基於持續性深度學習檢測釣魚網頁並防護

Detect and protect against phishing sites based on continual deep learning

國立中山大學資訊工程學系

111 學年度大學部專題製作競賽

組員: B083022053 黃啟桓

B093040040 鍾名捷

B093040042 黃柏翔

指導教授 : 徐瑞壕教授

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，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

目錄

圖次

第一章 緒論

第一節 研究動機

一、釣魚網頁的危害

釣魚攻擊已經成為網路使用者面臨的重大安全威脅之一。釣魚攻擊通常是指詐騙者通過製作虛假網站或電子郵件，以欺騙網路使用者提供其個人敏感信息，如密碼、帳號、信用卡號碼等。釣魚攻擊的目的是盜取個人信息或進行其他惡意行為，這對個人隱私和金融安全構成嚴重威脅。

釣魚攻擊的方式愈來愈隱蔽，以至於許多網路使用者往往無法分辨真假網頁或電子郵件，因此，建立一種有效的檢測釣魚網頁的方法是非常必要的。此外，釣魚攻擊也對企業和組織的安全造成了嚴重威脅，因為釣魚攻擊通常針對企業和組織的員工，試圖從中獲取敏感信息。

因此，針對釣魚攻擊問題進行研究和開發釣魚網頁檢測工具是十分必要的。這將有助於保障網路使用者和企業的安全，減少釣魚攻擊對社會帶來的傷害。

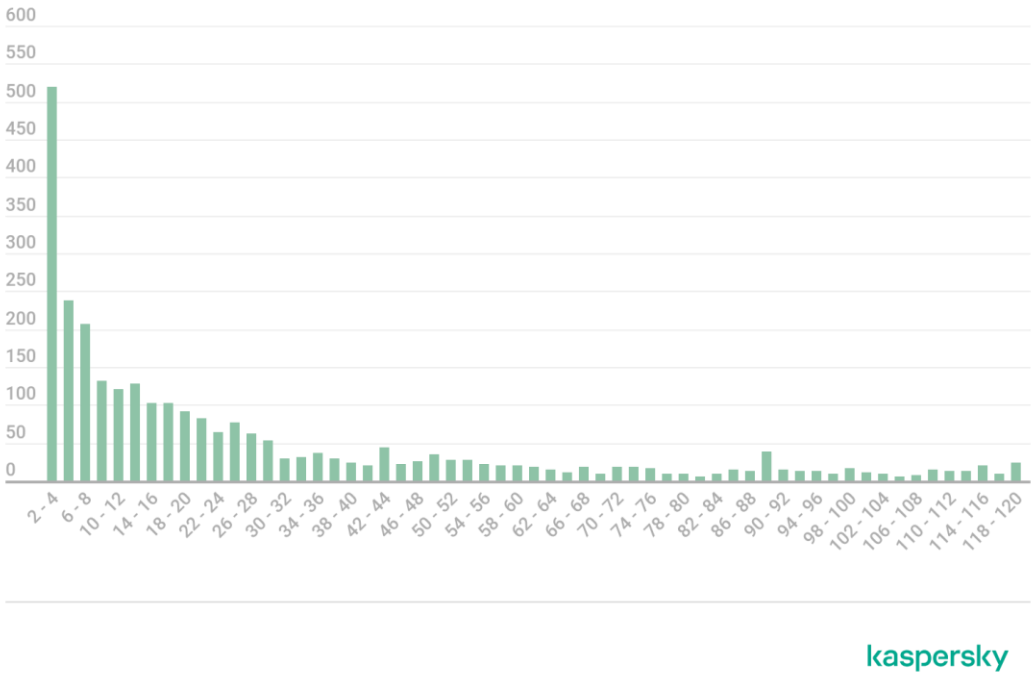
二、釣魚網頁發展

釣魚網頁的發展可以追溯到 20 世紀 90 年代，當時互聯網使用率開始普及，網路使用者逐漸增加，網路安全問題也開始受到關注。當時，黑客和駭客開始利用釣魚網頁來進行攻擊，盜取網路使用者的帳號和密碼等個人信息。

在 2000 年代初期，釣魚攻擊成為了一種主要的網路安全威脅，並且不斷發展和演進。釣魚攻擊的方式也變得更加隱蔽和精細，例如製作更真實的假冒網站、使用社交工程技術、利用電子郵件等途徑進行攻擊等。

隨著網路技術的不斷發展和安全技術的提升，釣魚攻擊的形式也越來越多樣化。現代釣魚攻擊不僅限於假冒網站和電子郵件，還包括社交媒體帳號、手機應用程式等。

根據釣魚網站(https://securelist.com/phishing-page-life-cycle/105171/)所提供的資料顯示，大部分的釣魚網站活躍的時間在 48 小時內。而要如何應對釣魚攻擊的不斷變化，就是我們本專題目標。



圖一、根據活躍小時數對釣魚鏈接分類圖。該圖顯示了每個鏈接生命週期前五天的數據。

三、釣魚網頁-惡意軟件分發

釣魚網站中的惡意軟件分發已成為當今網絡安全領域中一個嚴重的問題。釣魚網站通過欺騙手段和惡意意圖，試圖引誘用戶下載和執行惡意軟件，從而對用戶和組織的信息安全造成嚴重威脅。

這種釣魚網站中的惡意軟件分發行為對用戶和組織的安全和隱私構成了重大威脅。這些惡意軟件可能包含病毒、間諜軟件、勒索軟件等，可以竊取用戶敏感信息、破壞系統功能、散播廣告或進行其他惡意活動。此外，這些惡意軟件的分發也對網絡生態系統的穩定運行產生了不良影響。

因此，本論文的動機在於解決釣魚網站中的惡意軟件分發問題。我們希望提出一種有效的方法，能夠檢測和對抗釣魚網站中的惡意軟件分發行為，以保護用戶免受惡意軟件的威脅。

通過研究和解決這個問題，我們能夠提高用戶和組織的網絡安全水平，減少釣魚攻擊對用戶的損害。同時，我們的研究還有助於維護網絡生態系統的健康運行，減少惡意軟件對整個網絡社區的影響。

四、釣魚網頁-黑帽 SEO 技術

釣魚網站使用黑帽 SEO 技術已成為當今網絡安全領域中一個嚴重的問題。釣魚網站以其欺騙性手段和惡意意圖，試圖冒充合法和受信任的實體，引誘用戶提供個人信息或進行詐騙活動。其中一種常見的手法是利用黑帽 SEO 技術來提高釣魚網站在搜索引擎結果中的排名，增加其曝光度和吸引力。

這種釣魚網站使用黑帽 SEO 技術的行為對用戶和組織的信息安全造成了重大威脅。它們通過操縱搜索引擎算法，使得釣魚網站在搜索結果中排名靠前，從而誘導用戶訪問和受騙。這種技術不僅損害了用戶的隱私和財產安全，還對網絡生態系統的穩定運行產生了不良影響。

因此，本論文的動機在於解決釣魚網站使用黑帽 SEO 技術的問題。我們希望提出一種有效的方法，能夠檢測和對抗釣魚網站中使用的黑帽 SEO 技術，從而保護用戶免受釣魚攻擊的威脅。

通過研究和解決這個問題，我們能夠提高用戶和組織的網絡安全水平，減少釣魚攻擊對用戶的損害。同時，我們的研究還有助於維護搜索引擎的可信度和公平性，保護網絡生態系統的健康發展。

五、釣魚網頁-自動及惡意的網頁跳轉

釣魚網站作為一種網絡安全威脅，已經對用戶和組織的信息安全造成了重大影響。釣魚網站使用各種欺騙手段來引誘用戶提供個人敏感信息，從而進行詐騙、竊取身份信息或散播惡意軟件。其中，釣魚網站使用自動及惡意的網頁跳轉技術，進一步增加了攻擊的隱匿性和成功率。

這種自動及惡意的網頁跳轉行為使得釣魚網站可以將用戶引導到意想不到的網頁，進而進行欺騙、攻擊和數據竊取等惡意活動。這對用戶和組織的信息安全構成了嚴重威脅，並對網絡生態系統的穩定運行產生了不良影響。

因此，本論文的動機在於解決釣魚網站使用自動及惡意的網頁跳轉的問題。我們希望提出一種有效的方法，能夠檢測和阻止釣魚網站中的自動及惡意的網頁跳轉行為，以保護用戶免受釣魚攻擊的危害。

通過研究和解決這個問題，我們能夠提高用戶和組織的網絡安全水平，減少釣魚攻擊對用戶的損害。同時，我們的研究也有助於維護網絡生態系統的穩定運行，減少惡意行為對整個網絡社區的影響。

因此，本論文的動機在於研究並提出有效的方法來解決釣魚網站使用自動及惡意的網頁跳轉的問題，以提高網絡安全性並維護網絡生態系統的健康運行。

第二節 研究問題

一、 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。

二、 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 結構分析與 javascript 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。

三、 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。

四、 克服「災難性遺忘」

災難性遺忘是指人工神經網絡在學習新資料時突然而徹底地忘記先前學習的資料的趨勢。具體來說，這些問題指的是製作對新信息敏感但不受新資料干擾的人工神經網絡的挑戰。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。

第三節 文獻回顧與探討

一、 文獻回顧

閱讀數篇關於釣魚網頁檢測的論文，了解釣魚網頁的特徵。

釣魚網址的案例:

1. 網址混淆:

   例如: http://mail-google.com、http://www.goog1e.com (goog1e 的 l 是數字一)

2. 二級域名混淆:

   http://mail.google.com.xyz，也與 google 本身沒有關係

3. 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。

4. 網址嫁接:

   DNS 伺服器被入侵並趁機竄改設定。

5. 網頁偽裝:

   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。

二、 文獻探討

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。

第四節 研究方法及步驟

一、 研究方法

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。

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
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

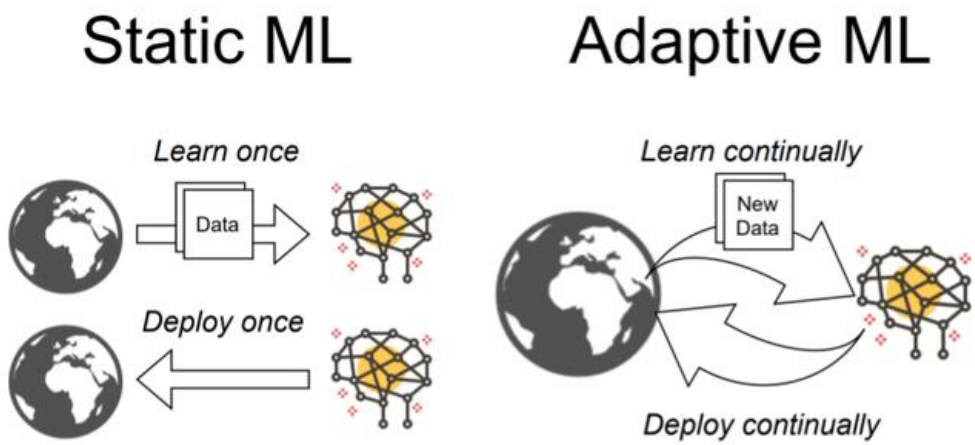
持續性深度學習[4]以上述的特徵來進行訓練。

持續性深度學習:

傳統機器學習通常首先使用數據中的所有可用示例學習模型，然後部署以供實際使用。這意味著無論何時模型完成學習，在實踐中使用時它都保持不變。該模型的靜態性質存在問題，因為它不適合我們不斷變化的世界。這讓許多現代機器學習方法無所適從，因為部署時的靜態模型無法使用永無止境的數據流。持續學習旨在通過仔細研究動態機器學習模型、靈活且持續地適應數據來解決這個問題。

「連續學習系統可定義為一種自適應算法，能夠從連續的信息流中學習，這些信息隨著時間的推移逐漸變得可用，而要學習的任務數量（例如分類任務中的成員類別）並不是預先定義的。重要的是，新信息的納入應該在不產生災難性遺忘或干擾的情況下進行。」[17]

因此，在持續學習場景中，隨著任務分佈在其生命週期內動態變化，學習模型需要逐步構建和動態更新內部表示。理想情況下，此類內部表示的一部分將具有通用性和不變性，足以在類似任務中重複使用，而另一部分應保留和編碼特定於任務的表示。

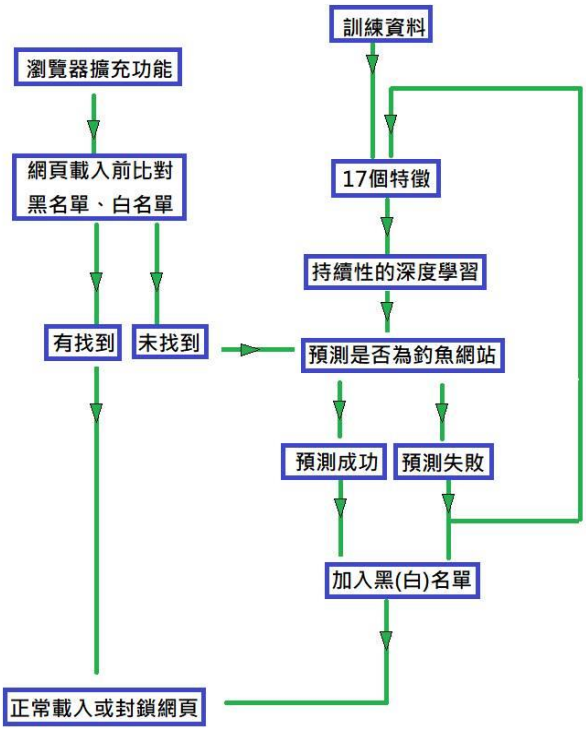


圖二、傳統機器學習與自適應性機器學習的差別

二、 研究步驟

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。



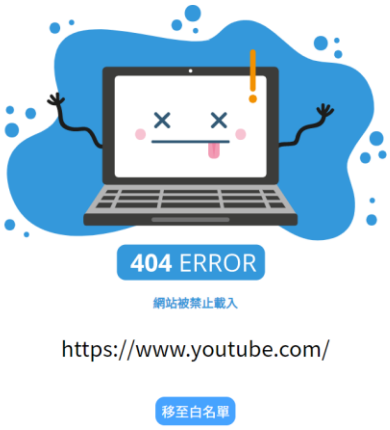
圖三、運作流程圖。左邊為客戶端的運作模式，右邊為伺服器運作模式

第五節 研究成果

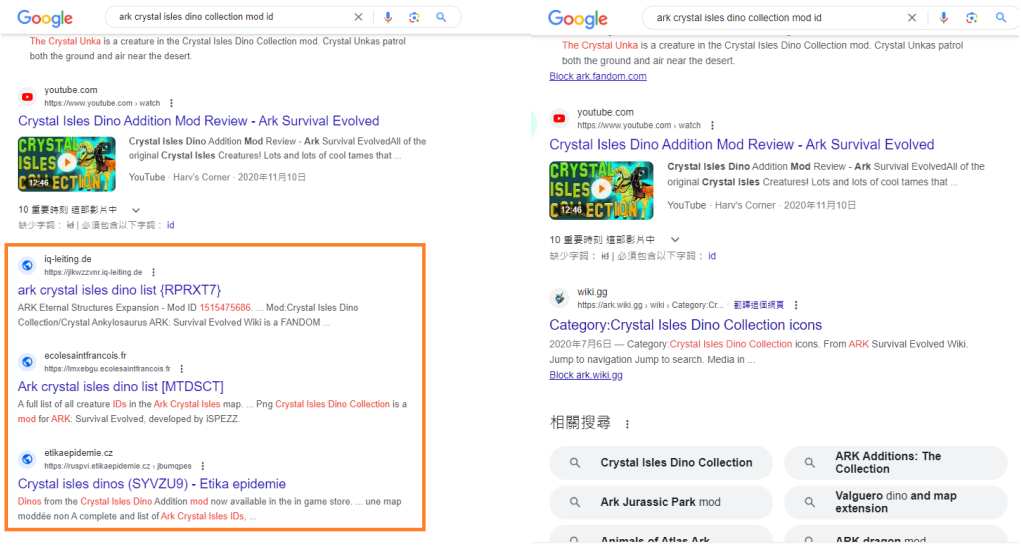
一、 Client 框架

框架採用 Chromium 基礎瀏覽器通用的擴充功能(extension)。

預測到該網站包含惡意軟件分發的時，將進行網頁的封鎖。以防止危害的發生。也可以自行針對個別網站加入黑名單。



圖四、嘗試將 youtube 網頁加入黑名單的結果



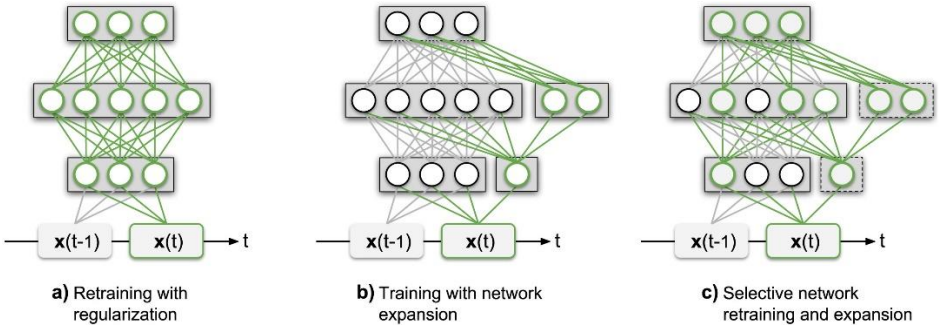
圖五、對濫用 SEO 的網站進行封鎖(左圖為封鎖前、右圖為封鎖後)

二、 Server 架構

使用 cookie 儲存每個使用者的最近期的訓練模型，在請求新的訓練資料時能透過持續性深度學習進行模組的更新，並且回饋給客戶端，使得客戶端能更新自己量身定做的模型。

三、 持續性深度學習

引用 Avalanche (https://github.com/ContinualAI/avalanche) 模組來進行持續性深度學習。期望能避免災難性遺忘、並且模型能與時俱進以應對多樣的釣魚網頁危害。



圖六、持續性深度學習的概念圖[17]。

第六節 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

[17] German I. Parisi , Ronald Kemker , Jose L. Part , Christopher Kanan , Stefan Wermter (May 2019) Continual lifelong learning with neural networks: A review.
https://www.sciencedirect.com/science/article/pii/S0893608019300231