即時網路流量異常檢測系統

專題編號:114-1-CSIE-S019
執行期限:113年第1學期至114年第1學期
指導教授:張世豪
專題參與人員: 111590024 楊宇詮
111590039 魏暐宸
111590040 賴政良
111590034 杜彥奇

一、摘要

本專題旨在應用人工智慧技術進行惡意網路流量之偵測，透過網路封包監聽工具擷取封包資料，並使用flow-based aggregation，以建立完整的網路流量紀錄。專題中採用非監督式學習方法，包含圖神經網路之圖注意力機制(Graph Attention Network, GAT)以及變分自編碼器(Variational Autoencoder, VAE)，對網路流量特徵進行嵌入與異常模式辨識，所構建之模型可於系統部署後進行再訓練，以提升對多變網路環境中潛在惡意網路流量之適應能力。本專題所提出之方法具備有效辨識異常流量之能力，為資安防護提供初步的防禦機制。
關鍵詞:異常偵測、網路流量分析、非監督式學習。

二、緣由與目的

在這個資訊量爆炸的年代，不論是企業伺服器或是個人電腦(尤其是自媒體行業)，都有受到網路攻擊的可能。近年來新聞報導因為網路攻擊並造成損失的數量增加，網路攻擊也有可能發生在你我身邊，防禦網路攻擊的第一步是檢測，但目前市面上的即時網路攻擊檢測系統都較為複雜且門檻(入手或使用)較高，因此我們期望撰寫一個能夠簡單入手且方便使用的惡意網路流量偵測系統。

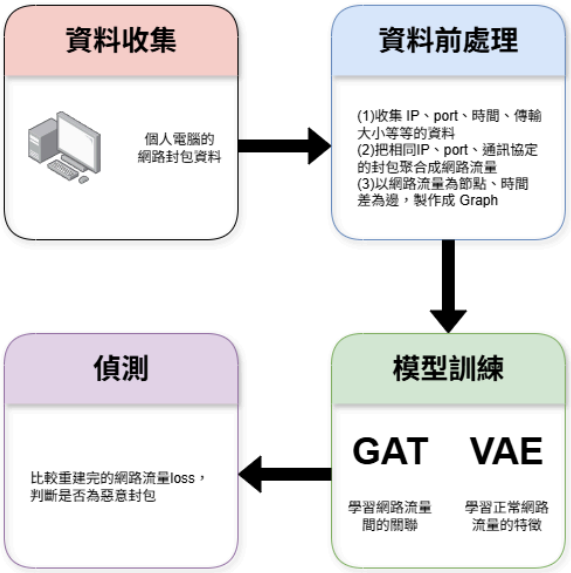
三、使用技術方法

(一)硬體環境:
CPU : Intel 13th i7-13620H
GPU : NVIDIA GeForce RTX 4060
(二)程式語言:Python
(三)擷取封包工具:Scapy
(四)人工智慧框架:PyTorch
(五)人工智慧模型架構:GAT+VAE

Graph Attention Network(GAT)是一種基於注意力機制的圖神經網路模型，透過為鄰接節點分配不同的注意力權重，使模型能自適應地學習節點間的關聯重要性。此特性使 GAT 在惡意網路流量這這類異質性較高的資料中表現出色。由於能有效整合局部與全域的結構資訊，GAT能在網路流量的圖結構中使惡意流量更加容易被發現。

變分自編碼器(VAE)是一種基於深度學習的生成模型，透過將輸入資料壓縮至潛在空間並重建原始輸入，學習出正常網路流量的分佈特徵。在惡意網路流量偵測中，VAE可利用重建誤差判斷輸入是否偏離正常模式，當重建誤差較大時，判定為可能的異常行為。由於VAE能有效捕捉資料的非線性特性與潛在結構，因此可降低傳統方法的誤報率，並顯著提升入侵偵測的準確度與穩定性。

四、系統架構圖



五、實作流程

(一)訓練資料蒐集:網路上的公開資料集(UNSW-NB15)、使用scapy配合Pcap擷取自己的主機平時的正常網路流量、並儲存為.csv檔供後續使用

(二)建立人工智慧模型:本專題使用GAT、VAE，利用GAT的注意力機制讓流量之間產生不同權重，利用VAE能夠學習輸入資訊的機制，讓VAE學會正常網路流量，利用重建誤差判斷惡意網路流量。

(三)訓練人工智慧模型:將資料正規化，經過「圖」轉換，以網路流量的統計特徵作為token，時間、SrcIP及DstIP等資訊為token間的關聯，產生尚未經過注意力關聯的圖資料集，再將此資料集輸入模型進行訓練。

(四)微調模型:嘗試各種數值的變化，找出最適合本專題所需的人工智慧模型數值。

(五)實作 UI 介面:讓使用者方便啟動並執行此分析系統，能夠隨時觀察自己電腦的網路流量是否異常，或是選擇微調模型，讓模型學習現有的網路流量。

六、實驗結果

先使用具有較多資料的UNSW-NB15的部分正常網路流量資料進行訓練，之後用組員平時擷取到的網路流量進行微調。

使用UNSW-NB15的部分網路流量(包含惡意網路流量)進行測試，透過人工智慧模型輸出的loss(VAE之重建誤差)值判斷該流量是否為惡意網路流量，由圖(1)可見，本專題的人工智慧模型能夠對網路流量進行分類，在紅線之左皆被判定為正常流量(loss<10)。

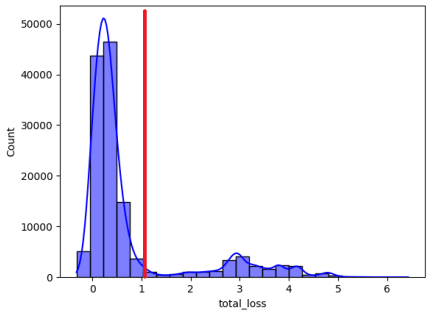


圖1.測試資料之loss分布，x軸的loss值經過log10計算。(此處loss皆為VAE之重建誤差)

七、結論

本專題擁有簡潔的圖像化界面可以讓使用者沒有任何門檻使用本專題提出之惡意流量偵測系統。專題中的人工智慧模型也擁有再訓練的能力，能夠根據不同網路環境進行調整，強化其適應能力。目前能夠偵測是否被攻擊，未來可著重於偵測出網路攻擊之攻擊模式，可以針對受到的攻擊模式配合其他防禦系統做出最好的反應，完成完整的網路攻擊偵測、防禦系統。

八、參考文獻

[1] CIC-DDoS2019
Iman Sharafaldin, Arash Habibi Lashkari, Saqib Hakak, and Ali A. Ghorbani,

"Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy", IEEE 53rd International Carnahan Conference on Security Technology, Chennai, India, 2019.

[2] UNSW-NB15
Moustafa, Nour, and Jill Slay. "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)." Military Communications and Information Systems Conference (MilCIS), 2015. IEEE, 2015.
Moustafa, Nour, and Jill Slay. "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 dataset and the comparison with the KDD99 dataset." Information Security Journal: A Global Perspective (2016): 1-14.
Moustafa, Nour, et al. "Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks." IEEE Transactions on Big Data (2017).
Moustafa, Nour, et al. "Big data analytics for intrusion detection system: statistical decision-making using finite dirichlet mixture models." Data Analytics and Decision Support for Cybersecurity. Springer, Cham, 2017. 127-156.
Sarhan, Mohanad, Siamak Layeghy, Nour Moustafa, and Marius Portmann. NetFlow Datasets for Machine Learning-Based Network Intrusion Detection Systems. In Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings (p. 117). Springer Nature.