國立台北大學資訊工程學系專題報告

HALA-Half Auto Log Analysis

-網路攻擊記錄檔分析系統

專題組員:林庭葦、王柏文

專題編號:PRJ-NTPUSCIE-107-008

執行期間:107 年 09 月至 108 年 05 月

1. 摘要

在現今網路愈來愈發達的時代，網路攻擊也日漸增多。而且在這個硬體效能迅速成長的時代，機器學習得到了速度上的改進，大家都開始把機器學習應用到各個領域上。但在資訊安全的領域上，機器學習的應用相較於其他的領域是較少的，但已經有駭客開始用機器學習的方式來做網路攻擊了。因此在資訊安全結合機器學習的這個領域上可以說是還處在一片混沌的狀態。但總是要做的，因此我們就有想要朝向機器學習結合資訊安全的方向，做一個可以減少人工鑑識攻擊的工具。

本專題研究該如何從大量的伺服器網站記錄檔，利用機器學習來辨識攻擊，並且使用資料視覺化，使人工鑑識記錄檔的時間大幅的縮短。透過刑事局所給的攻擊記錄檔來訓練機器學習之模型，預測網路攻擊。

2. 簡介

(一) 研究背景

現今資訊安全的議題愈來愈受到重視，駭客也不停的想辦法要做攻擊，尤其是以網頁攻擊為大宗，因為其容易入門，屬於攻擊中最好入門的一種。面對大量的網路攻擊，伺服器總是會留下 Log 記錄檔來記錄蛛絲馬跡，然而記錄下來的數量非常多，以人工的方式去觀察裡面是否有存在網路攻擊並且把它們全部都找出來，簡直是天方夜譚。有人使用 IDS，入侵檢測系統來檢測是否記錄檔中存在著網路攻擊，然而傳統的 IDS 都是用基於正規表達式來找尋攻擊，這種方式雖然能找出大部份的攻擊，但是誤報率極高，使得多出了一堆多餘的警報讓使用者沒辦法從大量的警報中找出真正的攻擊，另外面對正規表達式，駭

客只要聰明一些在攻擊做一些手腳，正規表達式不一定可以檢測出來，且面對未知的攻擊，正規表達式根本毫無作用，這使得使用 IDS 又有了一些壞處。因此我們往利用機器學習的方式來辨識網路攻擊的方向走。

(二) 研究目標

本專題把機器學習技術應用在資訊安全上面。運用機器學習來辨識記錄檔裡的攻擊，並且把資料給圖表化和統整，使使用者能更快的找到每個攻擊間的關聯，不必再像使用 IDS 一樣，必須承受高誤報率且也沒有工具可以統整記錄檔的結果。

(三) 主要預期效益

1. 使用機器學習能成功正確地辨識伺服器記錄檔且把誤報率給降低使誤報的數量不會干擾到尋找記錄檔裡的攻擊。
2. 架設網站使使用者可以把想分析的記錄檔傳上去交給我們分析。
3. 把資料做圖表化，並把資料統整的更清楚，讓重要的資料可以被使用者一瞬間就發現並找出它與其他資料的關聯。

3. 專題進行方式

(一) 實作平台

(a)前端

使用 Vue. js 當做前端的框架，並且利用 D3. js 框架來做資料圖表上的呈現方式，並利用 axios 來傳輸資料給後端。

(b) 後端

使用 Django 框架當做後端的框架，且搭配 python 做資料處理以及機器學習。

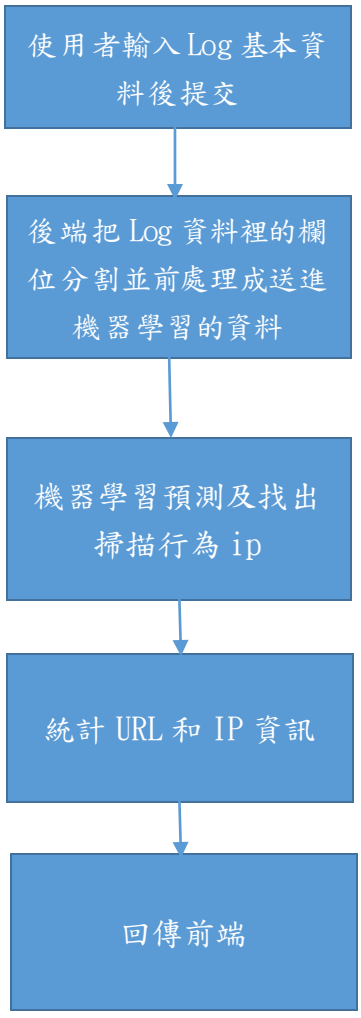
(c)資料庫

使用非關聯式資料庫 MongoDB，由於彈性佳，且速度快，適合用來處理大數據。

(d)機器學習

使用 SVM 做為機器學習的方法，由於其他機器學習的準確率最後結果都沒有 SVM 的結果來得準，故最後使用 SVM 做為機器學習的方法。

(二) 流程



（三）資料集來源

資料集來自刑事局的 Log 資料以及我們自己用 kali linux 做滲透測試所產生的資料。由於在記錄檔裡面因為攻擊數量極為的稀少，所以需要從 kali linux 做攻擊靶機來取得資料，否則會產生正負樣本數量不對等導致學習結果不彰。

（四）SVM

SVM 是一種監督式學習的機器學習演算法，基本概念就是找到一個決策邊界讓兩類之間的邊界最大化。在多維空間裡，一個特徵表示一個維度，一個資料有 n 個特徵就有 n 個維度。在 n 個維度裡用一個超平面(hyperplane)來分割資料，超平面則是多維空間裡的一個平面。使資料得到分類。大概如下圖 1。

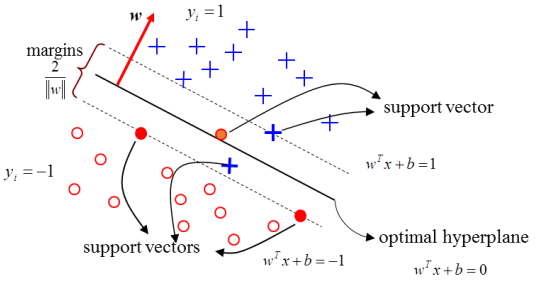


圖 1 SVM 分割示意圖

我們假設超平面 $wTx+b$ 可以完美的分割一個資料集的兩組資料， SVM 就是在找參數 w 和參數 b 來使這兩組資料跟超平面的距離最大化，也就是找最佳化的超平面。

（五）特徵選取

我們選用了 15 個特徵作為我們機器學習的特徵，機器學習需要好的特徵，不然就只是 garbage in gabage out 而已。我們的特徵如下表

關鍵字出現次數
Query 長度
URL 長度
總長度
數字在 Query 出現次數
大寫字母在 Query 出現次數
大寫字母在 URL 出現次數
特殊符號在 Query 出現次數
特殊符號在 URL 出現次數
空格出現次數
空格出現比例
大寫字母出現比例
HTTP 方法種類
HTTP Status Code 種類

上述特徵為我們 SVM 的特徵，每一個都跟攻擊有密不可分的關係。如在許多攻擊裡面，攻擊的長度都是很長的，且裡頭也有包含許多大寫字母來做編碼，數字也有可能被拿來做編碼。至於空格在 SQL injection 攻擊裡非常常見到，所以也列入，還有特殊符號也是因為許多攻擊會用到，例如注解掉使用者的程式碼，HTTP 方法則是看使用者是不是使用一般的 HTTP 方法，還是說是使用一些會影響到網站安全的 HTTP 方法，如 PUT，DELETE 等。HTTP Status Code 則是用來看訪問的狀態，是成功的，又或是失敗的? 這些特徵都是跟網路攻擊習習相關的特徵，算是整個專題比較核心的部份，都是在選特徵。

（六）URL 檢測

傳統的 IDS 沒有做到這點，導致

許多駭客反而都從 URL 的部份來做突破，使得許多的 IDS 都無法抓到 URL 裡的攻擊，導致漏報的問題發生。我們也有觀察到這個現象後，所以也才把 URL 的一些資訊納入我們的特徵，這樣漏報率才不會高。

(七) 掃描尋找

我們判定一個 request 為掃描的標準是以同一個 ip，在一秒鐘內如果有訪問超過三次的話，那我們就把那個 ip 視做為掃描行為。如果有掃描行為的話，我們就會在那個 request 上做標記。

(八) 整合顯示前端

我們把所有有被標記的 request 整合起來，包括它的掃描和攻擊，統計它們的資料，然後把它們分成 IP 和 URL 和流量三部份，送去給前端。實際上只會傳流量的部份去給前端而已，一來因為三個部份一起傳去給前端需要花費太長的時間，二來因為前端的記憶體有限，而且瀏覽器也不一定能承受那麼大的資料，D3.js 也不一定可以承受那麼大的資料運算。所以我們只有傳流量的部份到前端，其他的部份由前端動態來跟後端取資料。前端的部份會顯示流量及掃描和攻擊的小時分秒三個層次的流量圖，並進一步可以看到裡面的 request 詳細資料。

(九) 建議

我們還有建議系統，主要就是算出一天之中攻擊最多的一小時，藉由找出這一小時內攻擊最多的 ip，看它有沒有超過總攻擊占比的 20%，如果有的話，那我們把它視為集中式攻擊，反之我們視為分散式攻擊。而掃描也同樣有一模一樣的建議系統，使用者可以從建議系統去判斷到底是集中式攻擊及掃描或者是分散式攻擊及掃描。

(十) 200 攻擊

我們還會列出所有的 200 攻擊給使用者看。雖然 200 攻擊不一定代表攻擊是一定成功的，但是還是有著其參考性，我們會列給使用者看全部的 200 攻擊，讓使用者知道到底發生了什麼事。

4. 主要成果與評估

(一) 機器學習結果

我們的判斷為攻擊的標籤為 1，正常樣本的標籤為 0。

機器學習在訓練集上的訓練結果如表一

True Positive	225
True Negative	29763
False Postive	0
False Negative	13

表一

Accuracy 為 0.99956668

Precision 為 1

Recall 為 0.94537815

而在測試集上的預測結果如表二

True Positive	1947
True Negative	75035
False Positive	492
False Negative	4

表二

Accuracy 為 0.99359818

Precision 為 0.79827798

Recall 為 0.99794977

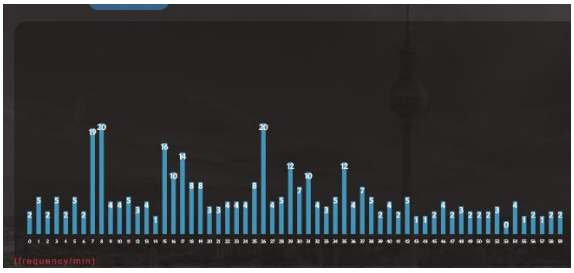
可以看出效果其實算是滿好的，加上其實我們透過一些誤報的資訊發現我們的資料集有一些資料沒有清洗乾淨。如圖二

圖二

就是一個典型的 HTTP Splitting 攻擊，它只會針對 URL 做攻擊，所以 IDS 完全檢測不到，我們已經找出大部分了，但是還有少部份沒有找到，如這個例子，但 SVM 幫我們找出來了，這就到達我們原本的目標了，也就是降低誤報率以及找出未被 IDS 偵測到的攻擊。

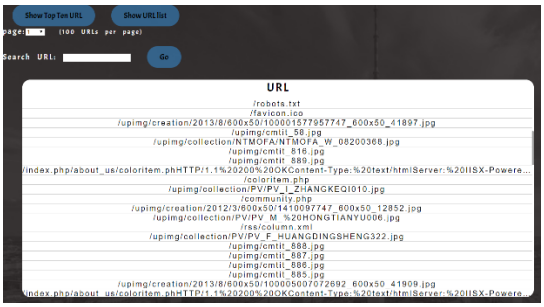
（二）網站結果

攻擊掃描流量的長條圖如圖三



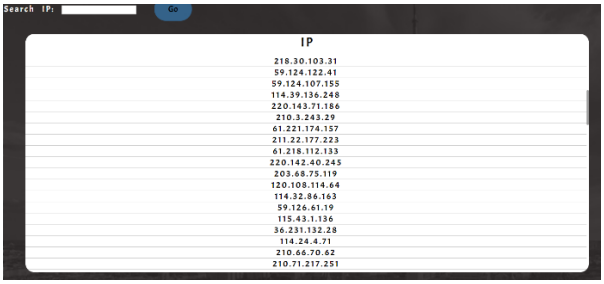
圖三

URL 的結果如圖四，另外有顯示前十大易受攻擊的 URL 功能，以及搜尋 URL 詳細資料的功能。



圖四

IP 的結果為圖五，另外說明 IP 的結果列表是由分數高到低排下來的，IP 打分數的根據是由他的攻擊次數以及掃描次數來做決定。另外 IP 也有搜尋功能。



圖五

建議功能如圖六

圖六

所有 status code 為 200 的攻擊列表如圖七

圖七

（三）Log 檔支援格式

我們的工具支援兩種 log 檔的格式，一種是 IIS 的，也就是 windows 伺服器記錄檔的格式，一種是 Access，也就是 Apache 記錄檔的格式。我們會提供讓使用者選取欄位順序，由於每個網管人員在設定記錄檔格式時，欄位都可以做調換或更改，如果不先讓使用者選擇順序，會導致我們不知道使用者的記錄檔到底長怎樣，造成無法分析。所以一定要給使用者選取欄位。

（四）未來可擴展方向

其實要找這種 request 是否為攻擊的問題，可以把它更進一步的簡化，把它看成 NLP 問題，用深度學習來解，只是在分詞上可能還要有待解決。以及資料集的大小，也是一大問題，因為深度學習就是需要大量的資料在撐，所以要怎麼取得這麼龐大的資料又是一個大問題。

另一個就是往時間方面去做分析，把各個 ip 及 url 攻擊的時間序列分析一下，找出其中的關聯，這也是一個可以擴展的方向。

還有一個最有潛力的擴展方向，便是利用生成對抗網路(GAN)，一種深度學習方式，來產生對抗樣本，自己產生攻擊樣本來解決攻擊樣本不足的問題，就如同應用在圖片上一樣，可以把每個 request 當作圖片的一列，然後把它們集合成一個圖片，利用 GAN 去找出哪個是攻擊哪個不是，這是一個非常好的想法，只可以礙於技術不足及資料量不夠和時間不夠，無法做出來。

5. 結語與展望

本次專題成功實踐了目前非常少人實作的把機器學習應用到資訊安全這一塊。雖然我們仍不知道這個系統能不能檢測出更刁鑽的攻擊，但我們的確使人工找出記錄檔裡的資訊的時間大大的減少了，不用像以前必需慢慢地看才可以。我們也有對許多重要的攻擊資訊都有把它篩選出來呈現給使用者看，如果使用者是個資訊安全專家，想必也可以一目瞭然所想要的資訊。最後期望有人可以遵循可擴展方向把這一塊做得更好，改善這個工具並且加強資訊安全對社會所造成的影響。

6. 銘謝

特別感謝指導教授在各方面對我們的指導以及有時候也能包容我們進度有所落後，我們最後還是把成品做出來了。

感謝刑事局所提供的資料。

感謝在機器學習應用在資訊安全領域的所有拓荒者。

7. 參考文獻

[1] https://medium.com/@chih.sheng.huang821/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-%E6%94%AF%E6%92%90%E5%90%91%E9%87%8F%E6%A9%9F-support-vector-machine-svm-%E8%A9%B3%E7%B4%B0%E6%8E%A8%E5%B0%8E-c320098a3d2e

[2]
https://zhuanlan.zhihu.com/p/2944
4990