Stojanović Marijana PR16/2019

Prodanović Katarina PR15/2019

Cyber Threat Intelligence

Cyber Threat Intelligence (u daljem tekstu CTI) su informacije koje organizacije koriste da bi bile upoznate sa prijetnjama koje su se dogodile, koje su trenutno aktulene ili će se u budućnosti desiti. CTI može pomoći organizacijama da steknu dragocjena saznanja o ovim prijetnjama, izgrade efikasne odbrambene mehanizme i ublaže rizike koji bi mogli da naruše njihovu reputaciju. Važno je razumjeti kako funkcionišu cyber prijetnje kako bi organizacije bile u mogućnosti da odaberu prave alate za zaštitu poslovanja.

Upotrebljene tehnologije

- PySpark
- MatPlotLib

Upotrebljeni programski jezici

- Python

Common Indicators of Compromise

IP adrese, URL-ovi, domain imena, linkovi, email adrese, DLL-ovi, file-ovi, file hash-vi itd... U kontekstu bezbednosti je bilo koja vrsta artefakta koji može ukazivati na narušavanje bezbijednosti i zlonamjernu aktivnost.

Threat Intelligence Lifecycle

Predstavlja proces transformacije sirovih podataka u gotovu inteligenciju za donošenje odluka i delovanja. CTI je izazovan jer se se pretnje stalno razvijaju - što zahteva da organizacije brzo reaguju i preduzmu odlučne korake. Cyber Threat Lifecycle se sastoji od šest koraka u kojima se podstične konstantno poboljšanje u pogledu sigurnosti.

1. **Requirements –** Ko su napadači I koji su njihovi motivi, koja je vrsta napada, koje su akcije koje se trebaju preduzeti da bi se zaštitili od napada.

2. **Collection –** Sakupljanje informacija od značaja, traženje javno dostupnih podataka, traženje stručnjaka u toj oblast i slično.

3. **Processing –** Nakon prikupljenih sirovih podataka, oni će morati da se obrade u format pogodan za analizu.

4. **Analyses –** Detaljna analiza prilikom koje se traže odgovori na pitanja postavljena u fazi Requirements.

5. **Dissemination –** Predstavljanje rezultata zainteresovanim stranama.

6. **Feedback –** Dobijanje povratnih informacija o dostavljenom izveštaju.

AlienVault

AlienVault je platforma namijenjena otkrivanju bezbjednosnih prijetnji. Open Threat Exchange (OTX) omogućava pristup globalnoj zajednici sa više od 100.000 učesnika u 140 zemalja, koji dnevno daju preko 19 miliona indikatora prijetnji. OTX omogućava svima u bezbednosnoj zajednici da aktivno diskutuju, istražuju, proveravaju i dele najnovije podatke o pretnjama, trendove i tehnike. Time pomažu jedni drugima da ojačaju sajber odbranu i podignu svijest o novim prijetnjama na globalnom nivou.
OTX zajednica izveštava i prima podatke o prijetnjama u obliku pulsa. OTX puls pruža rezime pretnje, povezane indikatore kompromisa (IOC), prikaz ciljanog softvera i druge vrijedne detalje koji mogu pomoći u otkrivanju prijetnje u određenom okruženju.
OTX puls se sastoji od jednog ili više IOC-a koji predstavljaju prijetnju ili definišu niz radnji koje se mogu koristiti za izvršavanje napada na mrežne uređaje i računare. Pulsni IOC-ovi uključuju IP adrese, domene, imena hostova, hashe datoteke (MD5, SHA1, SHA256, PEHASH itd), CVE brojeve, URL adrese, e-adrese, putanje datoteka, imena muteksa i još mnogo toga.
Indikatori pulsa mogu se preuzeti u obliku json fajla, CSV fajla itd.
Moguće je pretplatiti se na neki puls kao i kreirati puls na AlieVault platformi.
Skidanje pulseva sa AlienVault platforme omogućeno je povezivanjem preko API ključa koji se dobija prilikom logovanja svakog korisnika. Pulsevi se skidaju u formatu koji programer izabere. U ovom projektu izabran je json format.

PySpark

PySpark predstavlja interfejs za ApacheSpark u Python-u. Ne samo da omogućava pisanje Spark aplikacija korišćenjem Python API-ja, već omogućava PySpark Shell za interaktivno analiziranje podataka u distribuiranom okruženju. PySpark podržava većinu Spark-ovih funkcija kao npr. Spark SQL, DataFrame, Streaming, MLib (Machine Learning) i Spark Core.

Spark SQL i DataFrame

Spark SQL je Spark modul za struktuiranu obradu podataka. Pruža programsku apstrakciju pod nazivom DataFrame i takođe može delovati kao distribuirani SQL mehanizam za upite.
DataFrame predstavlja distribuiranu kolekciju podataka grupisanih u imenovane kolone. DataFrame je ekvivalentan relacionoj tabeli u Spark SQL i može se kreirati pomoću različitih funkcija u SparkSession-u. Kolonama u DataFrame-u pristupamo stavljanjem tačke posle naziva DataFrame-a i pisanjem naziva kolone koju želimo.

```
# Convert the list of pulses to a Spark DataFrame
df = spark.read.format("json").option("inferSchema","true") .load("Data.json")
df.createOrReplaceTempView("pulse")#pravim tabelu koja se zove puls
```

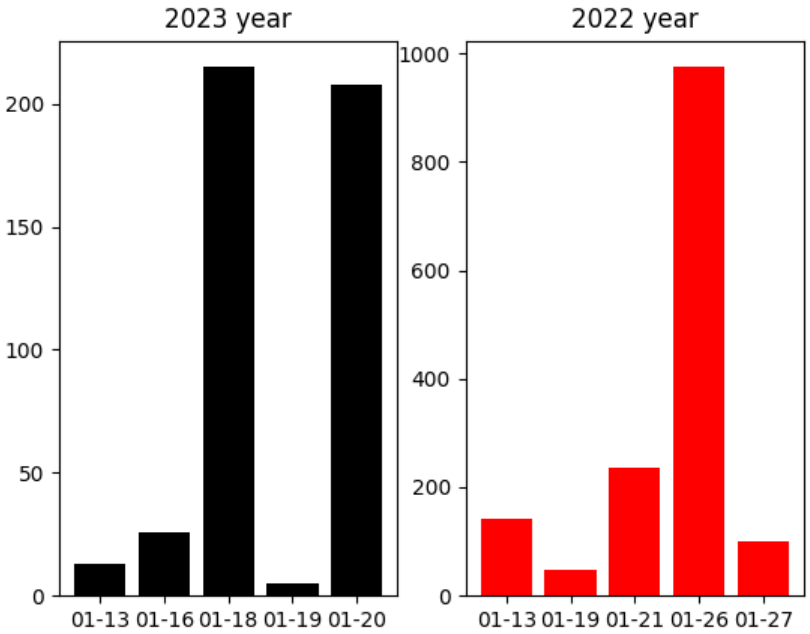
Na slici je prikazan kod gdje se učitava “Data.json” fajl u kom se nalaze svi pulse-vi, odnosno opisi pretnji koje su

skinute sa AlienVault platforme. Prikazano je kreiranje DataFrame-a “df” od učitanog json fajla. Nad tim DataFrame-mom kreiramo tabelu sa svim podacima o sigurnosnim prijetnjama.

```
# Create a SparkSession
spark = SparkSession.builder.appName("AlienVault").getOrCreate()
# Set your AlienVault API key
api_key = "8efd6a55d157ee2cdaee567709846929faaa8e2620f67f7abf6e34c1db260649"
# Set the base URL for the AlienVault API
base_url = "https://otx.alienvault.com"
# Set the endpoint for the AlienVault API
endpoint = "/api/v1/pulses/subscribed"
# Set the headers for the API request
headers = {
    "X-OTX-API-KEY": api_key
}
# Set the parameters for the API request
params = {
    "limit": 300 # Set the number of pulses to retrieve
}
# Send the API request
response = requests.get(base_url + endpoint, headers=headers, params=params)
```

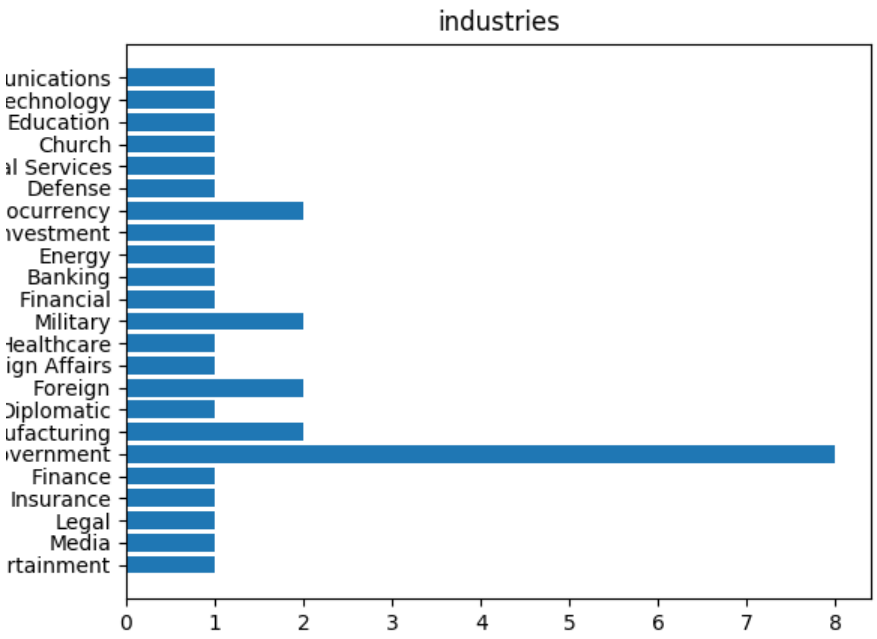
Na prikazanoj slici se može vidjeti kreiranje Spark sesije, povezivanje sa AlienVault platformom i postavljanje limita pulseva koji će biti preuzeti sa AlienVault-a. Limit je na ovoj slici postavljen na 300, ali se preporučuje skidanje većeg broja podataka radi tačnije analize. Zatim se šalje API request i ako je uspiješno povezivanje sa platformom, moguć je pristup skinutim podacima, njihovo učitavanje u DataFrame pomoću PySpark-a i dalja analiza.

U projektu je urađeno nekoliko analiza.



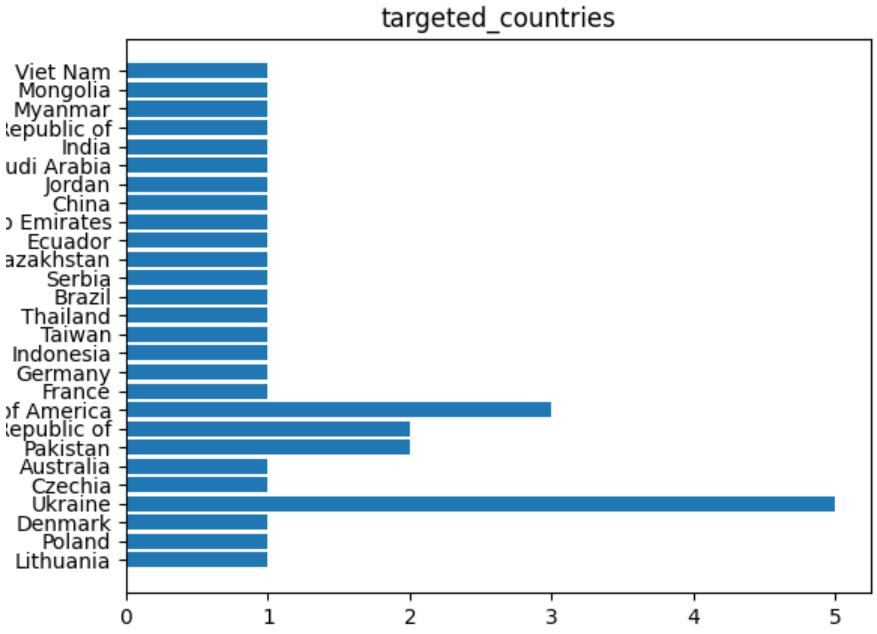
Na slici je prikazana analiza cyber napada u toku 2022. i 2023. Godine. Može se vidjeti učestalost napada po

datumima te godine.

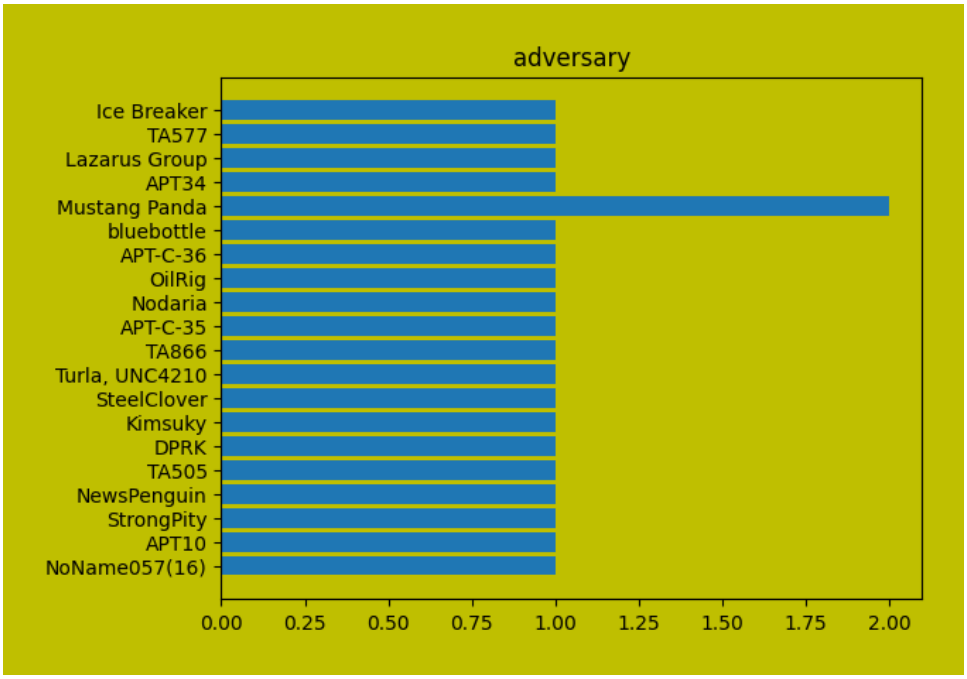


Grafikon prikazuje najčešće pogođene industrije napadima, gdje vidimo da je vlada najpodložnija, a iza nje su vojska, kripto industrije, industrije proizvodnje.

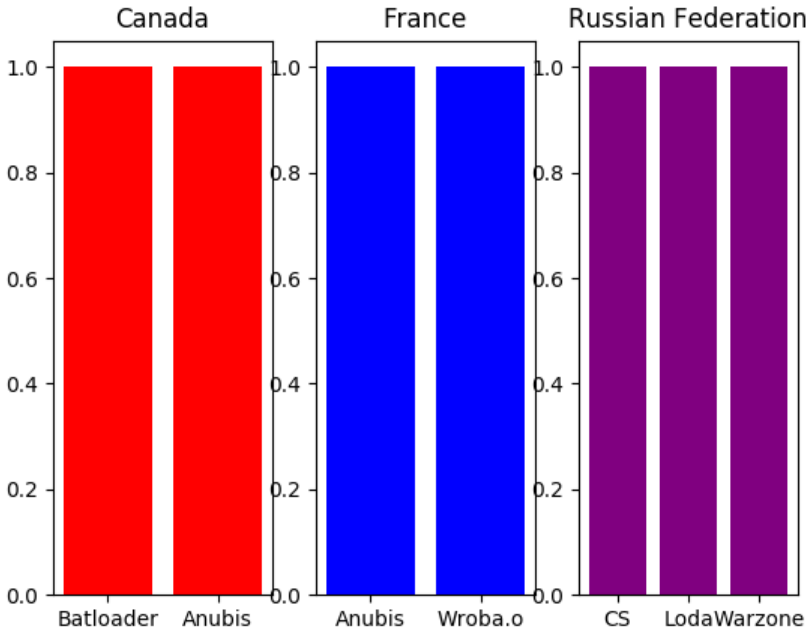
Polja “tags” se odnose na oznaku ili kategorizaciju dodeljenu indikatoru kompromisa. Ove oznake omogućavaju cyber analitičarima da lakše pretražuju i organizuju podatke, da brzo kategorišu i razumiju vrstu prijetnje koju indikator kompromisa predstavlja. Neki najučestaliji tagovi su porodica malvera, zemlja porijekla, uticaj...



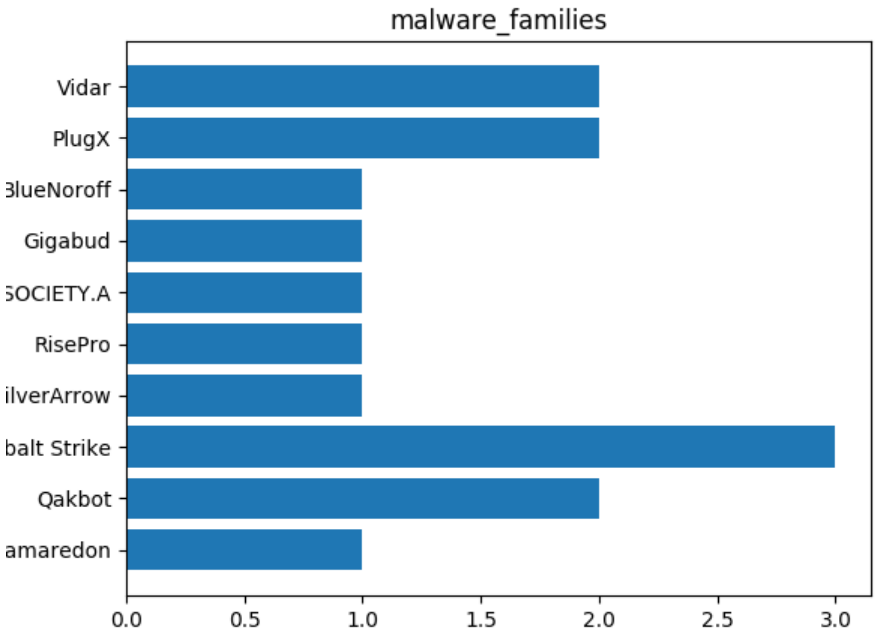
Grafikon prikazuje zemlje koje su najviše bile izložene napadima.



Polje “adversary” u kontekstu CTI-a, odnosi se na osobu, grupu ili organizaciju koja aktivno pokušava da nanese štetu organizaciji ili njenoj imovini putem zlonamerne cyber aktivnosti.



Na ovom grafikonu su prikazane tri zemlje i malware-i koji se najčešće pojavljuju u njima. Prikazuje se i broj pojavljivanja tih malware-a po zemlji. U toku analize vezane za ovaj projekat broj pojavljivanja je uvijek bio 1, ali može biti i više.



Grafikon “malware_families” prikazuje najučestalije familije malvera koje se pojavljuju u prikupljenim podacima.

Zaključak

Na osnovu urađene analize možemo vidjeti značaj CTI podataka za bezbjednost organizacija. CTI može pomoći organizacijama da steknu dragocjena saznanja o ovim prijetnjama, izgrade efikasne odbrambene mehanizme i ublaže rizike koji bi mogli da naruše njihovu reputaciju. Važno je razumjeti kako funkcionišu cyber prijetnje kako bi organizacije bile u mogućnosti da odaberu prave alate za zaštitu poslovanja. Kako bi analiza bila preciznija preporučuje se rad sa većom količinom podataka.