Privacy-enhancing proxy

Michael Koopman s1401335 en Sven Konings s1534130

# Inleiding

Tijdens de challenge van de eerste week hebben wij een privacybevorderende http-proxy gemaakt. In dit document beschrijven we de veranderingen die wij hebben gemaakt, ten opzichte van de versie die op Blackboard stond. De proxy werkt niet met https, omdat dat end-to-end encryptie gebruikt.

# onRequest

Op het moment, dat de client een webpagina wordt opgevraagd, worden de volgende headers aangepast:

* Accept-Encoding: Om ervoor te zorgen, dat we de tekst die teruggestuurd wordt , ook uit kunnen lezen, geven we aan, dat we alleen deflate accepteren als encoding. Dit werd ons door een studentassistent aanbevolen. Helaas werken sommige sites hierdoor niet. Mocht de webserver geen deflate ondersteunen, dan wordt opnieuw een request gedaan, met normale encoding. Scripts en iFrames en sommige advertenties kunnen door de nieuwe request niet meer geblokkeerd worden op die website.
* User-Agent: Wij sturen nu alleen mee, dat we een Mozilla/5.0 client hebben, met Trident/7.0 als layout-engine. Het eerste is niet privacygevoelig, aangezien vrijwel alle browsers dit sturen. Dat laatste is nodig voor Google, anders kunnen we niets googelen.
* UA-CPU: We geven nu niet langer door welke CPU we gebruiken. Voor de meeste websites is het absoluut niet van belang welke processor je hebt.
* Referer: Is nu leeg, websites kunnen niet langer volgen waar je vandaan komt.
* Cookie: Stuurt nu niets meer mee. We geven hier gemak op voor privacy, doordat sites nu niet meer weten of je ze al eerder hebt bezocht, of dat je al ingelogd was.
* Host: Filtert een aantal advertenties eruit. De gefilterde hosts zijn alles waarin de volgende Strings voorkomen: “google-analytics”, “googleads”, “googletagmanager”, “connect.facebook.net” en “shackle.nl”.

# onResponse

In deze methoden worden alle iframes eruit gefilterd. De methode die hiervoor gebruikt wordt, heeft de naam removeSubString. De nadelen hiervan zijn, dat sommige websites niet meer correct functioneren, maar ze worden vaak misbruikt voor advertenties en social media. Verder worden privacygevoelige javascripts, zoals die op de testpagina, eruit gefilterd. Het nadeel hiervan, is dat niet alle javascript wordt tegengehouden, dus er zou misschien wat doorheen kunnen komen. Aan de andere kant is het compleet wegfilteren van javascript ook niet fijn, aangezien heel veel websites dat gebruikt. Daarnaast worden sommige advertenties en trackingdiensten van Google geblokkeerd.