semestr letni 2020/2021

Politechnika Rzeszowska
Wydział Matematyki i Fizyki Stosowanej
Programowanie w R

Projekt w R - notatki do prezentacji

Vitalii Morskyi
Makarska Julia
Grupa L04
Inżynieria i analiza danych

Spis treści

1 Slajd 1.

Zjawisko Phishingu zachodzi już od wielu lat. Jednak między innymi ostatni rok pokazał nam jak ważne jest bezpieczeństwo w internecie. Od roku świat się zatrzymał i przeniósł wszystko do Internetu. Z uwagi na ten fakt podjęcie tematu Phishingu uznaliśmy za bardzo na miejscu. Chcemy pokazać jak łatwo można dać się okraść. Przedstawiony przez nas projekt obejmuje tylko niewielki kawałek tej metody oszustwa, jednak uznaliśmy, że temat jest ciekawy.

Phishing jest to atak oparty na wiadomościach e-mail lub SMS. Przestępcy internetowi próbują Cię oszukać i wymusić na Tobie działania zgodne z ich oczekiwaniami.
Wymowa nazwy tego oszustwa budzi skojarzenie z łowieniem ryb nie bez powodu. Przestępcy, tak jak wędkarze stosują odpowiednio dobraną "przynętę".

2 Slajd 2.

Korzystając z danych z roku 2019 można wyciągnąć prosty wniosek: Phishing to potężne zagrożenie. W roku 2019 zanotowano aż 65% wzrost ataków. Pod kątem wycieków newralgicznych danych phishing odpowiedzialny jest aż za 90% wszystkich przypadków. Wygenerował on 12 mld dolarów strat.

3 Slajd 3.

Korzystając z danych z roku 2019 można wyciągnąć prosty wniosek: Phishing to potężne zagrożenie. W roku 2019 zanotowano aż 65% wzrost ataków. Pod kątem wycieków newralgicznych danych phishing odpowiedzialny jest aż za 90% wszystkich przypadków. Wygenerował on 12 mld dolarów strat.

4 Slajd 4.

Głównym założeniem naszego projektu jest analiza adresu URL pod kątem niebezpiecznych kombinacji znakowych. Skupiamy się głównie na takich częściach URL jak:

- protocol (schemat)
- domain name (nazwa serwera)
- path (ścieżka do pliku)
- query (zapytanie)
- fragment (fragment)

Części te zostały wybrane z względu na to, że skupiamy się na analizie leksykalnej.
Następnie części te badamy pod kątem:

- długości
- statystyki długości:
  - nazwa domeny do adresu URL
  - ścieżka do adresu URL
  - argument do adresu URL
  - ścieżka do nazwy domeny
  - argument do nazwy domeny
  - argument do ścieżki
- ciągu znaków postaci litera-cyfra-litera
- ciągu znaków postaci cyfra-litera-cyfra

- połączenia powyższych ciągów
- liczby liter
- liczby cyfr
- liczby znaków interpunkcyjnych

Wybraliśmy dwa pliki i zmodernizowaliśmy tak by odpowiadały badanym cechom.

5 Slajd 5.

Przejdziemy teraz do prezentacji kodu.

6 Slajd 6.

Pierwsze z naszych wykresów przedstawiają porównanie ilości dobrych i złych domen w naszych plikach z danymi. Od razu widać, że jest ich podobna ilość.

7 Slajd 7.

Kolejny wykres przedstawia porównanie długości adresów URL. Widać, że adresy bezpieczne mają są dłuższe niż adresy niebezpieczne.

8 Slajd 8.

W następnym wykresie mamy podany rozkład symboli w hoście adresu URL. Odczytujemy to podobnie jak macierz.

9 Slajd 9.

Następnie poruszyliśmy kwestię porównania ilości znaków interpunkcyjnych w całości adresu. Widać, że w adresach złych występuje ich zazwyczaj więcej.

10 Slajd 10.

11 Slajd 11.

Kolejny wykres przedstawia rozkład znaków w domenach. Odczytujemy to podobnie jak macierz.

12 Slajd 12.

Patrząc na wykres gęstości ilości znaków interpunkcyjnych w linku widzimy, że gęstość ta jest większa dla adresów złych.

13 Slajd 13.

Ostatnie dwa wykresy poruszą temat kodu JavaScript. Na pierwszym widać gęstość tego kodu dla stron dobrych jest wyższa ale sam kod jest krótszy. Natomiast dla stron złych długość kodu jest większa ale gęstość mniejsza.

14 Slajd 14.

Ostatni wykres w naszej analizie przedstawia związek pomiędzy długością kodu JavaScript a bezpieczeństwem domeny. Bierzemy tu pod uwagę długość kodu zaciemnionego. W domenach bezpiecznych długość kodu zaciemnionego jest na poziomie zerowym. Natomiast w złych widać od razu, że tego kodu jest dość dużo.

15 Slajd 15.

Przeprowadzając powyższą analizę wiele się nauczyliśmy. Zauważyliśmy jakie związki leksykalne występują w przypadku domen złych. Wiemy jakie ”zamiany” występują najczęściej. Przyglądając się linkom nie zawsze da się to wszystko wyłapać od razu. Zatem najważniejszym, ale nie jedynym, wnioskiem płynącym z naszego projektu jest fakt, że trzeba uważać w jakie linki się wchodzi.

Dostałeś/łaś podejrzanego maila? Nie otwieraj żadnych linków!