑的特徵，很有可能就會被判斷成惡意的。例如，有些正常的網址訊息是當天首次傳送的，而且其中的網址若包含使用者名稱，或者網址是 IP 的形式，則就有機會被誤判為惡意的網址。另外，像有些合法的域名，例如“hkwinter.com.tw”和“163.to”利用 Google 搜尋卻找不到相同域名的網址，這樣的情形也很容易被判斷成惡意的。

而對於惡意網址訊息來說，如果沒有異常的傳訊者行為，且網址也無異常特徵的話，很有可能會被判斷成良性的。例如，有些惡意聊天機器人會有一些對話，最後以一句網址訊息結尾，若 entropy 偏高的話，就有可能造成誤判。或者當網址使用一些免費空間或轉址服務提供的域名，由於這些域名是合法的，所以利用 Google 搜尋可以找到相同域名的網址外，其網域註冊日也不會太新，因此造成誤判為良性的可能。像是 http://www.no-ip.com/就能申請免費的域名對應到某個 IP，或 http://www.dot.tk 可以將某個網址對應到一個新的域名等，這些動作都會使得網址看起來像合法的，但其實網站可能含有惡意內容。

關於測試瀏覽器工具列的部份，由於沒有自動化的測試方式，只能逐一點選每個網址，使得測試過程相當沒效率且耗時，因此就沒有以執行多個回合取其平均值的方式來評估，不過本方法的測試結果仍在預期的範圍內，且其他瀏覽器工具列雖不能得到絕對的偵測效能，但基於同一組網址的測試資料，可以得到與本方法相對的偵測結果。

伍、結論與未來研究方向

目前對於 IM 惡意網址的擴散，仍無有效的即時偵測方案，因此，本研究提出一個於客戶端偵測 IM 惡意網址的方法，幫助使用者分辨接收到的網址是否屬於惡意的。

本研究提出的方法是基於檢查異常網址特徵及異常傳訊者行為，透過網址的外觀和域名提供的資訊，而不需查看網頁內容，可以快速地分辨可疑的網址；同時檢查傳訊者的傳訊行為，可以發現可疑的惡意行為。並且定義一組惡意行為模式，用於偵測步驟的第一階段，比對接收到的網址訊息是否符合某個已知的惡意行為，若有則可以直接判定該網址是惡意的，以達到較即時的偵測。而沒有滿足惡意行為的網址，再利用第二階段的計分演算法，以事先設定好的分數模型，預測該網址是否屬於惡意，以防範未知的惡意行為。

最後經由實驗評估，本方法在一定的訓練資料數之下（建議良性和惡意訓練資料各取 30 筆），平均的 FPR 約從 0.78%到 0.94%，平均 TNR 約從 99.51%到 99.72%，整體效能達到一個可接受的範圍，且相較於其他瀏覽器的偵測工具列，也能有不錯的偵測效果，而本方法完全不需要依賴黑、白名單的幫助，即可有效地即時偵測 IM 惡意網址。不過對於具有正常網址特徵且看不出有異常傳訊行為的網址訊息，就有可能躲避本方法的偵測，如 4.5 節所提到的一些缺失。或許未來能加入檢測網頁原始碼異常的方法，並能維持偵測的即時性，以對付更多種類的惡意網址。同時也可以試著實作於現有的 IM 軟體上（如：MSN），實際評估偵測方法的即時性。

另外重要的一點是，實驗資料的來源決定整個實驗的可信度。本研究雖然是收集一般使用者的對話記錄，但只局限於某個區域網路內的使用者，而且使用的樣本數仍不夠多，可能難以反應真實環境的情形，因此，未來可以努力收集更多、更具說服力的實驗資料，來得到更真實的偵測效能。

參考文獻

卿斯漢、王超、何建波、李大治（2006），『即時通訊蠕蟲研究與發展』，*軟件學報*，第十七卷，第十期，頁 2118–2130。

Basnet, R., Mukkamala, S. and Sung, A.H. (2008), 'Detection of phishing attacks： a machine learning approach', *Soft Computing Applications in Industry*, pp.373–383.

Barabasi, A.L. and Bonabeau, E. (2003), 'Scale-free networks', *Scientific American Magazine*.

Fette, I., Sadeh, N. and Tomasic, A. (2007), 'Learning to detect phishing emails',

Proceedings of the 16th international conference on World Wide Web, pp.649–656.Group, A.P.W.,'Global phishing survey(2009)： domain name use and trends in 1H2008.', Available： http://apwg.org/reports/APWG GlobalPhishingSurvey 1H2008.pdf (accessed January 2010).

Hindocha, N. and Chien, E. (2003), 'Malicious threats and vulnerabilities in instant messaging',Symantec Security Response White Paper.

Mannan, M. and Oorschot, P.C.van (2005), 'On instant messaging worms analysis and countermeasures', *Proceedings of the 2005 ACM workshop on Rapid malcode*, pp.2–11.

Morse, C. and Wang, H. (2005), 'The structure of an instant messenger network and its vulnerability to malicious codes', *Proceedings of ACM SIGCOMM*.

Pan, Y. and Ding, X. (2006), 'Anomaly based web phishing page detection', *Proceedings of the Computer Security Applications Conference,* 2006.ACSAC'06.22nd Annual, pp.381–392.

Pastor-Satorras, R. and Vespignani, A. (2001), 'Epidemic spreading in scale-free networks', *Physical review letters*, Vol.86, No.14, pp. 3200–3203.

Secunia (2009), 'ICQ predictable file location weakness', Available： http：//secunia.com/advisories/10970/ (accessed January 2010).

Smith, R. (2002), 'Instant messaging as a scale-free network', Arxiv preprint cond-mat/0206378.

Williamson, M.M., Parry, A. and Byde, A. (2004), 'Virus throttling for instant messaging', *Proceedings of Virus Bulletin Conference*, pp. 38–48.

Xie, M., Wu, Z. and Wang, H. (2007), 'HoneyIM: fast detection and suppression of instant messaging malware in enterprise-like networks', *Proceedings of the 2007 Annual Computer Security Applications Conference* (ACSAC '07).

Zhang, Y., Hong, J. and Cranor, L. (2007), 'CANTINA：a content-based approach to detecting phishing web sites', *Proceedings of the 16th international conference on World Wide Web*, pp.639–648.