Inside the Walled Garden: Deconstructing Facebook's Free Basics Program

Rijurekha Sen† Sohaib Ahmad△ Amreesh Phokeer◇ Zaid Ahmed Farooq△
Ihsan Ayyub Qazi△ David Choffnes‡ Krishna P. Gummadi†
†Max Planck Institute for Software Systems (MPI-SWS) ‡Northeastern University
△LUMS ◇University of Cape Town

ABSTRACT

Free Basics is a Facebook initiative to provide zero-rated web services in developing countries. The program has grown rapidly to 60+ countries in the past two years [14]. But it has also seen strong opposition from Internet activists and has been banned in some countries like India [4, 11, 12, 22]. Facebook highlights the societal benefits of providing low-income populations with free Internet access, while detractors point to concerns about privacy and network neutrality.

In this paper, we provide the first independent analysis of such claims regarding the Free Basics service, using both the perspective of a Free Basics service provider and of web clients visiting the service via cellular phones providing access to Free Basics in Pakistan and South Africa.

Specifically, with control of both endpoints, we not only provide a more detailed view of how the Free Basics service is architected [13], but also can isolate the likely causes of network performance impairments. Our analysis reveals that Free Basics services experience 4 to 12 times worse network performance than their paid counterparts. We isolate the root causes using factors such as network path inflation and throttling policies by Facebook and telecom service providers.

The Free Basics service and its restrictions are designed with assumptions about users' device capabilities (e.g., lack of JavaScript support). To evaluate such assumptions, we infer the types of mobile devices that generated 47K unique visitors to our Free Basics services between Sep 2016 and Jan 2017. We find that there are large numbers of requests from constrained WAP browsers, but also large fractions of high-capability mobile phones that send Free Basics requests. We discuss the implications of our observations, with the hope to aid more informed debates on such telecom policies.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

1. INTRODUCTION

Facebook started the Free Basics program in 2015 in collaboration with cellular providers in some developing countries [9]. Subscribers of these telecom providers can access a set of web services on their mobile phone browsers, or via an Android app [5], without incurring data charges. Over the last two years, the program has grown to 60+ countries across Asia, Africa, South and Central America [14], with 25 new countries added since May 2016. Facebook claims that their goal with Free Basics is to bring more people online, in an effort to curb the digital divide [2].

The program has been strongly opposed by Internet activists, who have raised concerns about (i) lack of data privacy from Facebook, which maintains the proxies through which all Free Basics web requests and responses flow [7] and (ii) unfairness against paid web services losing users to the Free Basics services [6]. Such debates have paved the way for telecom regulators to ban this program in India [4, 12]. In countries with Free Basics, there are additional concerns about whether the program is actually used [8], and whether the users are really first-time Internet users as Facebook claims [3].

While there is no shortage of bluster, there is a paucity of independent, empirical analysis to evaluate the above claims. In our prior work [27], we made initial, preliminary observations about available Free Basics services and their network quality of service (QoS), using client-side measurements in Pakistan and South Africa. Importantly, we did not conduct a data-driven analysis of Facebook's claims about Free Basics users, data privacy, or network neutrality. In this paper, we develop new methodologies and conduct new analysis to address these topics.

Specifically, we implement our own web services and deploy them as part of the Free Basics program. Leveraging this additional vantage point of the web server under control, we perform careful experiments to identify multiple causes for the network performance gap between paid web services and zero-rated Free Basics services. We find there is (i) substantially higher latency along the paths through the Facebook proxies than via the direct paths, (ii) throttling at Facebook proxies that limit Free Basics traffic to 150 Kbps and (iii) different traffic differentiation policies by individual cellular providers, causing 6 times lower client-side throughput in two Pakistani providers. Section 4 discusses these experimental observations, along with details of the Facebook proxy network, caching, and data encryption policies.

Using the deployed services, we also characterize the mobile device capabilities of 47K unique visitors who visited our services between Sep 2016 and Jan 2017 (Section 5). Such analysis gives some indication of the socio-economic background of these visitors, whom Facebook claims to be poor first time Internet users in the developing countries.

We have communicated these observations to the Facebook Free Basics team. Further to help inform the public debate and allow others to repeat our experiments, we make all of our code and data publicly available[1].

We discuss these interactions with Facebook and the implications of our observations in context of the Free Basics debate in Section 6, and conclude the paper in Section 7.

2. RELATED WORK

Several related studies evaluate network performance in developing countries. Some found that DNS servers and a lack of good caching infrastructure are the primary causes of poor performance in some regions [19,32]. Another study has shown CDN server placements and routing protocols as primary performance bottlenecks [28]. To provide Internet connectivity in these environments, there are several efforts that include building low cost network infrastructure (e.g., using long distance WiFi [25] and software cellular base stations [18]), developing low cost data communication channels (e.g., using SMS or voice) [21], deploying specialized web proxies for developing countries [30], and customizing applications for low-end feature phones [24, 26].

Unlike previous work, this paper measures network QoS in developing regions *in the context of Facebook's Free Basics* program. Given that this particular program has been deployed in 60+ developing countries, some of them with the world's highest population densities [15], identifying such limitations potentially impacts large numbers of users.

Molavi et al. [23] measured traffic differentiation by cellular providers for normal web services; we apply similar methodologies to identify such practices for Free Basics proxies and Free Basics cellular providers. Similarly, our approach of using clients and web servers that we control was explored previously for debugging middlebox and proxy behaviors [31].

The topic of how users interact with zero-rated services in developed and developing regions was the focus of work by Chen et al. [20], which used client-side measurements and surveys to inform their analysis. Unlike this study, we use a web server as our vantage point and focus specifically on understanding whether Free Basics users are in fact low socioeconomic status populations based on the cost of mobile devices that they use to access our service. We use the same methodology as in our previous work [17] for mapping devices to cellular capabilities using online phone databases.

[1]https://bitbucket.org/rijurekha/freebasics_ccr

3. SERVICE DEPLOYMENT

To deconstruct Free Basics from within its walled garden, we developed our own web services and deployed them on the Free Basics platform. This deployment serves two purposes. First, we can use the services as vantage points to study the server-side architecture of Free Basics. As we show in Section 4, this helps us to identify the root causes for the differentiated network performance observed in our preliminary prior work [27]. Secondly, we can also study the Free Basics users who visit the site (which was done with IRB approval), espcially to analyze their mobile device characteristics as a possible indicator of the users' socio-economic backgrounds.

Bugle News. We built an RSS aggregator service called Bugle News[2] that fetches RSS feeds from news organizations including BBC, CNN, and Reuters, and provides users with corresponding headlines/ledes. The news stories are organized by topic and country. The service was offered in English between September 17th and December 15, 2016 and has been available in English, French and Spanish since December 16, 2016.

Learn Basics. We built an educational service called Learn Basics[3] that publishes free English-language and Mathematics educational material made available under the Creative Commons license. This service has been offered in English since July 2, 2016.

4. NETWORK CHARACTERIZATION

Net neutrality has been one of the primary points of opposition of the Free Basics program [6]. There have been concerns that the Free Basics services would have an unfair advantage over normal paid web services. In addition, the network quality of service (QoS) afforded to Free Basics services has important implications. Low QoS might reduce the appeal of the free content and cause users to disengage with certain services. It can also create a poor Internet experience for users coming online for the first time.

We made some preliminary observations about the network QoS differences between Free Basics and paid versions of the same service like BBC in our prior work [27]. However, whether it was caused by Facebook's Free Basics architecture or throttling policies of the cellular provider bearing the expenses of the Free Basics program, remained open questions. In this section, we isolate the root causes of Free Basics service performance by measuring the Facebook proxy architecture, caching policies, and network performance across different cellular providers.

4.1 Data Collection Methodology

In the Free Basics architecture [13, 27], mobile clients send requests to web services, via Facebook's proxies. In our experiments to examine network QoS, we control two vantage points in this architecture.

[2]http://buglenews.mpi-sws.org/

[3]http://learnbasics.mpi-sws.org/

Mobile clients: On the client side, we build the same experimental testbed as described in our prior work [27]. We use mobile phones with Free Basics SIM connections in Pakistan and South Africa. The phone is set up as a Wi-Fi hotspot and a laptop is tethered to it. The laptop has a separate ethernet connection for remote access.

Scripts are run on the laptop, to crawl Free Basics and paid versions of the same web service. The crawler uses the laptop's Wi-Fi connection, which in turn uses the phone's cellular connection for Internet access. No other devices connect to the phone hotspot. The tethered connection can support up to 14 Mbps download and 2.5 Mbps upload speeds, as tested with speedtest.net. The cellular connection data rates on the phones are much lower than this, hence the tethered connection does not form a bottleneck in our testbed. We log packet traces, for offline analysis of client side network performance.

We also use SIM cards from different cellular providers, namely Telenor and Zong in Pakistan. The goal is to compare network performance for the same server-client locations, for different service providers. We will refer to our two mobile clients as PK (client in Pakistan) and SA (client in South Africa) respectively.

Web server: Our Bugle News and Learn Basics servers are the other type of vantage points in our experimental setup. Client scripts crawl the content hosted at these servers, predominantly Learn Basics. Learn Basics has static content, which helps us rerun the same experiment at various times of the day over different days. This is necessary to get statistically significant results, but would be difficult in Bugle News with its dynamic news content.

We host the server primarily in Germany, but also move it periodically to other locations using Amazon EC2 hosting facilities and DNS redirection. The goal is to understand how the Facebook proxy path changes, based on different server-client geographical locations.

4.2 Facebook Proxies

Hosting Learn Basics in Virginia, Sao Paolo, Mumbai, Tokyo and Sydney, we crawl the site 20 times over 2 days, from each mobile client. This generates 1,000 HTTP requests for each of the ten client-server pairs. We collect traces at both client and server side and extract the IP addresses for the HTTP requests. We geo-locate these IP addresses by inferring information gathered from *whois* and geographic information encoded in PTR records for IP addresses near the server found via traceroute.

Architecture. Our first observation about the Facebook proxy architecture is that the IP address to which our mobile client sends HTTP requests is not the same as the IP address which sends HTTP requests to our web server. Both these IP addresses are owned by Facebook, indicating that Free Basics requests traverse at least two Facebook proxies between the client and server.

Table 1 shows the geographical locations of the different Free Basics network entities. We use "FB C-Proxy" to indicate the Client-side proxy that receives requests from mobile clients. Similarly, we refer to the proxy that contacts our Web server as the "FB S-Proxy." We find that each of our mobile clients in a given location sends 95% of the requests to a single FB C-Proxy, which is labeled as "primary" in the table. For the PK client, this is in Frankfurt and for the SA client, it is in London. FB S-Proxy IPs are geolocated to Facebook data centers, either in Lulea, Sweden or at Prineville in Oregon, USA.

Network entity	Geographical locations
Mobile clients	Pakistan, South Africa
Web servers	Germany, Virginia, Sao Paolo, Mumbai, Tokyo, Sydney
FB C-Proxy	London (primary for SA), Frankfurt (primary for PK), Marseille, Paris, Singapore, Los Angeles
FB S-Proxy	Lulea (Sweden) Prineville, OR (USA)

Table 1: Geographical locations of network entities.

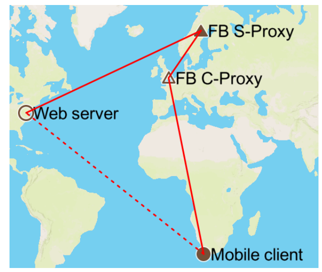

Figure 1: An example of the Free Basics proxy locations for a given pair of client/server location in our experiments.

The server locations used in the Facebook proxy architecture remained stable during the measurement period and were independent of the location of the web server. Figure 1 shows an instance of the Free Basics network entities during our experiment. The Free Basics path through proxies is shown in solid lines. The direct path, used for non Free Basics web requests, is shown with dotted lines. We measure ping delays along both paths and analyze network path inflations and associated latencies next.

Path latencies and inflation. We compared direct-path RTT latency from our client to our server with the measureable latencies along the Free Basics proxy paths. Figure 2 and Figure 3 show the CDF of ping delays for the PK and SA clients respectively. The solid lines denote the sum of the ping latencies from (i) the mobile client to FB C-Proxy and (ii) FB S-Proxy to the web server. There may be more machines between these two identified proxies. Thus this partial path latency is a lower bound on the actual Free Basics path latency. We call this lower bound latency as "FRB" in the figure. The dotted lines denote the ping delays between the mobile client and the web server and we call this "direct".

For the PK client with web servers at Sydney and Sao Paolo, direct latencies are larger than FRB in 30% to 50% cases. Recall that the FRB path is *lower bound* on the end-to-end latency through the Free Basics proxies because we

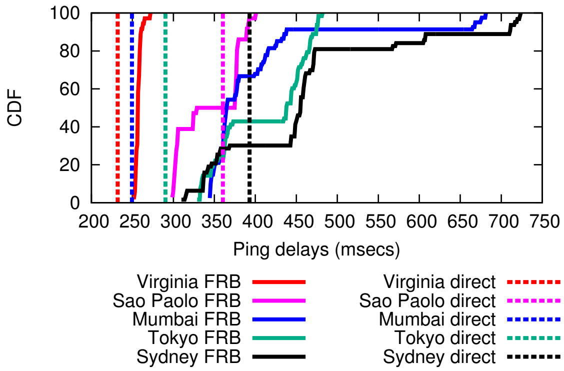

Figure 2: Ping delays for different web server locations for PK client.

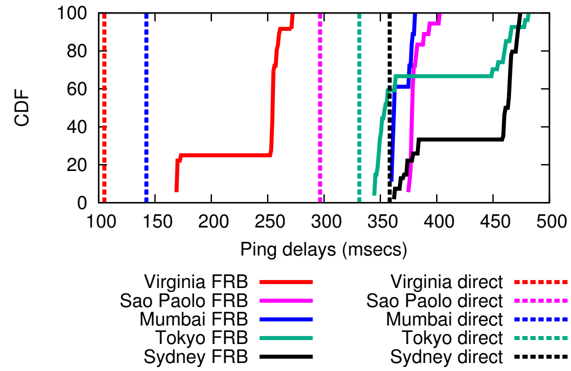

Figure 3: Ping delays for different web server locations for SA client.

cannot measure the segment between the C-Proxy and S-Proxy.[4] For the specific cases where direct latencies are large, we find that the FB S-Proxy is in the US and the FB C-Proxy is in Frankfurt. The latency between the Frankfurt and the US proxies ($\approx$190 ms), if added to FRB, will make the end-to-end Free Basics latency higher than direct.

In all other cases, direct latencies are smaller than the FRB latencies. The difference varies based on the server location. For the PK client, the difference is small for Virginia and Sao Paolo, moderate for Sydney and high for Mumbai and Tokyo. For the SA client, Tokyo and Sydney see lower differences, while Sao Paolo, Mumbai and Virginia see increasingly higher differences. *In summary, path inflation is pervasive in Free Basics and Facebook makes no observable attempt to optimize it based on the relative locations of clients and servers.*

An important question is whether such path inflation explains the reduced performance on Free Basics we observed in our prior work [27]. In the next section, we control for the proxy path latencies such that they are identical to those along the direct path. As we show, path inflation alone does not explain performance differences.

4.3 Throttling Policies

We now determine whether page-download performance is limited by anything other than end-to-end delay. We use the Learn Basics server running in Germany and the PK client. As the Facebook proxies are in Frankfurt and Sweden for this server-client pair, the latency difference between direct and proxy paths is minimal.

We create a 3 MB HTML file and include it using a hidden link on the Learn Basics web page. Facebook periodically crawls sites to determine which URLs to whitelist for availability via Free Basics. If a URL is not linked directly from a page already publicly available from the homepage, then it is not allowed. Thus, to use a large file to test bandwidth, we need to make it linked from a public page. However, we do not want users to download such a large file. To avoid this, we make the link hidden to the user, but visible for our crawler.

We fetch this object 300 times each for paid and Free Basics connections, using a Telenor SIM, interleaving the Free Basics and paid requests using identical HTTP headers. We run the same experiment for another cellular provider Zong, on the same day. We repeat the whole experiment the next day in reverse order, with Zong used first, followed by Telenor.

We calculate throughputs using packet traces gathered both at the server and the client. We plot the following CDF curves for Telenor (Figure 4) and Zong (Figure 5):

- **FRB client:** between client and FB C-Proxy
- **FRB server:** between FB S-Proxy and server
- **NFRB client:** between client and cellular provider
- **NFRB server:** between cellular provider and server

The first observation is the gap between the red solid and blue solid lines, showing that FRB clients see substantially lower average throughput compared to NFRB clients. This is what we also observed in our prior work [27]. *When using the same SIM card, requests for paid content gets 4x-12x higher throughput than for zero-rated content.*

The second interesting observation is the gap between the red dotted and blue dotted lines, which are the FRB and NFRB server side throughputs. The FB S-Proxy appears to self-throttle throughput at 150 Kbps, while the NFRB throughput peaks at 450 and 550 Kbps for Zong and Telenor respectively. Thus server-side throughput is lower for FRB than paid, giving one cause of throughput differences perceived by clients.

Another interesting observation comes from comparing the red solid lines between the two graphs. It indicates that, in addition to the server side self-throttling by FB S-Proxy, there can be additional throttling along the path from the FB S-Proxy to the client. Specifically, the client-side median throughputs are 120 Kbps for Telenor and only 20 Kbps for Zong. In this case, the bottleneck can be between the FB S-Proxy and FB C-Proxy, and/or between the FB C-Proxy and the client. It is highly unlikely that the network is the

[4] Note that this segment may itself contain one or more additional proxies, but we do not have sufficient visibility to determine whether this is the case.

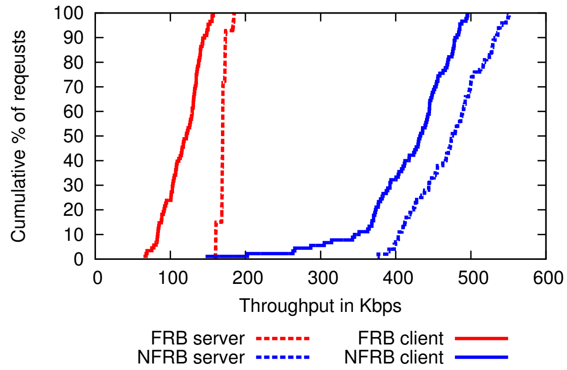

Figure 4: Telenor throughput along alternative paths to our service, indicating throttling near 120 Kbps (solid red line).

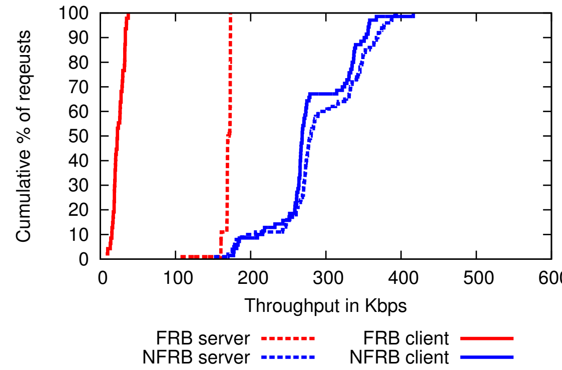

Figure 5: Zong throughput along alternative paths to our service, indicating throttling near 20 Kbps (solid red line).

bottleneck between the proxies. Further, the bottleneck rates are different from the FB C-Proxy to the clients in different ISPs. *Thus, we believe it is likely that the cause of the low bandwidth is ISP throttling* because the NFRB performance from the same clients is high, and the FB C-Proxy is unlikely to have ISP-specific throttling policies.

Thus though Free Basics services are at an advantage due to zero rating (potentially violating net neutrality), the network QoS each service gets depends on Facebook-imposed and the cellular-provider-imposed limits (another potential net neutrality violation). Further, different clients can see different performance for the same service, depending on which cellular provider they use.

4.4 Caching Policies

We observed that Facebook proxies also cache content, so we evaluated whether they respect server-specified caching policies. We find cache headers are respected for HTML and PHP files, but violated for images. We detect this as the Learn Basics content are textbook pages included as PNG images in HTML. We find requests for these images never come to our web server, and we are unable to measure server side throughput for experiments mentioned in the last section. Since traffic between client and Facebook C-Proxy is encrypted, we use *mitmproxy* at the client to look at the cache headers. We see our server policy of "Cache-Control: max-age=0, no-cache, no-store, must-revalidate" to be overwritten to "Cache-Control: public, max-age=0" for the image files.

5. DEVICE CHARACTERIZATION

There have been discussions about the target population of Facebook's Free Basics program [3]. Facebook claims that Free Basics brings millions of poor people online [1]. In this section, we analyze how well Facebook is able to reach the target population by analyzing the capabilities of user devices that send requests to our Bugle News and Learn Basics servers. Though only a subset of the Free Basics users visit these services, our analysis can give useful insights on the distribution of mobile devices for this user sample.

User agent string	% all requests	% PK requests
WAP browser	28.15	51.12
Generic Android	10.90	7.27
Device-specific string	55.46	39.52
Unidentified device	5.48	2.08

Table 2: Percentage of requests mapped to device or browser category.

Along with the device analysis of all web requests, we also present a separate analysis of requests coming from only Pakistan. Our goal is to analyze how the Free Basics devices in our dataset compare to the devices seen by a cellular provider in Pakistan [17].

5.1 Data Collection Methodology

As mentioned earlier, the Free Basics users' requests come to our web server through the Facebook proxies. The proxies set the “User-Agent” HTTP header to the original user-agent string from the requesting mobile browser, which can provide useful clues about the browser and device type being used. We assume that when a device type is specified, it is done so correctly, as we are unaware of any reason why this might be untrue.

We collect 706K non-empty strings across the two services (51K for Pakistan, which is ∼7.2% of overall requests). Table 2 shows the fraction of user-agent strings belonging to different categories. We find that significant fractions are mapped to WAP browsers or Generic Android, for which the mobile device model cannot be inferred. For the remaining 60.94% of requests, we successfully map 55.4% to their specific mobile-device model using the WURFL [16] open-source tool. For Pakistan, 39.5% requests are mapped to devices.

For the devices mapped using WURFL, Figure 6 shows a CDF of the fraction of requests made by each device type (log-scale x-axis). While we find a large diversity of devices accessing the service overall, there is a small number of devices that are substantially more popular than the oth-

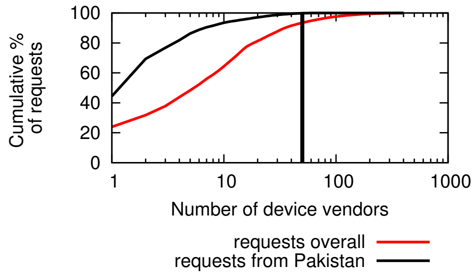

Figure 6: Devices from 50 vendors send 90% of requests.

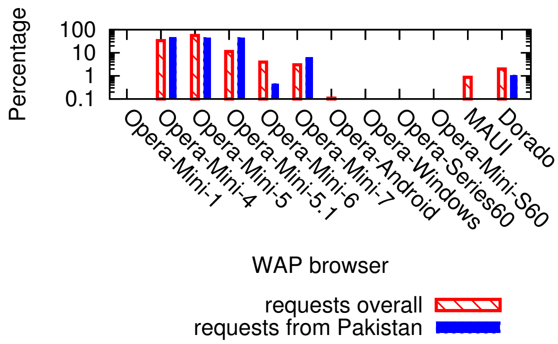

Figure 7: Proportion of WAP browsers

ers. Specifically, we find that devices from the top 50 vendors send 90% of the overall requests (shown with a vertical line), while a long tail of other 400 vendors' devices send the remaining 10% of the requests.

Using the same methodology as in [17], we crawl information from an online mobile phone database [10]. This gives the supported cellular interfaces (GSM, GPRS, EDGE, HSDPA and LTE) for devices from the top 50 vendors.

5.2 Prevalence of WAP Browsers

Table 2 indicates that a large percentage of requests use WAP browsers (~28% of the requests coming from all countries and ~51% from Pakistan). This high percentage of WAP requests shows there is a significant portion of constrained-capability browsers in our sample. Thus, it makes sense to follow the Free Basics technical restrictions of removing Javascript and rich multimedia from the web services, to support these browsers.

While we cannot determine the device type in these scenarios, we can extract the browser type based on the user-agent string. Figure 7 shows the percentage of requests from different WAP browsers. Opera Mini versions 4 and 5 dominate the requests. A small percentage of requests comes from others like MAUI and Dorado. Thus, when deploying services to Free Basics it is important to consider how a page will render in Opera Mini, since it is likely to impact a large fraction of users.

5.3 Capabilities of non-WAP Devices

We next analyze the set of requests that come from non-WAP browsers. As seen from Table 2, these constitute ~70%

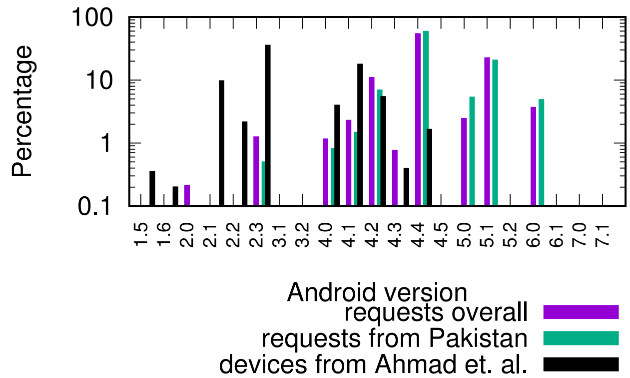

Figure 8: Proportion of Android versions

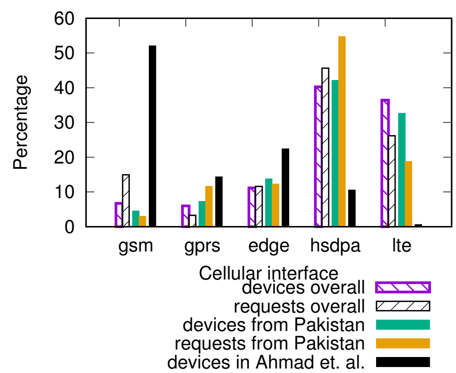

Figure 9: Proportion of cellular interfaces

of the requests from all countries and ~50% of the requests from Pakistan. These contain either (i) "Generic Android" user agent strings or (ii) device specific user agent strings, that we map to particular mobile phone models and extract their supported cellular technology information.

Figure 8 shows the proportion of Android OS versions among the requests. We also include the corresponding percentages seen by Ahmad et al. [17] as the last bar for each Android OS version. We find lower OS versions like Android 2.1, 2.3 and 4.1 in [17] have been replaced by higher OS versions like Android 4.4 and 5.1 in our dataset.

Figure 9 shows the percentage of devices and requests supporting each of the five cellular technologies, separately for all requests and those for Pakistan. Here also we include the corresponding percentages seen by Ahmad et. al., in [17], as the last bar for each cellular technology. The device distribution for our Free Basics users differs substantially from those in previous work [17]. Instead of lower data rate technologies like GSM, GPRS and EDGE dominating device proportions, our dataset is dominated by devices with higher data rate technologies like HSDPA and LTE.

These results suggest that a large fraction of requests (~40% from all countries and ~31% from Pakistan) are from users with modern smartphones providing full Web browsers. As high socioeconomic status is highly correlated with sophisticated smartphone ownership [29], this indicates that a substantial fraction of Free Basics users do not match the target audience that the service aims to reach (i.e., users with feature phones).

5.4 Conclusions

There can be two ways to explain the shift towards higher capability phone models, as seen for our non-WAP requests: (i) the dataset in [17] was from Jhelum district in Pakistan, which is semi-urban and hence had lower penetration of high-end smart phones. Most non-WAP requests coming to our Free Basics services are from high-end smart phones. This implies that these devices are potentially from urban areas. (ii) Since the dataset in [17] was from Dec, 2014, the distribution of phones in semi-urban areas in Pakistan has changed over the past two years. However it seems unlikely that it would change to the observed extent. It is difficult to differentiate these two cases, without more fine-grained location information for our Free Basics users or from a more recent cellular dataset.

While the devices support higher data rate cellular technologies, we can not infer whether the user actually uses a higher data rate SIM card or whether the user's cellular provider has support for HSDPA/LTE where he/she lives. Moreover, Facebook proxies throttle throughput on the server side (as shown in Section 4). Hence, we cannot reliably infer the cellular technology actually in use by looking at peak traffic rates at our web server.

Given these caveats, the conservative conclusion to draw from these analyses is the presence of a good proportion of high end mobile devices among Free Basics users, along with good proportion of WAP users. Thus, we see a mixed population of high and low capability devices among the Free Basics requests to our web servers. This indicates Facebook is reaching some of the constrained devices it targets, while other users with better phones are also using this program. Our methodology is also useful for reasons other than informing the target population debate. We have a scalable way of measuring a sample of users and the devices they have, which is useful to inform designs of future Free Basics services.

6. IMPLICATIONS

In this section, we list some of the implications of our measurement data in the context of the Free Basics debate [4, 12]. We also describe our interaction with Facebook to communicate our observations and their feedback.

Data privacy from Facebook. A significant point of concern against Free Basics is the flow of web requests and responses through Facebook's proxies [7]. Our measurements validate Facebook's advertised architecture [13]. We found at least two proxy machines belonging to Facebook, between the mobile client and our web server. We also found traffic between client and the first proxy to be encrypted. The segment between the second proxy and our servers was unencrypted, as our services did not support HTTPS.

Net Neutrality. A second point of concern against the Free Basics program is the unfair advantage free web services have over their paid counterparts, violating net neutrality [6]. Adding data to this debate requires measurements of services within and outside the Free Basics program, and possibly comparisons of their temporal growth in user base, keeping other factors constant. In addition, surveys of users to understand how, why, and how often they use Free Basics vs. paid services would help inform this debate. Both these approaches are beyond the scope of this paper.

Here, we ask a related question on the topic of net neutrality: whether there is any caveat to the advantages enjoyed by the free web services. As shown in our experiments, Free Basics services can see 4-12 times worse network performance than their paid counterparts and there are multiple factors contributing to this performance gap. This implies that the net neutrality debate should be simply focus on the advantage of zero-rated services, but also consider the constraints imposed on such free services.

Matching the target population. A third concern against Free Basics has been validity of the claim to bring millions of poor and first time Internet users online [3]. We used a server-side approach to characterize the devices that sent us requests. Our observations indicate there is no simple answer to this question, given that there are large fractions of high-capability mobile phones, but also large numbers of requests from constrained WAP browsers. We will complement our server side analysis of socio-economic background with client side user surveys as part of our future work.

Discussions with Facebook. We communicated our measurement methodology and empirical observations to the Facebook Free Basics team. We did this in the form of a research talk followed by a Q/A session at the Facebook headquarters. The most interesting outcome was Facebook's surprised reaction at the arbitrary bandwidth throttling policies of the different cellular providers. This highlights the importance of third party independent audits of a complex, global program like Free Basics, as presented in this paper. The mobile users, telecom operators, Facebook and web service providers form a complex eco-system in Free Basics. Additionally, there are regional variations in operator policies and locally relevant web content across the 60+ developing countries where the program is deployed. Under such circumstances, a key participant like Facebook can also find it difficult to keep track of all aspects, making transparency studies like ours a necessity.

7. CONCLUSION

In this paper, we analyzed the network architecture and policies responsible for observed network QoS in Free Basics, and also the mobile device capabilities from which the service requests come. These observations, along with our ongoing work in harnessing Free Basics to target useful applications to developing region populations, can increase the transparency of this program and assess its impact. This can be vital for more informed public debates on allowing or banning such a program in future and more nuanced dialogues between telecom regulatory authorities, cellular providers and Facebook.

8. REFERENCES

[1] Brought more than 25 million people online who wouldn't be otherwise. https://info.internet.org/en/blog/2016/05/10/announcing-the-launch-of-free-basics-in-nigeria/.

[2] Digital divide. https://en.wikipedia.org/wiki/Digital_divide.

[3] Facebook lures africa with free internet - but what is the hidden cost? https://www.theguardian.com/world/2016/aug/01/facebook-free-basics-internet-africa-mark-zuckerberg.

[4] Facebook's free basics service has been banned in india. http://www.theverge.com/2016/2/8/10913398/free-basics-india-regulator-ruling.

[5] Free basics by facebook - android apps on google play. https://play.google.com/store/apps/details?id=org.internet&hl=en.

[6] Free basics: Creating digital equality or divide? watch at 5:56 or 6:13 minutes for net neutrality concerns. https://www.youtube.com/watch?v=fZpO31LqNUM.

[7] Free basics: Creating digital equality or divide? watch at 8:03 or 17:08 minutes for privacy concerns. https://www.youtube.com/watch?v=fZpO31LqNUM.

[8] The impacts of emerging mobile data services in developing countries. http://a4ai.org/wp-content/uploads/2016/05/MeasuringImpactsofMobileDataServices_ResearchBrief2.pdf.

[9] internet.org by facebook. https://info.internet.org/en/.

[10] Mobile phone database from imei info. http://www.imei.info/.

[11] Opening internet.org and free basics: An open letter to facebook. https://openmedia.org/sites/default/files/openmedia-facebook-freebasicsletter.pdf.

[12] Prohibition of discriminatory tariff for data services regulations, 2016. http://www.trai.gov.in/WriteReadData/WhatsNew/Documents/Regulation_Data_Service.pdf.

[13] Technical guidelines - free basics. https://developers.facebook.com/docs/internet-org/platform-technical-guidelines.

[14] Where we've launched - internet.org. https://info.internet.org/en/story/where-weve-launched/.

[15] World databank. http://databank.worldbank.org/data/.

[16] Wurfl device detection and intelligence. http://wurfl.sourceforge.net/.

[17] 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 16th ACM Internet Measurement Conference (IMC)*, 2016.

[18] A. Anand, V. Pejovic, E. M. Belding, and D. L. Johnson. Villagecell: Cost effective cellular connectivity in rural areas. In *ICTD*, 2012.

[19] Z. S. Bischof, J. P. Rula, and F. E. Bustamante. In and out of cuba: Characterizing cuba's connectivity. In *IMC*, 2015.

[20] A. Chen, N. Feamster, and E. Calandro. Exploring the walled garden theory: An empirical framework to assess pricing effects on mobile data usage. In *Communications Policy Research South (CPRSouth)*, 2016.

[21] A. Dhananjay, A. Sharma, M. Paik, J. Chen, T. K. Kuppusamy, J. Li, and L. Subramanian. Hermes: Data transmission over unknown voice channels. In *MobiCom*, 2010.

[22] A. Gurumurthy and N. Chami. Internet governance as 'ideology in practice' - india's 'free basics' controversy. *Journal on Internet Regulation*, 2016.

[23] A. M. Kakhki, A. Razaghpanah, H. Koo, A. Li, R. Golani, D. Choffnes, P. Gill, and A. Mislove. Identifying traffic differentiation in mobile networks. In *Proceedings of the 15th ACM Internet Measurement Conference (IMC)*, 2015.

[24] A. Mathur, S. Agarwal, and S. Jaiswal. Exploring playback and recording of web-based audio media on low-end feature phones. In *ACM DEV*, 2013.

[25] R. Patra, S. Nedevschi, S. Surana, A. Sheth, L. Subramanian, and E. Brewer. Wildnet: Design and implementation of high performance wifi based long distance networks. In *NSDI*, 2007.

[26] A. A. Raza, M. Pervaiz, C. Milo, S. Razaq, G. Alster, J. Sherwani, U. Saif, and R. Rosenfeld. Viral entertainment as a vehicle for disseminating speech-based services to low-literate users. In *ICTD*, 2012.

[27] R. Sen, H. A. Pirzada, A. Phokeer, Z. A. Farooq, S. Sengupta, D. Choffnes, and K. P. Gummadi. On the free bridge across the digital divide: Assessing the quality of facebook's free basics service. In *Proceedings of the 16th ACM Internet Measurement Conference (IMC)*, 2016.

[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 *DEV*, 2015.

[29] S. F. Sultan, H. Humayun, U. Nadeem, Z. K. Bhatti, and S. Khan. Mobile phone price as a proxy for socio-economic indicators. In *Proceedings of the 7th International Conference on Information and Communication Technologies and Development (ICTD)*, 2015.

[30] X. S. Wang, A. Krishnamurthy, and D. Wetherall. Speeding up web page loads with shandian. In *NSDI*, 2016.

[31] X. Xu, Y. Jiang, T. Flach1, E. Katz-Bassett, D. Choffnes, and R. Govindan. Investigating transparent web proxies in cellular networks. In *Passive and Active Measurement Conference (PAM)*, 2015.

[32] Y. Zaki, J. Chen, T. Pötsch, T. Ahmad, and L. Subramanian. Dissecting web latency in ghana. In *IMC*, 2014.

APPENDIX

A. CODE AND DATA

To make our results reproducible, we have created a public repository of the code and data used in the project. As described in the paper, the experimental setups for this work have been non-trivial. On the client side, we needed a laptop, tethered on Wi-Fi to a mobile phone, that acted as a Wi-Fi hotspot. This mobile had a SIM card with Free Basics support for that country and also had credit balance to conduct the non Free Basics normal cellular connection experiments. We had these client setups in LUMS, Pakistan and UCT, South Africa. In Pakistan, we conducted experiments using SIM cards from two cellular service providers: Zong and Telenor.

In addition to the client side measurement setup, we implemented and controlled two web servers for Bugle News and Learn Basics, hosted at MPI-SWS in Germany. We registered these services to be part of Free Basics service list, going through Facebook's online application and approval process. We also moved the physical hosting of the Learn Basics server across Virginia, Sao Paolo, Mumbai, Sydney and Tokyo, using Amazon EC2 instances, for deconstructing the proxy architecture and network path inflation details, as described in Section 4.

If someone wants to conduct the experiments again, please get in touch with us. Both our web services are live on Free Basics, continuously accumulating visits from the ever growing number of Free Basics countries. Our collaborators in Pakistan and South Africa, who helped us conduct the client side measurements, are our co-authors in this paper. Thus they can help to conduct similar client side measurements in future, in these two countries.

In the public repository [5], we make available all the data collected in our experiments. We also include the code used to run th experiments, and the analysis code used in processing the data and generating the results and the graphs. Below we give a description of the repository, in connection to the different sections in the paper.

A.1 Network Characterization

The code and data for the discussions in Section 4 are in two folders: *4_ec2_experiments* and *5_throttling_pakistan*.

Path latencies and inflation. The *4_ec2_experiments* folder has a python script frb_crawl.py (with comments describing what it does). This script is run on the client side laptop, to fetch content from our Learn Basics server. The client side laptop further runs mitmproxy, to record the requests and the responses and the IP addresses with which it communicates. The server side runs tcpdump to store all incoming requests in pcap files, from which the requests coming from our clients are filtered using the string "Amreesh" in the user-agent field.

The *4_ec2_experiments* folder has two subfolders, one for Pakistan and the other for South Africa. Within each subfolder are three subfolders *clientside/* (contains the client side logs from mitmproxy), *serverside/* (contains the severside pcaps using tcpdump) and *server-client-matched/* (where our client outgoing requests are matched with server side incoming requests). Both the *clientside/* and *serverside/* folders have python scripts to process the mitmproxy and tcpdump outputs and generate text files. The *server-client-matched/* folder has a script times.sh, which computes the ping latency from mobile to C-Proxy and that between S-Proxy and our web server. The cdf.pl scipt computes the CDF of these latencies and cdf.gnu script plots the graphs (Fig. 2 and Fig. 3).

Throttling Policies. *5_throttling_pakistan* contains similar python crawler scripts frb_crawl.py and nfrb_crawl.py, to fetch a large HTML file from the Learn Basics server, over Free Basics connection and normal paid cellular connection respectively. The script trace.py computes throughputs from pcap files generated using tcpdump at the client and the server sides. The Telenor and Zong folders contain the CDF of the throughputs, which the plot.gnu script plots into graphs in Fig. 4 and Fig. 5. We do not upload all pcap files because of their large sizes. In case someone needs the raw pcap files for regenerating our throughput numbers or for conducting some other data analysis, please contact any of the authors for an immediate data exchange.

A.2 Device Characterization

The code and data for the discussions in Section 5 are in three folders: *1_pcaps_to_useragents*, *2_useragents_to_devices* and *3_devices_to_capabilities*.

The script country_useragents.py in *1_pcaps_to_useragents* takes the raw pcap files (example file trace_20160802.pcap) as input, to extract the country and the user agents and generate learnbasics.txt and newsbugle.txt as output.

The user agents are taken in *2_useragents_to_devices*, to produce the output sorted-freq-mobile-devices.txt, after mapping the user agents into mobile devices and sorting and enumerating unique mobile devices. The auto_device_detect.py script inside each subfolder for all and Pakistan specific user agents, processes the user-agents.txt files and maps the user agents to mobile devices using the Scientia Mobile website[6].

In *3_devices_to_capabilities* folder, list contains the names of 34 mobile device vendors, predominant among the mapped devices. The *links/* subfolder contains 34 text files, one for each vendor. These vendorlinks.txt files were created using https://magic.import.io/. We manually gave import.io any website (e.g. http://www.imei.info/phonedatabase/2-phones-alcatel/). import.io automatically crawled all pages under that page to get all the links, which we saved in the text file.

The script fetcher.py takes each link in vendorlinks.txt and downloads the webpage in a temporary folder. fetcher.sh calls fetcher.py for each vendor. The script filler.py takes each webpage (contains details for a particular device model) and extracts specified capability features. Sample outputs are in the *device_capability/* subfolder. Filler.sh calls filler.py for each vendor. The graphs in Section 5 are drawn using the capabilities information of the devices thus extracted.

[5] https://bitbucket.org/rijurekha/freebasics_ccr

[6] http://tools.scientiamobile.com/