國立中山大學資訊工程學系

專題計畫書

基於持續性深度學習檢測釣魚網頁並防護

專題組員

B083022053 黃啟桓

B093040042 黃柏翔

B093040040 鍾名捷

指導教授 : 徐瑞壕 教授

二、研究計畫內容（以 10 頁為限）：

(1) 摘要

隨著網路的蓬勃發展，網路詐欺、竊取資料等攻擊層出不窮，不僅是利用人性弱點，而擅長去偽裝釣魚網站的手法也非常逼真。無形中讓受害者交出自己的個人資料、財產安全、裝置權限。這些隱私資料對受害者的影響非常重大，可能面臨財產或身分遭盜用的結果。不只是個人用戶，企業方也深受其擾，並損失慘重。使得網路安全成為一個不可忽視的課題。

基於人工智慧的釣魚網頁複合式檢測，利用包含 URL 地址比對和網頁特徵提取進行複合式的，期望以持續學習的模型，並同時兼顧防禦以及用戶隱私問題，高度的彈性以檢測多樣與多變的釣魚手段，我們也希望可以將「持續性深度學習(continual-learning)」加入此專題中，學習新的技能或是任務時不會將過去學習而來的知識忘記，就像人類可以不斷學習新知識，並同時利用舊的知識，以適應多變及多樣的釣魚網站。本專題將重點聚焦於網頁的檢測，以簡易的工具來面向廣大的瀏覽器用戶。

關鍵字: 釣魚檢測、釣魚防禦、Phishing detection、人工智慧

(2) 研究動機與研究問題

1. 研究動機

釣魚網站是一種歷史十分悠久的詐騙手段，興取於 1990 年代中期，一開始一群年輕人線上的聊天室功能假冒系統管理員，竊取用戶的信用卡卡號。而現代駭客將網站包裝成各種生活之中常見的銀行網站、購物商城等等，已取得用戶的個資後做不法的用途。而釣魚網站近年也影響到了企業方面，[13]讓企業損失慘重的商業電子郵件詐騙（Business E-mail Compromise，BEC)，就是一例，先是透過釣魚郵件，取得公司郵件帳密並掌握交易資訊，再伺機假冒公司 CIO 或合作廠商，寄送詐騙電子郵件，要求匯款至指定帳戶或變更匯款帳號，讓使用者誤以為是對方而上當。而根據 FBI 旗下的網路犯罪申訴中心統計，近 3 年來，全球商業郵件詐騙案超過 4 萬起，詐騙金額更是高達 50 億美元。去年第 3 季日本航空 JAL 也有歹徒假冒出租波音客機的業者，詐騙約 1 億 176 萬元，顯示 BEC 詐騙事件不斷。

根據美國反網路釣魚工作小組(APWG)2014 年上半年的統計，釣魚網站平均的壽命已經由 6.1 天大幅縮短到 32 小時 32 分鐘[5]; APWG 在 2018 年第一季度報告二月 URL 檢測顯著的增加，一直持續到 3 月，但釣魚網站的域名保持不變“URL 的增加主要歸因於一次性 URL。 這些唯一的 URL 由網絡釣魚者自動生成，以允許受害者一次性訪問一個唯一的網絡釣魚 URL。”[6]。顯示傳統的釣魚檢測技術難以識別、防禦。

同時我們也希望處理人工神經網路可能會遇到「災難性遺忘」的問題，建立可以持續學習且具有高度彈性的釣魚識別防護系統，並將各種釣魚網頁比對的工具進行整合，提高釣魚識別防護模型的魯棒性。以達到安全上網、保護資料、保護裝置的效果。並以瀏覽器的擴充功能(extensions)打造便捷、立即、有效的防禦功能。

2. 研究問題

i. 哪一種應用是對於大眾而言最便利的

由於此專題著重於網頁的安全，所以相較於開發新的應用程式來保護使用者使用瀏覽器時面對的釣魚威脅，不如配合瀏覽器自帶的擴充功能使得安裝、卸載更加方便，並且只要瀏覽器啟動，其擴充功能也會跟著啟動，使得用戶在使用瀏覽器時全程受到保護。

ii. 如何建立高度彈性的防護

對於 URL、域(domain)、主機(hosts)在網頁載入前與黑名單進行比對，黑名單來源每天更新一次。

對於未在黑名單的釣魚網頁在網頁載入中進行 html 結構分析。

載入完成後以網頁圖像進行深度學習預測是否為釣魚網頁。

預測釣魚網站的深度學習模型以強人工智能的概念，使得模型可以不斷的更新，以面對多變的釣魚網頁。

iii. 人工智慧如何使檢測釣魚網站的準確率及效率更加提升

傳統的檢測方式仰賴龐大的資料庫，需要將標記為釣魚網站的資料紀錄到黑名單資料庫中。而檢測過程關係到搜索黑名單，將目標網站比對黑名單判斷是否為釣魚網站，執行速度將隨著黑名單的增加變得遲緩。且因釣魚網站推陳出新，在第一時間無法透過黑名單檢測成功。透過人工智慧的協助，只需要將標記過的資料訓練出能夠偵測釣魚網站的模型，即可讓訓練模型舉一反三地辨別出釣魚網站，提高準確率。而且只需要將目標網站代入模型算出結果，比傳統的資料庫檢索更加快速。

iv. 加強防護並兼顧隱私問題

考慮到 web server 容易會被駭客攻擊，而 web client 和 mail client 端可能會有隱私的問題，所以由 mail server 端將 phishing mail 直接擋下是更好的選擇。

v. 「災難性遺忘」

藉由 Synaptic Intelligence(SI)的模型，能夠更長期持續學習新的訓練數據，避免未來的訓練模型變得無法解決先前的訓練數據。[16]。

(3) 文獻回顧與探討

閱讀數篇關於釣魚網頁檢測的論文，了解釣魚網頁的特徵。

i. 釣魚網址的案例:

(1) 網址混淆:

例如: http://mail-google.com、http://www.google.com (數字一)與 google 本身沒有關係。

二級域名混淆: http://mail.google.com.xyz，也與 google 本身沒有關係。

(2)縮網址:

縮網址服務會遮蔽連結的網址，使用者無法有效的判斷是否為釣魚網址。

(3)網址嫁接:

DNS 伺服器被入侵並趁機竄改設定。

ii. 跨網站請求偽造 csrf:

(1) 偽造使用者請求:

例如: 偽造 facebook 使用者發布色情廣告的請求。

iii. 網頁偽裝:

(1)使用相近的網頁 html 結構及圖像來混淆視覺。

傳統的網頁釣魚檢測主要基於用戶的回報，依賴於黑名單。近年來，由於釣魚網站的快速生成，短壽命的特性，使得傳統的網頁釣魚檢測缺乏立即性，未能提供有效的防護功能。

而相關的網頁釣魚檢測技術隨著機器學習蓬勃發展，開枝散葉。本計畫希望結合數種網頁釣魚檢測工具。以增強面對釣魚網站的防禦力，比較單一網頁釣魚檢測工具與本計畫的檢測效果進行比對。並且以實際的行為防禦來自釣魚網頁的攻擊。

查閱大量來自 github 的原始碼，大多專案對於釣魚網站的偵測及防禦沒有整合在一起。而使用瀏覽器擴充功能的軟體大多使用黑名單的機制來判斷釣魚網頁，缺乏彈性。因此此研究希望透過將這些功能進行整合，以保護使用者使用瀏覽器上網的安全。

[9]駭客可以規避 Server 的防禦手段，如果從 web server 端去執行，使得防禦變得較為困難。如果從 client 端去設計防範釣魚網站的工具，效率也會提升，但也由於是從 client 端執行的緣故，可能有觸及隱私問題。 但如果是從 mail server 來做防禦，可以將 phishing mail 從源頭就阻擋下來，不會讓使用者接收到此類信件，而且因為在 Server 端可以蒐集到較多的 mail pattern，讓整體預測的正確性也會提升，但也會因此而牽涉到使用者的隱私

權問題。如果從 mail client 端做防禦的動作，雖是處於信件傳送的下游且樣本會較少，但同樣也牽涉到使用者隱私問題。

[7]使用機器學習來判斷釣魚網站中的眾多特徵值，以權重及分類找出特徵值。並使用決策樹、隨機森林、類神經網路、倒傳遞類神經網路等演算法，但是因為在演算法中當訓練集資料比例達到一定的比例或參數過多時，便產生「過度配適」的風險，可以使用「交叉驗證」解決，將訓練資料進行分組，一部分做為訓練子集來訓練模型，另一部分做為驗證子集來評估模型。

[12]CANTINA[10]採用 TF-IDF (term frequency-inverse document frequency)演算法，而 TF 指的是詞頻，IDF 為文件頻率，接著計算出網頁中各個詞語的頻率，並製作出頻率表，將詞頻最高的前 M 個送往搜尋引擎若網頁網域和前 N 個搜尋結果的網域相同，則視為正當網站。

[12]Moghimi[11]等人提取網頁元素所引用的網址，比較每個資源元素的網址與 URL 相近性，超連結為提取 html 中的 href 屬性，圖片、CSS 則提取 src 屬性，並計算出網頁中這些元素或是網址的相似程度。且整理網頁中每個資源元素所計算出的距離，加以判斷。

[12]先進行特徵的抽取後(網址列特徵、網頁異常特徵、 網域特徵、TF-IDF)，接著計算出 URL 間的萊文斯坦距離(文字間的轉換需要幾個編輯步驟)，然後進入支援向量機(SVM)中分類，並將樣本加入資料庫中，避免重複比對，並產生預測結果。

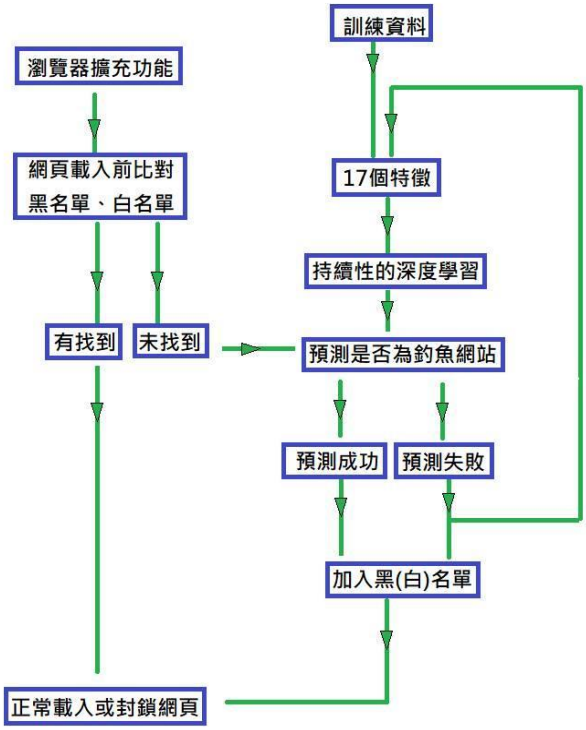
[16]如果直接將訓練好的模型再次以新的資料訓練，則有可機會發生災難性遺忘，再也無法解決舊的資料。於是採用 Synaptic Intelligence(SI)的模型，能夠更長期持續學習新的訓練資料，避免未來的訓練模型變得無法解決先前的訓練資料。SI 的原則是學習新的資料前，先計算目前模型中節點的重要性，若改變節點的值越不會改變準確性，則代表節點的重要性越低。訓練新的資料時只要改變重要性較低的節點，就能訓練出不但能解決新的資料，也不會遺忘舊的數據的資料。

而我們打算利用這個方法實作專題中提到的持續學習。

(4) 研究方法及步驟

透過閱讀數篇關於釣魚網頁檢測的論文，並進行實踐，分析各種檢測方法的優缺點及效果，取長補短以達到檢測效果的提升。並建立一套防禦機制，阻止釣魚網頁的運作，並回報給釣魚網頁回報的機構，以降低釣魚網頁的危害，增加使用網路的安全性。

以下是我們設計出來系統流程圖



釣魚網頁分析的特徵[14]:

1. URL 的特徵:
    a. URL 的 host(/domain)
    b. URL 的 IP 地址
    c. URL 中的 "@"符號
    d. URL 的長度
    e. URL 的深度
    f. URL 中的重定向 (Redirection)
    g. 域名包含 “http/https”
    h. 域名包含 “-”
    i. 縮網址服務

2. 域的特徵:
    a. DNS 的紀錄
    b. 網路流量
    c. 域名年齡(age)

d. 域名的結束時間

e. DNS 投毒問題

3. HTML 和 JavaScript 的特徵:

a. IFrame 重定向

b. 狀態欄 onmouseover 事件

c. 禁用右鍵點擊

d. 網站重定向次數

持續性深度學習[4]以上述的特徵來進行訓練。

本計畫將透過以下步驟來達成:

1.可以先利用瀏覽者的常用網站加入白名單，方便比對判斷。也建立一份黑名單，當判斷出可能為釣魚網站時，也將此網站加入黑名單，預測的效率及正確性也將增加[9]。

2.對於論文中釣魚網頁檢測的檢測方法進行實踐，例如:分析 URL、檢測網頁相似度、數據分析，對照論文的結果與自身實作的結果，並進行修正，以提升單個檢測技術的檢測能力。

3.結合多個釣魚網頁檢測技術，取長補短以達到檢測效果的提升。例如:傳統的釣魚網站檢測缺乏保護機制

4.建立一套釣魚網頁防禦機制的軟體，以阻止釣魚網頁的運作，自動暫停載入數據、自動清空 cookies、自動封鎖網站使用 JavaScript、自動封鎖網站背景執行。

5.計畫在檢測到新的釣魚網址時，能主動回報給釣魚網頁回報的機構，提醒更多人/裝置/程式，將被動防禦的效果提升到主動防禦。以降低釣魚網頁的危害，增加使用網路的安全性。

6.分析軟體的優缺點，包含檢測的速度、準確性，阻斷釣魚網頁的執行速度等方面，盡可能的達到最好。

7.計畫結合人工智慧的技術，並利用 LLL(selective synaptic plasticity)作為深度學習的工具，使其具有持續學習的特徵，形成強人工智慧。

(5) 預期結果

1. 建立一套面相於瀏覽器的用戶的軟體，例如: 瀏覽器的擴充功能(extensions)。
2. 利用持續學習避免「災難性遺忘」的問題[15]。
3. 擁有釣魚網頁檢測技術。包含:分析 URL[12]、檢測網頁相似度、數據

分析、MITM 工具[1]、DOM-tree 結構[2]、圖像匹配演算法[3]及。

4. 網頁防禦功能，目的是防止損害的持續及擴大。自動暫停載入數據、自動清空 cookies、自動封鎖網站使用 JavaScript、自動封鎖網站背景執行。
5. 主動回報給釣魚網頁回報的機構的功能。

(6) 參考文獻

[1] Kondracki, B., Azad, B. A., Starov, O., & Nikiforakis, N.. (2021). *Catching Transparent Phish: Analyzing and Detecting MITM Phishing Toolkits*. https://doi.org/10.1145/3460120.3484765

[2][3] Nguyen, L. D., Le, D.-N., & Vinh, L. T.. (2014). *Detecting phishing web pages based on DOM-tree structure and graph matching algorithm*. https://doi.org/10.1145/2676585.2676596

[4] Lomonaco, V., & Rish, I. (2021, July 19). *Continual Learning with Deep Architectures*. Continual Learning with Deep Architectures. https://icml.cc/virtual/2021/tutorial/10833

[5] Aaron, G. (2014, September 25). *Global Phishing Survey: Trends and Domain Name Use in 1H2014*. APWG. https://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_1H2014.pdf

[6] Aaron, G. (2018, July 31). *APWG report*. APWG report. https://docs.apwg.org/reports/apwg_trends_report_q1_2018.pdf

[7] 洪慕藍（2022 年 1 月 17 日）。*以機器學習演算法探討網路釣魚網站之特徵值*。南臺科技大學。https://hdl.handle.net/11296/ga7s7r

[8] Di Leom, M. (2023, January 16). *phishing-filter*. GitLab. https://gitlab.com/malware-filter/phishing-filter

[9] 曾黎明、黃克仲、陳天豪（2007 年 5 月 8 日）。*以 URL 資訊為基礎之網路釣魚偵測系統*。TANET2007 臺灣網際網路研討會論文集。http://itech.ntcu.edu.tw/tanet%202007/5/430.pdf

[10] Zhang, Y., Hong, J. I., & Cranor, L. F.. (2007). *Cantina: a content-based approach to detecting phishing web sites*. https://doi.org/10.1145/1242572.1242659

[11] Moghimi, M., & Varjani, A. Y. (2016). *New rule-based phishing detection method.* Expert Systems with Applications, 53, 231-242. https://doi.org/10.1016/j.eswa.2016.01.028

[12] 蔡淑靜（2017 年 1 月 17 日）。*基於支援向量機與整合式特徵抽取方法之釣魚網站偵測機制之研究*。國立高雄第一科技大學。https://hdl.handle.net/11296/ed3gke

[13] 羅正漢（2018）。*重新認識釣魚郵件威脅*。iThome。

https://www.ithome.com.tw/news/120507

[14] Sundari, S. G. (2020, May 11). *Phishing Website Detection by Machine*

Learning Techniques. GitHub. https://github.com/shreyagopal/Phishing-Website-Detection-by-Machine-Learning-Techniques

[15] von Oswald, J., Henning, C., Grewe, B. F., & Sacramento, J. (2022, April 11). *Continual learning with hypernetworks*. arXiv. https://arxiv.org/abs/1906.00695

[16] Lee, H. Y. (2021, June 5). *【機器學習 2021】機器終身學習(Life Long Learning, LL) (一) - 為什麼今日的人工智慧無法成為天網？*災難性遺忘(*Catastrophic Forgetting*). YouTube. https://www.youtube.com/watch?v=rWF9sg5w6Zk

表 C802

三、耗材、物品、圖書及雜項費用：

(1)　凡執行研究計畫所需之耗材、物品、圖書及雜項費用，均可填入本表內。

(2)　說明欄請就該項目之規格、用途等相關資料詳細填寫，以利審查。

(3)　依研究計畫實際需求擇優補助，每一計畫最高以補助新臺幣 20,000 元為限。

金額單位：新臺幣元

項 目 名 稱	說明	單位	數量	單價	金額	備註

合　　計						

表 C803

大專學生研究計畫指導教授初評意見表

一、學生潛力評估：

二、對學生所提研究計畫內容之評述：

三、指導方式：

四、本人同意指導學生瞭解並遵照學術倫理規範；本計畫無違反學術倫理。

指導教授簽名：

年　　月　　日

表 C804