Analiza obecności skryptów śledzących ruch sieciowy na stronach internetowych

Sebastian Jamroziński

https://github.com/fredyshox/ONOS-project

Kacper Rączy

Piotr Błoński

Pomysł

Kto śledzi nasz ruch w sieci?!

Celem projektu jest przeanalizwanie sieci trackerów popularnych stron internetowych na które wchodza polacy i znalezienie strategii na zablokowanie danej firmie zbierania o nas danych. Niektóre strony nie pozwalają wyświetlać zawartości, gdy mamy Adblock'a. Chcąć mieć dostęp do treści i dbać o naszą prywatność, chcemy w największym możliwym stopniu blokować kanały zbierania informacji.

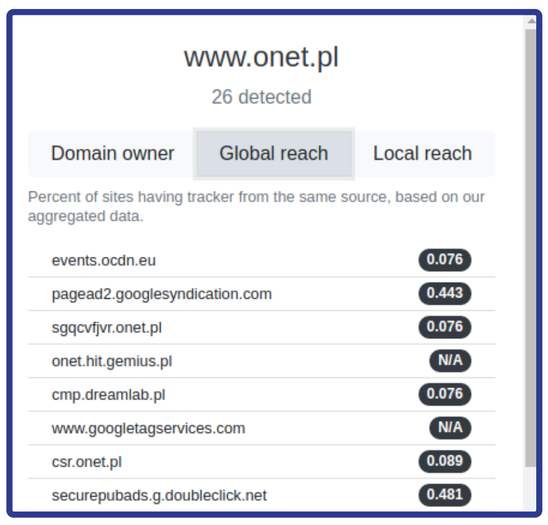
lightbeam

Prorotyp

Wtyczka do przeglądarki

W ramach prototypu/produktu, wykorzystującego wnioski z wykonanych analiz, powstaje wtyczka do przeglądarki, która ma pomóc w doinformowaniu się o różnych tracker'ach.

Nasza wtyczka potrafi:
- *śledzić żądania wychodzące w trakcje sesji przeglądarki*
- *filtrować żądania wykonywane przez potencjalne tracker'y*
- *prowadzić na ich temat statystyki*
- *prezentować informacje w przystępny sposób*



Zbieranie danych

Jak zebraliśmy dane?

Dane zbieraliśmy za pomocą wtyczki **lightbeam** *i* **OpenWPM**, czyli zautomatyzowanego narzędzia do przechwytywania ruchu sieciowego, zoptymalizowanego pod analizę prywatności w internecie.

Strony, które zdecydowaliśmy się odwiedzać, pochodzą z rankingów najpopularniejszych stron w Polsce, czyli indeks Alexa oraz Kazada.

Przetwarzanie danych

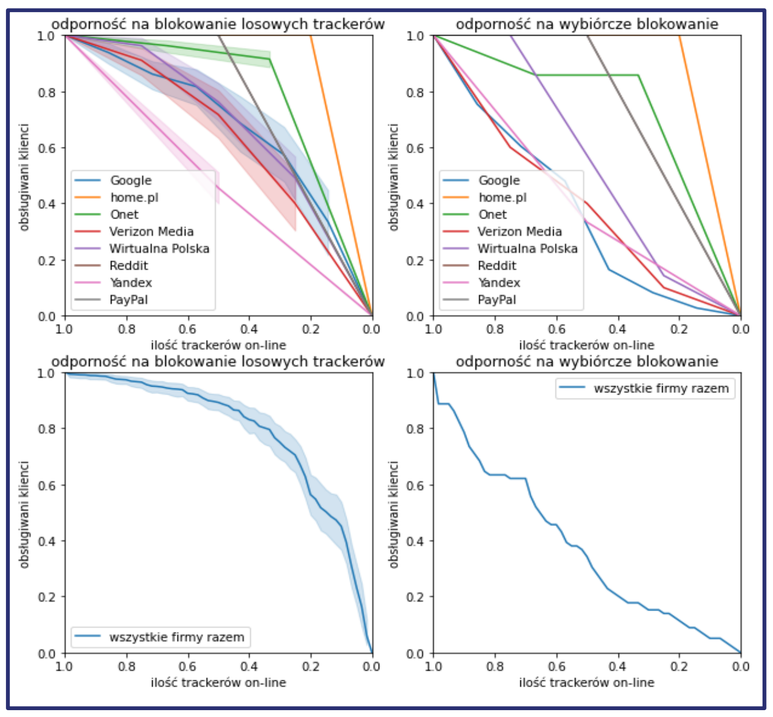
Jak przetwarzaliśmy zebrane dane?

Zebrane przez nas dane wymagały porządnej obróbki. Po pierwsze musieliśmy przefiltrować dane, by zawierały jedynie interesujące nas akty śledzenia. Badania ograniczyliśmy do analizy wychodzących żądań HTTP. Potencjalne tracker'y w żądaniach HTTP, odsiewaliśmy wykorzystując listy reguł easylist, wykorzystywanych przez oprogramowanie typu AdBlock.

Następnie uformowaliśmy model sieci, w której wierzchołkami były unikalne domeny, a o połączeniach między nimi decydowała relacja: ***domena strony źródłowej ~ domena trackera.***

Najważniejszym narzędziem na tym etapie, był Python (+ biblioteki) oraz SQL.

Wnioski



Google zbiera mnóstwo naszych danych

Większość trackerów które znaleźliśmy należały właśnie do tej super korporacji. Używane są nie tylko do zbierania danych ale także udostępniania źródeł różnym serwisom np. reklam. Udało nam się także zrobić analizy ataków na sieci trackerów w zależnościi od firmy!