Wydział Fizyki, Matematyki i Informatyki	Zarządzanie projektem informatycznym	4.02.2019r.
Informatyka Stosowana II semestr	Projekt zaliczeniowy	Autorzy: • Kurdziel Kordian • Maciejak Mateusz • Mikołajczyk Mariusz • Milanović Aleksander • Uryga Marcin

1. Cel projektu

Zadaniem projektu było spreparowanie plików z logami w celu zasymulowania wystąpienia ataków na bezpieczeństwo systemu lub sieci. Ponadto, kolejnym zadaniem było zbudowanie sztucznej sieci neuronowej, której zadaniem było odczytywanie przygotowanych plików z logami i próba wykrycia jak największej liczby ataków na podstawie tych danych.

2. Przygotowanie plików logów

W naszym projekcie wybraliśmy dwa proponowane pliki z logami:

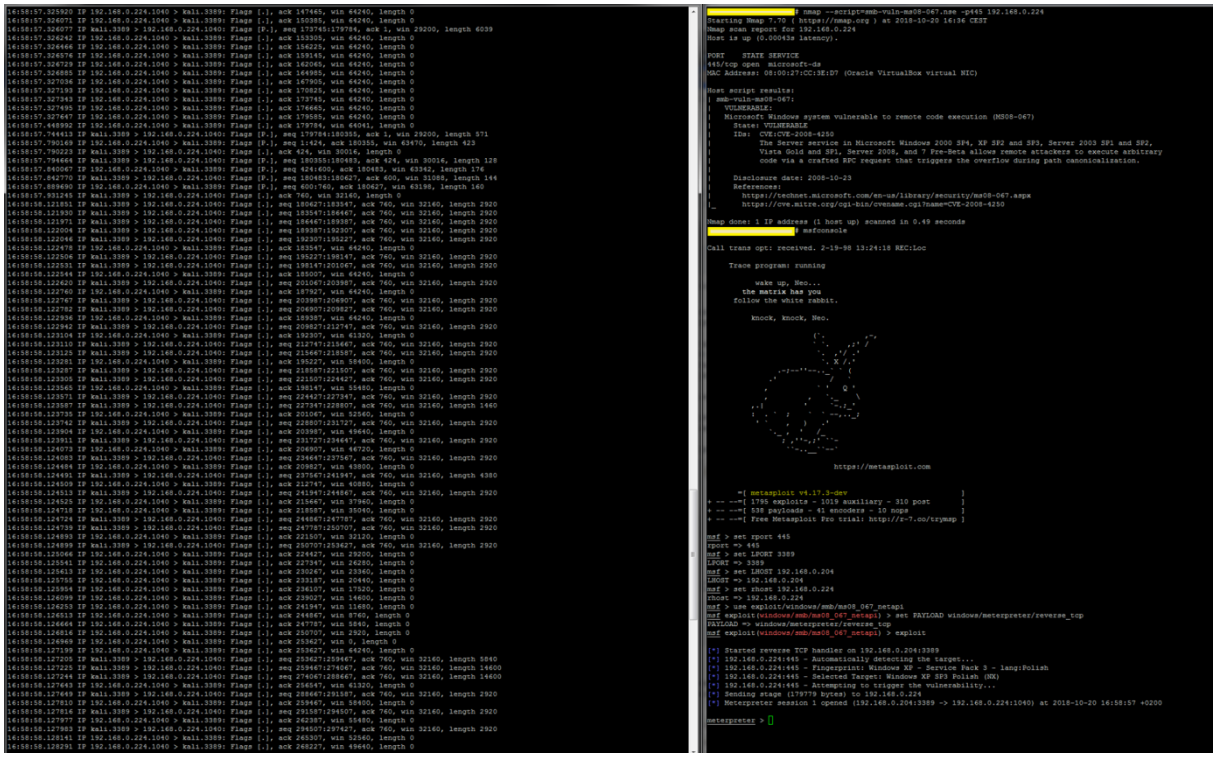
1) Plik flows.txt – zawierający dane nt. przesyłanych danych pomiędzy komputerami w danej sieci. Plik ten zawierał informacje nt. czasu wystąpienia, czasu trwania transmisji, komputera źródłowego, portu źródłowego komputera, komputera docelowego, portu komputera docelowego, rodzaju protokołu, liczby pakietów oraz liczby przesyłanych bajtów.

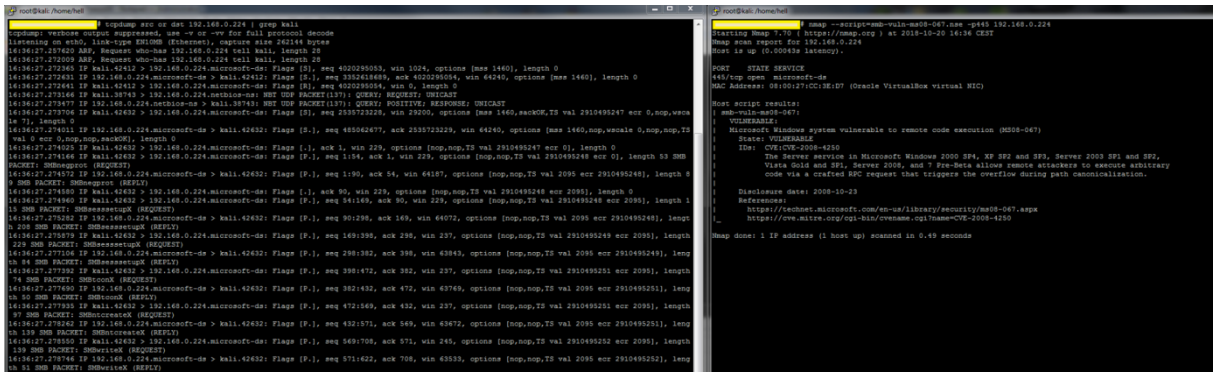
   W pliku tym ukryliśmy dwa ataki:
   a) **Man-in-the-middle (MITM)** – atak polegający na podsłuchu i modyfikacji wiadomości przesyłanych pomiędzy dwiema stronami bez ich wiedzy
   b) **MS08-067: Vulnerability in Server Service Could Allow Remote Code Execution –** luka w bezpieczeństwie systemu operacyjnego Windows pozwalająca na zdalnym przejęciem kontroli na komputerem za pomocą wysłania spreparowanego żądania RPC.

2) plik procs.txt – plik zawierający logi ze zdarzeń uruchomienia i zakończenia wykonywania procesów na komputerach w danej sieci. Plik ten zawierał informacje nt. czasu wystąpienia, użytkownika oraz nazwy komputera, na którym był uruchomiony proces, nazwy procesu oraz informacji, czy dany proces był wówczas uruchomiony czy zatrzymany, reprezentowaną za pomocą ciągów znaków *Start* oraz *End.*

   Plik ten został spreparowany tak, aby zawierał logi jak podczas ataku **Remote Code Execution.** Atak tego typu pozwala atakującemu na zdalne kontrolowanie komputera swojej ofiary bez względu na lokalizację jej urządzenia. Atak tego typu jest możliwy poprzez wykorzystanie różnych luk bezpieczeństwa w systemie operacyjnym.

Do spreparowania ww. plików zostały wykorzystane dwie maszyny wirtualne z systemami operacyjnymi Kali Linux oraz Windows Server 2003. Kali Linux zawierał w sobie potrzebne narzędzia do testowania ataków oraz do podsłuchu ruchu sieciowe podczas wykonywania wybranych ataków. Dzięki takiemu podejściu przygotowane przez nas pliki logów były bardziej precyzyjne i wiarygodne, ponieważ spreparowane dane były dodawane na podstawie wyników rzeczywistych eksperymentów. Podsłuchany ruch sieciowy był podstawą do odwzorowania pierwszego etapu ataku w logach. Poniżej umieszczono kilka zrzutów ekranu prezentujące proces testowania ataków za pomocą ww. narzędzi:





3. Wykrywanie ataków za pomocą sztucznej sieci neuronowej

Kolejnym etapem wykonania projektu było przygotowanie i utworzenie sztucznej sieci neuronowej, której zadaniem było wykrycie wystąpienia ataków na bezpieczeństwo w systemie poprzez wykonanie analizy plików z logami. W tym celu wykorzystaliśmy bibliotekę języka Python – Tensorflow, która zawiera pełny zestaw gotowych implementacji narzędzi do budowania sieci neuronowych.

Przed przystąpieniem do prac zostały przygotowanie dane trenujące z plików logów. Dane trenujące były pewnym wycinkiem logów, zawierającym badany atak. Dane te były konieczne do procesu trenowania sieci, aby podczas analizy całych plików z logami mogła poprawnie wykrywać ataki.

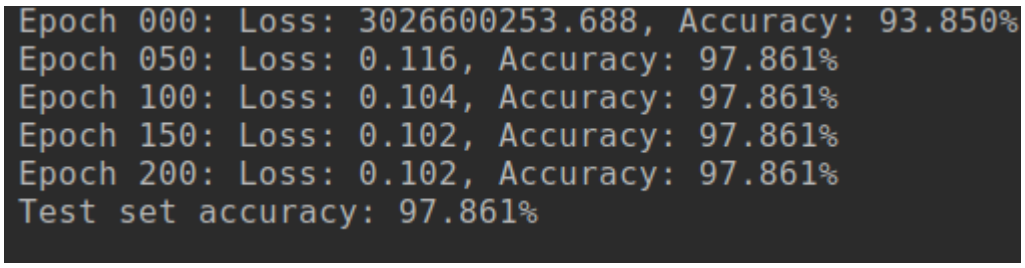
Przygotowana przez na sieć neuronowa została zbudowana wraz z następującymi parametrami:

- Liczba warstw ukrytych: 2
- Algorytm uczenia: stochastic gradient descent (SGD)
- Funkcja aktywacji: Rectified Linear Unit (ReLU)
- Liczba epok trenowania: 201

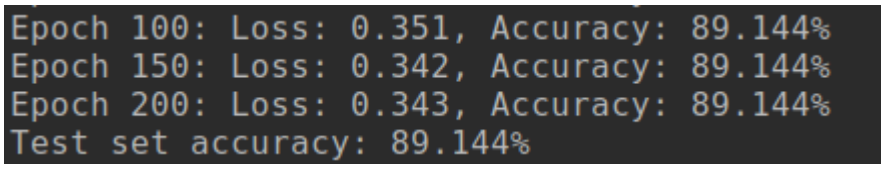
4. Wyniki wykrywania ataków z wykorzystaniem SSN

Podczas testowania utworzonej przez nas sieci neuronowej nie udało się wykryć ataków. Niska wydajność wykrywania była spowodowana specyfiką ataków zawartych w plikach logów. Ataki te były zawarte w kilku rozdzielonych od siebie wpisach, co utrudniało skuteczną ich analizę. Ponadto, jako dane trenujące sieć został wybrany mały podzbiór dużego pliku logów. Powodem doboru takich danych trenujących była ograniczona wydajność komputera, na których trenowano sieć; wybór zbyt dużego pliku do trenowania mógłby pochłonąć bardzo dużo czasu przy jednoczesnym wykorzystaniu słabej jednostki obliczeniowej.

Poniżej zamieszczono zrzuty ekranu prezentujące osiągniętą dokładność podczas testowania sieci:



Rysunek 1. Zrzut ekranu prezentujący osiągniętą dokładność podczas testowania sieci na pliku logów procs.txt



Rysunek 2. Zrzut ekranu prezentujący osiągniętą dokładność podczas testowania sieci na pliku logów flows.txt

5. Wnioski

Podsumowując wykonane prace udało się przygotować pliki z logami, w których zawarto kilka ataków na bezpieczeństwo systemu. Dzięki wykorzystaniu dwóch maszyn wirtualnych oraz profesjonalnych narzędzi do wykrywania i testowania luk bezpieczeństwa spreparowane pliki z logami zawierały jak najlepsze dane, odzwierciedlające możliwe, rzeczywiste ataki.

Dzięki wykorzystania narzędzi sztucznej inteligencji możliwe jest zbudowanie model sztucznej sieci neuronowej, której zadaniem jest wykrywanie ataków na bezpieczeństwo na podstawie analizy dostarczonych plików z logami. Dzięki oddzieleniu procesów trenowania sieci i jej rzeczywistego działania na danych możliwe jest jej wcześniejsze przygotowanie do właściwego wykrywania ataków. Narzędzie to efektywnie i szybko może posłużyć do detekcji i monitorowania sieci w celu wykrycia szkodliwego działania przez niepowołane osoby.