Web Performance Bottlenecks in Broadband Access Networks

Srikanth Sundaresan
Georgia Tech, Atlanta
srikanth@gatech.edu

Nazanin Magharei
Georgia Tech, Atlanta
nazanin@cc.gatech.edu

Nick Feamster
Georgia Tech, Atlanta
feamster@cc.gatech.edu

Renata Teixeira
CNRS and UPMC, Paris
renata.teixeira@lip6.fr

Sam Crawford
SamKnows, London
sam@samknows.com

ABSTRACT

We present the first large-scale analysis of Web performance bottlenecks as measured from broadband access networks, using data collected from extensive home router deployments. We analyze the limits of throughput on improving Web performance and identify the contribution of critical factors such as DNS lookups and TCP connection establishment to Web page load times. We find that, as broadband speeds continue to increase, other factors such as TCP connection setup time, server response time, and network latency are often dominant performance bottlenecks. Thus, realizing a "faster Web" requires not only higher download throughput, but also optimizations to reduce both client and server-side latency.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network Management*; C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network Monitoring*

Keywords

Broadband Networks, Web performance, Bottlenecks

1. INTRODUCTION

Broadband speeds are getting faster: the OECD reports that speeds are increasing by about 15–20% every year; average advertised speeds are now about 16 Mbits/s in the U.S. and 37.5 Mbits/s across OECD areas [1]. As speeds continues to increase, one might expect the Web to get faster at home; yet, ISPs and application providers are increasingly finding that *latency* is becoming a critical performance bottleneck [2]. The Bing search engine experiences reduced revenue of 1.2% with just a 500-millisecond delay [3]. Forrester research found that most users expected online shopping sites to load in two seconds or fewer [2].

This paper presents initial results from a large-scale, longitudinal study of the effects of broadband access network performance on Web page load times. We find that throughput is the dominant bottleneck for access links with downstream throughput rates of less than 16 Mbits/s. In the case of the increasing number of access links with higher downstream throughput, latency has become the bottleneck for Web performance. This paper offers insight into the extent to which performance characteristics of the underlying network introduces bottlenecks in a Web download and this in turn can help understand the limits of optimizations such as caching and prefetching in mitigating these bottlenecks.

SIGMETRICS'13, June 17-21, 2013, Pittsburgh, PA, USA.
ACM 978-1-4503-1900-3/13/06.

Metric	Type	Description
Page fetch time	Total	The time to set up TCP connections and retrieve all objects.
Page load time	Total	Page fetch time, plus the DNS time.
DNS lookup time	Per Domain	The DNS time for different domains.
Time to first byte	Per Object	The time from the initiation of the TCP connection to the arrival of the first byte of the requested object (including server processing time).
Object fetch time	Per Object	The time to download an object, exc. DNS time and time to first byte.

Table 1: *Performance metrics. For per-object metrics, the test measures the maximum, minimum, and average for each object.*

2. MEASUREMENTS AND METHOD

Modern browsers and servers perform various optimizations to improve page load time. Independent of these optimizations, the effects of network parameters such as the access network throughput and latency on various components of Web page load time are worth understanding. Ideally, we would measure Web performance from each access link using a variety of browsers. However, logistically, and due to significant differences between browsers, this is not feasible. Therefore, we perform active measurements from routers in a large number of homes using a simple tool to characterize the network properties of Web performance in home networks. Understanding how network effects affect Web performance can better inform designers of both Web sites and browsers about bottlenecks and how to mitigate them.

Active measurements from the home router We measure Web performance by periodically requesting the home page of nine popular Web sites. We use measurements from 5,667 US participants in the SamKnows/Federal Communications Commission study from September 1–31, 2012 [4]. We include only ISPs with more than 100 users and users who have reported more than 100 measurements during the duration of the study.

Measuring Components of Page Load Time We call the process of downloading all objects for a Web site a *transaction*. Our tool measures the time for each component of a transaction, as shown in Table 1. It periodically fetches the home page of a Web site, determines the objects required to render the page, performs DNS lookups, and downloads all the objects. We thus separate the page load time into DNS time and the fetch time, and also tease out individual component times per object, including the DNS lookup time

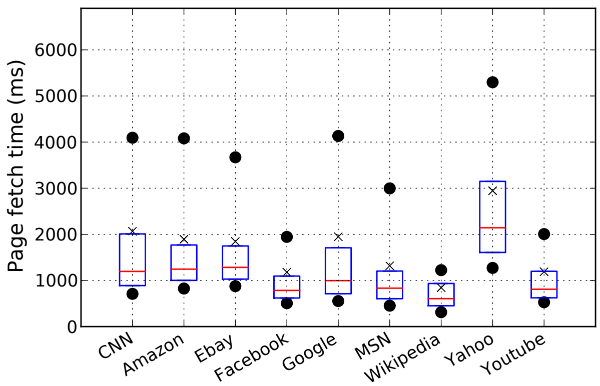


Figure 1: *Page fetch times for popular sites. We fetch the home page except for Google (http://www.google.com/mobile) and CNN (http://edition.cnn.com). Each box shows the inter-quartile range of the fetch times. The middle line is the median, the cross represents the average, and the dots the 10th and 90th percentile page fetch times.*

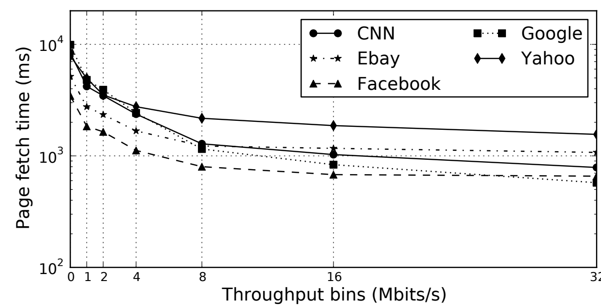
and the time to first byte. The tool uses persistent TCP connections if the server supports them and up to eight concurrent TCP connections to download objects. The tool does not attempt characterize page load times that a real Web browser might experience, but it reflects the same underlying network characteristics as the browser.

3. ACCESS-LINK BOTTLENECKS

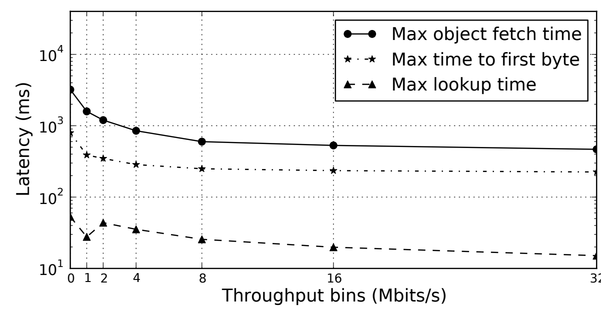
We characterize page fetch times from clients in the deployment and evaluate the effect of throughput on page fetch times. Major takeaways are that page load time for even very popular sites are high, and that as downstream throughput exceeds 16 Mbits/s, page load times cease to improve; latency increasingly becomes a performance bottleneck for Web performance at higher throughputs.

Page Fetch Times of Popular Web Sites We study the page fetch times to nine popular Web sites. Figure 1 shows the fetch time for each site. We see that the fetch times can exceed one second for even popular sites. As expected, the fetch time varies both by site and the location of the access network. Some variability results from differences in page size and design; the largest four sites (CNN, Yahoo, Amazon, and Ebay) also have the largest fetch times (*e.g.*, the median fetch time for CNN is more than one second).

Effect of downstream Throughput We study how page fetch time and its components vary with downstream throughput. We group access links according to downstream throughput into bins that reflect common ranges of access plans: 0–1 Mbits/s, 1–2 Mbits/s, 2–4 Mbits/s, 4–8 Mbits/s, 8–16 Mbits/s, 16–32 Mbits/s, and 32–64 Mbits/s. Figure 2a shows how the median page fetch time decreases as throughput increases, up to 16 Mbits/s. Beyond that, page fetch times decrease only modestly. For example, the median time for CNN is 8.4 seconds for links with throughput 0–1 Mbits/s and 1.3 seconds when it is 8–16 Mbits/s. When throughput exceeds 32 Mbits/s, the page fetch time is 790 ms, only slightly better than for links with 8–16 Mbits/s. Figure 2b shows how the maximum object fetch time, maximum time to first byte, and DNS lookup time decrease as throughput increases. For each group with a particular downstream throughput range, we plot the median of each of these values. As downstream throughput increases to 32–64 Mbits/sec, the object fetch time decreases from 3.2 seconds to 530 ms; in contrast, the time to first byte decreases from 800 ms to 230 ms and DNS lookup time decreases from about 50 ms to



(a) *Page fetch times*



(b) *Components of page load time*

Figure 2: *Page fetch times decrease with downstream throughput, but only up to 16 Mbits/s. X-axis labels denote the start of each throughput bin (*e.g., *"0" is the set of users with downstream throughput up to 1 Mbits/s.)*

about 15 ms. Thus, as downstream throughput increases beyond 16 Mbits/sec, time to first byte and DNS lookup times become a significantly larger component of page fetch time.

4. CONCLUSION AND FUTURE WORK

We present the first large-scale, longitudinal study of Web performance bottlenecks in broadband access networks. We characterize performance from more than 5,000 broadband access networks to nine popular Web sites and identify factors that create Web performance bottlenecks. Our results show that as broadband access speeds continue to increase, latency becomes a performance bottleneck. Page load times stop improving as throughput rates increase beyond 16 Mbits/s, and latency components such as DNS and time to first byte become more important. Last-mile latency could become a significant overall contributor to both DNS lookup times and the time to first byte [4]. Therefore, *even when caches are close to users, implementing optimizations in the home to reduce the effects of last-mile latency can offer significant performance improvements*. We are currently investigating techniques to mitigate the impact of such overheads on Web performance.

5. REFERENCES

[1] *OECD Communications Outlook*. OECD Publishing, July 2011.

[2] S. Lohr. For Impatient Web Users, an Eye Blink Is Just Too Long to Wait. `http://www.nytimes.com/2012/03/01/technology/impatient-web-users-flee-slow-loading-sites.html`, Mar. 2012.

[3] S. Souders. Velocity and the bottom line. `http://radar.oreilly.com/2009/07/velocity-making-your-site-fast.html`, July 2009.

[4] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè. Broadband internet performance: A view from the gateway. In *Proc. ACM SIGCOMM*, Toronto, Ontario, Aug. 2011.