Study the effects of ads in loading of Web pages

Team: Yash Shah (111485295), Vishwatej Reddy Anugu (111446995)
Supervised by - Aruna Balasubramanian

1. Introduction

Web page loading and rendering is a very complex process. It involves downloading, parsing and rendering of various components in a particular fashion. There have been various studies to improve metrics related to Page Load Time (PLT) like document complete time and fully rendered time. Advertising networks are an important part of publishers' business plans. Modern rich websites contain many ads in different formats which overall impacts on page load time on both desktop and mobile browsers. A single redirect causes browser to process an additional HTTP request before rendering the page. The page load time has a direct correlation with customer satisfaction. Google claims that 74% of people will abandon a mobile website that takes more than 5 seconds to load. The survey results of top news websites shows that time required to load advertising content was upto 4 times more than editorial content. The above factors push people for the adaption of ad blocking tools globally in order to control their online browsing experience.

2. Problem Statement

Since the web page load time is very critical to customer satisfaction and as advertising content on web page is becoming heavier than ever, we need to study effects of advertisements in loading of web pages. We need to analyze and perform empirical evaluation comparing different metrics to understand effects.

3. Approach

3.1 WebPagetest: WebPagetest is used for measuring and analyzing the performance of web pages. Developer Interfaces (Restful APIs and Batch Libraries) and flexible environment options are available for retrieving useful performance metrics.

We gathered different performance metrics to perform thorough evaluation to compare with / without adblock in different operating environment. We have classified performance metrics in two subcategories. We compared using the first view obtained from Chrome and Firefox browser for Desktop and only Chrome for mobile.

3.1.1 Page-level metrics: We acquired page-level metrics to compare performance impact of having ads. This metrics include first byte, start render, load time, visually complete and fully loaded.

3.1.2 Request-level metrics: We collected different comparison oriented metrics for evaluating impact of ads on mobile vs desktop environment as well as ads vs no-ads in desktop environment. These metrics include comparison of document complete time and fully loaded time in terms of time, number of requests and number of bytes. The comparison is done including and excluding our own custom adblock. Further, we have gathered breakdown of each resources (objects like html, css, js, flash, font, video etc.) by its MIME type. By gathering statistics like number of requests and number of bytes for each of the resource type in present / absence of adblock allows us to analyze interesting measures.

3.2 EasyList: The EasyList provide filter lists containing sets of rules for Adblock. Processed General advert blocking filters for individual request urls substring match. Webpagetest have the functionality to pass the list of urls (substrings) as formdata to block the specific subdomains or redirects. Here, we faced challenge since easylist doesn't straightforwardly provides list of urls. Instead, they provide set of rules to filter out specific objectives. It includes blocking of address by parts, blocking of domain name, blocking exact address. Further, it provides additional element hiding options using optional loading of required scripts, css or images.

To fully achieve the potential of webpagetest, in order to compare performance impact of ads, we Integrated processed Easylist on top of WebPagetest wrapper to block urls.

3.3 WebPagetest Metrics Definitions:

Page-level Metrics:

Load Time: Time from the start of the initial navigation until the beginning of the window load event (onload).

First Byte: Time from the start of the initial navigation until the first byte of the base page is received by the browser (after following redirects).

Start Render: Time from the start of the initial navigation until the first non-white content is painted to the browser display.

Fully Loaded: Time from the start of the initial navigation until there was 2 seconds of no network activity after Document Complete (including any activity that is triggered by javascript after the main page loads.

Visually Complete

First View: The First View row is a test that was done with a browser that had its cache and cookies cleared out and represents what a first-time visitor to the page will experience.

Requests: This is the number of requests that had to be made by the browser for pieces of content on the page (images, javascript, css, etc).

Bytes In: This is the amount of data that the browser had to download in order to load the page. It is also commonly referred to as the "Page Size".

4. Evaluation Set up (what metrics are you evaluating, what parameters are you varying, what is your hypothesis)

4.1 Evaluation Set up:

We are trying to evaluate the page load time taken by a website in different browsers and different devices with and without adblock. For evaluating our results we have used the chrome web browser and the mozilla web browser in the windows 10 environment and Lenovo yoga 720 laptop. For testing in the mobile version we have used Moto G generation 4 as our device. For computing the different metrics of the web page we have used the API's of webpagetest in order to get the fair estimations which otherwise if used our web browser would give incorrect results because of the cache stored. For implementing the feature similar to the popular ad blockers we used filters to filter out all the ads. For filters we used the easylist which provides us with the data for the ad web pages which includes unwanted images, frames and objects. We used the data and parsed it to form a list of required filters and sent it as a form element in the API request made to the webpagetest API.

4.2 We are evaluating the following metrics:
a) Time taken for the document to load.
b) Time taken to receive the first byte.
c) Time taken to start rendering the page.
d) Time taken for web page to become visually complete.
e) Time taken for the web page to load completely.
f) Number of requests for the document to load.
g) Number of requests for the page to load completely.
h) Number of bytes downloaded to load the document.
i) Number of bytes downloaded to load the webpage.
j) Requests made by different resources like html, css, js etc
k) Number of bytes download by different resources like html, css, js etc.

4.3 We are varying the following parameters:
a) Chrome Web browser in desktop
b) Mozilla Web browser in desktop
c) Mobile web browser
d) Including filters for ads
e) Excluding filters for ads

Our hypothesis is that the ads have a significant impact on the web page load times and if we remove all the ads then the web page loads significantly faster than the web page with ads. Also the page load time is affected when viewing the same web page in different browsers and different devices over different networks.

5. Results:

First we present the comparison between the page load times of a web page with and without ad blocker in the chrome desktop web browser for **www.flipkart.com**.
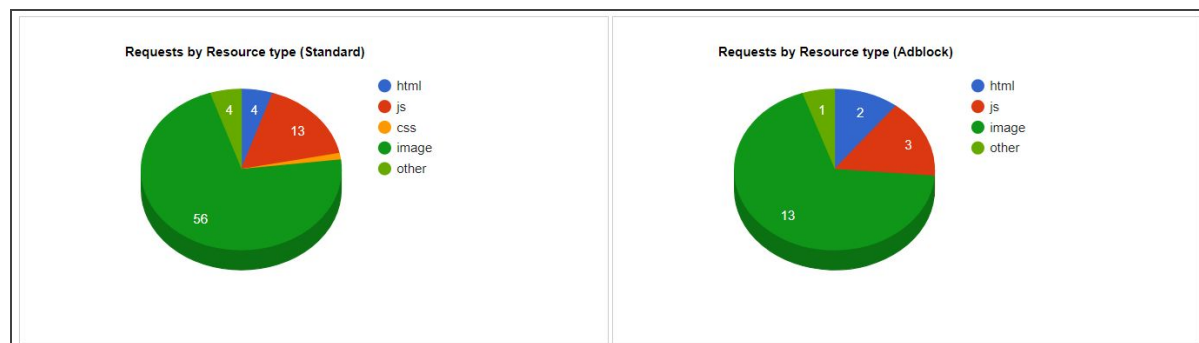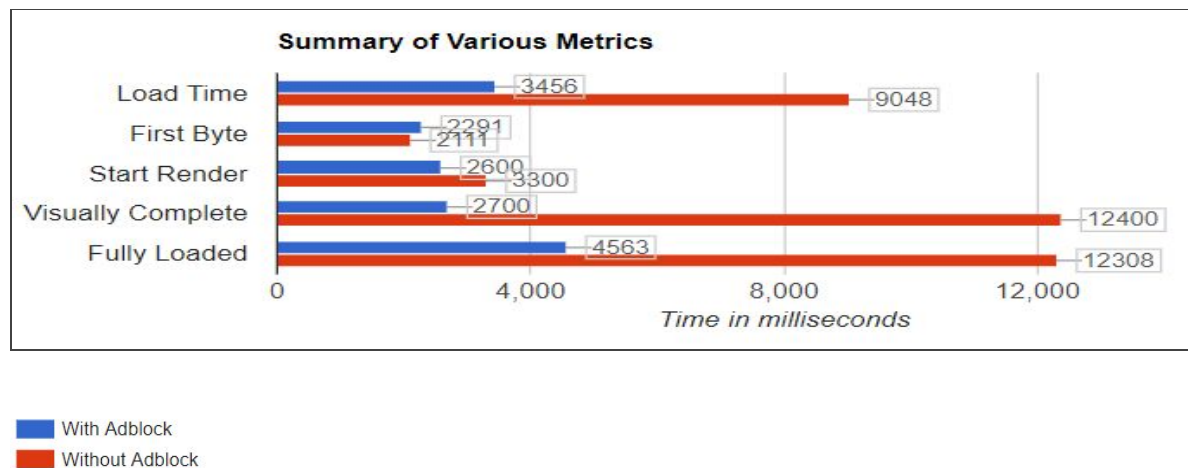
From the above results we observe that the time taken to load the document as well as the time taken to load the entire page is almost 3 times less when used the adblocker than without. There isn't any significant difference in the times at which the page starts rendering. We observe a vast decrease in time when the page becomes visually complete, it takes 6 times less time with adblock. We also observe a significant decrease in the number of request made when we use ad blocker than without.

Next we present for the **www.aol.com**

From the below statistics we observe that the time taken to load the document

only is 4 times less when used with adblock. The time taken to load the document is more compared to flipkart because there are more components in the web page. Similarly is the case with the rest of the metrics as well. The number of requests for this website is more compared to the previous website.
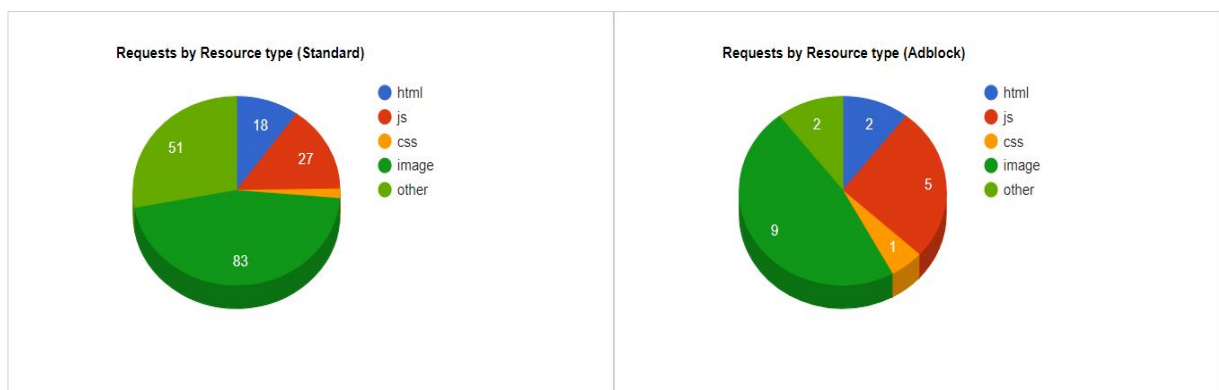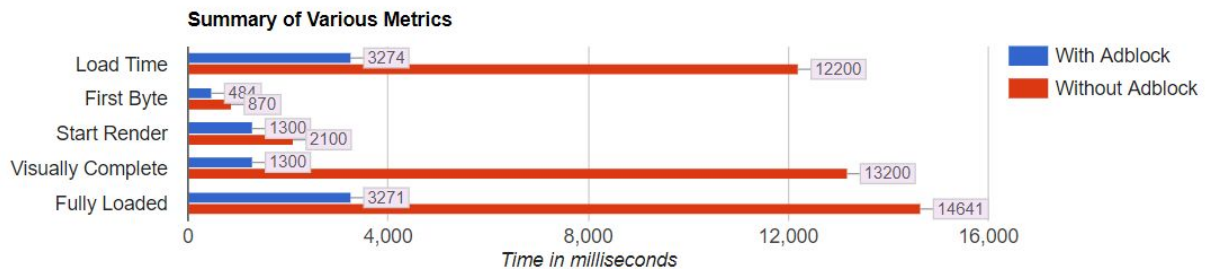
Next we present **www.microsoft.com**

Requests by Resource type (Standard)
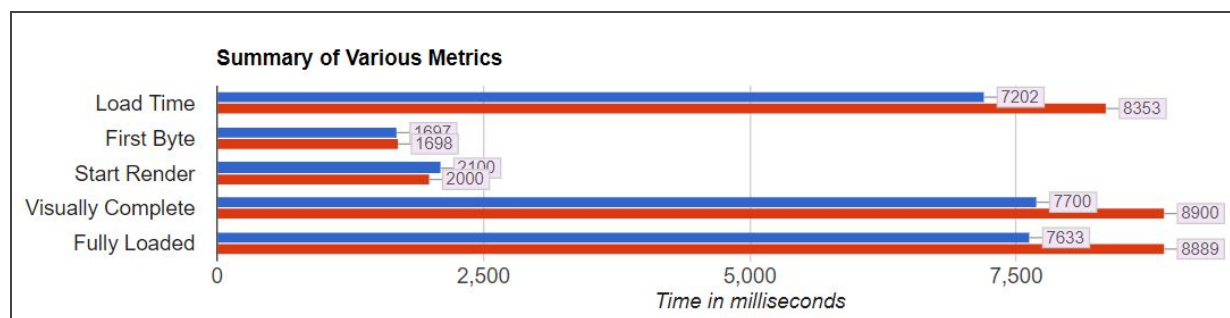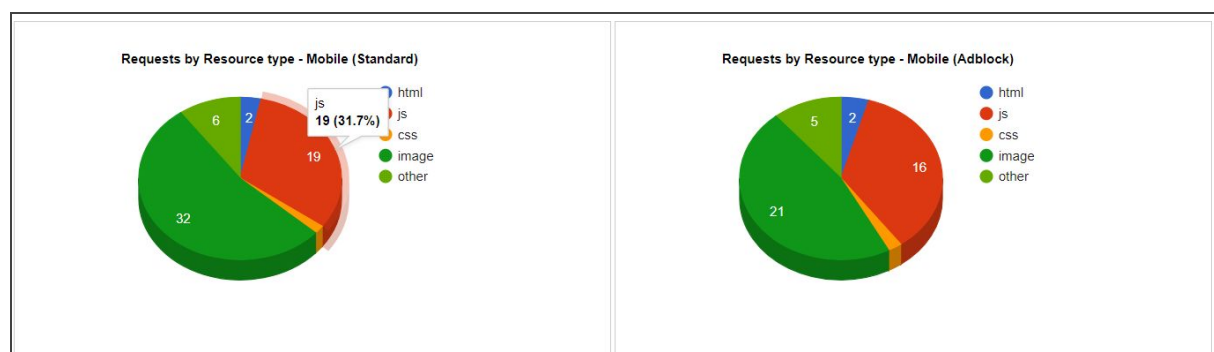
Requests by Resource type (Adblock)
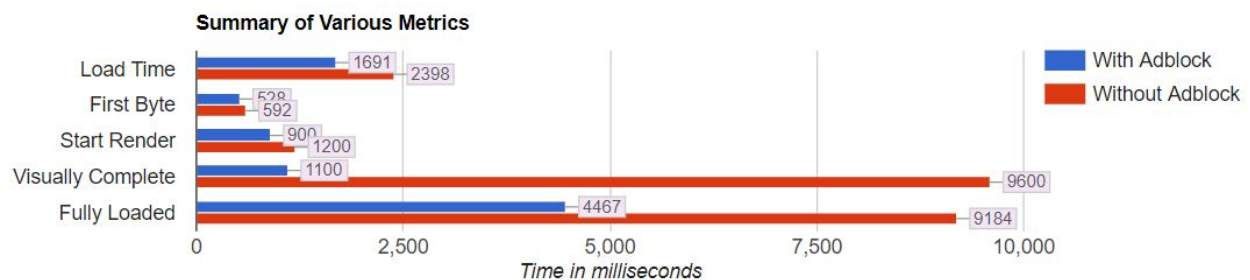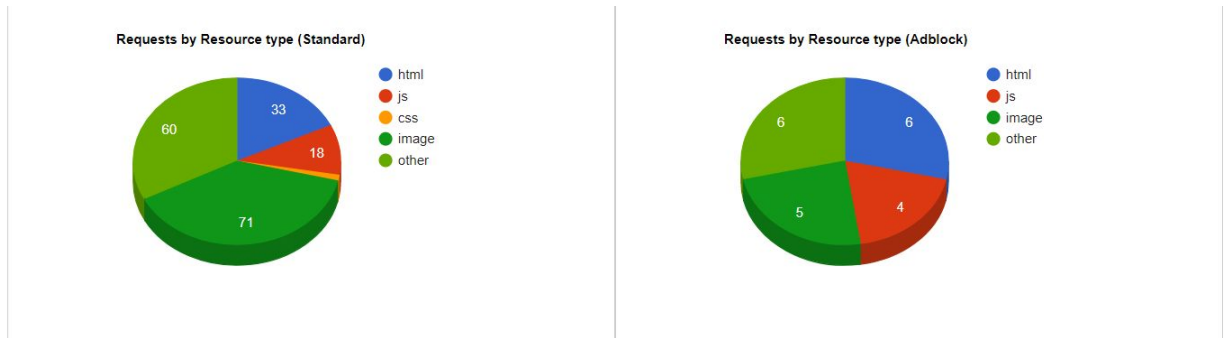
In this website we observe that there are ain't may ads as we see that there is no significant different in the page load times when used adblocker and when not used adblocker. This is true for the time taken to only load the document. But the time taken to load the complete page is 2 times less with adblocker compared to without it. We also observe a significant decrease in the number of requests indicating that there is a lot of javascript.

Then we make a comparison between the page load times of different websites in mobile browser with and without ad blocker.

Here we make a comparison for the **www.flipkart.com.**

Summary of Various Metrics

Metric	Mobile (Adblock)	Mobile (Standard)
Load Time	7202	8353
First Byte	1697	1698
Start Render	2100	2000
Visually Complete	7700	8900
Fully Loaded	7633	8889

Time in milliseconds

Mobile (Adblock)
Mobile (Standard)

Requests by Resource type - Mobile (Standard)

js
19 (31.7%)

Requests by Resource type - Mobile (Adblock)

We observe that there is no significant difference in the page load times in the webpages when used with and without ad blockers in the mobile browser with both just loading the document and also loading the entire web page, indicating that there are not more ads incorporated for the mobile version of the browsers. We also observe that there is no enough decrease in the number of requests as well.

Next we compare for **www.nba.com**

We observe that the mobile version of website also has a lot of ads and there is a significant decrease in the page load times with ad blocker when compared to without ad blocker. We can also observe that there is a significant decrease in the number of javascript requests and also image requests indicating that most of the ads were images and javascript requests.
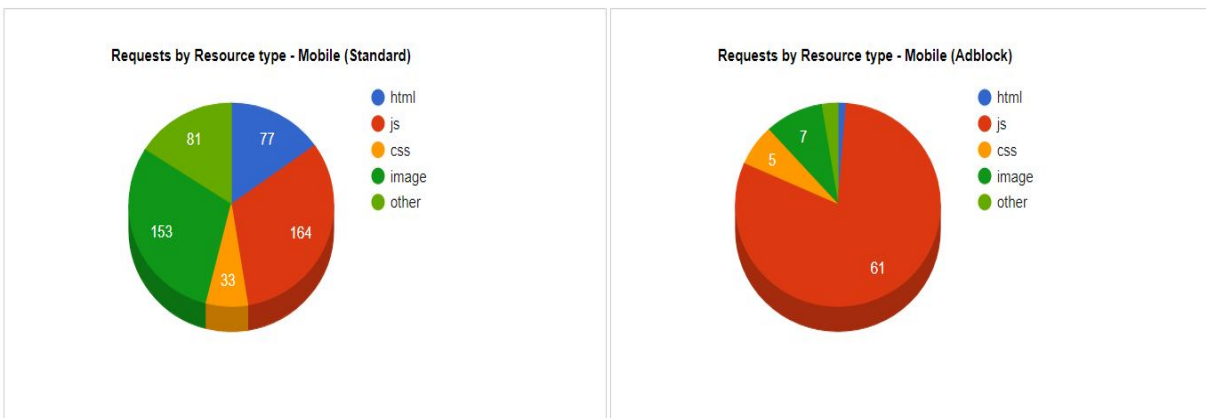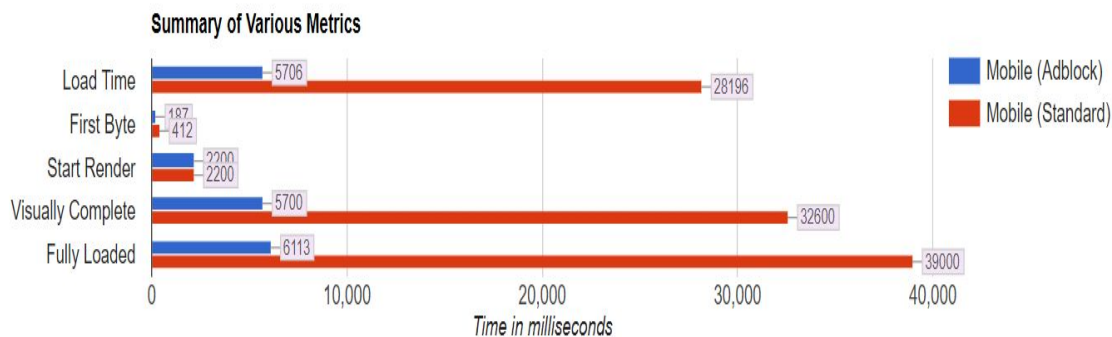

Now we present the comparison between the page load times of a webpage with adblock in different browsers (here we have taken chrome and firefox)

Summary of Various Metrics
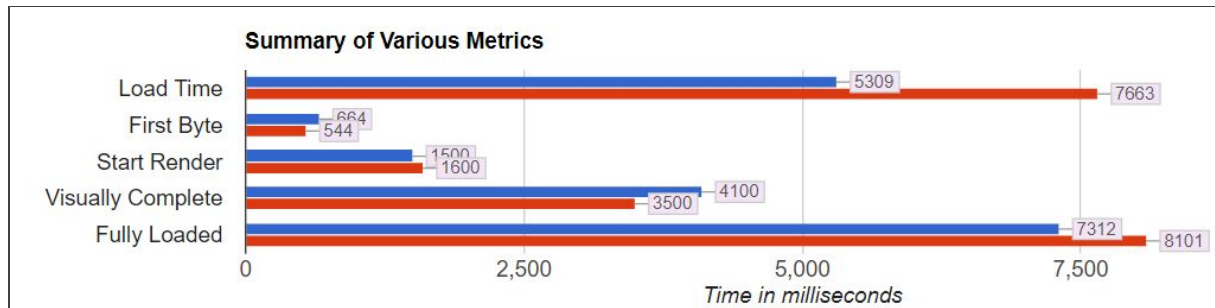
Chrome (with Adblock)
Firefox (with Adblock)

This data is for the website **www.news.google.com** . We observe that the document only page load time is 1.5 times less for the Chrome browser compared to the firefox and there isn't much difference for the time taken to load the webpage completely. We also observe that the time taken for the web page to become visually complete and the time taken to render the first byte is a little more for the chrome browser compared to the firefox. There isn't any significant difference observed in the number of request by resource type.
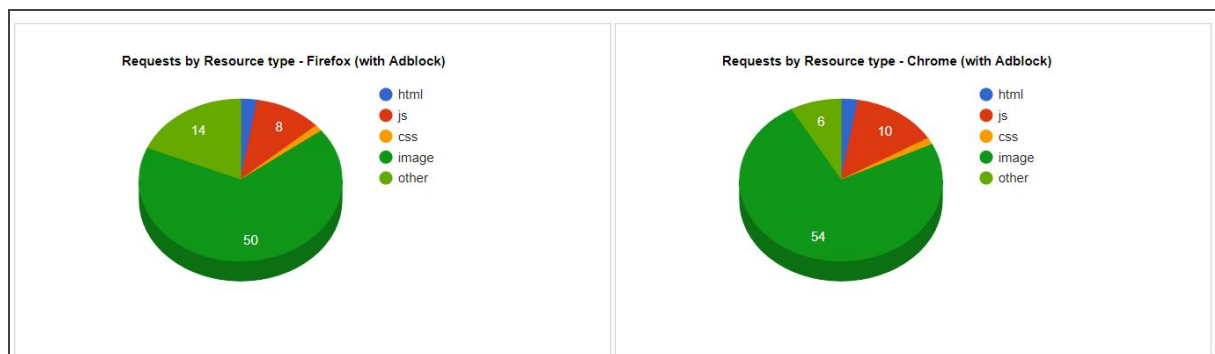
Next we present the data for **www.aol.com**


Summary of Various Metrics
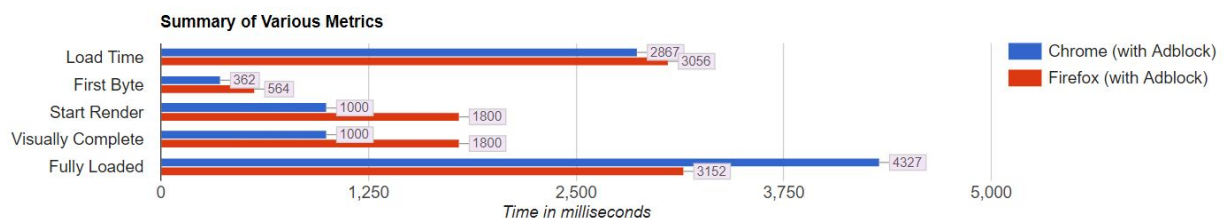
We observe similar inferences as the previous example.

We then present the overall cumulative distribution function of the page load time for different websites

We observe that for the median case, the websites take approx 10 sec for the webpage to load completely without adblock and approx. 3 sec to load with adblock and for the 99th percentile case it takes approx 14 sec to load without adblock and approx 8 sec to load with adblock. It takes almost 2.5 times more time to load the page completely without adblock compared to with adblock.
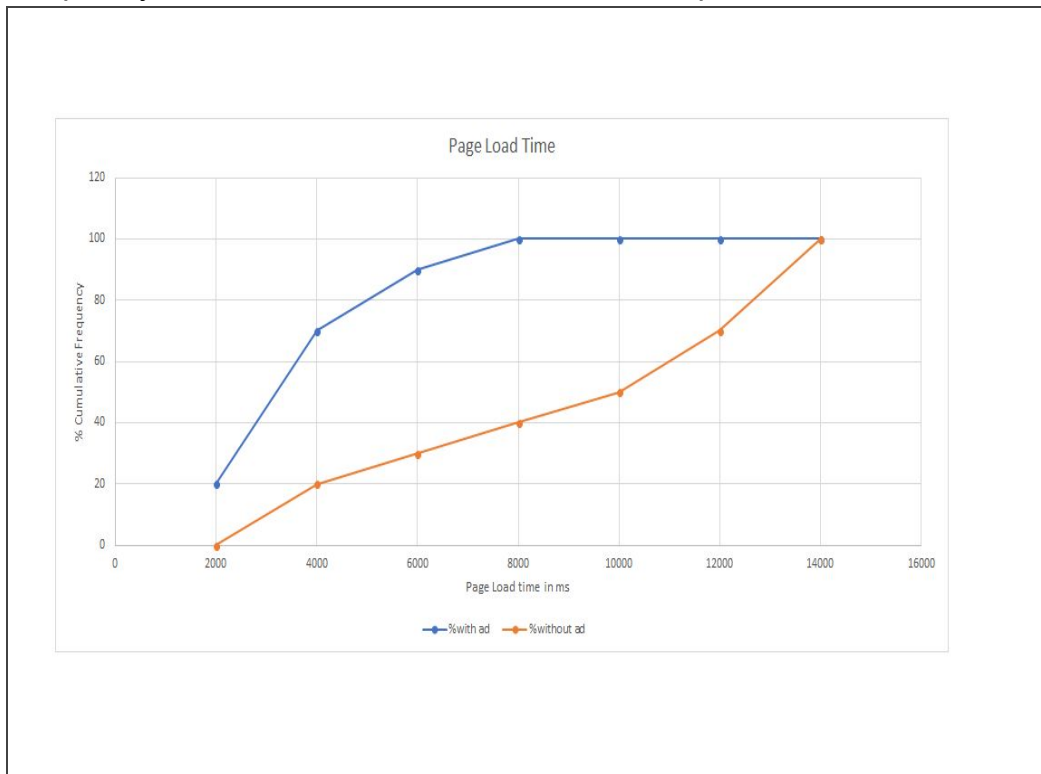

Next we present the cumulative distribution function for the number of requests made

In median case 125 requests were made without ad blocker and 30 requests

were made with ad blocker and in case of 75th percentile, 175 requests were made without ad blocker and 40 requests were made with ad blocker. This statistics indicates that 4 times more requests were fetched with ad blocker compared to without.
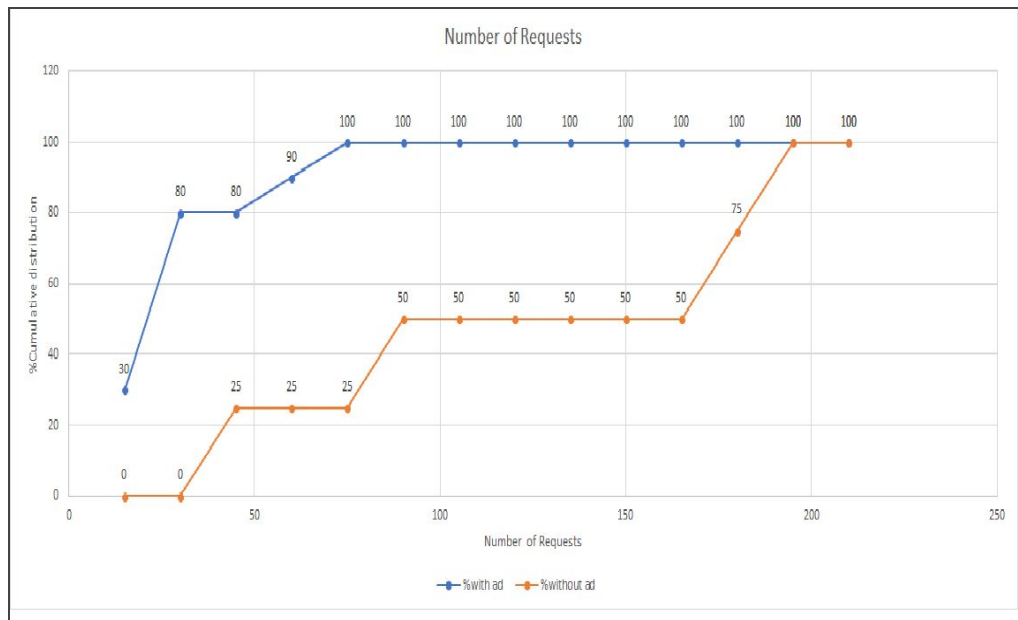
We then compare the cumulative distribution function of the Number of Bytes downloaded

Number of Bytes Loaded
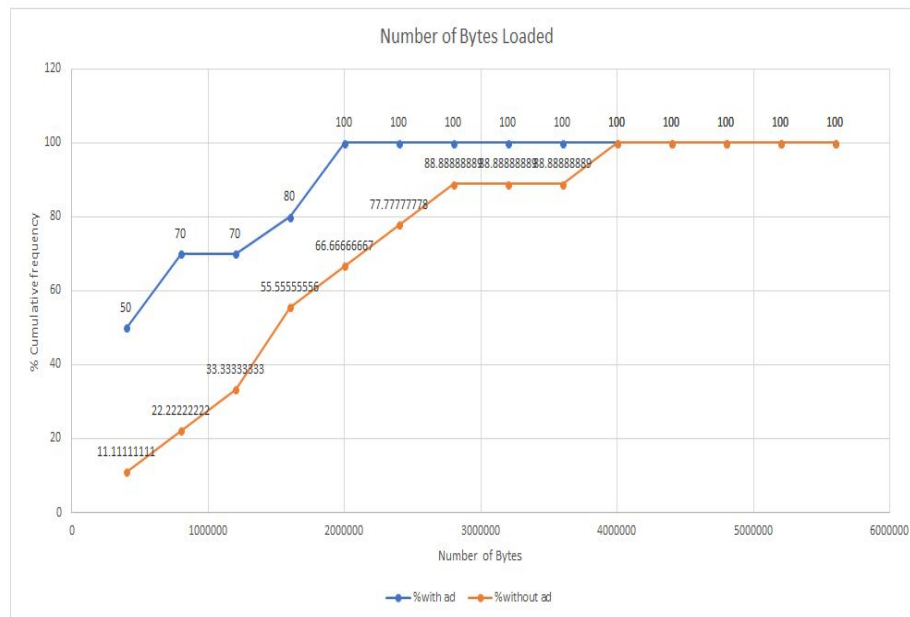
It has been observed that for the median case 1500 KB has been downloaded with ad blocker and 500 KB has been downloaded without ad blocker and for the 99th percentile case 3500 KB has been downloaded without ad blocker and 1600 KB has been downloaded with ad blocker. We observe that almost 2.5 times more data is downloaded for the web page with ads than the webpage without ads.

6. Clearly mark what you think you did for the addi/onal 30%

We have depicted our results in the form of bar charts and pie charts so that we can visually compare the various metrics with each other and also under various conditions or situations. We also presented the cumulative distribution charts of various metrics for different websites and pointed out the median cases, the 75th percentile cases and the 99th percentile cases in each of the metrics. This enables the viewers to understand the statistics better.

The interesting perspective of the problem is to apply the filters to block the ads from loading and the approach we took is to parse and get all the filters that could possibly belong to the ads and then pass all those filters as a form element to the Webpagetest API which would prevent those ad websites from loading and then compute the resultant time.

7. Conclusion:

From the above results we conclude that there is no significant difference in the web page load times for different browsers with adblock. Yet we observe a significant

difference in the page load times of web pages with and without ad blocker when viewed in the chrome browser. In case of mobile browsers for a few websites there wasn't any significant effect of the adblock for for a few other websites there was a significant effect. No surprising results were observed. All the analysis was near to what was expected.

Github link:

https://github.com/yash199/FCN-CSE-512-Project

References:

https://research.hubspot.com/why-people-block-ads-and-what-it-means-for-marketers-and-advertisers
https://blog.kissmetrics.com/loading-time/
https://www.nytimes.com/interactive/2015/10/01/business/cost-of-mobile-ads.html
https://adblockplus.org/filter-cheatsheet

https://sites.google.com/a/webpagetest.org/docs/using-webpagetest/quick-start-quide

https://easylist.to/