#supervision_hack

CATCH'EM ALL

KONTEKST ROZWIĄZANIA

Phishing jest obecnie najpopularniejszą metodą oszustwa. Przestępcy podszywają się pod osoby lub instytucje, takie jak banki, domy maklerskie czy najpopularniejsze korporacje. Oszuści wysyłają do użytkowników bankowości linki w wiadomościach mailowych lub sms, wykupują reklamy w mediach społecznościowych lub wysoko pozycjonują swoje fałszywe strony w wyszukiwarkach. Nieświadomi, często zastraszani lub mamieni obietnicą zysku internauci otwierają niebezpieczne strony, które swoim wyglądem zbliżone są do tych prawdziwych. Zamieszczone są na nich logotypy instytucji, pod którą się podszywają, utrzymany jest również jej wygląd i charakter. Oszuści na masową skalę generują fałszywe witryny – w 2020 CERT POLSKA zablokował 7459 stron, w 2021 już 33823, zaś tylko do września 2022 pojawiło się kolejne 30 tys. nowych stron. Duża część z nich została zgłoszona przez nasz zespół – tylko w tym roku zgłosiliśmy do blokady już prawie 12 tys. domen!

Chcielibyśmy zautomatyzować i ulepszyć proces oceny stron, tak by zwiększyć prędkość ich weryfikacji, a także dać gotowe narzędzie operatorowi.

OPIS I CHARAKTERYSTYKA ZADANIA

Waszym zadaniem jest stworzenie systemu wspomagającego pracę operatora pierwszej linii i ułatwiającego mu rozpoznanie strony jako bezpiecznej / potencjalnie oszukańczej / oszukańczej. API, na wejściu będzie przyjmowało adres url.

OCZEKIWANE REZULTATY

Waszym wyborem jest, w jaki sposób projekt będzie rozpoznawał i oceniał to czy strona jest prawdziwa/ potencjalnie oszukańcza lub fałszywa. Jednym z pomysłów może być analiza treści strony lub obrazków wykorzystujących logo innych marek. Można w. Być może sami wymyślicie coś jeszcze sprytniejszego, lub zdecydujecie się na połączenie kilku elementów. Rozwiązanie zostawiamy Wam w pełni swobodne.

JAKIE SĄ KORZYŚCI Z TEGO ROZWIĄZANIA?

Główną korzyścią, jaką niesie za sobą niniejszy projekt to zautomatyzowanie procesu rozpoznawania stron phishingowych oraz usprawnienie pracy operatora pierwszej linii, tak by zminimalizować czas potrzebny na prawidłową weryfikację strony. Chcielibyśmy móc blokować niebezpieczną stronę, nim nieświadomy użytkownik na nią wejdzie.

CO OD NAS DOSTANIECIE?

- Przykłady plików html faktycznych stron phishingowych z listy CERT Polska https://hole.cert.pl/domains/domains.txt
- W niedziele o godzinie 8:00 dostaniecie listę domen testowych, wśród których znajdować się będą domeny phishingowe wykorzystujące wizerunki różnych marek. Waszym dodatkowym zadaniem będzie sprawdzenie tych domen i przedstawienie listy domen, które uznacie za phishingowe.
- Lista marek i stron internetowych, pod które podszywają się przestępcy znajdują się na dole dokumentu.

CO MUSICIE DOSTARCZYĆ?

- Kod działającego rozwiązania opublikowany na otwartym githubie zawierający pliki readme.txt, install.txt, requirments.txt.
- Prezentację przedstawiającą wasze rozwiązanie – maksimum 10 slajdów w formacie PDF.
- Dokumentację rozwiązania, która powinna zawierać w szczególności dokładny opis algorytmów i metod detekcji stron phishingowych – **To dla nas bardzo ważne, najciekawsze pomysł będą punktowane extra.**
- Wyniki działania Waszego algorytmu na podstawie listy testowych domen, które od nas otrzymacie. Wyniki mają zawierać informacje, które ze wskazanych serwisów są według Was stronami phishingowymi.

Całość wyników wrzućcie na platformę Challenge Rocket, ułatwi nam to ocenę Waszych rozwiązań.

SPECYFIKACJA TECHNICZNA ZADANIA

Celem zadania jest stworzenie API, które będzie przyjmowało na wejściu adres url. Wynikiem działania będzie odpowiedź w postaci liczby, która procentowo określa czy domena może zostać zakwalifikowana jako Phishing (0% jeżeli jesteście pewni, że strona jest „czysta” i 100% jeżeli jesteście pewni, że strona jest phishingowa. Po przesłaniu na API wystawione przez drużynę oczekujemy informacji o przyjęciu domeny do weryfikacji (kod http 200).

Przykład danych wejściowych:

```
{ „url”: „https://www.example.com/” }
```

Przykład odpowiedzi:

```
{ „error”: „0” }
```

```
Odpowiedź powinna być wysłana jako form_data z następującymi polami:
```

```
{
 „domain”: https://www.example.com,
 „phising_estimate: „35”,
}
```

W projekcie nie powinniśmy korzystać z zewnętrznych API, które mogą być odpłatne.

W przypadku wątpliwości – pytajcie na Discordzie.

CO MOŻECIE UWZGLĘDNIĆ PRZY ANALIZIE

Co możecie uwzględnić przy analizie? Weryfikację adresów serwisów, wykorzystywane loga i wizerunku innych marek, favicony czy też ślady po klonowaniu serwisu popularnymi narzędziami. To tylko kilka z możliwych reguł detekcji. Liczymy na Waszą kreatywność, im ciekawsze metody detekcji wymyślicie tym bardziej punktowaną będzie Wasze rozwiązanie.

MARKI, POD KTÓRE PODSZYWAJĄ SIĘ OSZUŚCI

Oszuści na stronach phishingowych wykorzystują loga i marki innych instytucji. Poniżej lista stron internetowych, które są często kopiowane przez przestępców. Możecie uwzględnić w swoich algorytmach wykorzystywanie wizerunku tych marek na stronach phishingowych:

https://www.pkobp.pl/
https://www.pekao.com.pl/
https://www.aliorbank.pl/
https://www.getinbank.pl/
https://www.pocztowy.pl/
https://www.santander.pl/
https://www.ing.pl/
https://www.mbank.pl/
https://www.bnpparibas.pl/
https://www.bankmillennium.pl/
https://www.credit-agricole.pl/
https://facebook.com/
https://gmail.com
https://www.gaz-system.pl/pl
https://www.lotos.pl/
https://www.orlen.pl/pl
https://www.gkpge.pl/
https://pgnig.pl/
https://www.revolut.com/pl-PL/
https://www.tauron.pl/dla-domu
https://www.tesla.com/pl_pl