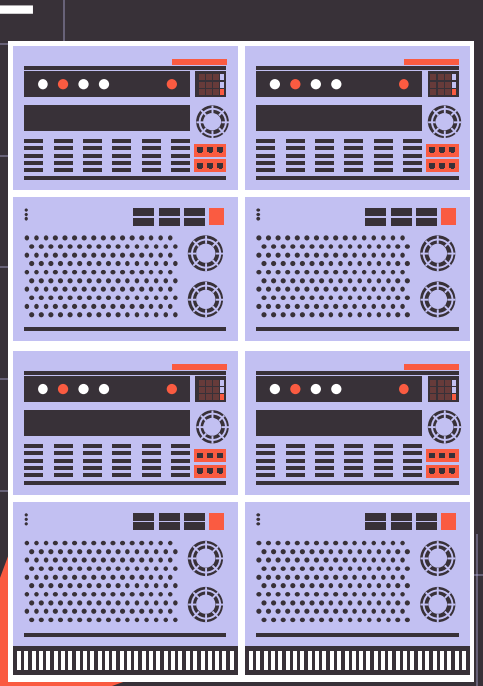
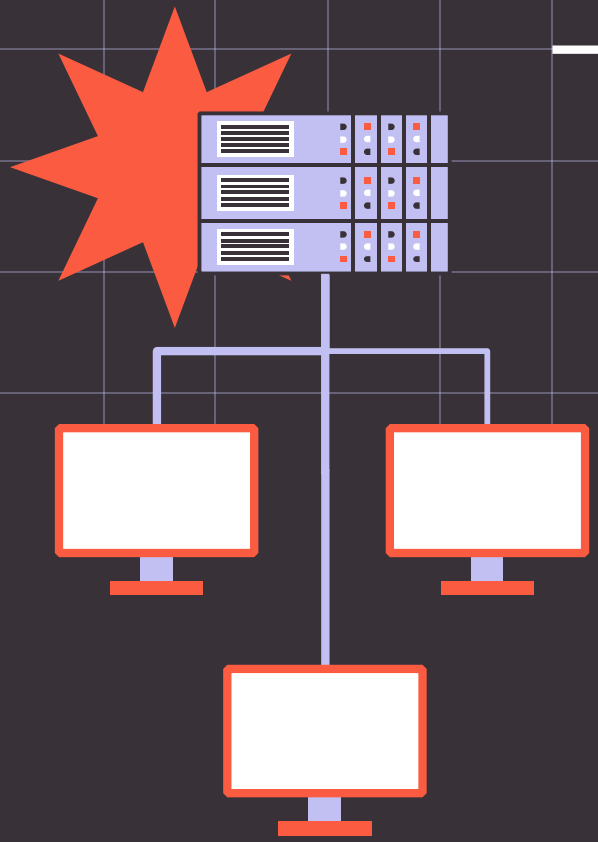
Sveučilište u Zagrebu
Fakultet organizacije i informatike



Web Tracking Forensics: Detecting and Analyzing How Cookies and Scripts Track Users Across the Web

Kolegij: Sigurnost informacijskih sustava

Članovi tima: Blažek Ilan, Babić Lovro, Puklek Dino, Biškup Dorian



Cilj projekta

Dizajnirati i implementirati framework koji omogućuje:

1. Praćenje mrežnog prometa radi otkrivanja praćenja putem kolačića i skripti.
2. Klasifikaciju različitih tipova trackera (kolačići, beaconi, fingerprinting).
3. Pohranu i analizu prikupljenih dokaza u bazi podataka.
4. Vizualizaciju odnosa između web stranica i trackera.

Teorijski pregled web trackinga

Web tracking predstavlja skup tehnika koje web stranice koriste za praćenje korisničkog ponašanja.

Najčešće metode:

- Kolačići (cookies): pohranjuju podatke o korisniku i sesijama.
- Beaconi i pikseli: male slike ili skripte koje šalju informacije poslužiteljima trećih strana.
- Fingerprinting: identifikacija korisnika putem tehničkih karakteristika preglednika i uređaja.

Cilj trackinga: personalizacija, oglašavanje, analitika, ali i potencijalni rizik za privatnost.

Prednosti i rizici web trackinga

Prednosti (za pružatelje usluga)

- Poboljšano korisničko iskustvo (personalizacija sadržaja i preporuka).
- Ciljano i učinkovitije oglašavanje
- Detaljna analitika posjeta i performansi web stranica
- Optimizacija dizajna i funkcionalnosti weba na temelju ponašanja korisnika
- Mogućnost prilagodbe ponuda i usluga specifičnim skupinama korisnika.

Rizici (za korisnike)

- Gubitak privatnosti i anonimnosti na internetu.
- Praćenje bez informiranog pristanka korisnika.
- Potencijalna zlouporaba ili curenje prikupljenih osobnih podataka.
- Nedostatak transparentnosti – korisnici često ne znaju tko ih i kako prati.

Podjela rada

Analiza mrežnog prometa

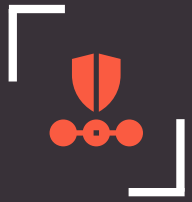
Dino Puklek

Baza i backend

Dorian Biškup

Detekcija i klasifikacija

Ilan Blažek



Vizualizacija i izvještaji

Babić Lovro

Plan rada - praktični dio

1.Postavljanje testnog okruženja:

- Virtualna mreža s više preglednika i testnih web stranica.
- Konfiguracija alata za snimanje prometa (Wireshark).

2.Prikupljanje podataka:

- Automatizirano posjećivanje stranica pomoću Puppeteera.
- Ekstrakcija kolačića i zaglavlja

3. Analiza i detekcija:

- Identifikacija skripti i kolačića koji komuniciraju s trećim domenama.
- Korištenje poznatih lista trackera (Disconnect, EasyList).

4. Klasifikacija i vizualizacija:

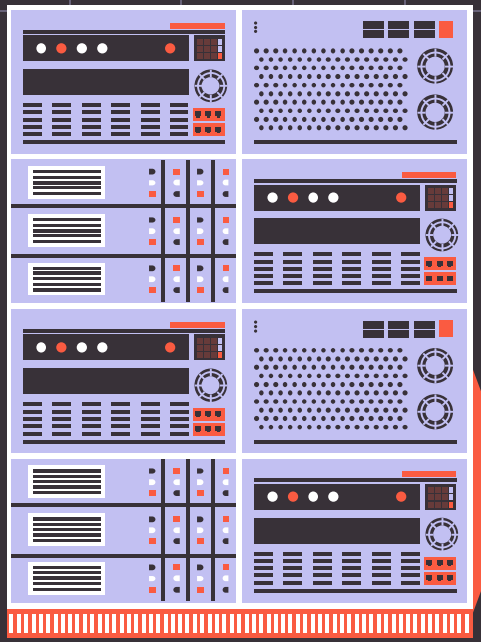
- Kategorizacija prema svrsi (analitika, marketing, funkcionalni).
- Izrada grafa povezanosti u Gephi-u.

5. Evaluacija i dokumentacija:

- Usporedba razine praćenja između različitih stranica.
- Izrada izvještaja.

Očekivani problem

- Velika količina mrežnih podataka i teškoća pri filtriranju relevantnih zahtjeva.
- Promjenjivi obrasci praćenja - tracker domena se često mijenjaju.
- HTTPS enkripcija otežava analizu sadržaja.
- Ograničenja pri automatiziranom pregledavanju (CAPTCHA, dinamični sadržaji).
- Lažno pozitivne/negativne detekcije pri klasifikaciji trackera.
- Potreba za pažljivim etičkim pristupom u analizi stvarnih podataka.



Zaključak

Projekt omogućuje uvid u stvarne mehanizme praćenja korisnika na internetu.
Prikazani alati i metode pokazuju kako se može detektirati i analizirati tracking.

Rezultati mogu pomoći u razvoju alata za zaštitu privatnosti i transparentnost web trackinga.
Dugoročno - doprinos edukaciji i razumijevanju digitalne privatnosti.