WHAT SLOWED DOWN MY APP - A MEASUREMENT STUDY OF RESPONSE TIMES OF SMARTPHONE APPS IN INDIA

BY

ANJALI | MT20082
SHRADHA SABHLOK | MT20069

Under the supervision of
Dr. Mukulika Maity
Dr. Arani Bhattacharya

INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY DELHI
December, 2021

What Slowed Down My App - A Measurement Study of Response Times of Smartphone Apps in India

By

Anjali | MT20082
Shradha Sabhlok | MT20069

Submitted in partial fulfillment of the requirements for the degree of
Master of Technology

To

Indraprastha Institute of Information Technology Delhi
December, 2021

Certificate

This is to certify that the work titled *What Slowed Down My App - A Measurement Study of Response Times of Smartphone Apps in India* being submitted by *Anjali (MT20082)* and *Shradha Sabhlok (MT20069)* to the Indraprastha Institute of Information Technology Delhi, for the award of the degree of Master of Technology, is an original research work carried out by them under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

December, 2021

Dr. Mukulika Maity
Dr. Arani Bhattacharya
Indraprastha Institute of Information Technology Delhi
New Delhi 110020

Acknowledgements

We would like to thank our advisors Dr. Mukulika Maity and Dr. Arani Bhattacharya for guiding us, providing us with insights and expertise and giving us the opportunity to work on this project.

Abstract

India has seen a rapid increase in the use of smartphones over the past few years with a majority of Indians accessing internet through smartphones. Smartphones have become one of the most important mediums of providing essential services like entertainment, news and even payment services in India. Thus when such a vast population depends on smartphones it is essential to study the difference in quality of experience of individual users, especially when there is huge diversity of the smartphone models, network conditions, and regions from which they are used. We quantify the Quality of Experience of using the android apps using response time i.e., the time needed to reflect UI changes corresponding to a user's action. This can be, for example, the time needed to reflect an item has been added to cart for Add product to cart action on Amazon app. In this work, we design a tool called EvalApp which uses automation to record the response times of a total of 30 actions for 12 apps popular in India. We then crowdsource this desktop app to a total of 51 users working from home from across north and central India and perform a causal analysis of the factors that affect the response times of actions. We find that the response time is correlated with the ping round-trip time to the nearest Google server, thus indicating that the network is the most common bottleneck. We further identify that the distance from a major city, phone's CPU frequency, and RAM are other factors that strongly affect latency.We observe that 3 out of 51 users face a consistent bias in their response times. We also do a few controlled experiments to observe that having a higher app version does not always ensure better response times. Apart from these tests, we also checked the effect of free available memory, size of transmitted/received bytes for various actions, time of day when experiments were performed, network types (like wifi, mobile data and hotspot) and ISP over the response times for app actions. We found no significant effect of these parameters over the action response times. Our observations are likely to help designers of apps to better identify techniques of improving the response times for users.

Contents

Chapter 1

Introduction

Over the last five years, there has been an increasing adoption of smartphones in India. For example, the number of smartphones sold in India increased by over 2 times from 2014 to 2019, with over 150 million smartphones being sold in 2019 [1]. Smartphone services are increasingly being used for payment, messaging, social networking, reading news as well as entertainment like watching videos and playing games. The number of apps downloaded in India is the highest in the world [2]. Thus, it is vital for such app developers to understand the quality of experience (QoE) observed by users in India.

The quality of experience of smartphone apps is primarily governed by the response time observed by the users on interaction [3, 4]. For example, a user watching videos over YouTube would have a worse experience if the video takes longer to load. A key concern in countries like India is that there is a wide divergence in the response times observed by different users. This concern about response time is especially important in the context of India because the vast majority of the population of India accesses the Internet through smartphones [5]. Since this divergence in latency can lead to poorer access to essential services among users, it is essential to quantify the amount of divergence seen by actual users while interacting with the apps. A second question that is currently not understood is the reason behind the divergence in response times. The divergence in response times can be due to multiple reasons, such as the type of smartphone used, location of the users, the version of the apps, nature of backhaul network as well as the type of network used. However, which of these causes dominates is currently not understood.

In this work, we conduct a large-scale crowdsourced measurement of the response times for different user actions such as "Add product to cart" on Amazon, “Play video” on Youtube, “Send message” on Whatsapp etc., on the 12 most common apps. For each of these apps, we choose the most common types of user actions and use an automation tool Appium [6] to activate specific actions automatically. We then wait until the UI element shows that the action has been completed, and compute the time elapsed between the change in the UI and the time of activation. Using this technique, we compute the response time of each application. We then crowdsource this automated tool to over 51 users after suitable Institutional Review Board (IRB) approval, and request them to run the automated tool by connecting it to a personal computer. Apart from logging the response times observed, we also send ping requests to 3 servers i.e., Amazon. Google, and Mobikwik from the smartphone to measure the network condition. We collect over 26000 such response time values, from across north and central India. We also request the users to specify their Android version, phone RAM, CPU cores & frequency, and distance from a major city [1] using a separate form.

We observe huge divergence in response times across the users. For example, Youtube’s “Play Video” action has a response time of $0.5s$, $2.3s$, $3.6s$ and $5.5s$ at 0^{th}, 25^{th}, 75^{th} and 100^{th} percentiles respectively. Similarly, Hotstar’s “Open Trending Page” has a response time of $0.05s$, $2s$, $3.9s$ and $6.5s$ at 0^{th}, 25^{th}, 75^{th}, and 100^{th} percentiles respectively. In most of these cases, the differences are higher than $1s$, and in some cases, higher than $6s$. Such perceptibly high divergence in response times is known to hurt the quality of experience [8]. To understand the impact of network, we perform controlled experiments in our campus network that has a high bandwidth connection (1Gbps). We perform these experiments in the middle of the night to ensure low network congestion. Further, to understand the impact of phone configuration we conducted these experiments on two different phone models: they are Samsung Galaxy S10e [9] with 6 GB RAM, 8 cores, max clock frequency 2.73 GHz and Infinix Hot 9 Pro [10] with 4 GB RAM, 8 cores, max clock frequency 2 GHz. We observe that the response times observed in the controlled experiments have lower divergence on all actions across all apps compared to uncontrolled data. We further note that the response times experienced on the Samsung phone is lower for majority of the actions compared to the uncontrolled data. The only exceptions are for Amazon’s “Add product to cart” and “Go to cart” actions, that have higher response times likely due to festive sale during controlled experiment’s timing. We further observe that the response times experienced with the Infinix phone is higher than the uncontrolled data for certain actions and lower for certain actions.

[1] We identify cities with data centers [7] as a “major city” in this work.

For instance, Flipkart's "Product Profile" action for uncontrolled data is $0.75X$ of the Infinix's response times. Similar behavior is observed for "Person Profile" and "Post in a Group" actions for Facebook. Whereas, for "Play Video" action of Hotstar on the Infinix, the median response time is $4X$ better than uncontrolled data.

We then perform a causal analysis to identify the underlying reasons behind such high response times. We first compute the correlation amongst the parameters such as ping RTT (Round Trip Time), distance, network type etc. We compute Spearman correlation coefficient. We find that median and minimum ping RTT values have moderately high correlation with the distance (> 0.3, but < 0.5). We further find out that the ping to three destinations i.e., Amazon. Google, Mobikwik have strong correlations amongst themselves (≥ 0.7). Hence, we choose Google's ping values as representative of ping RTT values. Next, we compute the correlations between the response time and the input factors i.e, median & minimum ping RTT, distance, CPU frequency, RAM, network type, Android version. We find that in general the median ping RTT to Google is correlated with the response times. 6 actions have significant correlation with minimum ping RTT. Interestingly, we find that the geodesic distance from the city with a data center is more strongly correlated with the response time than the ping RTT values. A total of 9 actions have significant positive correlation with distance (with p-values < 0.001). Further, we observe that phone's CPU frequency and RAM also have a significant correlation for a total of 11 and 9 actions respectively. This confirms the observation by a few studies that hardware is a major bottleneck on smartphones [4, 11]. We observed that factors like free available memory on device, transmitted/received bytes per action, performing experiments at different times of day, different network types and different ISPs had no impact on the response times of the app actions. Finally, we further note that some users face a consistent bias in their response time i.e., their median response times exceed the 80^{th} percentile of the overall response time. We found out the causes behind such behaviour are old phone model and large distance from major city.

We summarize our contributions as follows:

- We create an automated tool EvalApp to compute the response times of 12 common Android applications for a total of 30 common interactions.
- We crowdsource EvalApp to over 51 users spread across north and central

India. We further collect a total of over 26000 data points with each user performing an IRB-approved experiment for over 5 days.

- We note the divergence in response times of the apps. We compare the response times observed and perform a causal analysis to identify that the quality of the network, distance, CPU frequency, and RAM are the main bottlenecks for most users. We find that the response time is correlated with the ping RTT times to a few standard servers. Furthermore, we find that a better phone in a campus network with low ping RTT values can reduce the response times of most actions by 1.3 times. Interestingly, distance from a major city is also moderately correlated with the response times for most apps.

- We observe that the divergence in response times is significantly high for 3 out of 51 volunteers. We find out its root causes. We perform a controlled experiment to study the affect of app versions on the response times. We show that having a higher app version does not always ensure better response times, as higher versions tend to download more data.

- We also studied other factors like available free memory on the android device, size of transmitted/received bytes for various actions, performing experiments at different times of day, using different network types (like wifi, mobile data and hotspot) and also the service provided by different ISPs. It was observed that all these factors reproduced the same behavior as the median response times seen were almost the same in all apps for all actions.

Chapter 2

Design and Implementation

We first discuss the design of EvalApp. We then discuss our technique of crowdsourcing and the collected dataset.

2.0.1 Design of EvalApp

The goal of EvalApp is to create a user-friendly framework that can be utilized by non-technical users to perform tests on Android apps to measure the response time. It uses a GUI-based desktop application written in Java programming language, with the GUI written using JavaFX library [12]. The test code for each app is written using Appium library [6], which is an open-source automation tool to test apps.

The design of EvalApp is shown in Figure 2.1 Appium starts a server on a personal computer (PC) and requires the smartphone to be connected to it via a USB port.

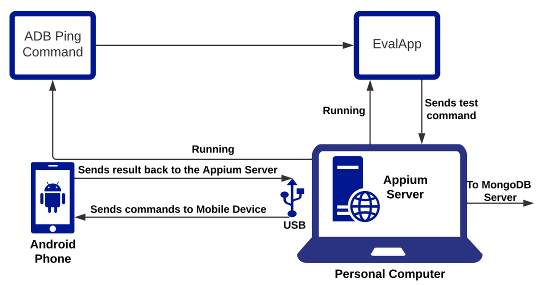


Figure 2.1: Design Flow Chart of EvalApp

The automation scripts for each individual app first need to be installed on the PC. When the user connects their smartphone to the PC using a USB cable and launches the tests from EvalApp, these automation scripts activate some specific user actions on the apps. This script is designed in such a way that it waits for a specific change on the User Interface (UI) to occur, and then records the time elapsed between the request sent and the change on the UI. The exact set of UI actions used for each app is shown in Table 2.1.

Table 2.1: A list of 12 apps and 30 actions for which we have measured the response times.

	App	App Type	Action	UI Response
1	YouTube	Streaming	Search channel	Search results displayed
			Open channel Page	Channel logo shown
			Search video	Search results displayed
			Play video	Video player box appeared
2	Hotstar	Streaming	Search video	Search results displayed
			Play video page	Video title displayed
			Open trending page	Video player box appeared
3	LinkedIn	Social	View profile	Profile picture shown
			Check connections	Connections page appeared
			Search Person	Search result displayed
			Open person profile page	Profile picture shown
4	Facebook	Social	Post a message	Message box appeared
			Search person	Search results displayed
			Open person profile page	Profile picture shown
5	Google News	News	Search news	News page title displayed
6	DailyHunt	News	Search news	News box appeared
7	Amazon	Shopping	Search a product	Search results displayed
			Open product page	Product rating stars shown
			Add product to cart	Cart items increased
			Go to cart	Cart items shown
			Remove product from cart	Product removed shown
8	Flipkart	Shopping	Search a product	Search results displayed
			Product profile	Product rating stars shown
			Add product to cart	Cart items increased
			Go to cart	Cart items shown
			Remove product from cart	cart items decreased
9	WhatsApp	Messaging	Send message	Ticks shown
10	Telegram	Messaging	Send message	Ticks shown
11	Paytm	Payment	Send 1 payment	payment success shown
12	Google Maps	Navigation	Search a location	Street Thumbnail displayed

The data collected is sent to the Appium server, which in turn forwards it to a remote MongoDB server for storage. Simultaneously, we collect ping data from the smartphone running *adb ping* commands in the connected mobile device to `google.com`, `mobikwik.com` and `amazon.com` once every five seconds throughout the experiment to collect at least 10 ping RTT values per server.

Table 2.2: Range of parameters of the experiments. Many of these parameters depend on the individual settings used by the users.

Parameter	Value/Range
Number of Users	51
Range of Android versions	7 - 11
Range of Android RAM	2 GB - 8 GB
Maximum physical distance from major city	397 km
Times of Day (Morning)	9:00 - 15:00
Times of Day (Evening)	17:00 - 23:00
Number of identical actions repeated by each user	5-10
Ping probes sent per app experiment	10

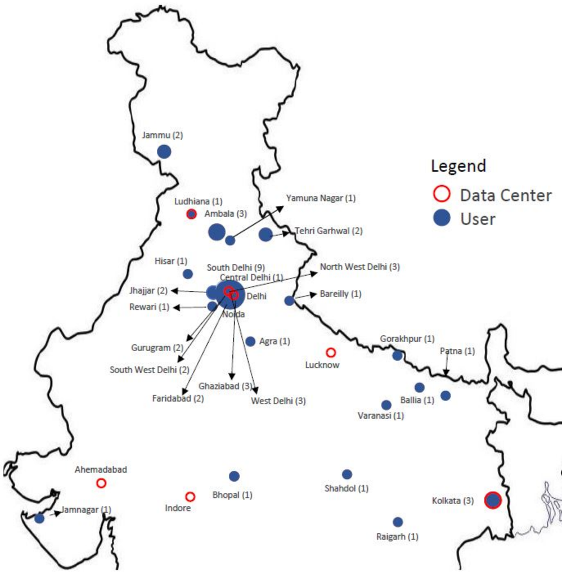


Figure 2.2: Locations of the volunteers who run EvalApp and cities with data centers (major cities). The number in parenthesis denotes the number of users in each location.

2.0.2 Our Dataset

We requested a total of 51 students of our department to serve as volunteers. The volunteers had to install and login to these 12 apps. Due to the COVID pandemic, most of the students attended their classes from home. Thus, the students were spread over north and central India, with a small majority of students coming from the National Capital Region (see Figure 2.2). We requested them to run the experiments twice a day for at least five days to reduce the chances of outliers affecting our conclusions. The volunteers used either WiFi, cellular network, and/or WiFi hotspot depending on their convenience. The parameter ranges seen in our experiments are shown in Table 2.2.

We collected a total of over 26000 data points from these volunteers. Among the volunteers majority users(i.e. around 80%) had 4 GB and 6 GB of RAM. 15% of the users performed experiments with Android phones having 8GB of RAM. Whereas, only 1 user had RAM as high as 12 GB. Considering the Android versions on volunteers' phones, around 43% were having highest Android version of 11, 31% had Android 10 and rest of the 26% carried Android versions 7, 8 and 9. Experiments were performed twice a day by each volunteer; once with WiFi Network and once with Mobile Data switched on (ensuring WiFi is switched off). Those volunteers, without WiFi access connected to the mobile hot-spot to perform the experiments. The data has been collected for different service providers with varied frequency bands over a span of 7 months from April, 2021 till October, 2021.

Collecting this dataset required approval from the Institutional Review Board (IRB), informing the volunteers about the ethical issues involved and getting their consent.

Furthermore, we performed controlled lab experiments for analyzing the app response times in a high bandwidth network during low network congestion and on two different phone models with one being a high end phone. Our campus is connected to the National Knowledge Network (NKN) via a Gigabit link, run by the Government exclusively for educational institutions. The network latency of NKN is known to be significantly smaller than the commercial home networks, with the ping RTT to Google server having a median of just $9.1ms$. To ensure low network congestion, we performed the experiments in the night at $12-3$ am. We performed these experiments on Infinix Hot 9 Pro and Samsung Galaxy S10e having 4 GB and 6 GB of RAM respectively. The number of CPU cores is 8 in both the testing devices having maximum clock speed of 2 GHz and 2.73 GHz respectively. The Android versions running on both these devices is 10. In addition to response times, we measured other statistics such as the bytes transmitted (Tx) and received (Rx), number of TCP and UDP connections used.

We also measure response times across app version upgrades. For that, we took a Lenovo phone with model number *A6020a40* that has a RAM of 2 GB and runs Android Oreo *8.0.0*. We analyze the response time of the actions across the three most recent versions of Google Maps, YouTube, Flipkart, and Telegram. We could not run this analysis for other apps as 1) the forced update for some apps did not allow us to run their older versions, and 2) authentic APKs of older versions were not available for some of the apps.

Chapter 3

Distribution of App Response Times

We discuss the response times along with their data sent and received and the number of TCP/UDP connections of each type of app separately.

Streaming Apps:

Figure 3.1 shows the response times of each action as well as the data consumption of Youtube and Hotstar. We first note that the response times have appreciable divergence, with the 0^{th} and 100^{th} percentiles (shown using the boxes) having a significant difference. For example, Youtube's "Play video" action has a response time of $0.5s$ and $5.5s$ at 0^{th} and 100^{th} percentiles. The 25^{th} and 75^{th} percentiles for the same has values of $2.3s$ and $3.6s$ respectively. Similarly,"Search channel", "Open channel" on Youtube, and "Open trending page" on Hotstar have high divergences, with the 0^{th} percentiles and 100^{th} percentiles being equal to $0.7s$ and $1.7s$, $1.3s$ and $2.6s$, and $0.05s$ and $6.5s$ respectively. The 25^{th} and 75^{th} percentiles of these being equal to $1s$ and $1.4s$, $1.7s$ and $2s$, and $2s$ and $3.9s$ respectively. We also note that these divergences are lesser for controlled experiments on both the phones. Next, we observe that for most actions except for "Play video" action of Youtube the response times experienced in controlled experiment are lesser than in uncontrolled settings. Further, the response time observed with Samsung phone is lesser than the Infinix phone. For example, the median response time for "Open channel" action on Youtube results in a median response time of $1.9s$, $1.3s$, $1.5s$ respectively for uncontrolled, Samsung phone and Infinix phone. We note that the actions which have very high divergence in response times have a huge divergence in the number of Rx bytes too. For example, the Rx bytes for Youtube's "Play video" are $0.9MB$

and $3.5MB$ at 0^{th} and 100^{th} percentiles respectively. The values at 25^{th} and 75^{th} percentiles for same are $1.6MB$ and $2.4MB$ respectively. Similarly, the Rx bytes for Hotstar's "Open trending page" are $0.1MB$ and $7.1MB$ at 0^{th} and 100^{th} percentiles respectively. The Rx bytes for at 25^{th} and 75^{th} percentiles are $1MB$ and $3.8MB$ respectively. Note that we could not perform controlled experiments for Hotstar's "Open trending page" action, as this feature was discontinued when the controlled experiments were performed. We observe that Youtube opens around 12 TCP connections and 13 UDP connections. While Hotstar opens around 14 TCP connections and no UDP connections.

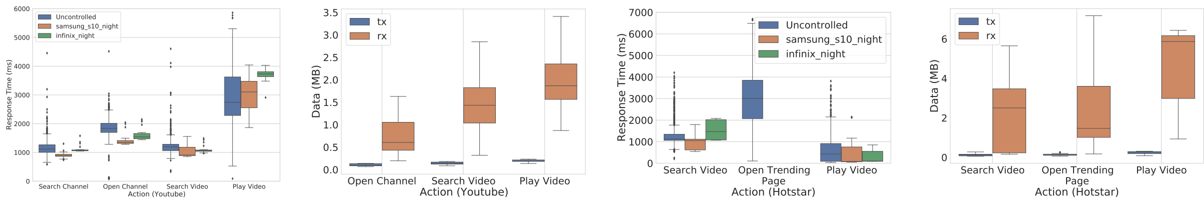


Figure 3.1: Streaming app: Distribution of Response Times for a) Youtube c) Hotstar. Distribution of Tx/Rx bytes for b) Youtube d) Hotstar.

Social Apps:

Figure 3.2 shows the response times of each action as well as the data consumption of LinkedIn and Facebook. Once again, we note here that response times have divergences for uncontrolled data and this divergence is lesser for controlled data. We also observe that for LinkedIn, Samsung mostly experiences better response times compared with uncontrolled data. However unlike Samsung, the Infinix phone has very similar median response time compared with the uncontrolled data. Further, we note here that Rx bytes is small i.e., less than 1MB. On the other hand, the response times experienced with Facebook is very similar for uncontrolled and Samsung phone. Further, the response times observed with Infinix phone are higher than experienced in uncontrolled and Samsung phone. We observe that LinkedIn opens around 4 TCP connections and no UDP connections. While Hostar opens around 6 TCP connections and 3 UDP connections.

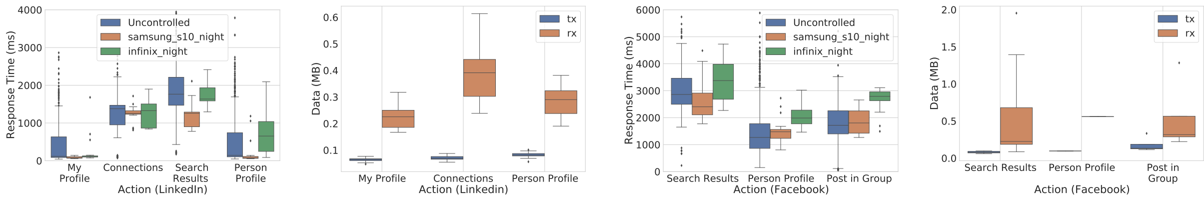


Figure 3.2: Social Apps: Distribution of Response Times for a) LinkedIn c) Facebook. Distribution of Tx/Rx bytes for b) LinkedIn d) Facebook.

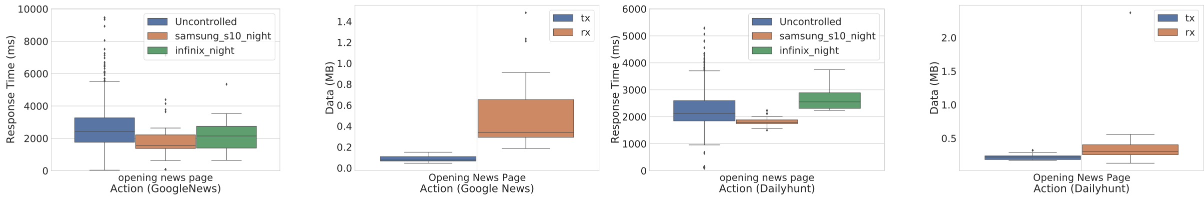


Figure 3.3: News Apps: Distribution of Response Times for a) Google News, c) Daily Hunt and of Tx/Rx bytes for b) Google News, d) Daily Hunt.

News Apps:

Figure 3.3 shows the response times of each action as well as the data consumption of Google News and Dailyhunt. Here, we have similar observations as LinkedIn. The Samsung phone mostly experiences better response times compared with uncontrolled data. However unlike the Samsung phone, the Infinix phone has similar median response time (in case of Google News) and slightly higher response time (in the case of Dailyhunt) compared with the uncontrolled data. We observe that Google News opens around 8 TCP connections and 3 UDP connections. While Dailyhunt opens around 10 TCP connections and no UDP connections.

Shopping Apps:

Figure 3.4 shows the response times of each action as well as the data consumption of Amazon and Flipkart. We have a very interesting observation for Amazon app, specifically for these two actions “Add product to cart” and “Go to cart”. The timelines of uncontrolled and controlled experiments were different. When we performed the controlled experiments, there was a festive sale. As an artifact of that, we observe that the response time of “Add product to cart” and “Go to cart” have very high response times, the median is $45s$ and $25s$ respectively. Also, we observed that these tests were failing on the Infinix phone, probably because Amazon has changed its UI design. For Flipkart, we have similar observations like other apps. The response time of controlled settings

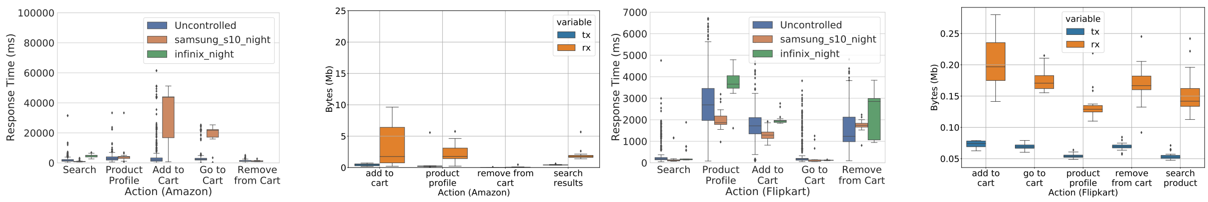


Figure 3.4: Shopping Apps: Distribution of Response Times for a) Amazon c) Flipkart. Distribution of Tx/Rx bytes for b) Amazon d) Flipkart.

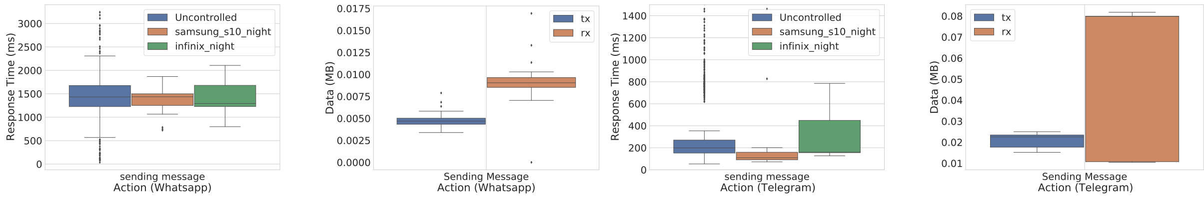


Figure 3.5: Messaging Apps: Distribution of Response Times for a) Whatsapp, c) Telegram and of Tx/Rx bytes for b) Whatsapp, d) Telegram.

have lesser divergence than uncontrolled data. Also, the response times observed with Samsung phone is always less than the uncontrolled data. However, the response times observed with Infinix phone is higher than uncontrolled data for these three actions “Product profile”, “Add product to cart”, and “Remove from cart”. We observe that Amazon opens around 22 TCP connections and no UDP connections. While Flipkart opens around 10 TCP connections and no UDP connections.

Messaging Apps:

Figure 3.5 shows the response times of each action as well as the data consumption of Whatsapp and Telegram. Next we note that, both the messaging apps experience a similar response time for all the settings. We observe that the Rx bytes is very small < 0.1 MB. In comparison to Telegram, Whatsapp downloads $10X$ lesser bytes, yet experiences a $10X$ higher response time. This is likely due to mandatory end-to-end encryption enabled by Whatsapp. We observe that Whatsapp opens around 3 TCP connections and no UDP connections. While Telegram opens around 2 TCP connections and no UDP connections.

Payment Apps:

Figure 3.6 shows the response times of each action as well as the data consumption of Paytm. Here again, we note that response time observed in controlled experiment with both Samsung and Infinix phones are lesser compared to uncontrolled data. The Rx bytes is relatively small $< 0.22MB$. We observe that Paytm opens around 10 TCP connections and no UDP connections.

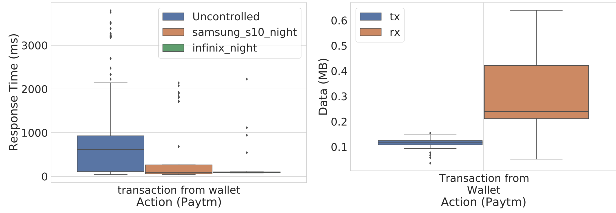


Figure 3.6: Payment apps: a) Distribution of Response Times, b) Distribution of Tx/Rx bytes for PayTM.

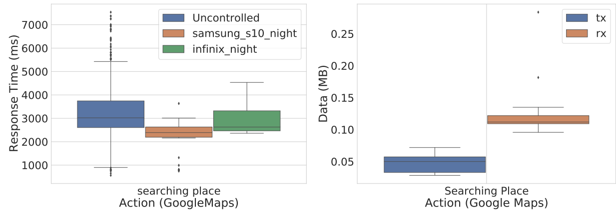


Figure 3.7: Navigation apps: a) Distribution of Response Times, b) Distribution of Tx/Rx bytes for Google Maps.

Navigation Apps:

Figure 3.7 shows the response times of each action as well as the data consumption of Google Maps. Here again, we note that response time observed in controlled experiment with both Samsung and Infinix phones are lesser compared to uncontrolled data. The Rx bytes is also small $< 0.15MB$. We observe that Google Maps opens around 3 TCP and 3 UDP connections.

Chapter 4

Causal Analysis of Response Times

We first discuss the relationship between the parameters that we have collected. We then compute the correlation between these parameters and the response times. We also compute correlation between transmitted/received byte per action and the action response times. We also study the effect of network types, ISPs and time of day when the experiments were performed on the response times. We then take note of some interesting observations and try to confirm them through additional analyses or controlled experiments.

4.0.1 Relationship between Parameters

For each user, we collect the data about their Android version, the amount of RAM on their smartphone, and their distance from a major city. We specifically consider distance from a major city, because India is known to have more developed network and data center infrastructure in a few major cities of the country [13]. For each experiment, we also log the type of network used (WiFi/-cellular), time of day, and the Round Trip Time (RTT) of the ping probes to Amazon, Google and Mobikwik servers. These parameters act as potential causal factors and we observe their correlation with the response times.

We first confirm whether distance from major city actually has an impact on the ping RTT's. We compute the Spearman rank correlation coefficient between the median and minimum ping RTT observed by a user during one particular experiment and the air distance reported. Note that we do not utilize Pearson correlation coefficient, even though it is simpler to compute, because it is more vulnerable to outlier values. The correlation values are shown in Table 4.1. We find that in each case, we have a moderately high correlation value of greater

Table 4.1: Spearman Correlation Coefficient between distance and ping RTT RTT's to Amazon, Google and Mobikwik servers.

	Median RTT	Min RTT
Amazon	0.382	0.478
Google	0.345	0.323
Mobikwik	0.377	0.392

than 0.3.

We further check the relationship between the median and minimum ping latencies among the three different servers. We do this by again computing the Spearman coefficients among the ping median and ping minimum values. We find that each of the six possible pairs have Spearman coefficients greater than 0.7, indicating a strong correlation between them. Thus, we utilize only Google's median and minimum ping RTT times to further identify the cause of divergence in app response times. Also in addition to ping median, ping variance was considered to check if it significantly affects response times. But the results were not different from what we had already established.

We next check whether the ping RTT's depend on the type of network used (WiFi, 4G or mobile hotspot). We find that 4G and mobile hotspots have a significantly higher ping latency than WiFi, with a correlation coefficient of 0.63 and a median difference of $40ms$. Interestingly, while WiFi remains better than 4G even farther (>200km) from the major cities, the gap between their ping RTT's reduces to a median value of only $23ms$. This indicates that congestion in 4G is one key factor that is increasing the ping values in the major cities. We also find that the ping RTT's to the three servers has 1.2-1.6 times higher standard deviation using 4G than using WiFi, with the values ranging from 23ms to 43ms.

4.0.2 Correlation Between the Parameters and Response Time

We again compute the Spearman coefficients between the individual parameters and the response times. We omit the time of day and ping variance among the parameters because their correlation coefficients never exceed 0.1 for any of the actions, indicating that they do not have a significant effect. We plot the coefficients in the form of a parallel coordinates plot in Figure 4.1. We also show the entire table of correlation coefficients in the Table 4.2. Note that we plot the negative values of the coefficients for RAM and Android version since higher RAM and higher Android versions are more common in newer hardware and thus we expect them to lead to lower response times.

To further understand the impact of each parameter, we also specify the app

Table 4.2: Values of Spearman Rank Correlation Coefficients between each parameter and response time of action.

Network Type	Ping Min	Ping Median	Distance	RAM	Freq	Android	Action
0.067	0.062	0.053	0.111	−0.014	0.083	0.041	Amazon add to cart
0.088	0.092	0.087	−0.003	−0.235	−0.126	−0.074	Amazon go to cart
−0.018	−0.0	−0.031	−0.125	−0.139	−0.006	−0.237	Amazon product profile
−0.04	0.0	−0.034	0.173	0.037	0.243	−0.084	Amazon remove from cart
0.019	0.035	0.021	0.147	−0.132	0.067	0.108	Amazon search results
−0.057	−0.08	−0.047	0.026	−0.043	0.047	0.219	Dailyhunt opening news
−0.117	−0.006	−0.058	−0.105	0.104	0.175	0.053	Facebook person profile page
−0.041	0.064	−0.01	−0.194	0.12	0.265	−0.2	Facebook post in a group
−0.073	0.08	−0.007	−0.108	0.24	0.214	−0.007	Facebook search results
0.005	−0.088	−0.062	−0.018	−0.05	−0.064	0.145	Flipkart add to cart
0.062	0.188	0.092	−0.045	−0.022	0.121	0.08	Flipkart go to cart
0.034	0.191	0.109	−0.074	0.006	0.203	0.0	Flipkart product profile
−0.06	0.062	0.017	0.143	0.047	0.017	−0.04	Flipkart remove from cart
−0.058	0.054	0.006	−0.037	0.068	−0.041	−0.169	Flipkart search results
−0.109	0.085	0.034	−0.119	0.126	0.223	0.1	GoogleMaps searching place
0.076	0.107	0.099	0.047	0.06	0.043	−0.041	GoogleNews opening news
0.047	0.069	0.022	−0.054	−0.014	0.013	−0.063	Hotstar open trending
0.167	0.009	−0.007	−0.072	−0.209	−0.221	−0.172	Hotstar play video
−0.013	0.054	0.045	0.286	0.186	0.246	0.169	Hotstar search video
−0.193	−0.137	−0.142	0.162	−0.002	0.053	0.148	LinkedIn my connections
−0.038	−0.011	−0.041	0.137	0.107	0.121	0.199	LinkedIn my profile
−0.047	−0.017	−0.019	0.12	0.125	0.072	0.21	LinkedIn person profile
−0.041	−0.063	−0.072	−0.062	0.011	0.042	−0.024	LinkedIn search results
0.102	0.174	0.213	−0.052	0.032	−0.003	−0.231	Paytm wallet transaction
−0.083	0.045	0.01	−0.069	0.207	0.142	0.012	Telegram send message
−0.067	0.009	−0.032	0.051	0.142	0.046	0.022	Whatsapp send message
0.165	0.185	0.149	0.107	−0.022	0.106	−0.136	Youtube open channel
−0.088	0.034	−0.051	−0.056	−0.015	0.004	−0.031	Youtube play video
0.012	0.118	0.12	0.003	−0.136	0.014	0.172	Youtube search channel
−0.001	0.048	0.1	0.047	−0.129	0.006	0.218	Youtube search video

actions with the three highest values of correlation for each parameter and the ones with the three lowest values in Table 4.3.

We first confirm that the correlations are statistically significant by using the p-values of the Spearman coefficients. We find that our dataset is sufficiently large to ensure that a correlation coefficient greater than 0.1 has a p-value of less than 0.001, indicating that there is less than 0.1% chance of noise showing a positive correlation.

We now make a few observations from Figure 4.1 and Table 4.3. First, we observe that the correlations with the type of network is usually either smaller or close to the correlations with the median RTT. This indicates that the higher

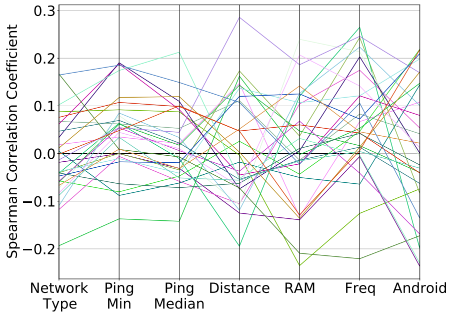


Figure 4.1: A parallel coordinates plot showing the Spearman Correlation Coefficients between the parameters and the response times of each action.

response time using 4G can be explained by the ping RTT values. On the other hand, a total of 6 actions have positive correlation values > 0.1 with minimum ping RTT to Google server, indicating that it is a much better metric to predict response time than the type of network.

A second issue is that there is a very wide divergence of correlation with Android version, both positive and negative. This is counter-intuitive, because we expect newer phones to have higher Android versions and thus faster response times. We hypothesize that this is due to newer app versions downloading more data. We confirm this hypothesis in §4.0.4 by performing controlled app experiments with different app versions on a few apps.

Our third observation is that there are also negative correlation values with RAM for a total of few actions. This is also a counter-intuitive observation. We first rule out the impact of any bias factor like distance or ping RTT's by confirming that the RAM has negligible correlation (< 0.1) with these parameters. We first note the actions that have this property are "Facebook Post in Group", "Flipkart Product Profile", "Facebook Search Results" and "Flipkart Go to Cart" with correlation values of -0.180, -0.137, -0.120 and -0.105. We note that these actions also have small negative correlation values with the ping median RTT's. Facebook is known to support only weak consistency [14], thus indicating that these actions are reported as finished before the data is sent to the server.

Finally, we note that there is a significant positive correlation with distance from a major city as well as with the amount of CPU frequency for many of the actions. While there is also a significant positive correlation with RAM, this is usually weaker than the correlations with CPU frequency. Note that prior studies [11, 15] had shown such correlation with RAM and CPU frequency for mobile web response times, but we confirm that it holds for mobile applications

Table 4.3: App actions that have the highest Spearman correlation for each feature. The numbers in brackets denote the correlation coefficients. In the last two rows, we show the number of features which have significant positive correlation with p-values < 0.001.

Min RTT	Median RTT	Distance	RAM	Frequency	Android Version	Network Type
PayTM Wallet Transaction (0.200)	PayTM Wallet Transaction (0.240)	LinkedIn My Profile (0.241)	Flipkart Add to Cart (0.240)	Hotstar Search Video (0.223)	DailyHunt Opening News (0.435)	Youtube Open Channel (0.18)
Youtube Open Channel (0.174)	Youtube search video (0.187)	Hotstar Search Video (0.241)	LinkedIn Person Profile (0.224)	Amazon Remove from Cart (0.243)	Flipkart Add to Cart (0.264)	GoogleNews Open News (0.146)
GoogleNews Open News (0.171)	Youtube open channel (0.170)	Amazon Remove from Cart (0.237)	LinkedIn My Profile (0.143)	GoogleMaps Search Place (0.223)	LinkedIn My Profile (0.248)	Hotstar Play Video (0.120)
Correlation $>$ 0.1:6	Correlation $>$ 0.1:5	Correlation $>$ 0.1:9	Correlation $>$ 0.1:9	Correlation $>$ 0.1:11	Correlation $>$ 0.1:10	Correlation $>$ 0.1:3
Correlation $>$ 0.15:4	Correlation $>$ 0.15:1	Correlation $>$ 0.15:3	Correlation $>$ 0.15:3	Correlation $>$ 0.15:7	Correlation $>$ 0.15:6	Correlation $>$ 0.15:2

too. Furthermore, we note that distance from major cities have stronger values of correlation than the amount of RAM, thus indicating that it has a stronger effect on response times. We further validate this observation by selecting the 17 volunteers who are over 100km from a major city. We find that in this case, the correlations with distance significantly increase, with a total of 8 out of 30 actions having a correlation higher than 0.2. This validates the recent claims of regional divide that have been made by studies on development economies [16].

4.0.3 Correlation between data size and action response times

In this section we discuss the correlation between data size (number of received bytes per action) and action response time.

We only discuss the relationship of response times with received bytes as transmitted bytes are very less in size, hence not significant. To get an understanding of how these two parameters are correlated, we first plot scatter plots (Figure 5.6-5.12) with regression line, between the data size and the response times to get an idea of how strong or weak the correlation is. We can infer this from the confidence interval of the scatter plots. For this, we removed all outliers from the dataset to get better inferences from scatter plots. Then we compute Spearman Coefficients between data size and action response times which are listed in the Table 4.4. From the p-values and the Correlation Coefficient value we infer that out of all 30 actions, Youtube Search Video, Facebook Post in a Group actions show a significant positive correlation between received bytes and response time.

Table 4.4: Values of Spearman Rank Correlation Coefficients between data size and response time of action.

Data Size(in MBs)	Action
−0.201	Amazon add to cart
−0.204	Amazon go to cart
−0.527	Amazon product profile
0.303	Amazon remove from cart
−0.153	Amazon search results
0.012	Dailyhunt opening news
−0.412	Facebook person profile page
0.583	Facebook post in a group
−0.162	Facebook search results
0.144	Flipkart add to cart
0.155	Flipkart go to cart
0.071	Flipkart product profile
0.044	Flipkart remove from cart
0.040	Flipkart search results
−0.136	GoogleMaps searching place
0.182	GoogleNews opening news
0.050	Hotstar play video
0.148	Hotstar search video
−0.056	LinkedIn my connections
−0.222	LinkedIn my profile
−0.025	LinkedIn search results
0.058	Paytm wallet transaction
0.225	Telegram send message
0.096	Whatsapp send message
−0.080	Youtube play video
0.071	Youtube search channel
0.512	Youtube search video

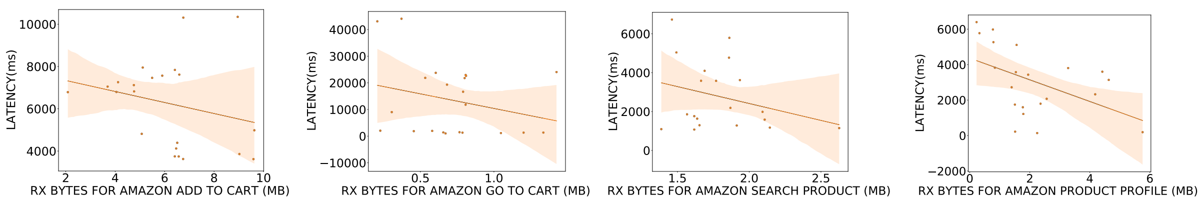


Figure 4.2:

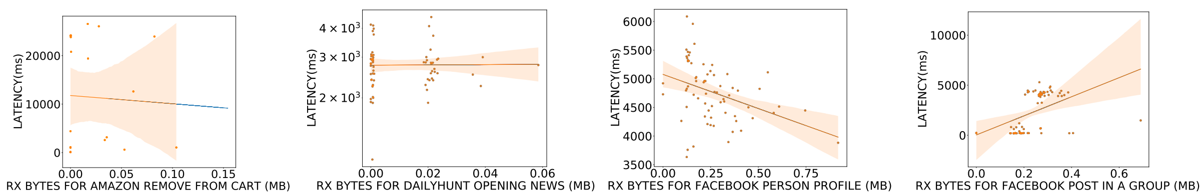


Figure 4.3:

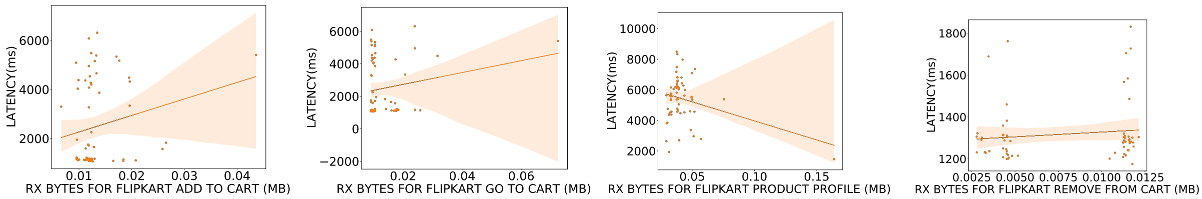


Figure 4.4:

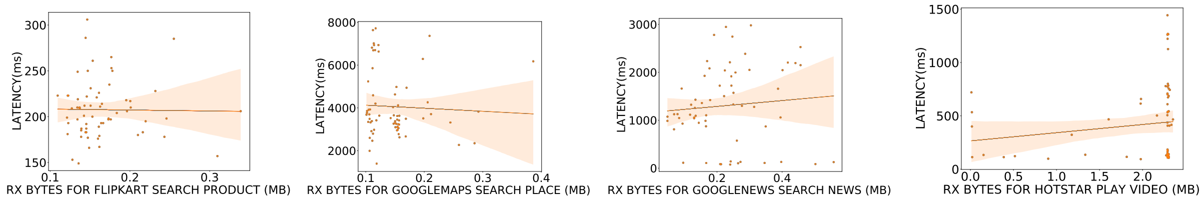


Figure 4.5:

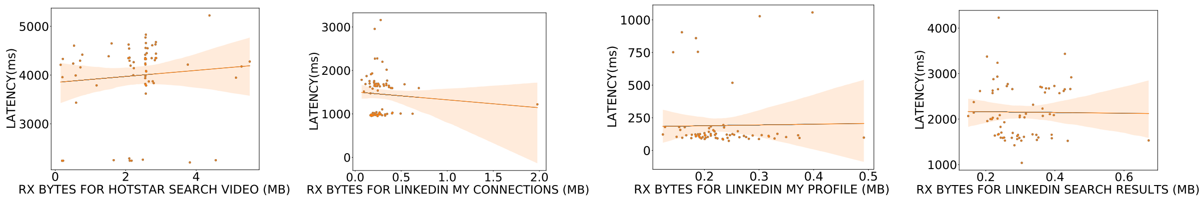


Figure 4.6:

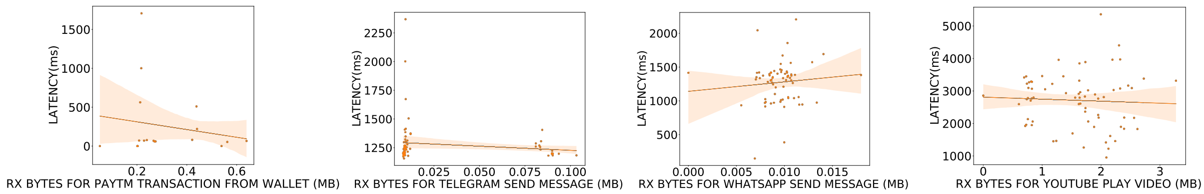


Figure 4.7:

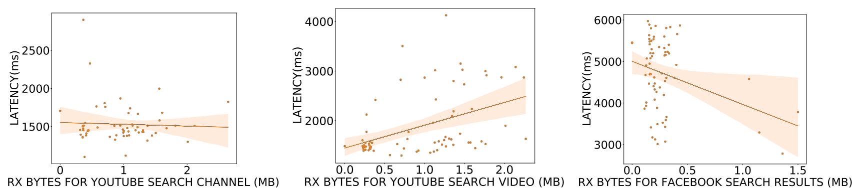


Figure 4.8:

4.0.4 Effect of App Version Upgrades

We further perform controlled lab experiments to analyze the changes in response times with version upgrades. As described in Sec. 2.0.2, we present the results for 4 apps: Google Maps and Flipkart, Youtube, and Telegram. Figure 4.10a) plots response times of searching a location on Google Maps versions *V 9.67.1*, *V 10.8.1* and *V 10.32.2*. We see that both median and 75^{th} percentiles response times have increased significantly from around 3.8s to 6s, with version update from *V 9.67.1* to *V 10.8.1*. With *V 10.32.2*, it is around 4s. To identify the cause, we looked at the transmitted (Tx) and received bytes (Rx) across the versions. We see that both Tx and Rx bytes have increased significantly with version updates; Rx bytes from 2.8MB to around 7.4MB and Tx bytes from 80 KB to 200KB. Figure 4.10 b) shows the response times for viewing the details of a product (product profile) on Flipkart for versions *V 6.10*, *V 6.15* and *V 7.15*. We see that even though the response time values are similar for *V 6.10*, *V 6.15*, the values are significantly higher for *V 7.15*. For example, the median latency has increased from 0.5s to around 2.8s. Similarly, we see that both Tx and Rx bytes have increased significantly with version updates; Rx bytes from 4MB to more than 12MB from V6.10 to V7.15. For YouTube (shown in Figure 4.9 (a)), we have observed a similar relationship i.e., response times have increased with Tx and Rx bytes across version upgrades. Whereas for Telegram (shown in Figure 4.9 (b)), the response times have improved with version upgrades. Hence, we conclude that having a higher app version does not always ensure lower response times. We further note that this is also reflected in the correlation values (shown in the Table 4.2), where Flipkart's search result action has a negative correlation with Android version, but Telegram' send message action has a positive correlation.

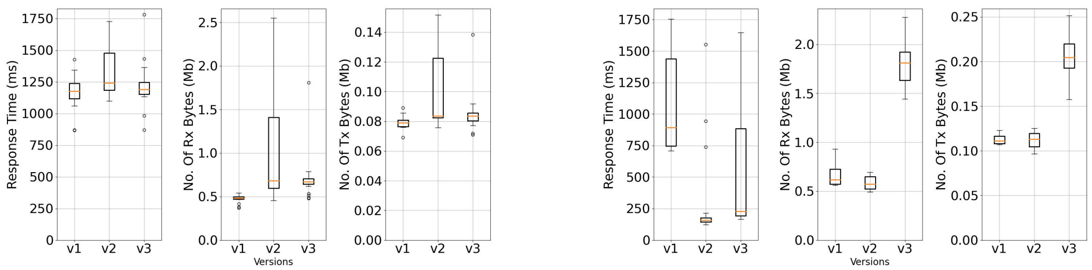


Figure 4.9: Response Times for a) Youtube "Play video" action b) Telegram "Send message" action for three app versions

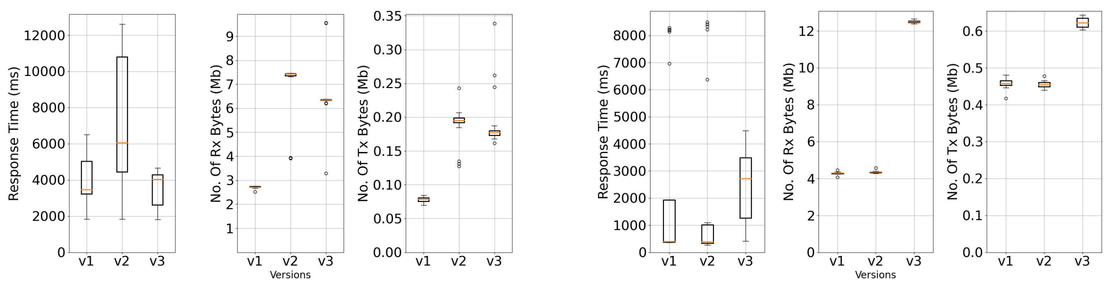


Figure 4.10: Response Times for a) Google Maps "Search a location" action b) Flipkart "Product profile" action for three app versions

We further show the response times, amount of bytes received and sent by Youtube and Telegram for different versions in Figures 4.9(a) and (b). We note that both Youtube have a trend of increase in data consumption with version upgrades. For Youtube, the amount of increase is significant enough to affect the response time, whereas it is not so for Telegram. This is also reflected in the fact that two Youtube's actions, "Search Channel" and "Search video" have significant amount of negative correlation with increase in Android version. On the other hand, Telegram's response time is not strongly correlated with Android version, as its response time does not change significantly.

4.0.5 Effect of memory on response time

In this section we investigate if the device memory impacts the response times. We ensured to eliminate causes like different internet conditions or different time of day at the time of performing experiments by doing the experiments for two consecutive days at the same time of day. Then we first measured the response times in controlled settings on a Lenovo phone with model number *A6020a40* that has a RAM of 2 GB, internal memory of 16 GB and runs Android Oreo *8.0.0* which had some preinstalled apps like email, phone, FM radio, settings, clock, calculator and the apps that we required to conduct the experiments. The device in this setting had almost 1.5 GB free space available on it. Next we installed some more memory intensive apps like Facebook Messenger, Microsoft Teams, Spotify, Adobe Scan, Opera Browser etc., so that 15.5 GB (96.8 %) out of 16 GB of the device memory is utilised and performed the experiments again on the same device to record response times. We only filled it to 15.5 GB as beyond that point a warning message is displayed stating that some functions may not work properly.

Our findings (shown in figure 5.1-5.5) for some actions are very intuitive and for other actions are counter-intuitive. For eg. the response times for actions like Youtube Play Video (Figure 4.15) and Linkedin My Connections (Figure 4.14) increase, as we install more apps. But for actions like Youtube Search Video (Figure 4.15), Hotstar Play Video and Google News Open News (Figure 4.13), the response times decrease. For linkedin my connections and youtube play video median response times are less when less apps are installed and higher when more apps are installed on the device. On the contrary, for google news and hotstar play video and search video the reponse times are lesser when more apps are installed. Thus only 2 actions show decrease in response times when the device has more free memory and 3 actions show the opposite, rest all actions show no change. Hence, there's no fixed pattern as to how the available memory in the device impacts response times. We could not close down the cause for the case where the response times decreased with increase in the number of installed apps.

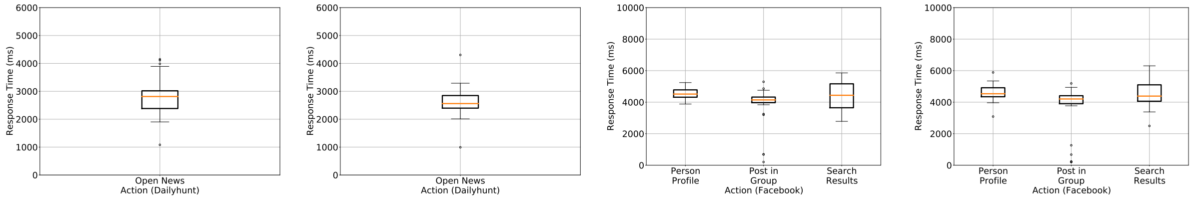


Figure 4.11: Distribution of Response Times when less apps installed for a) Dailyhunt c) Facebook. Distribution of Response Times when more apps installed for b) Dailyhunt d) Facebook.

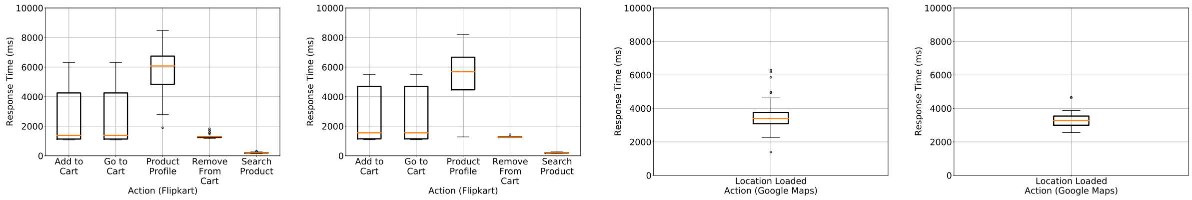


Figure 4.12: Distribution of Response Times when less apps installed for a) Flipkart c) Googlemaps. Distribution of Response Times when more apps installed for b) Flipkart d) Googlemaps.

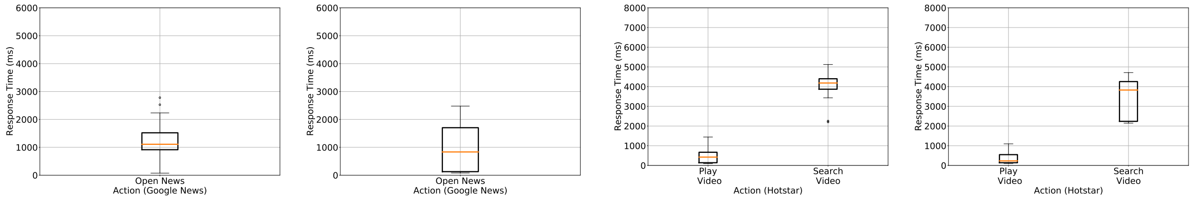


Figure 4.13: Distribution of Response Times when less apps installed for a) GoogleNews c) Hotstar. Distribution of Response Times when more apps installed for b) GoogleNews d) Hotstar.

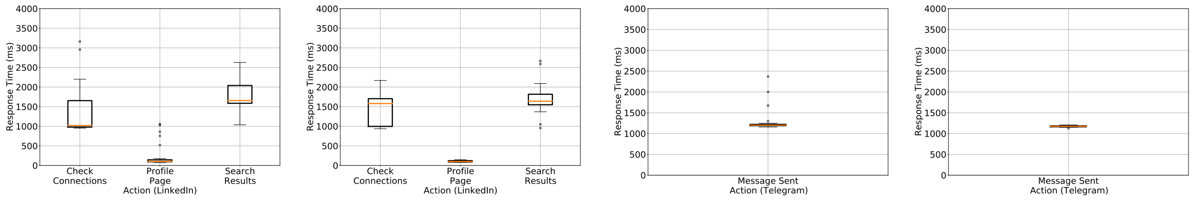


Figure 4.14: Distribution of Response Times when less apps installed for a) LinkedIn c) Telegram. Distribution of Response Times when more apps installed for b) LinkedIn d) Telegram.

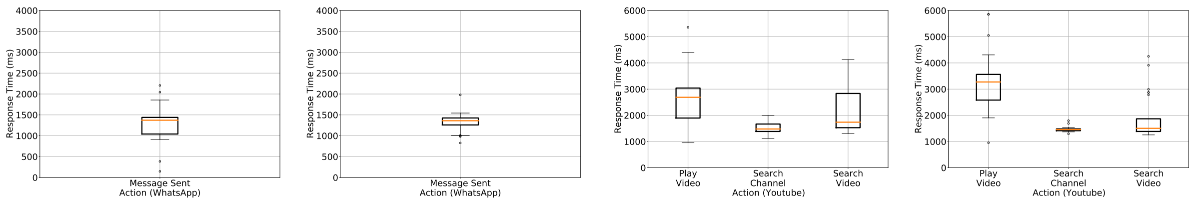


Figure 4.15: Distribution of Response Times when less apps installed for a) Whatsapp c) Youtube. Distribution of Response Times when more apps installed for b) Whatsapp d) Youtube.

4.0.6 Effect of Time of Day on response time

To study the effect of time of day on the response times we performed two sets of controlled experiments, once in the day time(9:00-15:00 IST) and once in the

Table 4.5: Number of Data Points for Morning and Evening Experiments

Time of Day	Number of Data Points
Morning	10,045
Evening	10,817

night time(17:00-23:00 IST). And then we compared response times from the two different datasets we obtained. The number of datapoints for each dataset are shown in Table 4.5. It was seen that only two actions - Flipkart add to cart and Linkedin profile page had less response times when the tests were performed in the morning, than when tests were performed in the evening. Hence, we came to a conclusion that time of day did not impact response times. We say this because there was not a clear pattern in all the apps across all the actions. The related graphs have been shown in figures 5.13-5.18. Thus various app servers have come to give good performance irrespective of the time of day.This also helped us study the effect of congestion during peak hours on response time.

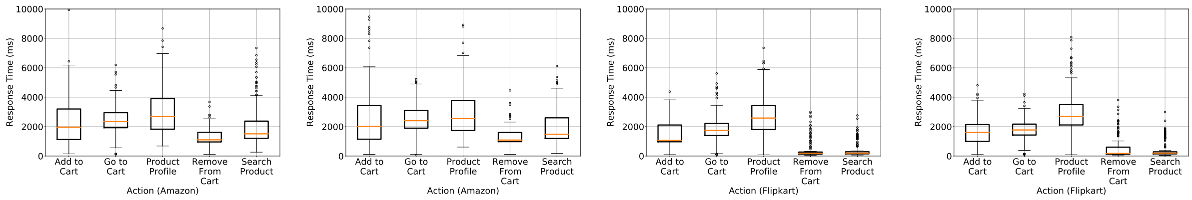


Figure 4.16: Distribution of Response Times for morning run a) Amazon c) Flipkart. Distribution of Response Times for evening run b) Amazon d) Flipkart.

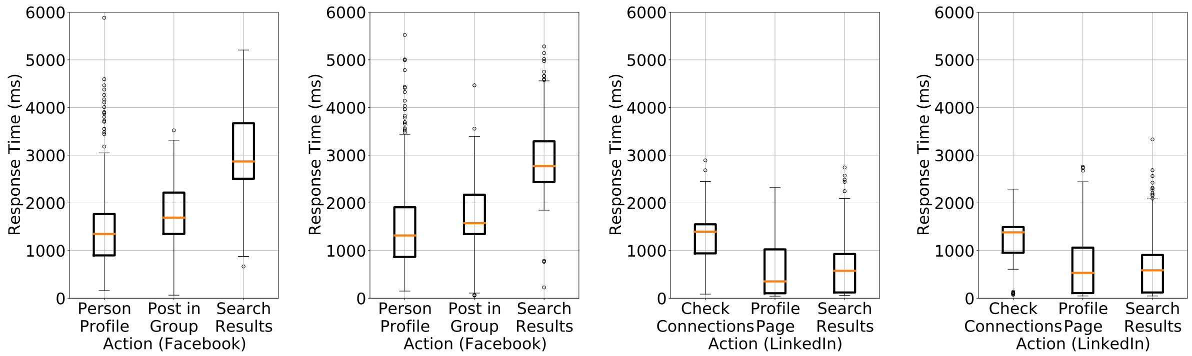


Figure 4.17: Distribution of Response Times for morning run a) Facebook c) Linkedin. Distribution of Response Times for evening run b) Facebook d) Linkedin.

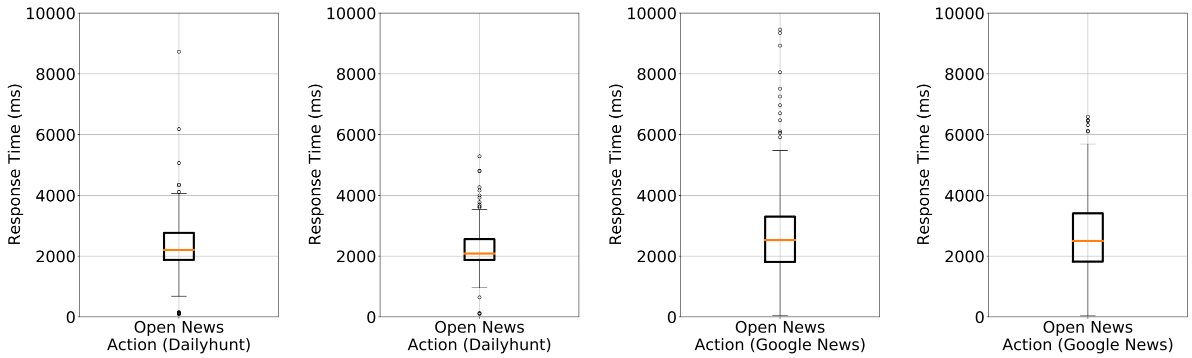


Figure 4.18: Distribution of Response Times for morning run a) Dailyhunt c) Googlenews. Distribution of Response Times for evening run b) Dailyhunt d) Googlenews.

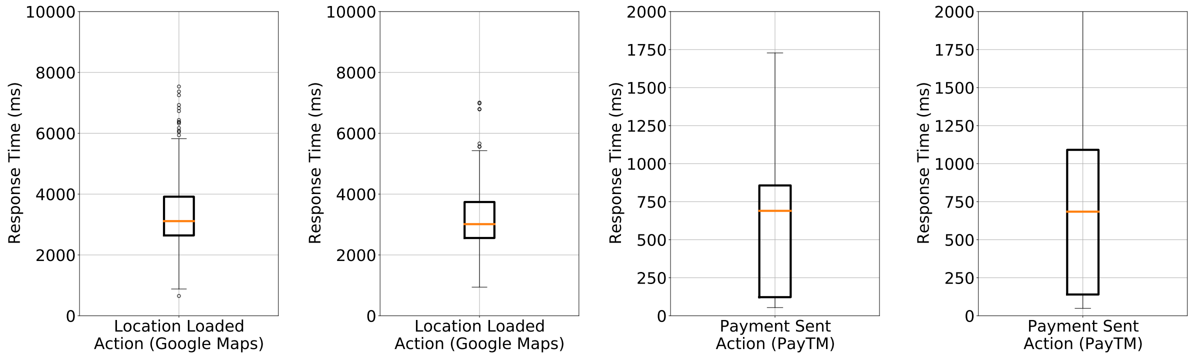


Figure 4.19: Distribution of Response Times for morning run a) GoogleMaps c) Paytm. Distribution of Response Times for evening run b) GoogleMaps d) Paytm.

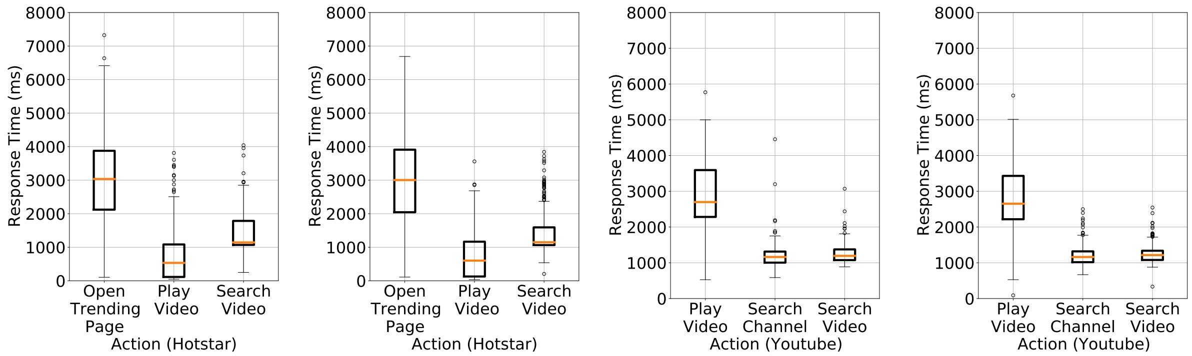


Figure 4.20: Distribution of Response Times for morning run a) Hotstar c) Youtube. Distribution of Response Times for evening run b) Hotstar d) Youtube.

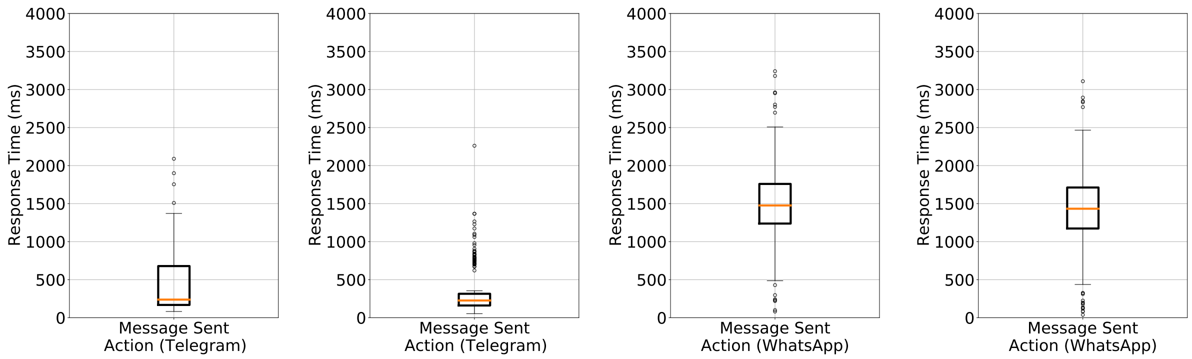


Figure 4.21: Distribution of Response Times for morning run a) Telegram c) Whatsapp. Distribution of Response Times for evening run b) Telegram d) Whatsapp.

Table 4.6: Number of Data Points for different Network Types.

Network Type	Number of Data Points
WiFi	9,550
Mobile Data	9,758
Hotspot	1,554

4.0.7 Effect of Network Type on response time

As mentioned before, users performed the experiments using wifi, mobile network and also using other device's hotspot as a substitute for wifi. We separated the data in these three different categories and plotted response times to see if the network type impacted response times significantly. The number of datapoints for each network type are shown in Table 4.6. We found out that for network type hotspot, the Add to Cart action of Amazon had higher response times compared to WiFi or mobile data. Similarly, for google news the mobile data and hotspot type had a slightly higher median response times with higher variance when compared to WiFi network. In case of Linkedin WiFi outperformed both Hotspot and Mobile Data in having lower response time medians. On the other hand, we have seen that Facebook person profile have lesser median response times in case of Hotspot and mobile data in comparison to WiFi networks. Only slight improvement in response times for wifi network was observed for 3 actions as above, when compared to other two network types. Thus, we could not draw any major conclusions from change in 3 actions out of 28 actions. Hence, we can say that the deployment of 4G has been done very well in the country, giving comparable results to wifi. The related graphs have been shown in figures 5.19-5.24.

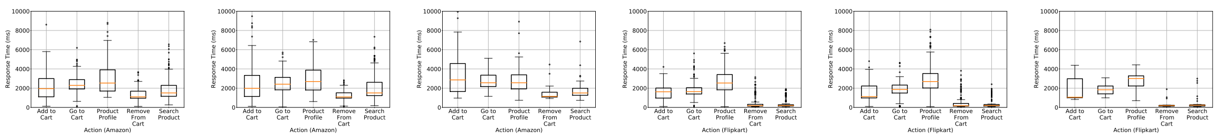


Figure 4.22: Distribution of Response Times for network type wifi a) Amazon d) Flipkart. Distribution of Response Times for network type mobile data b) Amazon e) Flipkart. Distribution of Response Times for network type hotspot c) Amazon f) Flipkart.

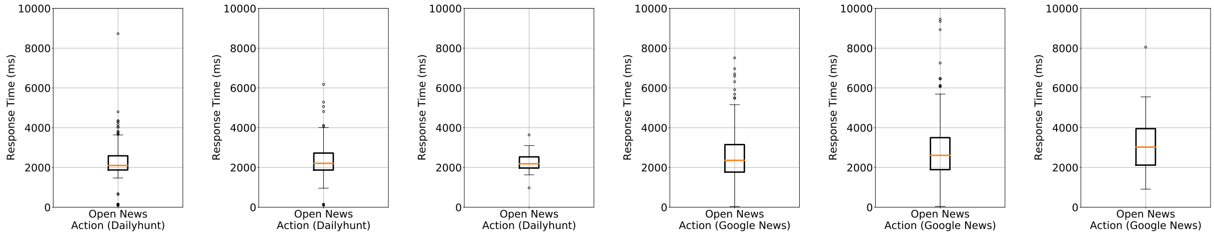


Figure 4.23: Distribution of Response Times for network type wifi a) Dailyhunt d) Google news. Distribution of Response Times for network type mobile data b) Dailyhunt e) Google news. Distribution of Response Times for network type hotspot c) Dailyhunt f) Google news.

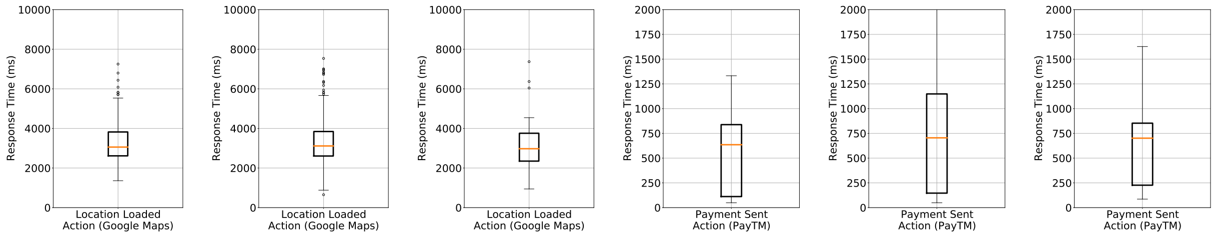


Figure 4.24: Distribution of Response Times for network type wifi a) Googlemaps d) Paytm. Distribution of Response Times for network type mobile data b) Googlemaps e) Paytm. Distribution of Response Times for network type hotspot c) Googlemaps f) Paytm.

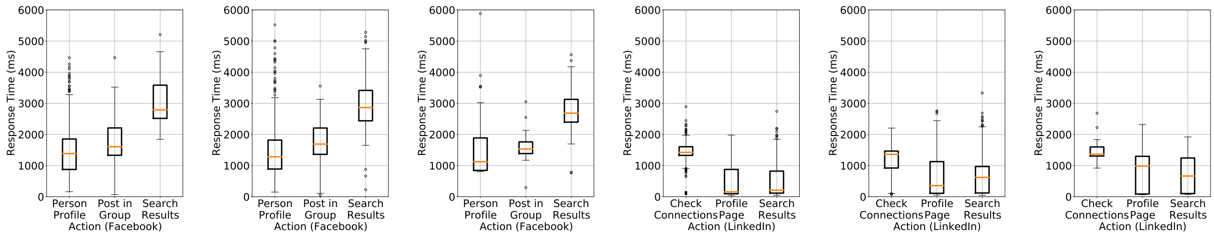


Figure 4.25: Distribution of Response Times for network type wifi a) Facebook d) Linkedin. Distribution of Response Times for network type mobile data b) Facebook e) Linkedin. Distribution of Response Times for network type hotspot c) Facebook f) Linkedin.

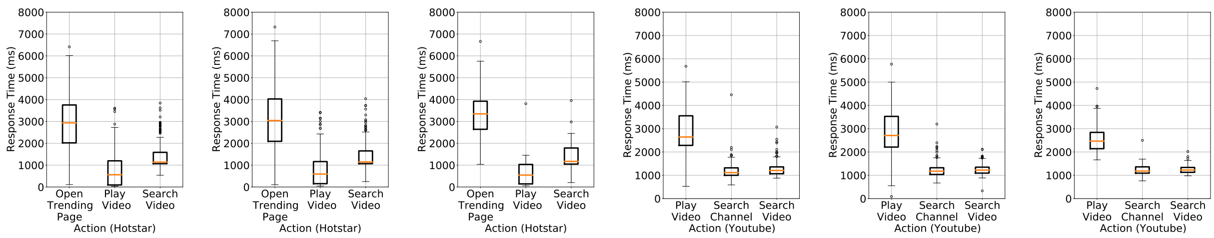


Figure 4.26: Distribution of Response Times for network type wifi a) Hotstar d) Youtube. Distribution of Response Times for network type mobile data b) Hotstar e) Youtube. Distribution of Response Times for network type hotspot c) Hotstar f) Youtube.

4.0.8 Effect of ISPs on response time

The users who performed the experiments had subscribed to various ISPs for cellular connections as well as wifi. The number of datapoints for various ISPs is as mentioned in Table 4.7. We did action wise analysis of all the ISPs to discover

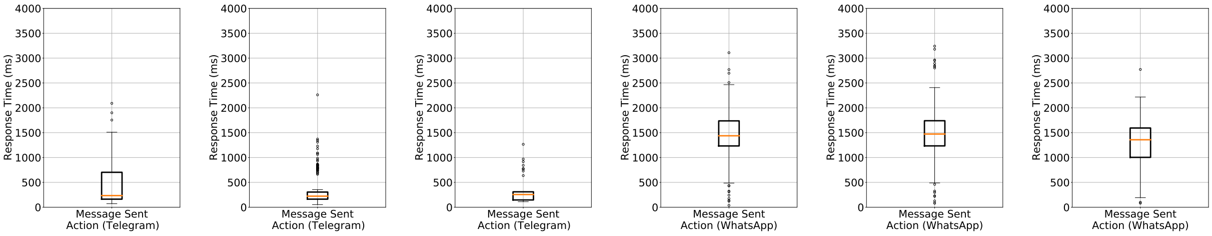
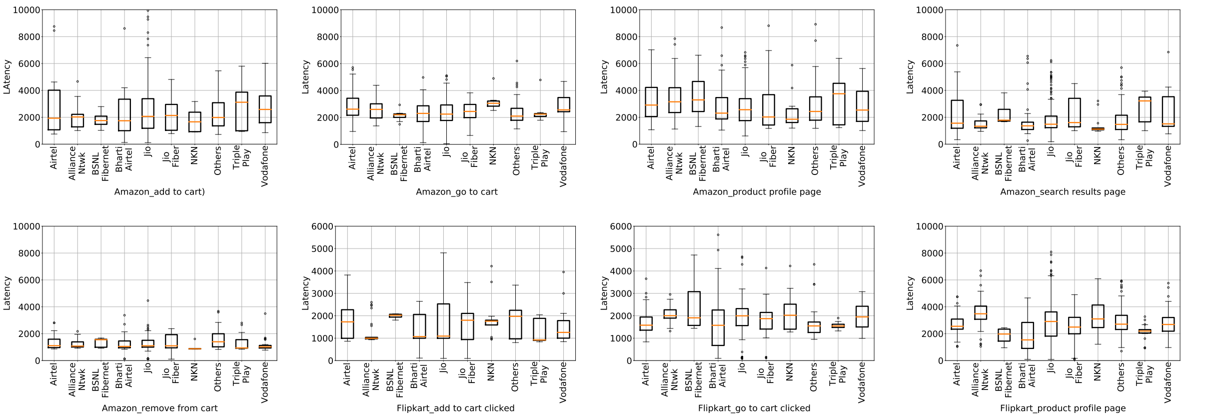


Figure 4.27: Distribution of Response Times for network type wifi a) Telegram d) Whatsapp. Distribution of Response Times for network type mobile data b) Telegram e) Whatsapp. Distribution of Response Times for network type hotspot c) Telegram f) Whatsapp.

Table 4.7: Number of Data Points for different Network Types.

Network Type	Number of Data Points
Airtel	2,632
Allianze Network	1,751
BSNL Fibernet	215
Bharti Airtel	2,043
Jio	6,841
Jio Fiber	1,685
NKN	664
Others	3,250
Triple Play	664
Vodafone	1,117

if any ISP performed significantly better over others in terms of response times. We see that out of 30 actions, in 12 actions NKN performs better than the other service providers. But there are actions like FLipkart add to cart and go to cart where the median response time of NKN is higher than other service providers. The results (shown in figure 5.25) that we got indicate that all ISP perform almost the same. Thus, all The companies are very competitive in providing good services to users.



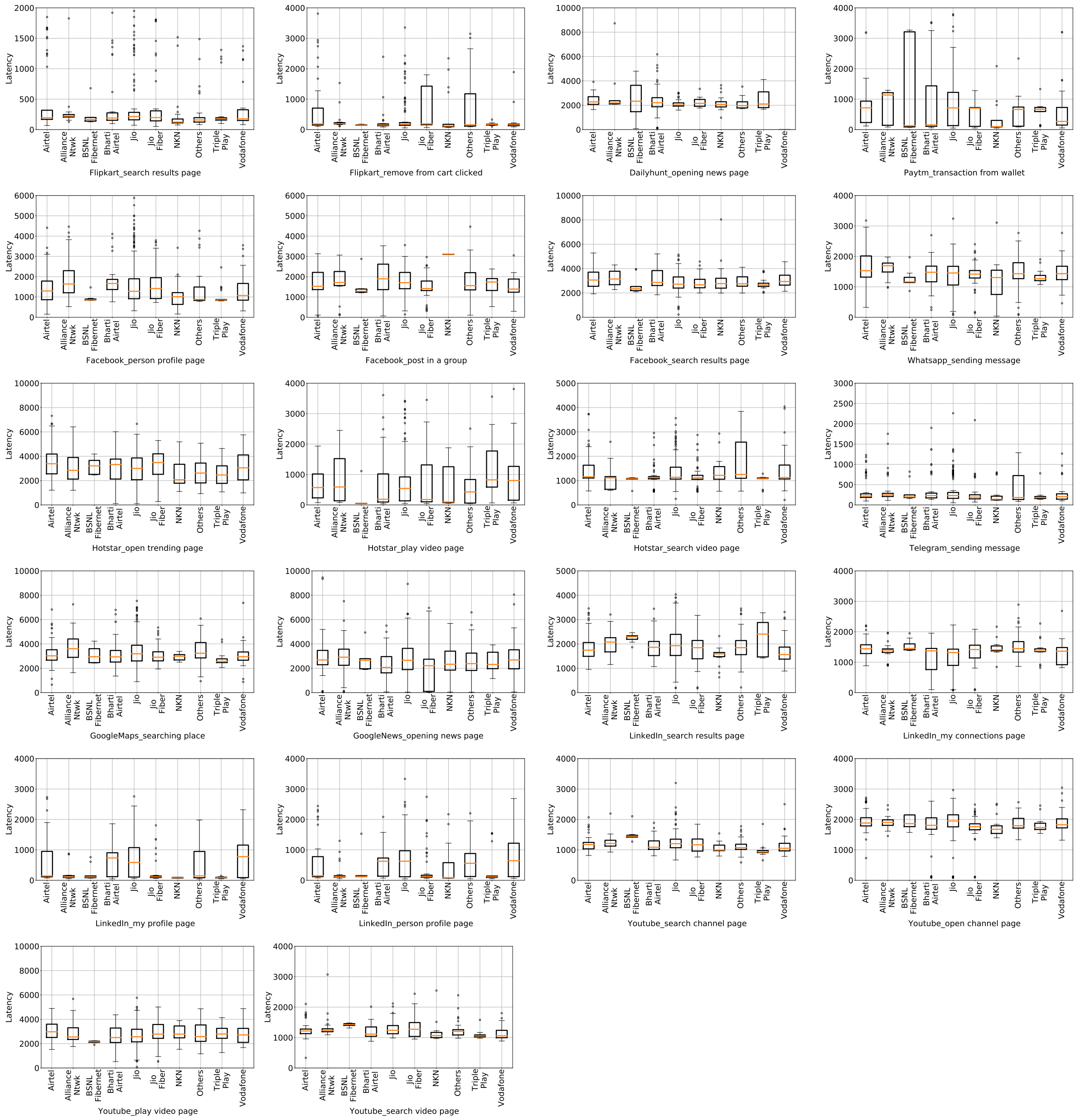


Figure 4.28: Distribution of Response Times according to ISPs for various actions

4.0.9 Identification of Consistent Bias

We have seen in the discussion of §3 that the response time for most apps has a significant amount of divergence. A natural question to ask is that if this divergence is seen equally by all users. We evaluate this question by first computing the 80^{th} percentile of the response times for each action. We then check if the median of any user's actions consistently exceed the 80^{th} percentile of the overall response times. We find that only a total of 3 among the 51 volunteers find their median response times exceeding the 80^{th} percentile for 8 or more ($> 25\%$) actions.

We investigate the reason behind the consistently higher response times of these three users. We found that one of them uses the smartphone Motorola G5, which was released in 2017 and is much older than the rest of the volunteers. A second volunteer who faced consistent higher response times lives far from any major city, with the distance from the nearest major city being 260km. The third volunteer was in a major city and used a relatively higher quality smartphone with RAM of 8GB, but still faced consistently high response times. On further analysis, we found that the user had low ping RTT values than the median, thus ruling out problems with the cellular tower and/or WiFi router. Thus we could not reach a conclusion as to why this user experienced high response times. This shows that large distance and older smartphone can also affect response times. Interestingly, all these users obtained these high response times using both cellular and WiFi network.

Chapter 5

Related Work

Latency Optimization: A number of studies aim to measure and/or optimize the latency of mobile web and apps [3, 11, 17–20]. AppInsight [17] and WProf [11] identify the critical path in the execution of mobile apps and web respectively. The work [21] compares the differences in latency between mobile apps and web apps if they provide the same functionality, and shows that mobile apps tend to have lower latency than web apps in most cases. MopEye [18] identifies the network round-trip time observed by different smartphone apps, and utilizes crowdsourcing to collect large-scale data. The work [19] utilizes MopEye's dataset to identify network anomalies that lead to poor application performance. QoE-Doctor [3], PerfProbe [20] and AppInsight [17] use UI automator provided by the Android framework for analyzing the latency of different apps. Such an approach requires modification of the APK, which makes the users uncomfortable in performing the experiments as they are unaware of the logging time period. Finally, SIF [22] and PUMA [23] design automated techniques of analyzing the app properties and testing the user interface similar to EvalApp, but do not collect crowdsourced data.

Disparities in Global Network Latency: M-Lab [24] performs active tests to identify regional and longitudinal variations in Internet latency across the world. The work [25] identifies the geographical and protocol-level differences that lead to large diversities in the end-user Internet latency. The work [26] quantifies the latency observed from different regions while playing videos over the streaming platform YouTube. Also, [27] identifies the disparities in last-mile latency depending on the time of day, ISP used, and geographical location. We do not focus on network latency directly, but on the response times observed by app users.

Network Characterization in Developing Regions: A few works have focused explicitly on characterizing network performance in developing regions, in places like South Asia and Africa [28–32]. The works [29] and [31] look at the web latencies within Africa and specifically within Ghana respectively. Sharma et al. [28] study the latency observed by Indian cellular networks before the deployment of LTE. WebMedic [32] characterizes the webpages visited by users in developing countries and the smartphones used, and identified memory as the bottleneck. Another work [30] characterizes the smartphone hardware used in a specific district in Pakistan. Our focus is on apps, with the same goal of identifying the problems of smartphone users in developing countries.

Chapter 6

Conclusion

In this paper, we analyzed the response time of popularly used apps in India. We designed a tool EvalApp to automate the collection of response times. We collected response time of a total 30 actions on 12 popularly used apps from 51 volunteers. We observed that response times have high divergence. We performed causal analysis and found that ping round-trip-time to a Google server, distance from a major city, CPU frequency, and RAM of a phone are correlated with the response times. We further observed that 3 volunteers consistently faced higher divergence in response time for most actions. We also observed that the response time of apps do not always improve with version upgrades. Through this study we also concluded that there is no effect of free available memory on the android device, size of transmitted/received bytes for various actions, time of day when experiments were performed, network types (like wifi, mobile data and hotspot) and ISP over the response times for app actions.

Our volunteers share the same demography, drawn from our university's students, living in urban areas. Furthermore, they enjoy good Internet connectivity having attended online classes for over a year. We have focused only on north and central India due to our limited ability to gather volunteers from other regions. Conducting this study across different demography and other regions including remote regions to remove potential bias is left for future work.

Bibliography

[1] I. D. C. (IDC), "India smartphone market registered 42% yoy growth, record october shipments of 21 million units, says idc india," 17 Dec 2020. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prAP47181020

[2] 42matters, "India app market statistics in 2021 for android," May 25, 2021. [Online]. Available: https://42matters.com/india-app-market-statistics

[3] Q. A. Chen, H. Luo, S. Rosen, Z. M. Mao, K. Iyer, J. Hui, K. Sontineni, and K. Lau, "Qoe doctor: Diagnosing mobile app qoe with automated ui control and cross-layer analysis," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC '14, 2014, p. 151–164.

[4] B. Choi, J. Kim, D. Cho, S. Kim, and D. Han, "Appx: An automated app acceleration framework for low latency mobile app," in *Proceedings of the 14th International Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 27–40. [Online]. Available: https://doi.org/10.1145/3281411.3281416

[5] S. Keelery, "Mobile phone internet users in india 2015-2023," Oct 16, 2020. [Online]. Available: https://www.statista.com/statistics/558610/number-of-mobile-internet-user-in-india/

[6] O. Foundation, "Appium mobile automation," 2021. [Online]. Available: https://appium.io/

[7] "Data center map," 2021. [Online]. Available: https://www.datacentermap.com/india/

[8] M. Maity, B. Raman, and M. Vutukuru, "Tcp download performance in dense wifi scenarios: Analysis and solution," *IEEE Transactions on Mobile Computing*, vol. 16, no. 1, pp. 213–227, 2017.

[9] “Samsung galaxy s10e – full phone specifications,” 2019, [accessed 09-Jan-2021]. [Online]. Available: https://www.gsmarena.com/samsung_galaxy_s10e-9537.php

[10] “Infinix hot 9 pro – full phone specifications,” 2020, [accessed 09-Jan-2021]. [Online]. Available: https://www.gsmarena.com/infinix_hot_9_pro-10272.php

[11] J. Nejati and A. Balasubramanian, “An in-depth study of mobile browser performance,” in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW ’16. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016, p. 1305–1315. [Online]. Available: https://doi.org/10.1145/2872427.2883014

[12] O. Foundation, “Javafx library,” 2021. [Online]. Available: https://openjfx.io/

[13] “India’s digital divide: How broadband speed splits the nation,” 2018. [Online]. Available: https://www.speedtest.net/insights/blog/india-state-city-broadband-speeds-2018/

[14] H. Lu, K. Veeraraghavan, P. Ajoux, J. Hunt, Y. J. Song, W. Tobagus, S. Kumar, and W. Lloyd, “Existential consistency: Measuring and understanding consistency at facebook,” in *Proceedings of the 25th Symposium on Operating Systems Principles*, ser. SOSP ’15. New York, NY, USA: Association for Computing Machinery, 2015, p. 295–310. [Online]. Available: https://doi.org/10.1145/2815400.2815426

[15] M. Dasari, S. Vargas, A. Bhattacharya, A. Balasubramanian, S. R. Das, and M. Ferdman, “Impact of device performance on mobile internet qoe,” in *Proceedings of the Internet Measurement Conference 2018*, ser. IMC ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–7. [Online]. Available: https://doi.org/10.1145/3278532.3278533

[16] T. Agarwal and P. K. Panda, “Pattern of digital divide and convergence in access to ict facilities among the indian states,” *Journal of Infrastructure Development*, vol. 10, no. 1-2, pp. 37–51, 2018.

[17] L. Ravindranath, J. Padhye, S. Agarwal, R. Mahajan, I. Obermiller, and S. Shayandeh, “Appinsight: Mobile app performance monitoring in the wild,” in *10th USENIX Symposium on Operating Systems Design and Implementation (OSDI 12)*. Hollywood, CA: USENIX Association,

Oct. 2012, pp. 107–120. [Online]. Available: https://www.usenix.org/conference/osdi12/technical-sessions/presentation/ravindranath

[18] D. Wu, R. K. C. Chang, W. Li, E. K. T. Cheng, and D. Gao, "Mopeye: Opportunistic monitoring of per-app mobile network performance," in *2017 USENIX Annual Technical Conference (USENIX ATC 17)*. Santa Clara, CA: USENIX Association, Jul. 2017, pp. 445–457. [Online]. Available: https://www.usenix.org/conference/atc17/technical-sessions/presentation/wu

[19] S. Zhang, W. Li, D. Wu, B. Jin, R. K. Chang, D. Gao, Y. Wang, and R. K. P. Mok, "An empirical study of mobile network behavior and application performance in the wild," in *2019 IEEE/ACM 27th International Symposium on Quality of Service (IWQoS)*, 2019, pp. 1–10.

[20] D. K. Hong, A. Nikravesh, Z. M. Mao, M. Ketkar, and M. Kishinevsky, "Perfprobe: A systematic, cross-layer performance diagnosis framework for mobile platforms," in *2019 IEEE/ACM 6th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, 2019, pp. 50–61.

[21] Y. Ma, X. Liu, Y. Liu, Y. Liu, and G. Huang, "A tale of two fashions: An empirical study on the performance of native apps and web apps on android," *IEEE Transactions on Mobile Computing*, vol. 17, no. 5, pp. 990–1003, 2018.

[22] S. Hao, D. Li, W. G. Halfond, and R. Govindan, "Sif: A selective instrumentation framework for mobile applications," in *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 167–180. [Online]. Available: https://doi.org/10.1145/2462456.2465430

[23] S. Hao, B. Liu, S. Nath, W. G. Halfond, and R. Govindan, "Puma: Programmable ui-automation for large-scale dynamic analysis of mobile apps," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 204–217. [Online]. Available: https://doi.org/10.1145/2594368.2594390

[24] T. Høiland-Jørgensen, B. Ahlgren, P. Hurtig, and A. Brunstrom, "Measuring latency variation in the internet," in *Proceedings of the 12th International*

on Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '16, 2016, p. 473–480.

[25] I. N. Bozkurt, A. Aguirre, B. Chandrasekaran, P. B. Godfrey, G. Laughlin, B. Maggs, and A. Singla, "Why is the internet so slow?!" in *Passive and Active Measurement*, M. A. Kaafar, S. Uhlig, and J. Amann, Eds. Cham: Springer International Publishing, 2017, pp. 173–187.

[26] T. V. Doan, L. Pajevic, V. Bajpai, and J. Ott, "Tracing the path to youtube: A quantification of path lengths and latencies toward content caches," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 80–86, 2019.

[27] V. Bajpai, S. J. Eravuchira, and J. Schönwälder, "Dissecting last-mile latency characteristics," *SIGCOMM Comput. Commun. Rev.*, vol. 47, no. 5, p. 25–34, Oct. 2017. [Online]. Available: https://doi.org/10.1145/3155055.3155059

[28] A. Sharma, M. Kaur, Z. Koradia, R. Nishant, S. Pandit, A. Raman, and A. Seth, "Revisiting the state of cellular data connectivity in india," in *Proceedings of the 2015 Annual Symposium on Computing for Development*, ser. DEV '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 149–157. [Online]. Available: https://doi.org/10.1145/2830629.2830649

[29] R. Fanou, G. Tyson, E. L. Fernandes, P. Francois, F. Valera, and A. Sathiaseelan, "Exploring and analysing the african web ecosystem," *ACM Trans. Web*, vol. 12, no. 4, Sep. 2018. [Online]. Available: https://doi.org/10.1145/3213897

[30] S. Ahmad, A. L. Haamid, Z. A. Qazi, Z. Zhou, T. Benson, and I. A. Qazi, "A view from the other side: Understanding mobile phone characteristics in the developing world," in *Proceedings of the 2016 Internet Measurement Conference*, ser. IMC '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 319–325. [Online]. Available: https://doi.org/10.1145/2987443.2987470

[31] Y. Zaki, J. Chen, T. Pötsch, T. Ahmad, and L. Subramanian, "Dissecting web latency in ghana," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 241–248. [Online]. Available: https://doi.org/10.1145/2663716.2663748

[32] U. Naseer, T. A. Benson, and R. Netravali, "Webmedic: Disentangling the memory-functionality tension for the next billion mobile web users," in *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, ser. HotMobile '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 71–77. [Online]. Available: https://doi.org/10.1145/3446382.3448652

[33] S. Zhuang, J. H. Wang, P. Zhang, and J. Wang, "Understanding the latency to visit websites in china: An infrastructure perspective," *Computer Networks*, vol. 169, p. 107102, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128619308722

[34] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257

[35] M. Rajiullah, A. Lutu, A. S. Khatouni, M.-R. Fida, M. Mellia, A. Brunstrom, O. Alay, S. Alfredsson, and V. Mancuso, "Web experience in mobile networks: Lessons from two million page visits," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1532–1543. [Online]. Available: https://doi.org/10.1145/3308558.3313606

[36] V. Lalwani, "Tech savvy indians drive to villages for covid-19 vaccinations. those without smartphones lose out," May 17, 2021. [Online]. Available: https://scroll.in/article/994871/tech-savvy-indians-drive-to-villages-for-covid-19-vaccinations-those-without-smartphones-lose-out

[37] J. Nejati, M. Luo, N. Nikiforakis, and A. Balasubramanian, "Need for mobile speed: A historical analysis of mobile web performance," 2020.