1) Using the crawler I visited Tranco 100 websites in the “vanilla mode” and “ad blocking mode” and the associated network data is recorded in the vanilla_data and adblock_data folder.

2) **Top 10 Third-Party Domains for Request URLs**

douyin: 360
techradar: 337
huffpost: 335
asahi: 327
kotaku: 321
deadspin: 312
bonappetit: 273
slate: 266
msn: 255
foreignpolicy: 254

Comparisons

Vanilla mode has a total of 15,188 requests while adblock mode has 15,128. The difference shows that ad blocking does have some effort but it doesn’t seem large enough based on the sample size. This may not be because adblockers aren’t as effective but as a result of the number of ad blocking scripts being used interfering with the count. Also Douyin seems to be the domain with the highest request urls in both modes and of the domains that appear in both top ten lists like Douyin and techradar they have less requests using an adblocker

Fuctionalities

All the domains except douyin (a popular Chinese short video streaming and sharing platform) seems to be a news site for different domains like music, entertainment, global affairs, food e.t.c In addition, asahi and Kotaku seem to be Japanese new sites

3) **Top 10 Third-Party Domains for Cookies in Vanilla**

techradar: 129
huffpost: 89
bonappetit: 86
slate: 84
newyorker: 82
pitchfork: 80
cbslocal: 72
foreignpolicy: 72
fortune: 72
businessinsider: 69

Functionalities

The top ten websites, except douyin (a chinese short video streaming platform) seem to all be websites that provide news on different domains like entertainment, global affairs, culture, technology, finance e.t.c

Comparisons

There is a significant difference in total cookies in both modes with vanilla having 2934 cookies and the adblocker mode only having 185. This shows the effectiveness of ad blockers in blocking cookies. Also of the domains that appear in both top ten lists like there have significantly less cookies when crawled with the ad block mode. For example foreign policy goes from 72 to just 15 cookies. In addition HuffPost still has a lot of cookies irrespective of mode being top 2 on both list indicating it's pervasive cookie implementation

Manual Analysis

A manual analysis shows that a lot of the cookie domains are advertising and tracking platforms E.g(.taboola.com, .doubleclick.net, .3lift.com, .demdex.net, .amazon-adsystem.com, .cxense.com, and .everesttech.net e.t.c)Also the cookies contain information on the name of the cookie, the domain it belongs to, the path, its expiration date and whether or not it is session-based or persistent,

4) **Top 10 Third-Party Domains for Cookies in Vanilla**

rollingstone: 68
bonappetit: 67
variety: 59
newyorker: 57
thenation: 56
pitchfork: 55
arstechnica: 54
douyin: 54
nbcnews: 54
nypost: 54

Comparisons

In vanilla mode, Rolling Stone, Bon Appétit, and Variety are among the top domains with the highest number of JavaScript API calls. This shows that these domains rely heavily on JavaScript API Calls. Interestingly, some domains like NBC News and New York Post show up on both lists with only half of the API calls blocked showing that their Javascript calls may mainly be used for interactivity rather than tracking users. Lastly in AdBlock mode we have less than half of the API calls we see in vanilla mode showing the effectiveness of Ad blockers

Functionality

All the domains except douyin (a popular Chinese short video streaming and sharing platform) seems to be a news site for different domains like music, entertainment, global affairs, food e.t.c

5) **Top 10 Third-Party Domains for Custom Headers**

douyin: 4895
msn: 3846
techradar: 2421
kotaku: 1773
bonappetit: 1551

dailymail: 1521
gizmodo: 1474
deadspin: 1454
bloomberg: 1382
qq: 1290

Comparisons

Douyin has the highest number of Custom headers in both Vanilla and Ad block mode. Even more surprising, it has more custom headers in Ad block mode than in vanilla mode. The total number of custom headers in both modes are alarming with a Total for the Vanilla mode being 68610 and a total for the ads mode being 48110. In addition domains like qq, bloomberg, techradar, kotaku, deadspin, and daily mail appear in both Top ten lists

Description

HTTP headers in general are additional information sent between a client and a server during web communication. It's difficult to list all the response headers for each ot the top side because according to my data they each have over a thousand but as a use case the bloomberg domain which is last in the list had the following headers

Response Header: server, Count: 224
Response Header: date, Count: 254
Response Header: content-type, Count: 244
Response Header: content-length, Count: 227
Response Header: connection, Count: 5
Response Header: location, Count: 12
Response Header: strict-transport-security, Count: 178
Response Header: set-cookie, Count: 27
Response Header: x-powered-by, Count: 14
Response Header: referrer-policy, Count: 6
Response Header: cache-control, Count: 229
Response Header: content-security-policy, Count: 32
Response Header: x-content-type-options, Count: 54
Response Header: permissions-policy, Count: 22
Response Header: fastly-restarts, Count: 9
Response Header: content-encoding, Count: 136
Response Header: accept-ranges, Count: 171
Response Header: age, Count: 170
Response Header: x-served-by, Count: 156
Response Header: x-cache, Count: 183
Response Header: x-cache-hits, Count: 156
Response Header: link, Count: 10
Response Header: vary, Count: 132
Response Header: alt-svc, Count: 191
Response Header: p3p, Count: 23
Response Header: timing-allow-origin, Count: 26
Response Header: cross-origin-resource-policy, Count: 34
Response Header: expires, Count: 44
Response Header: etag, Count: 172

Response Header: access-control-allow-origin, Count: 180
Response Header: content-disposition, Count: 8
Response Header: x-xss-protection, Count: 36
Response Header: x-amz-id-2, Count: 94
Response Header: x-amz-request-id, Count: 95
Response Header: access-control-allow-methods, Count: 86
Response Header: x-amz-replication-status, Count: 78
Response Header: last-modified, Count: 114
Response Header: x-amz-server-side-encryption, Count: 89
Response Header: x-amz-version-id, Count: 92
Response Header: x-timer, Count: 146
Response Header: via, Count: 25
Response Header: x-amz-cf-pop, Count: 19
Response Header: x-amz-cf-id, Count: 19
Response Header: edge-control, Count: 53
Response Header: x-wss-client-request-id, Count: 53
Response Header: x-wss-server, Count: 53
Response Header: x-cache-status, Count: 13
Response Header: content-range, Count: 2
Response Header: access-control-allow-credentials, Count: 32
Response Header: access-control-allow-headers, Count: 18
Response Header: x-request-id, Count: 1
Response Header: x-runtime, Count: 1
Response Header: x-frame-options, Count: 12
Response Header: pragma, Count: 27
Response Header: allow, Count: 2
Response Header: x-envoy-upstream-service-time, Count: 2
Response Header: request-context, Count: 4
Response Header: x-azure-ref, Count: 2
Response Header: cf-cache-status, Count: 4
Response Header: cf-ray, Count: 4
Response Header: x-px-hash, Count: 1
Response Header: access-control-expose-headers, Count: 1
Response Header: active-cdn, Count: 1
Response Header: reporting-endpoints, Count: 2
Response Header: report-to, Count: 12
Response Header: document-policy, Count: 2
Response Header: permissions-policy-report-only, Count: 2
Response Header: cross-origin-embedder-policy-report-only, Count: 2
Response Header: cross-origin-opener-policy, Count: 4
Response Header: x-fb-debug, Count: 2
Response Header: accept-ch, Count: 9
Response Header: x-msedge-ref, Count: 6
Response Header: x-cdn, Count: 2
Response Header: nel, Count: 5
Response Header: x-datastream-cache-status, Count: 1
Response Header: cross-origin-opener-policy-report-only, Count: 3
Response Header: observe-browsing-topics, Count: 1

Response Header: google-lineitem-id, Count: 1
Response Header: google-creative-id, Count: 1
Response Header: google-mediationgroup-id, Count: 1
Response Header: google-mediationtag-id, Count: 1
Response Header: content-language, Count: 1
Response Header: x-rgw-object-type, Count: 1
Response Header: retry-after, Count: 1
Response Header: x-ms-request-id, Count: 1
Response Header: x-ms-version, Count: 1
Response Header: x-azure-ref-originshield, Count: 1
Response Header: x-fb-optimizer, Count: 1
Response Header: priority, Count: 1
Response Header: follow-only-when-prerender-shown, Count: 1
Response Header: critical-ch, Count: 1
Response Header: content-security-policy-report-only, Count: 2
Response Header: cf-bgj, Count: 2
Response Header: cf-polished, Count: 2
Response Header: x-server, Count: 1
Response Header: linkedin-action, Count: 3
Response Header: x-li-fabric, Count: 3
Response Header: x-li-pop, Count: 3
Response Header: x-li-proto, Count: 3
Response Header: x-li-uuid, Count: 3

The headers like date, server and content type are standard. Date tells the date when the response was generated by the server. Server specifies the software and version that is running on the server. And content-type indicates the type of the content sent in the response (HTML vs PNG). On the other hand the page also has over a 1000 custom headers like fastly-restarts, x-served-by, x-cache, Edge-control e.tc Fastly-Restarts provides information about the number of times the request has been restarted internally. x-served-by tells the identifier of the server that served the request. x-cache tells whether the response was served from cache or not. One thing of note is that a lot of the other custom headers in all the top ten domains start with a “x-” prefix . With a clear exception of the ones in the standard_headers.txt file

Functionality

All the domains seem to have different functionalities. Douyin is a popular Chinese short video streaming and sharing platform. MSN is a windows web portal with apps and service that provides users access to things like news, email, games. e.t.c . TechRadar is a technology news website. Kotaku is a video game blog. Bon Appétit is a food website. The Daily Mail is a British news website. Gizmodo is a Technology blog. qq is a Chinese messaging web portal. Deadspin was a sports news website. Lastly Bloomberg is a financial news website