Report Computer Software - Challenge 1 by Eline Brader (s2674483), Lars Ran (s1403192)

Exercises

A simple local site

The website shows information about the server, the IP address of the client and its used browser and device. Hiding the server name would not affect much of the client's own privacy. The IP address should not be excluded from the messages as otherwise the server would not know where to send the data to. The used browser and device are interesting to hide. This can be established by editing the header with the name “User-Agent”. The value could be changed to any random String, such as “ “, as long as this has nothing to do with the client’s identity. The return value should be changed to “return {requestHeaders:req.requestHeaders}” to assure the changes are done.

Whats my ip?

When fetching the page whatsmyip.org the browser starts by getting the mainframe of the page. After this a lot of extra GET statements are sent outward. Most of these GET statements pertain to the site. For example, images and text, as well as menu headers are loaded. However, there are also some GET statements to third parties, most prominent of these are for example twitter and facebook, this is not that strange as these symbols are scattered over the site already. Some others are “redIntelligence” and “google ads” and such. These are of course for showing ads. However, some of them seem for tracking your info too. Furthermore there are some bigger URLs that might be ways to send data without using the POST statement.

BBC

A lot of the sites loaded when loading this page is from an external site of the bbc (bbci). These contain not only pictures and text but also site layout and fonts. These are all marked as a third party site. Furthermore there is again a lot of trackers, analytics and adds hiding in plain sight. There are like 5 different (maybe more) third party analytics sites that collect information. From the GET statements we could gather at least 10 companies that were informed by our visit. However, this might be way more in the background. After every so many seconds a GET statement would be sent to ping.heartbeat to check if we are still online. When we close the tab this information is launched into the online world as well.
The way we tried to circumvent trackers is by looking at the `req.urlClassification.thirdParty` array. Most of the time when a url was “suspicious” this array was non-empty as well. Therefore we cancelled any request containing “tracking_ads”, “tracking_analytics”, “any_strict_tracking” or “any_basic_tracking” in this array. We tried some different approaches, such as checking the third party field in the http request. However, since the layout of bbc came from a (self-owned) third party site this worsened the browsing experience.

How far did we go?

Well, we actually fine-tuned the above approach. We checked the effects of removal of all the different thirdParty terms on the Canvas website: canvas.utwente.nl. Exclusion of all terms made the website unusable. When only messages containing terms with “tracking_ad” and/or “tracking_analytics” were excluded, the Canvas site would work without problems. It was found acceptable to allow messages to be sent if these would contain all other mentioned terms. Tracking by ads was considered to be least desirable when it comes to one's privacy.