Atvirų duomenų šaltinių analizė
Leak Search
26 komanda

Mantas Matūzas (mantas.matuzas@yahoo.com)
Ernestas Kardzys (ernestaslt@gmail.com)

https://github.com/ernestaskardzys/leak-search

Apie mus

- Mes esame programuotojai, specializuojamės Java, Kotlin, Spring Boot, Amazon Web Services, Google Cloud, Docker, mikroservisų technologijose.
- Mantas Matūzas turi daugiau nei 5 metus programavimo patirties, Ernestas Kardzys daugiau nei 10 metų programavimo patirties.
- Abu turime visiškai minimaliai Python patirties.

- Mūsų dalyvavimo hackatone pagrindinis tikslas - norėtume, kad mūsų hackaton’o idėją Lietuvos kariuomenė panaudos Lietuvos gynybai.

Problema ir Idėja

- Problema:
  - Kaip kiek galima anksčiau sužinoti, kad įmonės privatūs duomenys yra nutekėję (*data leak*)?
  - Jei apie nutekėjusius duomenis skaitome Delfi, vadinasi, jau per vėlu...
- Idėja:
  - Paieškos (Google, Bing etc.) varikliai indeksuoja interneto svetaines. Pasiimti paieškos rezultatus ir su *scraper* pagalba analizuoti svetaines - pasiimti ir analizuojanti svetainės turinį, ieškant raktažodžių.
  - Pirmas lygis:
    - Paieška ir įspėjimas klientui, jei rasta informacija apie įmonę internete
  - Antras lygis:
    - Informacijos paieška apie nutekėjusius duomenis žinomuose hackerių forumuose

Išbaigtumas ir techninis sudėtingumas

- Python programavimo kalba parašyta integracija su Microsoft Bing paieškos varikliu
- Parašytas *web scraper*, kuris aplanko interneto svetaines ir pateikia rezultatus vartotojui įvertinti
- Programa yra apie 300 kodo eilučių, jai sukurti nereikėjo šimtų tūkstančių eurų :)

Citybee leak pavyzdys


```
{
  "search_result": [
    {
      "query": "citybee leak",
      "urls": [
        "https://raidforums.com/Thread-CityBee-LT-Database-Leaked-Download",
        "https://raidforums.com/Thread-SELLING-Citybee-LT-Full-New-Price",
        "https://cybernews.com/security/110000-user-records-from-car-sharing-service-citybee-leaked-and-sold-on-hacker-forum/",
        "https://www.lrt.lt/en/news-in-english/19/1346403/hacker-who-leaked-citybee-user-data-tells-media-cyber-security-was-poor",
        "https://kernal.eu/posts/citybee-leak/",
        "https://www.hackread.com/citybee-database-login-credentials-leaked-online/",
        "https://www.delfi.lt/en/business/hacker-i-posted-everything-i-have-citybee-data-protection-was-poor.d?id=86499949",
        "https://news.ycombinator.com/item?id=26153873",
        "https://cybernews.com/news/the-domain-of-infamous-hacker-forum-raidforums-just-got-suspended/",
        "https://citybee.lt/lt/"
      ]
    }
  ],
  "website": "raidforums.com"
}
```

Informacija apie Robinhood leak’ą


```
{
  "key": "robinhood",
  "values": [
    {
      "score": "2",
      "url": "https://9to5mac.com/2021/11/08/popular-trading-platform-robinhood-reports-security-breach-and-user-data-leak/"
    },
    {
      "score": "10",
      "url": "https://money.com/robinhood-data-breach-whos-affected/"
    },
    {
      "score": "6",
      "url": "https://www.the-sun.com/tech/4034792/robinhood-security-breach-how-to-protect-yourself/"
    },
```

Pritaikomumas arba idėjos ateičiai

- Galima ieškoti ne tik informacijos apie nutekėjusius duomenis, bet, pakeitus raktažodžius, apie bet ką kitą
- Programą pritaikius analizuoti *dark web* duomenis būtų galima surinkti informaciją apie nusikaltimus
- Programą galima pritaikyti ieškoti paieškų *trends* konkrečioje šalyje - pavyzdžiui, prognozuojant nelegalią migraciją į ES

Limitacijos

- Reikalinga integracija su paieškos varikliu - Google, Bing ar panašiai - kuri yra mokama
- *Leak'ų* paieškos sprendimas yra labai paprastas, jį dar reikia tobulinti

Ačiū!

https://github.com/ernestaskardzys/leak-search