Web Connectivity in Ghana: A survey

Kai Lüke
Institute of Computer Science
Department of Mathematics and Computer Science
Freie Universität Berlin
Berlin, Germany
kai.lueke@fu-berlin.de

ABSTRACT
Starting from the research “Dissecting web latency in Ghana” of Zaki et al. the field of Internet connectivity in Ghana is explored and the different aspects like latency, DNS resolution times, complexity of web sites, caching and prefetching, TLS/SSL, TCP and cellular network performance, role of ISPs, Internet Exchange Points, are brought in relation. Measurement approaches and tailored solutions for areas with bad connectivity are listed. Possible ways of improvement are discussed.

Categories and Subject Descriptors
C.2.6 [Internetworking]: Standards; C.4 [Performance of Systems]: Reliability, availability, and serviceability

Keywords
Developing Countries, Internet, Latency, Performance, TCP, Mobile, Cellular

1. INTRODUCTION
The Millennium Development Goals, which were supposed to be reached at the end of this year, also include measurements in target 8F for “mak[ing] available the benefits of new technologies, especially information and communications”. By these numbers almost 40% of the world population have Internet access [26]. Despite the big gap between global south and global north in a variety of aspects, the huge number of 6.9 billion mobile-cellular subscriptions could be utilised to raise Internet connectivity quickly. As in Ghana the population number was outnumbered [17], because people tend to have more than one subscription, the distribution might not be equal, but has led to 40.7% mobile Internet penetration in 2013 [11]. Concerning the ICT Development Index Ghana ranks 6th among African countries [13] and has direct connection to major undersea cables [30]. Recently LTE is rolled out as 4G cellular standard and the licence requires coverage of half the country within five years [25].

But the complexity of web site construction and data amount per page load has grown faster than the Internet connectivity in developing countries, thus leading to slow page loading times while bandwidth has increased in Ghana [30, 15]. These are not only frustration in usability but can also eliminate certain use cases. The following overview will show that growing bandwidth alone can not solve the problem due to the impact of latency and package loss on common transport protocols and the nested dependencies of requests during page load.

2. SPEED OF DATA TRANSMISSION
Round trip times to servers on the other half of the world can not be faster than the theoretical limit which is the speed of light through glass fiber. By example taking \( s = 12767 \text{ km} \) as distance from Accra to Seoul and maximum speed because of the refractive index for glass is \( v = \frac{1}{n} \approx 200000 \text{ km s}^{-1} \) the round trip takes always more than \( t = \frac{2s}{v} \approx 128 \text{ ms} \).

The importance of low latency when interacting with web sites is shown at the example of usage reduction for search engines after introducing an artificial delay as shortly described in the detailed survey of Briscoe et al. on reducing latency [3].

The links between Europe and North America are quite close to the theoretical limit and even almost half a billion dollar is spend for few millisecond gains by establishing new undersea cables [18]. Developing countries are not close to optimum and so improvements are possible in many ways. But still it must be questioned if an access outside the continent is justified or whether the content and services can be located at infrastructure nearby.

While throughput can be increased by adding more links, the performance of TCP is affected by latency. Also the recursive dependencies in requests while loading web pages result, as described later, in connectedness of latency and throughput.

3. METHODS OF DATA COLLECTION
Statistical data and specially detailed recordings can be taken at various places. In 2006 the operator of a main Internet cafe in Accra allowed analysis of logs containing 13.7 million requests [9]. Also browser access to hundreds of websites ranked popular by Alexa has been profiled at Accra, Kumawu and Hohoe (all in the south) in 2012 and 2014 by Zaki et al. [30] using a similar approach to [27]. These ap-
proaches help in tracing down how the slow loading times experienced by users are composed by the single parts. But as they happen on application layer, the reasons for decreased throughput and higher round trip times for packages can not be examined. Regarding to Pötsch the charts in [30] do also only include web pages which were able to be received fully which was not the case for all websites. However the paper focuses rather on DNS latency and solution space in the browser. Other works referenced in the chapter here on TCP contain research for these incidents.

Others showed how measurement devices for broadband and cellular connections can be deployed at diverse places in India and South Africa [14, 7]. The special benchmark routers contribute to appropriate long term data and also eliminate possible disturbance to accuracy when the line is shared with other users or due to the local network setup (ibid. and [22]). Also applications on smartphones gain interest as they provide an easy way to reach data sources [7, 19]. All these works also note the faced challenges like unstable power supply and absence of connectivity or suspiciousness towards the smartphone application.

Measurement can help in empowering users which constantly see a much slower connection speed than advertised by the provider [7].

4. INTERNET SERVICE PROVIDERS AND INTERNET EXCHANGE POINTS

The contrast between urban and rural areas regarding (fast) internet access is not only present in developing countries, but can therefore significantly determine whether a certain limit is exceeded which makes usage of demanding services not worthwhile any more. Particular for Ghana is also the gradient from the south to the overall less prosperous north. As urban inhabitants tend to be a better income source, providers can be inveigled to privilege them in resource allocation e.g. as it was observed by which gateways were given to urban and rural cellular clients in India [14].

Studies where measurements were realised for more than one provider indicate differences not only in bandwidth and high load during popular usage hours but also latency, specially last mile latency, which is the time to get to the first hop in the provider’s IP infrastructure [22]. It might even be that just network configurations bring betterment for cellular access [14]. In addition to correlation of the previous findings [7] also notes that the interconnection of ISPs in the region including neighbour countries plays a huge role when routes outside of the continent can be avoided.

Internet exchange points bring not only ISPs together but also provide a place where content can be made available. The GIX in Accra operates for a decade now and contributed in reducing access times for those organisations who are interconnected there. On average the traffic is around 500 MB/s. Specially the common content distribution networks and popular web services need to take part by locating their servers there. With the now abandoned AIX there used to be an other IX in Accra run by different members. While basis is laid, further steps are to be taken [4, 2, 8].

5. WEB SITE ARCHITECTURE, LATENCY AND THE LOADING PROCESS

The work of Zaki et al. is the centre of this section. It gives insights on how assumptions within design decisions of web services on conditions much different from the realities of Internet access in developing regions cause avoidable round trips leading to huge delays. Web pages become more complex and featureful and the dynamically loaded content through AJAX requests implies that even more requests need to be done until the page is in a complete state. Utilising browser automation through the Selenium framework access to hundreds of Alexa’s TOP sites (global and local) has been recorded to HAR reports (HTTP Archive), allowing a deeper look on how loading times are constituted and which dependencies are in the critical path through the Wprof tool. In 80% of the cases an initial delay is incorporated because of HTTP redirects (like from name.tld to www.name.tld) [30].

Similar work has been done by Sundaresan et al. to identify bottlenecks [23, 24]. Like in [27] we can see in Zaki et al. how blocking - as all TCP connections are in use and computation time have their share. But specially the time until the connection is set up - which is name resolution and TCP handshake - stands out with DNS having around 40% of the time in 2012 and up to 14-24% in 2014 (although mind that the amount to be downloaded increased in total). Time to connect contributed 5% of page loading time in 2012 and 15-20% in 2014. All data according to the two clients in Accra via campus WiFi and 3G as well as one 3G client in a more remote area.

While it might be interesting trying to reduce the blocking and speeding up receiving time by allowing more concurrent TCP connections it is also observed that competing TCP connections can have negative impact on each other [5].

Often it might be worth to reflect on which part of the web site has the highest value to the user. Blocking of images or JavaScript as well as using text based browsers will reduce load time of media rich sites. It was found that filtering of advertisement for products which are anyway not available in developing countries should be considered [9]. There is also a difference in browsers on the ability to render half received HTML sites to get quick feedback. But as they - like ziplinks - are rather simple compared to established browsers, they might just be used in addition. In contrast the usage of lightweight REST APIs through AJAX could also end up in faster interaction because it is not any more the whole HTML site which has to be send by the server.

As web interfaces for mail services are commonly used Du et al. suggest usage of SMTP because of advantages in saving data transfer [9]. In addition it would also avoid slower interaction through the browser and prevent loss of the written mail in case of transmission failure.

HTTP 1.1 pipelining allows to fetch more than one resource through a TCP connection. Similarly, SPDY brings even more flexible multiplexing of requests and the possibility to mark priorities, but also adds in-built compression and works exclusively over encrypted channels (i.e. to outperform HTTP without TLS the gain of SPDY must amortise
the increased start up time). The interesting parts for connections with high round trip times are server push and server hint, so that the server can continue transferring the probably next wanted content without the need of a client request or just indicate that the client should probably start requesting some content which is often requested soon by other clients. This technique is handy as the manual flattening of the dependencies to avoid too many recursive requests (and shorten a long critical path) is not necessary any more.

On the downside usage of only one TCP connection results in an entire delay under conditions of high packet loss [29]. Yet in the tests in Ghana an overall speed up of 40% was observed and further refinements might be possible [30].

6. INTERNET ACCESS WITH LOW OR INTERMITTENT CONNECTIVITY

The global digital divide has also lead to solutions paying attention to the conditions in the global south. Aggressive caching and constructing an index for offline browsing take care that benefits of Internet access are also available during times of an unstable or non-existing connection and even setups are used with the absence of a direct line and data is physically transferred. As the approach on information exchange is quite different, challenges of these mechanical backhaul systems include how to break down the process of web searches to be one round only as the “latency” is now in the order of hours while much data can be transferred at once on storage devices [6, 12, 20].

A more conventional system which serves a customised interface utilising a proxy was tested in Hohoe [15]. To make best usage of the connection, requests are queued and can even be scheduled for less busy times. It also uses filters to reduce page size and a shared cache which can be accessed through an offline index.1

The other popular Internet access method which is raising analogical as in the global north is through second and third generation mobile Internet subscriptions. Due to speed and data volume costs being the regulative factors Sambasivan et al. developed an application enabling mobile users to be informed about the site prize (as by download size) and gives short links/bookmarks instead of going through search engines to reduce delay. They noted on the conditions of mobile Internet usage and show observations with their custom infrastructure which was used to monitor page size by a non-caching proxy [19].

7. CACHING AND PREFETCHING

Caching can play a major role in speeding up page loading time under the here discussed conditions [9]. The emersion of CDNs and common JavaScript libraries allow already received content to be used between different web sites. Serving the common HTTP redirections from a cache resulted in 20% faster page load in the test set of Zaki et al. [30]. But a caching mechanism which is more intrusive than the internal browser cache may result to unwanted side effects. Also in contrast to static pages dynamic web pages are altered and personalised according to the situation of the request and caching them is be not of much use because of their limited life time. New approaches like micro-caching appear to be promising as style and programme code of a web site are highly cacheable [28].

Not much applicable to cellular access, caching and prefetching in home routers has various aspects. From obvious DNS caching it reaches to caching of TCP connections and even limited content caching. By also introducing popularity based prefetching for DNS, TCP and HTTP requests the extensive studies of Sundaresan et al. end in cutting down page loading times to the half [23, 24].

Yet cellular access already has a longer path until reaching the point where such methods can be deployed and also non-cacheable encrypted traffic in the “post-Snowden era” is increasing and shipped by major web services. In 2014 15% of the Alexa sites accessed at the research of Zaki et al. where using TLS/SSL [30]. Mid of this year the initiative Let’s encrypt2 plans to give free of charge TLS certificates in an automatic manner and is backed by Mozilla, Cisco, Akamai and IdenTrust SSL CA. This will influence not only start up time for connections as already written in the study but also the effectiveness of caches.

8. LATENCY IN DNS RESOLUTION AND POSSIBLE IMPROVEMENTS

The bad performance of DNS request in Ghana was the outstanding part of the recent study of Zaki et al.. Specially when a page load incorporates 11 name resolutions, as observed during their research, the round trip times slow down the page load tremendously by making up 72% of the time. Thus they do not only propose caching3 but also find that a better DNS infrastructure is needed in Ghana. Concerning caching is has to be considered that DNS resolution is also used for load balancing and localisation.

The distance to root servers which are mainly located in North America and Europe is one experienced nuisance. But also the often occurring recursive requests to the local DNS server, root server, top level domain server to sub domain servers in combination with the high round trip times sum up to a big delay. It was shown that by assuming a fictive zero latency for root servers in Ghana the access times could meet with the ones from Europe and North America. But not only placing root servers closer is giving a solution, also routes have to be optimally chosen:

“We found that nearly all of the traceroutes’ first hops were outside Ghana and somewhere in Europe (Switzerland, UK, etc.), from where, the route sometimes diverts to the US or Asia. None of the routes went to DNS root servers that are geographically nearby.” [30]

Others say that “ISPs should be more proactive to deploy local root anycast instances to improve their DNS query

1 https://letsencrypt.org/
2 In the Accra WiFi case caching is done on university campus which leads to improvement from 2012 to 2014.

1 It should be considered to include access to a Wikipedia dump as well like provided by the Evopedia application.
9. PERFORMANCE OF TCP UNDER CONDITIONS OF HIGH LATENCY AND PACKAGE LOSS

The TCP slow start behaviour for congestion control is well known and the three-way handshake for opening a connection and the way reliability is achieved by receiving ACKs has constraints on throughput (e.g. through head-of-line blocking) when round trip times are high and/or package loss occurs. A study claims that under their conditions loss makes web page latency 5 times slower, transfers take 5 times longer on average and 77% of losses were recovered by timeout instead of fast recovery. It is up to the algorithms to reduce these effects and contributions of the proposed methods to the Linux kernel have been done in 2013 [10]. Also RFC 6298 wants to improve the situation, but also spreading of newer software often needs longer time.

While the focus lays on speed and latency, stability of TCP connections is a challenge as well. Traditionally package drop is considered to come from congestion of a router. But this assumption is not accurate for unstable links and wireless transfers. Instead of reducing sending rate because of timeouts it might even help to retransmit quickly to prevent further collapse of the connection. This is also the case when competing connections share one link and some flows take the whole bandwidth while others do not recover and experience, thus leading to unfairness. A middlebox solution by Chen et al. has been tested to reduce these issues by preventing timeouts through scheduling of TCP connections and tactical package drops [5]. An other aspect impacting TCP latency is the bufferbloat in cellular networks which is discussed in the last section.

Adjustments which, like changing initial TCP window size to 10 segments, might have helped in certain situations, but research in general solutions by more robust algorithms performing in any corner of the world appears to be more desirable.

9.1 Alternatives to TCP

Other protocols like SCTP could in fact rule out the shortcomings of TCP which dates back to beginnings in 1974. But changing away from such an established protocol is a long way and promoting improvements to TCP instead of new protocol are adopted.

But in the context of Internet connectivity in developing countries SCTP\(^4\) has some properties to note on. On one hand the four-way handshake which should prevent DOS attacks seems to go contrary to the efforts to reduce up time. On the other hand the flow control algorithms might match better with the conditions and thus reducing loading times - but this is to be investigated. Clearly the features of taking multiple paths and having multiple addresses (multi home) do fit to the unsteady nature of connectivity in remote areas.

9.2 SSL/TLS

Normal start up has almost 10 round trips and it is an ongoing field to search for optimisations. False start was one attempt, now caching of server parameters - called fast-track - is promising [3, 30].

10. THE MOBILE CASE, RELIABILITY AND LAST MILE LATENCY

Much can be learned from the studies by Koradia et al. on Indian cellular networks in urban and rural areas [14]. The gap between urban and rural network performance is not necessarily having technical reasons but also reflects provider decisions of resource allocation. Also influence of network configuration at the gateway on latency of 3G links and throughput has been observed. The occurrence of stalled connections has been investigated and it is related to bursty transmission and the number of packages in flight, but some questions remain open. Some packets were delayed by 30 seconds. It was seen that in general connections without stall were 1.25 to 3 times faster, 2G connections were affected 40%, 3G 30%. The 5 minute transfers spend 15% of the time in this state. The effects could be reduced by changing the receive window size, but on cost of throughput.

Bufferbloat is a phenomenon in cellular network where packets are not directly forwarded but kept in a per client buffer which can have impact on TCP performance [1, 22]. Under these conditions the congestion control algorithms NewReno, Westwood+ and CUBIC performed differently when more then one connection is in use, specially with mixed short and long staying connections.

When new cellular networks are installed it is thus not enough to only look on the higher bandwidth but also efforts must be taken towards optimal latency. As there are situations where TCP recovery does not work well, providers have to make sure that they rule out reasons contributing to unstable TCP connections.

11. CONCLUSIONS

The aspects of slow Internet access in Ghana are diverse. The total latency in web requests consists of the various ways to improve the situation, be it on the international level down to even improving the behaviour of one’s own device. Yet it needs people working on these issues, be it in the non-profit GIX, in rural communities and school LANs which are not connected to the Internet or in developing robust and efficient communication protocols. But after all the digital divide is connected to the economical situation and relation of the global south and north towards each other.

12. REFERENCES


\(^4\)http://tools.ietf.org/html/rfc7053


