二、研究計畫內容(以10頁為限):

(一)摘要

本計畫將透過anaconda所提供的編譯器spyder來開發，利用python所寫出的scapy網路封包分析函式來挖掘和監測封包，配合Nprobe來偵測網路流量，利用netflow v9的流量分析資料表格標準得到封包數據，找到有問題的封包或是不正常的數據流量，並將資料進行分類，將需要的資料儲存在SQLite資料庫中，使用Scikit-learn實作機器學習，進行使用者上網行為分析，並以圖表方式呈現，能記錄下上網情形，且達到即時發現和防範網路上的不安全因素的效果。

(二)研究動機與研究問題

近年python語言越來越受到歡迎，除了簡潔的語法與程式碼可讀性外，提供了非常多的套件可以使用，而在網路分析方面，不管是netflow的python版本還是scapy函式都非常強大且實用，所以才有了使用python擷取封包實作使用者上網行為分析的想法。

因應大數據潮流與網路安全管理日漸重要，因而有此計畫，透過封包分析，讓公司管理者清楚了解員工的網路使用情形，得知何者或何事佔用了頻寬，或是能夠了解公司網路的尖峰使用時間，做為頻寬分配或使用者限制的參考標準，能讓管理者更方便管理和做後續應用。

個人使用方面，能夠了解自己的上網習慣，將常使用的網站名稱或是IP鍵入資料庫內容，進行上網行為分析，當有與以往不同且不正常的上網行為發生時能及時發現，並找出問題，讓對網路沒有深入了解者也能大略知道網路的使用情形。

最簡單的分析方式就是建立port number的分析，如做出https,DNS,http幾種服務的比例圖，如果https比例最多，代表網路流量以網頁服務為主，此種方式能做粗略的上網行為分析。如果要做再更深入一點的紀錄，便是將造訪的網站和造訪次數鍵入資料庫內，將類似於http協定的網站，這種安全性較低的網站記下，讓人可以點擊查詢。有太多太多的資料及方式能進行網路管理，但是這也造成網路資料繁多且複雜，讓一般人難以了解自己流量到底被甚麼占用，即使電腦被攻擊或入侵也不知道。

本研究基於以上動機，希望設計一個能幫助不管是管理者還是使用者都能使用的程式，如果有一程式能將較為重要資料如流量、特殊封包種類等，先分類規劃好放入資料庫內，結合用戶流量分析和web訪問分析的方式，以圖表方式呈現，那這樣能讓網路管理者更方便進行網路管理，清楚得知使用者平時的上網行為，那將會方便許多，使用者也不會對自己網路情形一無所知。

(三)文獻回顧與探討

擷取封包用來分析網路狀況能找出異常情形和網路流量，也是學習各種網路協定的好方式，網路上能找到許多抓取封包的程式和方法可以提供我們做參考。

tcpdump是一款強大的網絡抓包工具，運行在 linux 平台上，允許用戶截取和顯示透過網路傳送或收到的[TCP/IP](https://zh.wikipedia.org/wiki/TCP/IP)和其他[封包](https://zh.wikipedia.org/wiki/%E6%95%B0%E6%8D%AE%E5%8C%85)，利用各式指令顯示出抓包結果，並能夠指定擷取某個主機所收到和發出的數據封包，但要清楚了解並利用tcpdump得到的封包資訊，必須要有一定的網路基礎。

Wireshark以前名為Ethereal，是一個免費開放原始碼的網路[封包](https://zh.wikipedia.org/wiki/%E5%B0%81%E5%8C%85)分析軟體，提供了各種不同作業系統所使用的版本，功能強大，提供良好的圖形介面，還有方便的過濾器功能，利用關鍵字和條件，能指定抓取封包種類，wireshark可以解析超過七百種的通訊協定，幾乎所有協定都可以解碼，配套工具包是npcap，npcap是由winpcap改進而來，提供更好的性能、速度和安全性。

Netflow也是監聽網路數據的一種方式，基本分為exporter和collector兩個部份，exporter負責收集資訊，collecter則是顯示出得到的資訊結果，最早由cisco研發，提供多種版本v5、v7、v8、v9，分析七個關鍵參數(Source IP address、Destination IP address、Source Port、Destination Port、Protocol Type、 Type Of Service、 Input Logical Interface)做出後續應用，提供了輸出格式方便觀看數據，但在使用netflow時，會對設備性能有一定的影響，在高負載情況下要謹慎使用。

Scapy是由python語言組成的網路封包分析函式，有與wireshark類似的功能，抓取封包後能詳細顯示封包的數據，也能自行製作封包，能配合python寫出想要顯示的封包和內容，在開發方面給予了更大的彈性，本研究將針對此函式監聽部分做應用，基本上只要下達sniff()函式就能開始抓取封包，scapy也提供了取多參數以供使用，像是iface能用來指定網卡，count用來限制抓取量，timeout限制抓取時間等。

Scikit-learn是python用來實作機器學習的一個套件，scikit-learn已將會需要用到的演算法大致分為六類，只要選好演算法，便能直接使用，scikit-learn提供了dataset機器學習資料集給使用者做練習，也能創建自己的資料參數，方便自己學習，結果的呈現方式也能自行定義。

本研究將斟酌使用以上提到的套件來實現使用者上網行為分析，希望達到能預測某台電腦被入侵的機率高低分析，亦或是對使用者平時上網行為做資料統整。

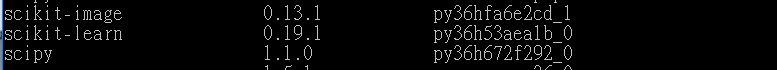
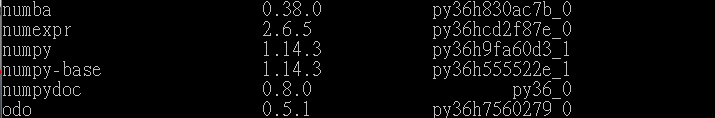
(四)研究方法及步驟

4.1研究方法

1.將以python語言開發為主，先將anaconda和python下載安裝完畢，接著將netflow環境建設完畢。

2.使用anaconda prompt 並下達 pip install scapy ，下載好scapy函式。

3.在下載scikit-learn前，先確認python版本是否在2.6或3.3版本以上，接下來要確認有沒有下載numpy，且版本要在1.6.1以上，scipy有在0.9版本以上，之後便可以開始下載scikit-learn。基本上有裝anaconda，這些套件都會一起安裝好，如果不確定以上套件有沒有安裝，可以在anaconda prompt 下達conda list指令。



4.以上環境建設完畢，便可以開始抓取資料，主要使用scapy的sniff功能，將資料分類處理好後建立資料庫，最後以圖表方式呈現。

4.2研究步驟

製作介面顯示結果

機器學習製作圖表

選擇資料加入資料庫

抓取封包資料

1. 抓取封包資料

(六)參考文獻

[1] 上網行為分析 - Huawei Technical Support

<http://support.huawei.com/hedex/pages/EDOC1000089429DZE1112R/03/EDOC1000089429DZE1112R/03/resources/vsm_cfg/analyze/vsm_cfg_analyze_0500.html>

[2] 網路行為分析 - 國立臺灣大學計算機及資訊網路中心C&INC, NTU

<http://www.cc.ntu.edu.tw/chinese/epaper/0044/20180320_4407.html>

[3] 抓包神器tcpdump 使用介紹- 掃文資訊

<https://hk.saowen.com/a/f40476f54e1940b74e43d181ab314892298002d4f04791a3d524d7164b9c70a9>

[4] 網路封包分析的好幫手—Wireshark 擷取分析、防範攻擊無所不包- 技術專欄

<https://www.netadmin.com.tw/article_content.aspx?sn=0808050013>

[5] Jan Ho 的網絡世界- NetFlow

<https://www.jannet.hk/zh-Hant/post/netflow/>

[6] Python 使用scapy 进行抓包 - xiaix's Blog

<https://xiaix.me/python-shi-yong-scapy-jin-xing-zhua-bao/>

[7] 機器學習：使用Python

<https://machine-learning-python.kspax.io/>

[8]