Mobile Infrastructure to analyze Page Download Time/speed/size over multiple Network Service Providers

Aditya Prakash (110350983), Alpit Gupta (110451714), Kumar Sasmit (110308698)[1]

*Supervised by - Aruna Balasubramaniam

ABSTRACT

Web page rendering is a very complex process. It involves downloading, parsing and rendering of various components in a specific fashion. Page Load Time (PLT) has been an area of concern for quite some time, and there have been various studies for its improvement. However, the time for downloading a page, called Page Download Time (PDT), has relatively been ignored. And, rarely, there has been any study that shows the factor which affects the page download time.
In this paper we present a detailed analysis of how page downloading gets affected under various network service providers. Our developed mobile infrastructure provides a real time statistics of the downloaded pages. Besides, we wrote a robust script to collect data using both Cellular and Wi-Fi networks. Based on the data collected over different places, we could analyze that given a website, page download time and total bytes download vary over different network service providers. Moreover, we showed that signal strength can be misleading at times. Higher signal strength might lead to slower speed, reasons being higher bandwidth, network traffic, tower misconfiguration and signal interference.

1 INTRODUCTION

Web page download time is a key performance metric that many techniques aim to reduce. Numerous studies and media articles report its importance for user experience and consequently to business revenues. For example, Amazon increased revenue of 1% for every 0.1 second reduction in PLT, and Shopzilla experienced a 12% increase in revenue by reducing PLT from 6 seconds to 1.2 seconds. We observe many factors deal with Web page download time including web page complexity, number of downloaded objects, browser computation, caching levels and many more. Many of these mentioned factors have been quite well addressed by some papers on HTTP and SPDY.

2 PROBLEM STATEMENT

Network Signal Strength is one of the key factors which significantly affect the page download time. Low signal strength might considerably reduce the page download performance even for simple web pages. Same web page might incur different PDT with different network providers for a given signal strengths. It might be the case for a web page to take lower PDT over lower signal strength. There might be a better bandwidth support for selective web pages by particular network. Similarly given a website, total bytes' download might vary with multiple service providers. Why do we see network download seed varying for different websites with same signal strength? For Example: www.google.com (homepage) has considerably less complex and high availability web page; still low internet signal strength can significantly affect its page download time.

Here in this project, we stressed more on variability of page download time, download speed and download size of landing pages with respect to varying network Service providers (NSPs).

3 ARCHITECTURE AND DESIGN

We used below hardware/software infrastructure to perform this real time experiment:
Time: We performed this data collection during night to measure the best provided network signal strength by providers and further avoiding any signal interference. Therefore, in-field experiments were performed in the midnight of April 17th 2016 between 10:00 PM to 04:30 AM.
Hardware: We used newly flashed Samsung Galaxy S6 Duos (rooted device) to run all our apps and scripts to measure the page download time along with the network signal strength, total bytes download and network download speed. We made sure to switch off the background data in the test phone.
Software: We designed and developed multiple apps and scripts to make the experiment more modular, robust and extensible for future use.

3.1 Network Signal Analyzer:

We designed and developed a location agnostic Android application platform to measure network signal strengths (Cellular & Wi-Fi). Following are the key features of this app:

- A researcher can plan to survey a location by loading any of the map.
- Use of android APIs (Cellular Manager/ Wi-Fi Manager) to measure network signal strengths of different network providers like AT&T, T-Mobile, Wi-Fi across multiple locations.
- It provides settings to select the number of scanning iterations over the place to accurately calculate the signal strength with chosen inter-scanning time interval to avoid conflicts.

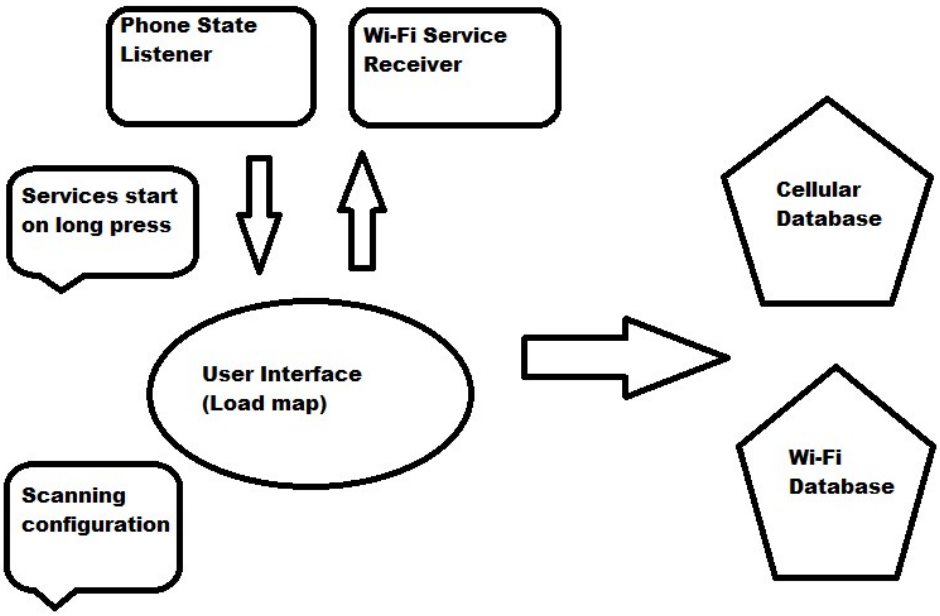

Figure 1. Functional Design of Network Signal Analyzer

- All required UI validation and error scenarios are handled. In case of any ongoing scan, researcher is not allowed to push the second scan. Proper toast message will be shown as “scanning is in progress”, on completion finished message will be displayed.
- Around 2000 lines of code have been written to make the app researcher friendly and fulfil all his/her required demands.

- In addition to this, we have also added a nice feature of loading our own survey map to perform experiments which can identify the already scanned locations by putting a red circular mark on it, hence researcher can easily remember the places which he/she has already scanned.

App Interface for measuring signal strength

Figure 2. Network Signal Analyzer

We could have used phone network signal strength or any third party app from the android play store to do so but we wanted to have our own controlled environment for the study and feature of repetitive multiple scanning to provide accurate measurement. As per internet sources, few of the phone vendors might display false signal strength values to lure their customers.

3.2 Interactive graphical Page download Time Analyzer:

We designed and developed a visual interactive android app which helps researcher in following ways:

- Quicker and detailed user interface to calculate the page download time, bytes downloaded and network speed of Top 100 Alexa websites.
- Graphical analysis (pie chart) of various page download time components (Using cUrl) along with network speed and bytes downloaded.

cUrl: A command line tool for getting or sending files using URL syntax. Curl offers many other features such as proxy support, user authentication, FTP upload, HTTP post, SSL connections, cookies, file transfer resume, Metalink, as well as various other features.

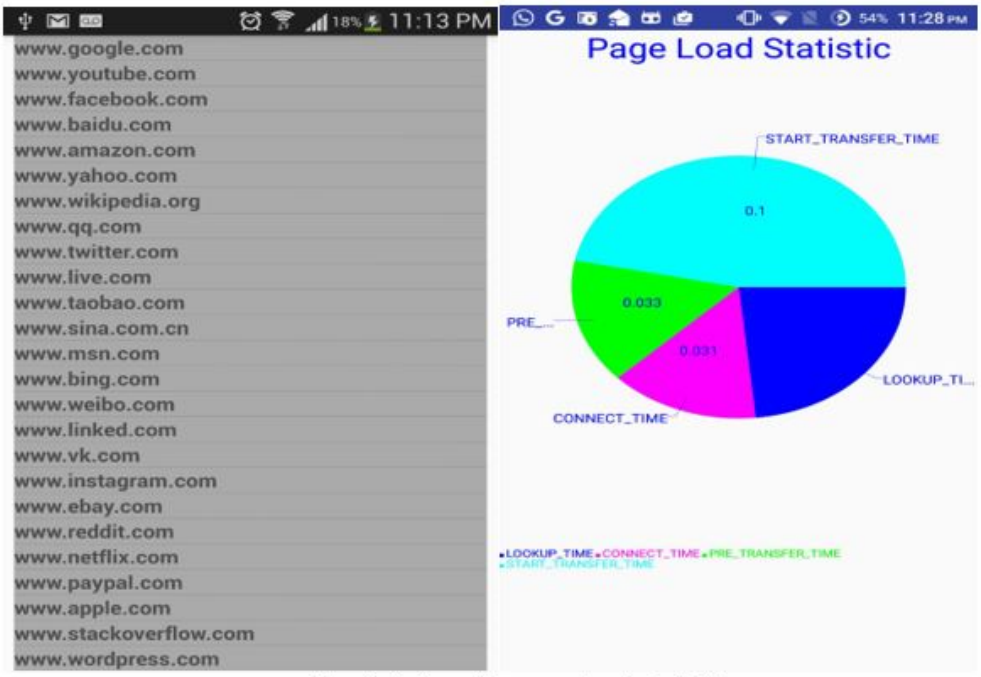

Figure 3. Interactive graphical Page download Time Analyzer

3.3 PDT Dump Tool Script:

We wrote a powerful shell script which downloads the source code of all the top 100 Alexa websites and dump many of the its information to the file including detailed page download time, network download speed and size of downloaded bytes.

4 Evaluation and Results

We performed real time experiments using both Cellular and Wi-Fi networks. We measured AT&T, T-Mobile and Wolfienet-Secure Wi-Fi signal strengths. In order to get the more accurate values, we carried out the experiments in the midnight. Below are the locations which were used for the in-field experiments using above mentioned network operators.

- Chapin Apartments, 700 Health Science Drive, Stony Brook
- Student Activity Center, Stony Brook University
- Stony Brook Hospital, Main Entrance, Stony Brook
- Stony Brook LIRR (Long Island Rail Road)

Finally, we dissected the collected data on below three key graphs and found many interesting behaviors along with our postulates:

	Chapin Apartments, 700Health Science Drive, Stony Brook	Student Activity Center, Stony Brook University	Stony Brook Hospital, Main Entrance, Stony Brook	Stony Brook LIRR (Long Island Rail Road)
AT&T	-97dBm	-85dBm	-95dBm	-85dBm
T-Mobile	-109dBm	-87dBm	-67dBm	-81dBm
Wi-Fi	-60dBm	-70dBm	-65dBm	NA

Network Signal Strength statistics

Figure 4. Network Signal Strength statistics

4.1 Total download time Vs Top 100 Alexa Websites

In most of the places, we can see AT&T and Wi-Fi have better signal strength as compared to T-Mobile (stated in the above table), hence taking less time too. But we have closely observed that some of the places where signal strength where quite close or reverse, still AT&T/Wi-Fi provided the better page download time. Even there is quite fluctuation in PDT with AT&T and Wi-Fi for measuring the performance of a particular website over the places. Following are the important take aways from our analysis:

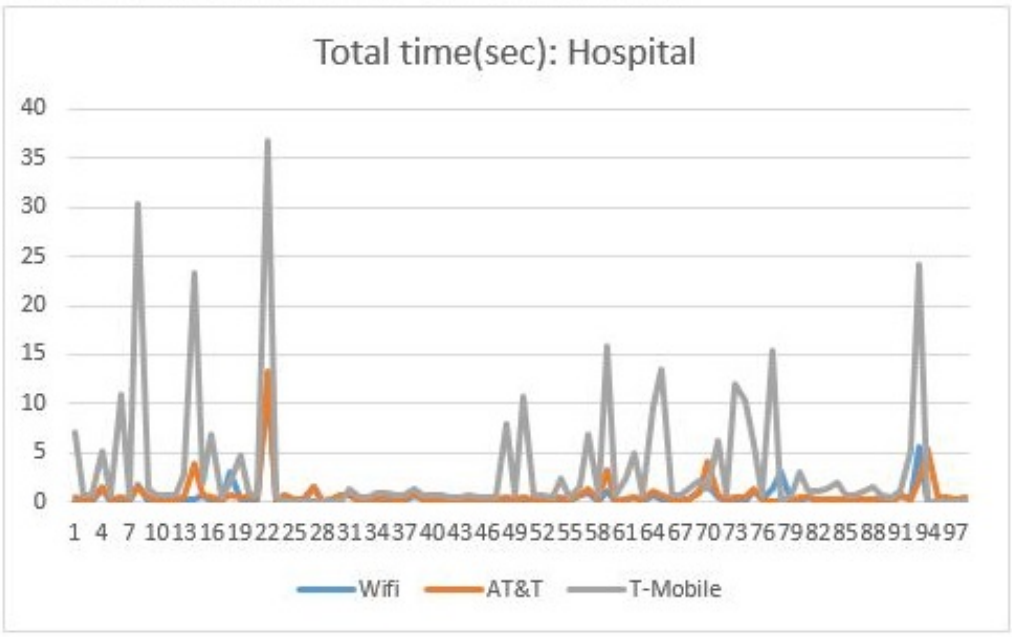

Figure 5. Page load time Vs Top 100 Alexa Websites

- The type of service (network protocols/technologies) being provided at a particular location. It is obvious that 4G will download the site pretty faster than the 3G service provided by the service providers, resulting in low latency. We have found 4G support with AT&T whereas T-Mobile providing Edge(E) network during our experiments.

- The routing path that is taken by the service providers need not be the same. One might follow the longer routing path and other might follow the shorter. (We are trying to analyze the packet route information)
- Customers and peer relationship might also affect. If the service provider is the customer of the other service provider where the site is hosted, then time taken to download the page for the former service provider will take longer time in comparison with that of later one.
- Other parameters like frequency of the band, network congestion, signal strength, provider introduced ads/scripts might also affect the page download time.

4.2 Total Bytes downloaded Vs Top 100 Alexa Websites

Different service providers providers allows some custom data along with the landing pages, this may be custom ads or may be some information regarding the headers specifics to the network providers. Following are the important takeaways from our analysis:

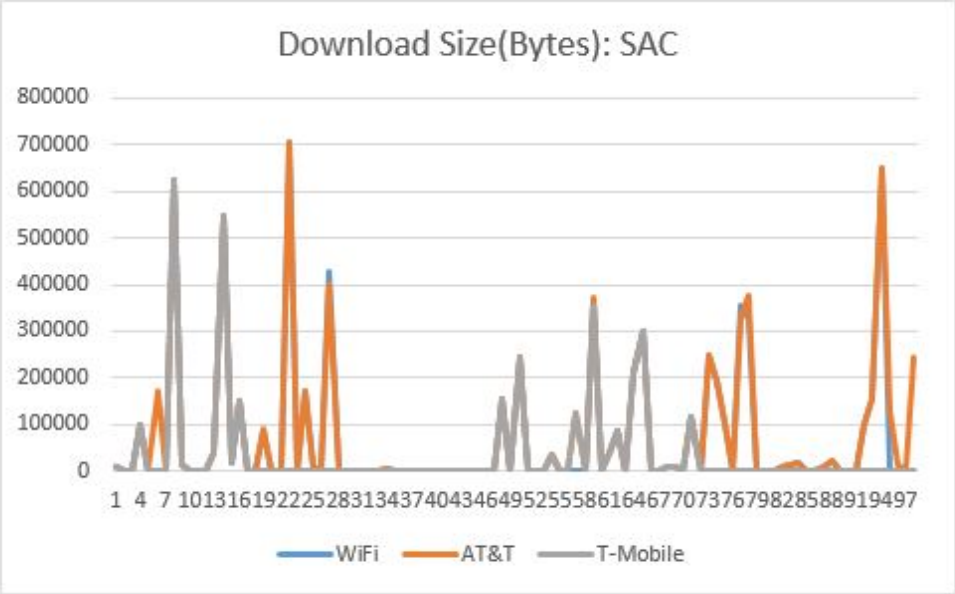

Figure 6. Bytes downloaded Vs Top 100 Alexa Websites

- Different service providers always embed different ads/monitoring scripts in the sites that will be accessed by the user through that providers. Hence, the size for downloading the websites differ under different service providers. (We have manually verified many of the websites to load in the browser and this argument valid)
- Involvement of the service providers would slap an extra information may be in the header or in the payload which might lead to the variance in the size of the data download. (We are trying to analyze the packet route information)
- Other factors can be retransmitted data for lost packets, Difference in packet header sizes if they use different transmission protocols(2G/3G/LTE), We have observed larger total bytes downloaded in case of 4G over Edge network, supporting this argument.

4.3 Downloading speed Vs Top 100 Alexa Websites

We have measured best signal strength with Wi-Fi which clearly shows better page download speed in most of the cases. We can also investigate and claim other reason which can be the reason for this, especially in between cellular networks. Following are the important take aways from our analysis:

- **Server Capacity:** Generally, the download speed depends upon the server speed. It is the bottleneck in downloading the data. If the connection speed is more than the server speed, then data will be downloaded in the maximum speed supported by the server.
- **Server Load:** A service request depends up on various factors to be served upon. Server load is one of the criteria that is very important. If the server is busy, then it is bound to take more time to serve the request and therefore the download speed will be impacted.

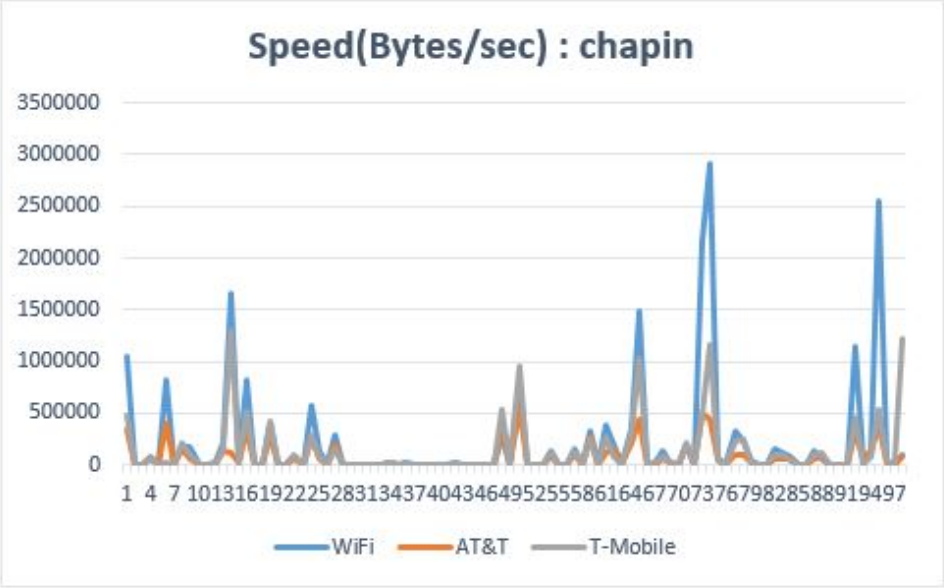

Figure 7. Download speed Vs Top 100 Alexa Websites

Other parameters like the number of servers on which the objects are hosted, their location, traffic on the network, Signal strength might significantly affect the download speed.

5 Other Components Affecting PDT

We performed further analysis by delving into other fine detailed time information obtained during our survey. Individually these values showed a similar behavior as shown by total download time which is already included in the analysis of our final result and evaluation. These times are:

- **time_appconnect:** The time from the start until the SSL/SSH/etc connect/handshake to the remote host was completed.
- **time_connect:** The time it took from the start until the TCP connect to the remote host (or proxy) was completed.

- **time_namelookup:** The time it took from the start until the name resolving was completed.
- **time_pretransfer:** The time from the start until the file transfer was just about to begin. This includes all pre-transfer commands and negotiations that are specific to the particular protocol(s) involved.
- **time_redirect:** The time it took for all redirection steps include name lookup, connect, pretransfer and transfer before the final transaction was started. time_redirect shows the complete execution time for multiple redirections.
- **time_starttransfer:** The time it took from the start until the first byte was just about to be transferred. This includes time_pretransfer and also the time the server needed to calculate the result.

6 Conclusion

While choosing the best channel for better page download time, should we only look for signal strength?
No. In some places, network speed can be fast even with low signal strength while in other places it can be slow even with the high signal strength. The reason being the dependency on bandwidth, network traffic, tower misconfiguration, tower deafness, signal interference and distance from the tower.

Given the importance of page download time/bytes/speed over different network service providers, we designed and developed a robust Mobile Infrastructure to collect and analyze the various network parameters. We performed real time experiments over multiple places and collected the ample amount of rich data. Our infrastructure can be used by the researchers to get the real time sense of page download time/bytes/speed of the mobile network and present those statistics in graphical representation. Moreover, we also provide an extremely robust script that helps in accumulating the contribution of various components involved. Finally, the results show that page download time/speed/bytes depends not only on network signal strength but also on the bandwidth, protocol used, server capacity and many other factors.

7 Future work

Although we analyzed a lot of different network parameters in details for the page download time/speed/bytes, we can still see plenty of research opportunities in this field. Analyzing non-landing pages for the websites is one of the main areas. Generally, landing pages i.e. home pages of the web sites cover most of the experiment's focus and non landing pages are relatively ignored. But in some cases like E-commerce, Media and Sports sites, non-landing pages do contribute to a lot of traffic. Hence collecting and analyzing the statistical behavior of various components like speed, time and bytes downloaded for non-landing pages will surely help in better design of websites.

Moreover, comparison between PLT and PDT for web experience can help researchers to combine the best of the both for making a new standard which can help in designing the sites for better response time. Finally, dependency on hardware, including radio antenna, NIC, memory r/w speed etc. can also help in understanding the effect of hardware on PDT.

8 Code Repo and References

8.1 Code base and Survey data:

https://github.com/waytoalpit/Mobile-Infrastructure-to-analyze-Page-Download-Time-spe

8.2 Reference:

http://www.ericmobley.net/measuring-performance-google-analytics/
https://en.wikipedia.org/wiki/CURL/
http://developer.android.com/index.html/
http://blog.livedrive.com/2012/08/why-is-an-internet-speed-of-125kbs-the-same-as-1mb
#sthash.7A7Uixc6.dpuf/
https://7labs.heypub.com/tips-tricks/download-speeds-explained.html
http://developer.android.com/reference/android/net/wifi/WifiManager.html
http://developer.android.com/reference/android/telephony/TelephonyManager.html
https://www.overleaf.com/latex/templates/
https://blog.kissmetrics.com/loading-time/
http://www.aresearchguide.com/1steps.html