

Characterizing Performance of Residential Internet Connections Using an Analysis of Measuring Broadband America's Web Browsing Test Data

by

Alexander M. Gamero-Garrido

Electronic Engineer, Simón Bolívar University (2011)
with Engineering Degree Project at Lund University (2010)

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

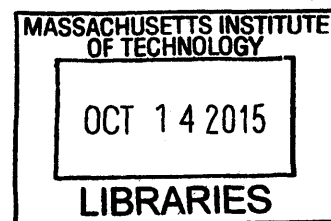
Master of Science in Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

ARCHIVES



© Massachusetts Institute of Technology 2015. All rights reserved.


Signature redacted

Author


Signature redacted

Technology and Policy Program
August 14, 2015

Certified by

David D. Clark
Senior Research Scientist

Computer Science and Artificial Intelligence Laboratory
Signature redacted Thesis Supervisor

Accepted by

Munther A. Dahleh

William A. Coolidge Professor of Electrical Engineering and Computer Science
Director, Institute for Data, Systems, and Society
Acting Director, Technology and Policy Program

Characterizing Performance of Residential Internet Connections Using an Analysis of Measuring Broadband America's Web Browsing Test Data

by

Alexander M. Gamero-Garrido

Submitted to the Institute for Data, Systems, and Society
on August 14, 2015, in partial fulfillment of the
requirements for the degree of
Master of Science in Technology and Policy

Abstract

This thesis presents an analysis of F.C.C.-measured web page loading times as observed in 2013 from nodes connected to consumer broadband providers in the Northeastern, Southern and Pacific U.S. We also collected data for multiple months in 2015 from the MIT network. We provide temporal and statistical analyses on total loading times for both datasets. We present four main contributions. First, we find differences in loading times for various websites that are consistent across providers and regions, showing the impact of infrastructure of transit and content providers on loading times and Quality of Experience (QoE.) Second, we find strong evidence of diurnal variation in loading times, highlighting the impact of network and server load on end-user QoE. Third, we show instances of localized congestion that severely impair the performance of some websites when measured from a residential provider. Fourth, we find that web loading times correlate with the size of a website's infrastructure as estimated by the number of IP addresses observed in the data. Finally, we also provide a set of policy recommendations: execution of javascript and other code during the web browsing test to more adequately capture loading times; expanding the list of target websites and collecting trace route data; collection of browsing data from non-residential networks; and public provision of funding for research on Measuring Broadband America's web browsing data. The websites studied in this thesis are: Amazon, CNN, EBay, Facebook, Google, msn, Wikipedia, Yahoo and YouTube.

Thesis Supervisor: David D. Clark
Title: Senior Research Scientist
Computer Science and Artificial Intelligence Laboratory

Acknowledgments

There is a famous joke by a former MIT president that compares being a student here to “taking a drink from a fire hose.” Cheekiness aside, I find the analogy very fitting to my own experience. Having pushed my intellectual—and, sometimes, physical—boundaries beyond what I ever imagined, there is simply no way I would have thrived in such an extenuating environment without the help of many.

Let me begin by thanking the staff at the three organizations with which I interacted the most: Ed, Frank and Dava at the Technology and Policy Program (TPP), Sue at the Computer Science and Artificial Intelligence Laboratory (CSAIL), and Kate at the Department of Political Science. I am especially grateful to Barb DeLaBarre, the Academic Administrator of TPP. As one of my fellow students said “if I did not meet Barb [and received what is practically pro bono therapy] on a regular basis, surely my life would fall apart and everything would go to pieces.”

Dr. David Clark -my thesis supervisor- has been a grounding stone of my research experience since 2013. I appreciate him taking the time to meet me and enlighten me with his profound understanding of the quirks and bolts of the Internet. Dr. Steve Bauer, another member of the Advanced Network Architecture group at CSAIL, has also been kind enough to guide me this past year. Steve’s advice on all matters technical has provided invaluable for a novice data scientist and Internet Explorer like myself, and this work certainly could not have been completed without his help. Finally, I cannot end this paragraph without mentioning my former advisors and mentors: Prof. Nazli Choucri, at MIT; Professors Juan Carlos Grieco, Juan Bogado and William Colmenares, at U. Simón Bolívar; Prof. Gunnar Bolmsjö, then at Lund U.; Dr. Jacob Svendenius, then at Haldex; and Miguel Velarde, at that long-time second home of mine, Guayoyo en Letras magazine.

I am thankful to the sponsors of my research assistantships during these past two years: the Office of Naval Research, the National Science Foundation and the Hewlett Foundation.

My fellow TPPers are the smartest, most fun people I have ever met. I understand this is a bold claim, but there is just no other way to describe them. Thanks for making the darkest, coldest of days warm, fun and fuzzy. Special mention to the Edgerton Control

Center (Corinne, Stacey and Lauren) and Henry in TPSS; my partner in crime in long lab nights and occasional cigarettes, Cecilia; and the rest of the regular "technology and party program" crowd: Tessa, Tiziana, Scott, Nora, Alejandro, Pierre, Julius, Erin, Ben and Jiyoung (that honorary TPPer that happened to go to Harvard.)

Two years ago now, a Texan, a Canadian, a Michigander and a Venezuelan set up a rather unusual social experiment. They decided to live together in one of the liberal-academic bastions of the world: Cambridge, Massachusetts. They founded 'Murica house, and about a year later a Mississippian-Floridian joined them in the arduous task of maintaining a "play-hard, party-hard" embassy.

Jokes aside, I will be forever grateful for my time in this house with Justin, Jacob, Chris and John. There is a fond place in my memory for the superb conversations we had in our kitchen, properly equipped with patio furniture! You guys are like my brothers.

I am at risk of writing the longest acknowledgements section in the history of MIT, but no such thank you note would be complete without mentioning Robert. Thank you for caring for me and for your patience, even when I had sunk into a self-absorbed PhD decision mania; being by your side made my life infinitely better during this stressful time. You rock!

Last but definitely not least, I thank my family (Arelis, Remigio and Wlady) for reminding me there is a home for me somewhere in this convoluted world, and for nurturing the academic in me from the very beginning. It has been six years of me living in foreign lands, and you have been one of the few constant sources of much-needed emotional support.

Thank you all!

Contents

1	Introduction	15
1.1	Background	16
1.2	Results	17
1.3	Thesis Organization	18
2	Web Browsing and Network Measurements: a Primer	19
2.1	Web Browsing: How Does it Work?	19
2.2	Network Measurements: Description and Relevance	21
2.2.1	Metrics	21
2.2.2	Congestion	22
2.2.3	Quality of Experience (QoE)	23
2.2.4	Factors Impacting Quality of Experience	24
2.2.5	Measuring Broadband America	25
3	Analysis of Web Browsing Test Data	27
3.1	Measuring Broadband America’s Web Browsing Test	27
3.1.1	Technical Overview	27
3.2	Analysis of F.C.C. Web Browsing Test	28
3.2.1	Data Included	29
3.2.2	Loading Time Boxplots	29
3.2.3	Localized Congestion	36
3.2.4	Diurnal Variation	37
3.2.5	Step Changes on Loading Times for a Specific Website	42

3.2.6	Drop in Facebook Loading Time and Bytes Transferred	45
3.3	Analysis of MIT-Collected Data	45
4	Implications for Quality of Experience Measurements and Public Policy	53
4.1	Relative Value of Web Browsing Data	54
4.2	Limitations of Loading Times Reported by the F.C.C. for QoE Measurements, and Policy Recommendations to Address them	54
4.3	Measuring the Impact of Content Delivery Networks	57
4.4	Diurnal Variation for Each Target Website	57
5	Conclusion	61
A	Loading Time Histograms	65

List of Figures

3-1	Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Comcast network in the Northeast, with an advertised speed of 20 Mbps.	32
3-2	Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Cablevision network in the Northeast, with an advertised speed of 15 Mbps.	33
3-3	Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Verizon network in the Pacific, with advertised speeds between 25 and 50 Mbps.	34
3-4	Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Cox network in the South, with advertised speeds between 12 and 25 Mbps.	35
3-5	Yearlong time-series of loading times from the Cablevision network (15 Mbps) in the Northeast.	38
3-6	Yearlong time-series of loading times from the Comcast network (20 Mbps) in the Northeast.	39
3-7	Yearlong time-series of loading times for Wikipedia, measured from units connected to the Internet through the Verizon network in the Northeast, with advertised speeds between 15 and 25 Mbps.	40
3-8	Time-series of loading times for Amazon and Wikipedia in January 2013, measured from units connected to the Internet through the Verizon network in the Northeast, with advertised speeds between 15 and 25 Mbps.	41

3-9	Yearlong time-series of loading times for CNN, measured from units connected to the Internet through the Comcast network in the Northeast, with an advertised speed of 20 Mbps.	42
3-10	Yearlong time-series of loading times for Yahoo, measured from units connected to the Internet through the Time Warner Cable network in the South, with advertised speeds between 15 and 20 Mbps.	43
3-11	Yearlong time-series of loading times for msn, measured from four different providers in three regions. Note that although loading times are different in nominal terms, a similar increase can be seen during the second quarter (days 90-120).	44
3-12	IP address and loading times for msn, measured from the Cox network in the South.	46
3-13	IP address and loading times for YouTube, measured from the Verizon network in the Pacific.	47
3-14	Total loading time for Facebook, measured from the AT&T network in the Pacific. The visible drop in loading times occurs on May 7th, 2013.	48
3-15	IP address and loading times for Facebook, measured from MIT. The two highlighted periods show variations in loading times that correspond to changes in the server's IP address blocks.	49
3-16	Total Loading Times for Google and Yahoo, measured from MIT.	50
3-17	Histograms of loading times of different websites measured from MIT.	52
4-1	Median loading time and number of unique IP addresses for eight target websites, measured from fiber and cable providers (all but Frontier and AT&T.)	58
4-2	Average diurnal change in loading times for each target website, including data for all 12 providers in the three regions studied.	59
4-3	Change in loading times for each target website, measured at MIT from March 13th to June 20th, 2015.	60

A-1 Histograms of loading times of different websites for all of 2013, measured from the Comcast network in the Northeast, with an advertised speed of 20 Mbps. 66

A-2 Histograms of loading times of different websites for all of 2013, measured from the Cablevision network in the Northeast, with an advertised speed of 15 Mbps. 67

A-3 Histograms of loading times of different websites for all of 2013, measured from the Verizon network in the Pacific, with advertised speeds between 25 and 50 Mbps. 68

A-4 Histograms of loading times of different websites for all of 2013, measured from the Cox network in the South, with advertised speeds between 12 and 25 Mbps. 69

List of Tables

- 3.1 States included in each region for the data analysis presented in this chapter. 30
- 3.2 Providers, advertised speeds, number of units (whiteboxes), and technologies studied for each region. 31
- 4.1 Total bytes transferred and loading times from Google Chrome’s developer tools. 55

Chapter 1

Introduction

No service should be stuck in a "slow lane" because it does not pay a fee. That kind of gatekeeping would undermine the level playing field essential to the Internet's growth. So, as I have before, I am asking [the F.C.C.] for an explicit ban on paid prioritization and any other restriction that has a similar effect [1].

President Barack Obama, 2014.

When speaking about "gatekeeping" in the above policy excerpt, the President follows a narrative on network neutrality that dates back to at least 2003 [2–4] and is prevalent in liberal media [5–7]: websites¹ need to be protected from undue discrimination by broadband providers (ISPs.) Given ISP's control over last-mile networks—those that reach the users' homes— websites can be blocked or "throttled" by the ISP if they offer competing services or simply refuse to pay a high fee. The quality of the experience (QoE) of the users with such websites could be degraded as a result.

In order to test the validity of such claims, measurements on the performance of websites (such as Google or Yahoo) as experienced by residential broadband subscribers are needed. However, collecting such data carries significant challenges, as the users—ordinary citizens—are distributed all over the United States, and the required time-scale and frequency of measurements adds considerable complexity. Further, many reports, including those by the Federal Communications Commission, tend to focus on performance of a single part of the

¹And other over-the-top services, such as video streaming platforms.

Internet²; since content of websites needs to be delivered from the target websites' servers to the users' devices at home, such partial-network analysis can only imperfectly capture loading times experienced by users.

We address the characterization of web browsing loading times by analyzing two datasets: a relatively unexplored set of measurements collected by the F.C.C. regarding web browsing performance for residential broadband subscribers; and our data collected from network probes at the Massachusetts Institute of Technology. In particular, we provide extensive temporal and statistical analyses on total loading times for nine top websites in the U.S., measured at actual subscribers' homes in three geographic regions and M.I.T. We use such analyses to: characterize diurnal variation in website performance; provide evidence of the impact of infrastructure beyond the ISP—particularly the target website's number of IP addresses—on web performance; and suggest changes to F.C.C.'s testing procedures to adequately capture website loading times as experienced by end-users.

1.1 Background

Much of the network neutrality debate, and most of what the F.C.C. reports, focuses on one section of the network: the access network, those that connect users' homes to the rest of the Internet. Although performance of the last-mile *is* critical for web browsing [8], other systems can greatly impact loading times experienced by the user—and, consequently, the quality of their experience with the web [9]. These non-ISP systems include specialized content delivery networks, transit providers and, importantly, the networking and hosting infrastructure of the target website itself.

In order to visualize the importance of systems beyond the access networks, it is useful to draw an analogy between web browsing and catalog mail orders. When a shopper wants to buy a product from a print catalog, she places an order in the mailbox, addressed to the retailer. Her neighborhood mailman would then bring the envelope with the order to a mail processing center, where it would be routed to subsequent processing centers, until it

²In the case of the F.C.C., the focus is on performance of the access networks; although this last-mile performance is critical for web browsing quality of experience, it is only one of the systems involved in delivering content.

reached the retailer’s neighborhood. There, a mailman would deliver the product order to the retailer’s building. Then, the retailer would need to process the request by examining the list of items the shopper ordered; prepare the packaging to ship the products; and drop the boxes at the post office, where it would follow a reverse route all the way to the shopper’s home. Delays at any point in this delivery chain, including the processing time inside the retailers facilities, increase delivery times due to the sequential nature of catalog mail orders.

To some approximation, that is also how web browsing works: the user tells her browser where she wants to go; the browser sends a request for a site, and that request packet is routed initially through the user’s ISP (the mailman); the packet then travels to either a delivery network, a transit provider, or both (mail processing facilities); and it is delivered to the target website, where the servers (analogous to the retailer) must process the request and prepare the objects to send back to the user’s device. The packets with this information then follow a route back. It follows that delays at any point in the network (delivery chain) would increase the user’s wait for her content—a.k.a. the total web loading time.

Characterizing end-to-end performance is, then, necessary to study web browsing Quality of Experience (QoE). To this end, we analyze a subset of web loading times measured by the F.C.C. from households in the Northeastern, Southern and Pacific U.S. Nine websites are included in our study: Amazon, CNN, EBay, Facebook, Google, msn, Wikipedia, Yahoo and YouTube. To validate our inferences, we also collect and analyze loading time data for these nine websites from M.I.T.

1.2 Results

We find that performance for different websites is diverse as a result of a myriad of factors, as we explain in the next chapters:

- Size of the target website’s infrastructure—including internal or hired content delivery networks—which impacts its ability to deliver content and respond to requests.
- Usage patterns, in particular diurnal variation of users’ browsing activity, which results in network loads that fluctuate throughout the day.

- Localized congestion along connection paths providing links from specific providers to target websites, resulting in reduced effective throughput and increased loading times.
- Architectural changes on the website’s server-end, including the allocation of additional IP address blocks and hosting capabilities.
- And congestion in the ISP’s internal network, which increases loading times for all websites.

We provide evidence showing the impact of all the above factors on web loading times. Our results are displayed in five forms. First, we provide histograms of web loading times from a single provider and region, showing difference in performance for various websites. Second, to our knowledge, we provide the first temporal analysis of F.C.C. data for twelve continuous months, showing incidences of localized congestion and persistent diurnal variation. Third, we explore the influence in performance of target websites’ infrastructure, by correlating time-series of loading times and IP addresses. Fourth, we aggregate loading times for each website from all providers and regions analyzed, and compare it to the apparent size of their networking infrastructure. Fifth, we compare overall diurnal variation for each target website, by looking at median loading times at different times of the day. Finally, we validate our inferences by repeating the previous five analyses on data collected at M.I.T.

Finally, we show the limitations of data collected by the F.C.C. regarding web browsing loading times, and explain why it underestimates those experienced by real users. We proceed to suggest changes to the web browsing test in order to address some of these limitations; and provide some considerations regarding measuring packet loss.

1.3 Thesis Organization

The rest of this thesis proceeds as follows. Chapter 2 presents an introduction to the technical aspects underpinning web browsing and measurements. Chapter 3 includes the bulk of the data analyses of this thesis, on both F.C.C. and M.I.T.-collected measurements. Chapter 4 presents high-level inferences of individual website performance. Finally, we present our conclusions in Chapter 5.

Chapter 2

Web Browsing and Network Measurements: a Primer

Quality of Experience (QoE) relates to how users perceive the quality of an application. To capture such a subjective measure ... is an art on its own [10].

Kuipers et al., 2010.

In this chapter, we introduce basic concepts related to web browsing and network measurements. We begin by describing the general process behind web browsing. Then, we present the network metrics used throughout this thesis. Further on, we present some considerations regarding network congestion and how it is defined by some of the actors involved in web browsing. Sections 2.2.3 and 2.2.4 present an overview and factors affecting Quality of Experience for web browsing. Finally, the last section briefly introduces Measuring Broadband America, the project of the Federal Communications Commission that produced the datasets we analyze in subsequent chapters.

2.1 Web Browsing: How Does it Work?

Although the applications built on the web are ever-transforming, the basic set of operations that occur when a Uniform Resource Locator (URL, such as `www.mit.edu`) is typed into a web browser has remained relatively constant over the years. This section will provide a

brief description of such operations to inform our exploration of QoE for web browsing.

There are four general actions executed when loading a website [8]:

1. DNS lookup (also called DNS resolution)
2. TCP connection establishment
3. Server response (on the target website's end)
4. Object download

We now provide brief descriptions for each of the above actions. DNS (Domain Name Server) resolution is usually the first step to load a website. During DNS resolution, a (usually) remote DNS server will tell the user's browser the IP address in which it can find the target website.

Once a target IP address is acquired, the browser attempts to initiate a (usually Transmission Control Protocol, or TCP) connection with the intended remote server¹. TCP connection establishment follows a process called a *three-way handshake*, in which the client (web browser) sends a synchronization message, the server (target website) returns an acknowledgement message, and finally the client acknowledges the receipt of the server's acknowledgement by sending another message. A total of three messages are exchanged, which adds some round-trip travel times (RTTs) to the overall connection time.

Once the TCP connection is established, the third step is for the server to prepare the content it will send to the client. Content is divided in *objects*. Once the server has prepared the content, the fourth step will be for the browser to download the initial objects associated with the URL, which is often a HyperText Markup Language (HTML) file. Time to First Byte (TTFB) measures the time from initiating DNS resolution to the beginning of this initial download, when the client—the end-user's browser—receives the first payload byte.

Then, depending on the website, further actions might occur at the browser (for example, executing Javascript code), and additional content might be downloaded. Among those extra resources to be fetched, there are two types of objects: "static, in which case the URL

¹Note that "connection" does not imply a direct link between the end-user's browser and the target website; rather, the connection includes anything from several to tens of mid-points or routers.

[indicating where to find the object] is in the [initial] homepage itself; or dynamic, in which case the URL is determined by running active scripts." [8]

Unsurprisingly, the process to load a web page varies greatly from website to website. Simple pages, such as `www.wikipedia.org`, will be completely loaded after a number of requests that is considerably lower than those required by more complex sites, such as `edition.cnn.com`.

2.2 Network Measurements: Description and Relevance

2.2.1 Metrics

Here we describe several metrics relevant to the web browsing test of the Federal Communications Commission (F.C.C.), one of the main datasets informing this thesis. The test itself will be described later in this chapter.

Throughput refers to the rate at which information is transferred between nodes of the network. In the context of residential broadband, it is typically expressed as the maximum download speed advertised by the Internet Service Provider (ISP), such as Comcast or Verizon. Throughput is normally expressed in megabits per second or Mbps, and has been the traditional indicator of network performance used extensively in reports by the F.C.C.

Time to first byte (TTFB) refers to the period that starts with a request to initiate a connection and ends with the reception of the first byte. TTFB is typically expressed in milliseconds (ms.)

Total download time is the time it takes the user's browser to download the contents of the target website, including those parts delivered from third party servers.

Internet protocol (IP) address is a number assigned to each device connected to a network, including website hosts. The majority of Internet traffic uses the IPv4 standard, which consists of four blocks of one byte each². As mentioned in the previous section (2.1), a DNS resolution is required to translate domain names (such as `example.com`) to IP addresses (such as `93.184.216.34`, or the corresponding network address of `example.com`.) Finally, a

²A complete description of IPv4's numbering system can be found in [11].

higher number of IP addresses for a single website typically signals a larger, more complex network (with more hosts that need addressing.)

2.2.2 Congestion

Definitions of congestion in computer networks are far from uniform. For a study to have any hopes of characterizing QoE, however, network congestion is a necessary consideration, given its impact on web browsing performance. Let us start with a simple definition: “Internet congestion occurs when a part of the Internet, [for example] a link or server, does not have the capacity to service all the demand presented to it.” [12]

Given this definition, congestion can be present at any point along the connection path from a user’s device to the target website³. A typical experience of browsing the web might require using network infrastructure of: a service provider, such as Comcast or Verizon; a transit provider or a content delivery network, such as Akamai; and the target website, for example Google or CNN. These actors provide disparate services to the user (some of them do not contract with the user at all), and may define congestion differently given their operational needs and constraints. Below we present some of such definitions, as a starting point for our later discussion of QoE.

From a technical perspective congestion is present when the data input “rate into a [link] exceeds the service rate of the [link] at a point in time” [13], usually resulting in the build up of an input queue in that link. In more practical terms, congestion on the Internet is signaled by dropped packets—once the queue (buffer) is full, the router at the congested link has no place to store another incoming packet and must discard it. A packet that is “dropped” will signal the sender to slow down⁴ and to retransmit that particular packet. Drops usually occur because a link (meaning the connection between two devices, including routers) is congested, and more specifically because the input queue is full [12].

From a service provider’s perspective congestion is usually defined as high utilization of a link (over 70 or 80 percent of its capacity) for a period of several or more

³In fact, Transmission Control Protocol (TCP), one of the central protocols of the Internet, operates under the assumption that the location of congestion is unknown.

⁴By reducing the size of the congestion window.

minutes [13]. This definition accepts that some packet drops are inevitable, given variations in user demands and the very large size of most provider's networks. Providers are concerned with persistent congestion, which can impair connection quality for multiple users.

Content providers usually care more about the performance of their application on the user's end. Netflix (a large video streaming service), for example, has displayed messages on users' screens warning them that their service provider's network is "crowded." [14] Location of congestion is somewhat irrelevant for content providers; if any link along the path is experiencing severe congestion, performance of the application in question will likely degrade.

Finally, **users** normally do not concern themselves with Internet congestion, so long as their connection and applications are working to their satisfaction. When some noticeable performance impair occurs, such as significant increases in website loading times, users might become aware of network congestion. From there on, they could complain to their network or content providers, or simply stop browsing in that particular moment; the quality of their experience (QoE) with the web is affected.

2.2.3 Quality of Experience (QoE)

According to the International Telecommunication Union (ITU), QoE "refers to the overall acceptability of an application or service, as perceived subjectively by the end user." [10] QoE is, then, inherently hard to measure, and depends on factors other than network connectivity. One could imagine a scenario in which a user's screen is damaged, and the visual quality of content displayed is therefore perceived as poor. No action taken at any location in the network could improve this aspect of the user's experience.

Most cases are not as extreme. Kuipers et al. provide three groups of parameters that affect QoE: quality of the content, human perception, and Quality of Service (QoS.) [10] Of these, only the last is directly measured by the data analyzed in this thesis; however, some QoE inferences can be drawn from changes in network metrics. Faster connection speeds, for instance, theoretically lead to lower loading times; these shorter waits for content generally improve user experience.

QoS is a "set of standards and mechanisms for ensuring high-quality performance for critical [networked] applications ... The goal of QoS is to provide preferential delivery service

for the applications that need it by ensuring sufficient bandwidth, **controlling latency** ... and reducing data loss” (emphasis added.) [15] Although all of these factors impact the quality of a user’s connection, only latency is perceived by the user; most people do not think about (or get frustrated by) how many millions of bits their computer has received in a second (what throughput measures.) Rather, users normally detect longer loading times (latencies), one of the main foci of our analysis on Chapter 3.

2.2.4 Factors Impacting Quality of Experience

QoE, as we have defined it, can be affected by the performance of all the systems involved in delivering content to the user. These include the infrastructure of the target website, the network of the service provider, (leased or owned) content delivery networks, transit providers, and even the quality of the user’s residential wireless connection. In this section, we will briefly discuss each of these factors impacting QoE.

One simple example can illustrate the diversity of QoE-affecting systems: if a target website’s server is not in service, and as a result the user’s browser displays an error message, QoE is severely impaired because *there is no website to experience in the first place*. In reality, QoE ranges from flawless performance to complete outage.

Infrastructure of the target website is critical for QoE as it is the source of much⁵ of the content delivered to the user. In particular, specific resources requested by the user—for example, a news article—need to be fetched and delivered by server-side systems. An explanation of the details of such requests go beyond the scope of this thesis, but suffice to say here that several types of computing infrastructure are involved, including data centers and interconnection links to third-party content providers (such as advertisement companies.) Finally, geographic distance (and, consequently, the distance in network terms) from the user’s home to the target website’s server can increase loading times.

Content delivery networks (CDNs) can greatly improve website performance as seen by the user. CDNs are a “collaborative collection of network elements spanning the Internet, where content is replicated over several mirrored Web servers in order to perform transparent and effective delivery of content ... by maximizing bandwidth, improving accessibility, and

⁵Some content, such as third-party advertisements, could come from other sources.

maintaining correctness" [16]. Further, CDNs "are typically deployed at the edge of ISPs to reduce the latency between the end-host and the content" [8].

Many big content providers (including Google and Netflix) have their own CDN. Many other websites rely on third parties, such as Akamai or Level 3, to improve the quality of their users' experience. **Network delivery techniques**, such as caching (creating local copies of content at the user's computer) or progressive video streaming, can also contribute to the betterment of performance.

Service providers' infrastructure have direct impact on QoE for all networked applications, since most subscribers are *only* connected to the Internet through their home access point. Therefore, congestion on their network, or at any interconnection point with a target website or transit provider⁶, can severely affect QoE.

Finally, **quality of the wireless connection at home** can degrade the performance of any networked application, since it is the final (and, often, only) link to the user's device.

2.2.5 Measuring Broadband America

This thesis is based largely on analysis of the web browsing test produced by Measuring Broadband America (MBA), a project of the F.C.C. aimed at studying broadband performance in the U.S. SamKnows, the Commission's contractor, deployed thousands of white-boxes that connect to the user's home network (either as a bridge or as a router), and perform connection tests to external websites every hour⁷.

MBA's web browsing test measures how long it takes to download the "HTML front page for each web site and all associated images, JavaScript, and stylesheet resources" [17] of nine popular websites: Amazon, CNN, EBay, Facebook, Google, MSN, Wikipedia, Yahoo and YouTube. In addition to loading time, the test provides TTFB for each measurement. These times should be affected by all systems along the connection path, as the download is made from the actual website.

⁶We use this term loosely to refer to any network not belonging to either the user's service provider or the target website. These include, but are not limited to, backbone providers, CDNs, and other service providers.

⁷Other tests are performed to centralized nodes of Google's M-Lab and Level 3's infrastructure; however, this thesis only uses the data from the web browsing test, which connects exclusively to external websites.

Chapter 3

Analysis of Web Browsing Test Data

An Open Internet ... means consumers will demand more and better broadband as they enjoy new lawful Internet services, applications and content.

Federal Communications Commission, 2015.

3.1 Measuring Broadband America's Web Browsing Test

3.1.1 Technical Overview

In this section we present an overview of the F.C.C.'s web browsing test, one of fourteen tests run by SamKnows whiteboxes. Every hour¹, the test node attempts to download the HTML frontpage and associated resources for each of the following URLs²:

- <http://edition.cnn.com>
- <http://www.amazon.com>
- <http://www.ebay.com>
- <http://www.facebook.com/policies>
- <http://www.google.com/mobile>

¹Tests were not run when "there was any [user-generated] Internet activity beyond a defined threshold value." [18]

²Extracted from the raw data released by the F.C.C.

- <http://www.msn.com>
- <http://www.wikipedia.org>
- <http://www.yahoo.com>
- <http://www.youtube.com>

In our analysis, we have looked at four variables in the web browsing test’s raw data, released to the public by the F.C.C.³: total download time, time to first byte (TTFB), resolved IP address, and number of bytes transferred. Relevance of some of these metrics for QoE is discussed below; definitions and relevance of TTFB and IP addresses is included in Chapter 2.

Total download time measures how long it took the whitebox to fetch all the initial resources of the website. Higher loading times generally lead to QoE degradation; a delay as small as 100 milliseconds can be noticed by humans [19]. Further, a page loaded in a full second or more will lose the user’s full attention, and after ten seconds the user might switch to a different task altogether [20].

An important limitation of the total download time as reported by the F.C.C. (discussed more extensively in Chapter 4) is its failure to account for javascript and other resource execution, which constitutes a critical portion of complex websites’ content. As a result, reported loading time is an underestimate of what users might experience.

Finally, **number of bytes transferred** indicates the size of the page fetched by the whitebox. This metric will be relevant mostly for our discussion of Facebook’s drop in loading times (see Section 3.2.6.)

3.2 Analysis of F.C.C. Web Browsing Test

We analyzed the MBA web browsing data for all of 2013 for units located in the Northeastern, Southern and Pacific United States, and this section will present the results⁴. We find

³As labeled in the data dictionary (available on the F.C.C. website): fetch time, bytes total, ttfb avg time, and address

⁴We present a small subset of the tables, charts and statistical analyses of the data. For the complete set, as well as the scripts and filtered data, please refer to the Thesis Supplemental Materials, available on

evidence of diurnal variation for most providers and targets, in the form of increased website loading times at peak hours (between 7 and 11pm⁵) and in the middle of the day (between 11am and 3pm). We also find significant variation of loading times to different websites from a single provider's network.

In the data presented in this section, there is also evidence of: variation in loading times of a single website across multiple providers; localized and persistent network congestion; and load-sharing schemes on websites' server-ends.

Finally, for the specific case of Facebook, we discuss a sharp drop in loading time and number of bytes transferred in May 2013.

3.2.1 Data Included

We include data from units located in the Northeast, South and Pacific regions of the U.S. Table 3.1 presents which states are included in each region⁶. Our analysis includes units in homes connected to the Internet using the three prevailing communication technologies in the country: cable, DSL and fiber.

Internet Service Providers to analyze were selected among those with the highest number of units present in the F.C.C. study, with the additional goal of including a diversity of advertised speed tiers. Table 3.2 shows the network operators included in our analysis for each region, as well as number of units and communication technology for each provider⁷.

3.2.2 Loading Time Boxplots

Our analysis shows similar distributions for loading times of a single website across multiple providers, highlighting the importance of server-side infrastructure, content delivery networks and peering interconnection agreements. Figures 3-1 and 3-2 show loading time distributions for all websites⁸ under study, measured from two providers in the Northeast: Comcast and

the web at dspace.mit.edu under the author's name.

⁵We use the same definition of "peak time" as the F.C.C. [17].

⁶Within a region, time zones are consistent, which facilitates parsing and filtering.

⁷Note that we analyze a **subset** of all data collected by the F.C.C. in 2013.

⁸Note that, due to a sharp and persistent drop in loading times for Facebook in May 2013 (discussed later in this chapter), only data for January through April is included.

Region	States
Northeast	CT MA ME
	NH NJ NY
	PA RI VT
South	AL AR LA
	MS OK TX
Pacific	CA NV OR
	WA

Table 3.1: States included in each region for the data analysis presented in this chapter.

Cablevision⁹.

Since all target websites for a single provider connect to the user’s browser using the same “last mile” infrastructure, differences in website performance can be attributed to systems outside the provider’s network. This is reinforced by the fact that websites have similar loading time distributions across different providers, as shown on Figures 3-1 and 3-2 for Comcast and Cablevision, respectively.

In these boxplots, Google and YouTube show smaller boxes that are also located closer to the x-axis, which indicates consistently good performance¹⁰. Conversely, EBay and Yahoo have wider distributions (bigger boxes) located farther away from the x-axis, both indications of more variability and worse performance.

In the Pacific and Southern U.S., the other two regions under study, similar patterns are seen on target website loading time boxplots. Figures 3-3 and 3-4 present the boxplots of data measured from the Verizon Pacific and Cox Southern networks, respectively. Note that our previous inference holds: YouTube and Google still show short distributions closer to the bottom of the plot, signaling better performance. On the other hand, Amazon, EBay and Yahoo show wider distributions located farther away from the x-axis, signaling worse

⁹Though the advertised speeds are not identical, differences in web browsing loading time attributable exclusively to throughput are small. For example, for a page size of 1,510 KB, the median page size of CNN measured from Cablevision—which is also the largest page size of all nine websites—the theoretical loading time on a 15 Mbps connection is 805 ms; on a 20 Mbps the loading time is 604 ms. For all other websites the difference in theoretical loading time is smaller, given their smaller page size.

¹⁰Wikipedia’s distribution is similar to Google and YouTube’s, but this is likely a consequence of its small page size (a median of just under 60 KB for both Comcast and Cablevision). Google and YouTube’s loaded page size is ten to twenty times higher than that of Wikipedia.

Region	Provider	Advertised Speed	Number of Units	Technology
Northeast	Cablevison	15	100	Cable
Northeast	Comcast	20	108	Cable
Northeast	Frontier	0.8 to 5	35	DSL
Northeast	Verizon	15 to 25	87	Fiber
Northeast Total	-	-	330	-
Pacific	AT&T	0.8 to 12	67	DSL
Pacific	Charter	30	81	Cable
Pacific	TWC	15	125	Cable
Pacific	Verizon	25 to 50	36	Fiber
Pacific Total	-	-	309	-
South	AT&T	0.8 to 12	67	DSL
South	Cox	12 to 25	50	Cable
South	TWC	15 to 20	53	Cable
South	Verizon	25 to 50	24	Fiber
South Total	-	-	194	-
All Regions Total	-	-	833	-

Table 3.2: Providers, advertised speeds, number of units (whiteboxes), and technologies studied for each region.

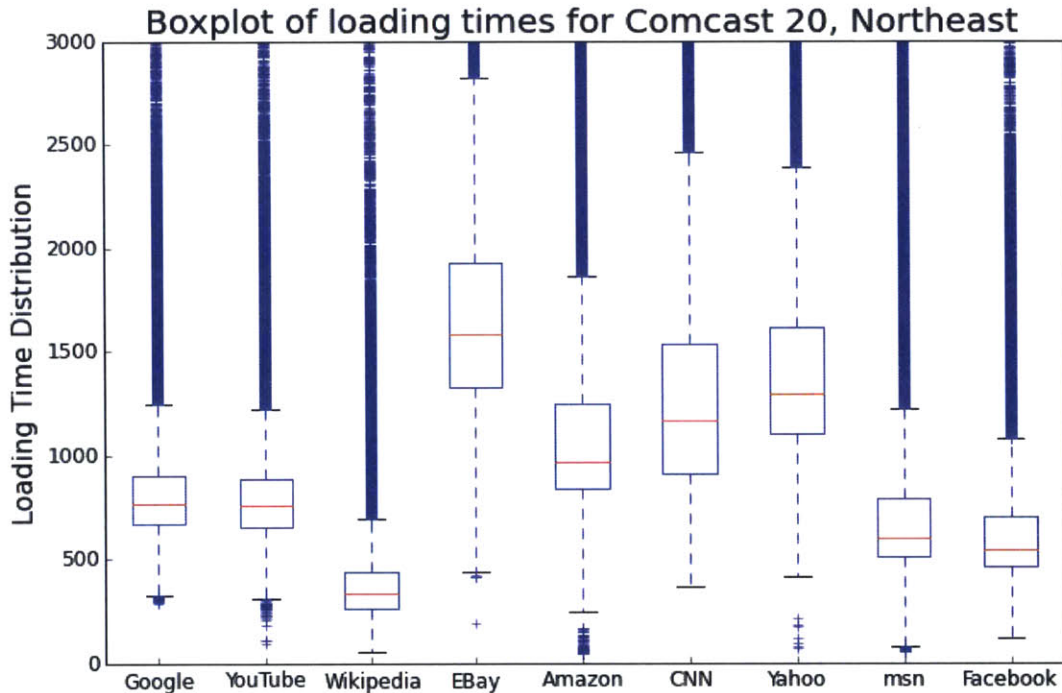


Figure 3-1: Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Comcast network in the Northeast, with an advertised speed of 20 Mbps. The boundaries of the blue boxes represent the first and third quartiles, and the red line is the median. The whiskers are outliers beyond 1.5 times the inter-quartile range, shown as a black bar. Longer distributions (signaled by bigger boxes) are an indication of higher variability in loading time, whereas boxes located closer to the x-axis indicate lower loading times. Facebook’s data is for Jan-Apr.

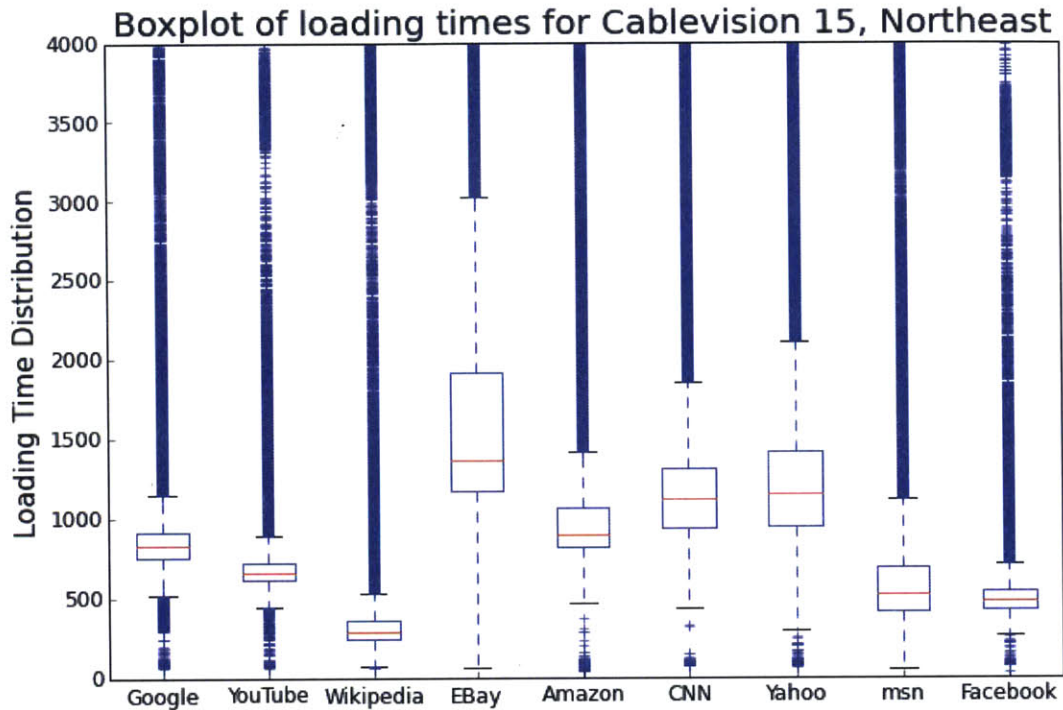


Figure 3-2: Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Cablevision network in the Northeast, with an advertised speed of 15 Mbps. The boundaries of the blue boxes represent the first and third quartiles, and the red line is the median. The whiskers are outliers beyond 1.5 times the inter-quartile range, shown as a black bar. Longer distributions (signaled by bigger boxes) are an indication of higher variability in loading time, whereas boxes located closer to the x-axis indicate lower loading times. Facebook’s data is for Jan-Apr.

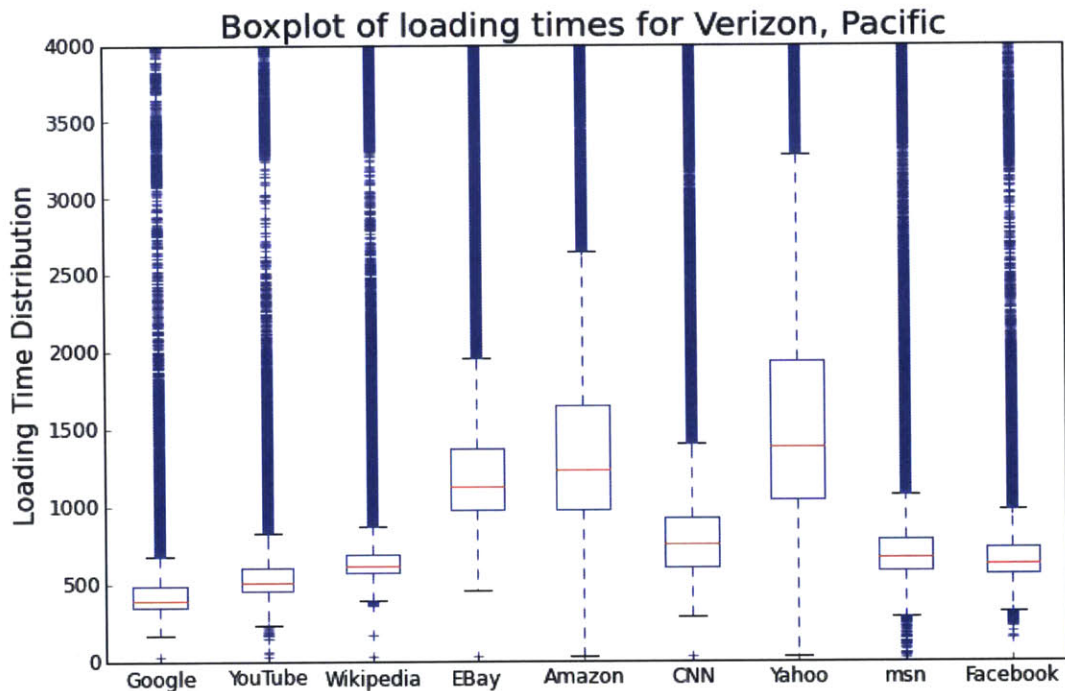


Figure 3-3: Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Verizon network in the Pacific, with advertised speeds between 25 and 50 Mbps. The boundaries of the blue boxes represent the first and third quartiles, and the red line is the median. The whiskers are outliers beyond 1.5 times the inter-quartile range, shown as a black bar. Longer distributions (signaled by bigger boxes) are an indication of higher variability in loading time, whereas boxes located closer to the x-axis indicate lower loading times. Facebook's data is for Jan-Apr.

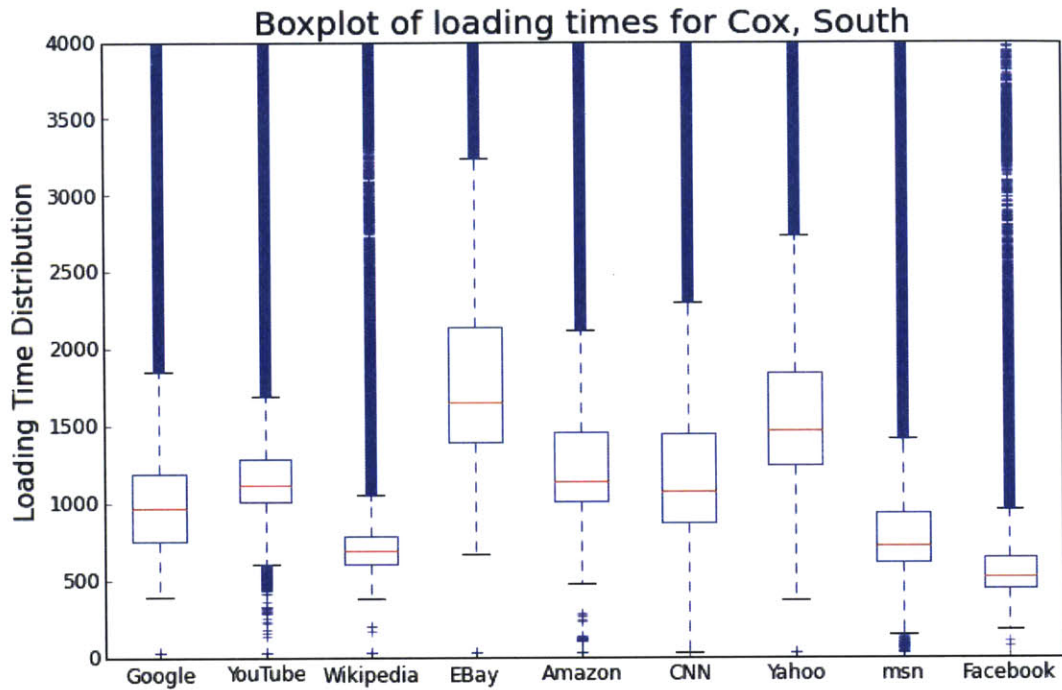


Figure 3-4: Boxplot of total loading times (in milliseconds) of different websites for all of 2013, measured from the Cox network in the South, with advertised speeds between 12 and 25 Mbps. The boundaries of the blue boxes represent the first and third quartiles, and the red line is the median. The whiskers are outliers beyond 1.5 times the inter-quartile range, shown as a black bar. Longer distributions (signaled by bigger boxes) are an indication of higher variability in loading time, whereas boxes located closer to the x-axis indicate lower loading times. Facebook’s data is for Jan-Apr.

performance.

A simple hypothetical example can illustrate the importance of network and computing infrastructure outside the user's Internet Service Provider (ISP.) If website X delivers its contents to users in the Northeast using CDN Y, and there is congestion on the peering interconnection point between Y and the user's provider Z, the measured performance of X from the user's home—what the F.C.C. measures—will be degraded. Similarly, if X is experiencing high demand on its server-side, such as a high number of database queries, measured performance can also degrade as a result. In reality, network congestion anywhere along the connection path¹¹ and higher server-side loads can constructively interfere to degrade the user experience.

3.2.3 Localized Congestion

We find evidence of network congestion and diurnal variation of loading times for at least some websites for all providers in this study. Some instances of congestion dissipate after a few days or weeks, whereas others are persistent, particularly towards the end of 2013. As can be seen in the time series in this chapter, performance across websites varies wildly; Google and YouTube (owned by Google) generally show the lowest loading times, while Wikipedia shows the highest.

Let us use this paragraph to explain the methodology employed to produce the time-series plots in this chapter that relate to F.C.C.-collected measurements. For each chart, the data presented is for all units in a particular ISP's network in a particular region. Then, some of the following raw data is plotted for the specified period: total loading time, time to first byte, and IP address.

A particularly striking and localized instance of congestion is shown on Figure 3-5 for Cablevision in the Northeast. The spike in early July (around day 190 of the year) is visible on the loading times of Wikipedia and Yahoo, but not on those of YouTube, nor any other of the target websites not shown on Figure 3-5. Such spike in loading times is not present for Wikipedia and Yahoo when accessed from other providers in the Northeast, as seen for Comcast in Figure 3-6; which implies higher loading times are unlikely to be a consequence

¹¹Including, of course, the ISP's internal network.

of slow responses at the website's server-end. All this evidence suggests that a problem exists outside both Cablevision's and Wikipedia's networks, perhaps at a peering interconnection point between either of them and a transit provider or CDN.

3.2.4 Diurnal Variation

Increased loading times at peak periods of each day for several weeks or months can signal underprovisioned interconnection links, as well as overloaded website server-end systems. We find evidence of diurnal variation for at least some websites for all providers under study.

Clear diurnal variation can be seen in Figure 3-7 for Wikipedia in the Northeast, when accessed from the Verizon network¹². From 7:00pm to 11:59pm, when there is the most load on residential networks, Wikipedia's median loading time increases by 78% from its all-day median, from 257 to 780 milliseconds. By comparison, Wikipedia's loading time on the Comcast network increases by 1% at peak times, which suggests that performance degradation is unlikely to be a consequence of server-side load. Rather, higher loading times at peak hours is likely to be the result of a congested link somewhere along the connection path from Verizon users' homes to Wikipedia's servers.

Wikipedia is not the only website showing diurnal patterns on the Verizon network for early 2013. As seen on Figure 3-8, zooming in to the month of January shows diurnal variation in loading times for Amazon as well.

We find several other instances of diurnal variation visible in our charts. One such incidence is shown on Figure 3-9, for CNN and Comcast in the Northeast. Dispersion in loading times greatly increases at peak times, particularly during the first half of 2013, and can be seen by comparing the plot on the left (peak time) to that on the center (3-7am) and that on the right (11am-3pm). During peak hours, the median loading time for CNN increases by 24.9% when compared to its 24-hour value (from 1,191 ms to 1,488 ms.)

Another provider and target website combination showing strong diurnal patterns is shown on Figure 3-10 for Time Warner Cable and Yahoo in the South. As can be seen in

¹²Although not shown here, the strong diurnal pattern is also present for Verizon and Wikipedia in the South and the Pacific.

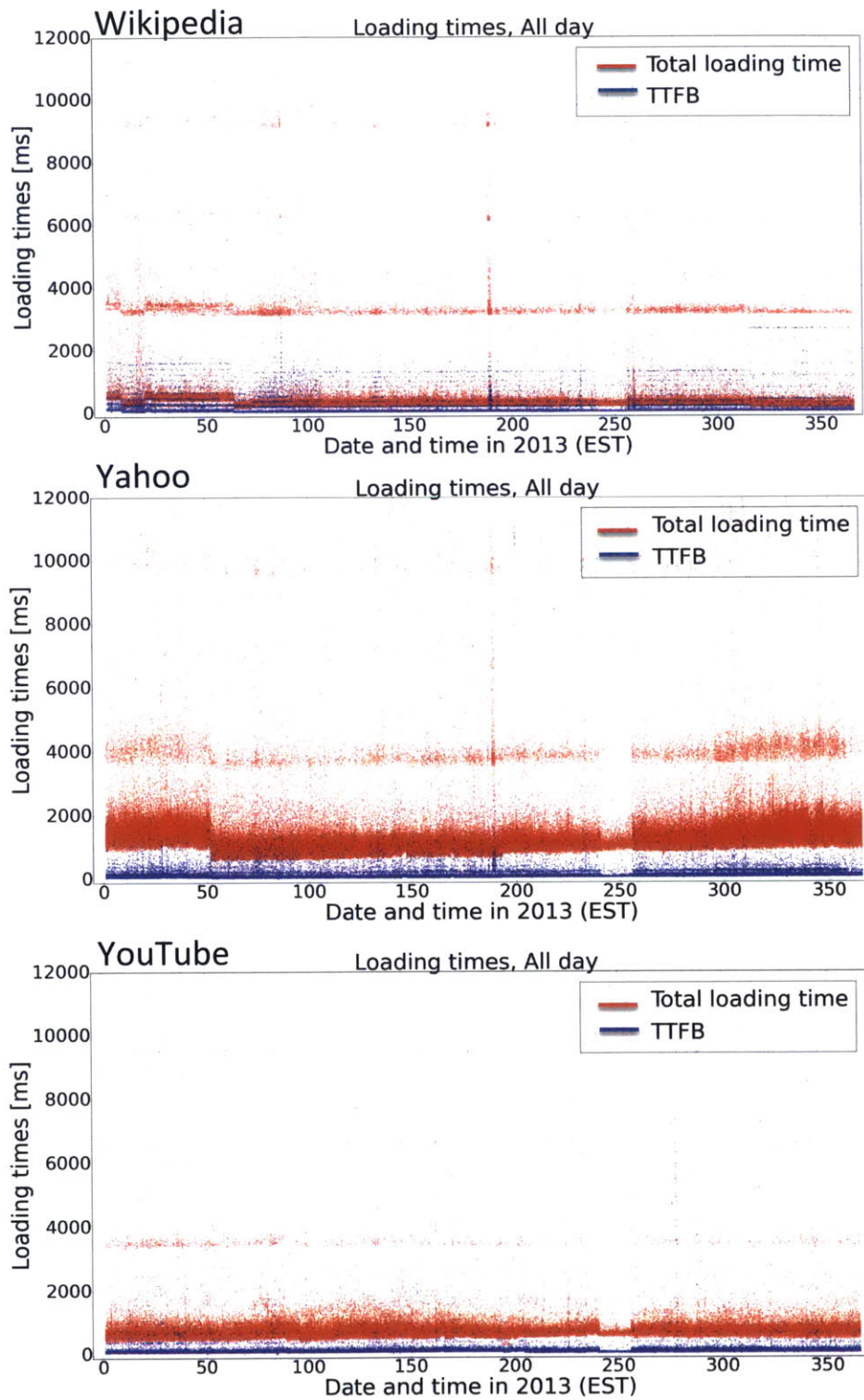


Figure 3-5: Yearlong time-series of loading times from the Cablevision network (15 Mbps) in the Northeast.

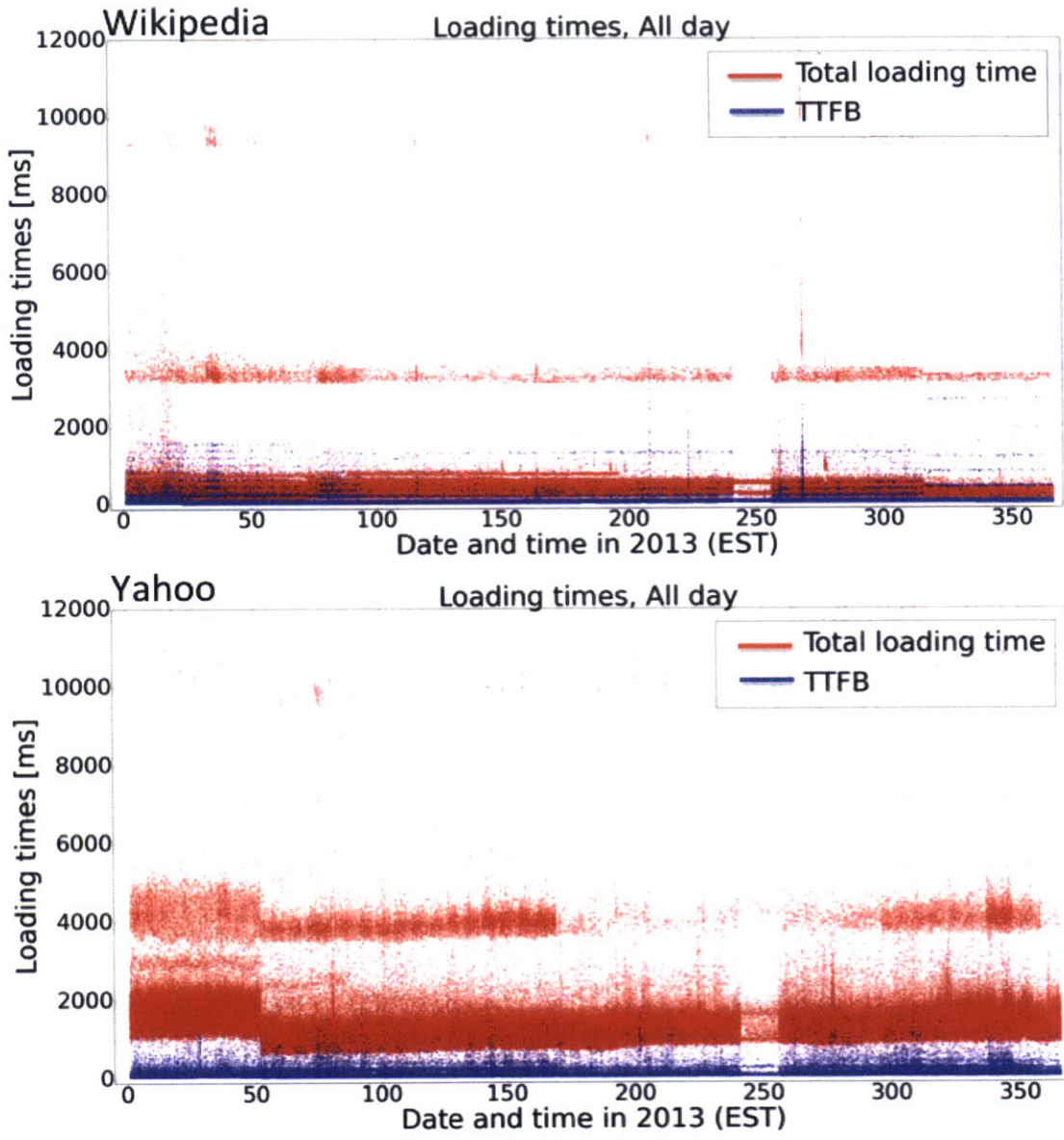


Figure 3-6: Yearlong time-series of loading times from the Comcast network (20 Mbps) in the Northeast.

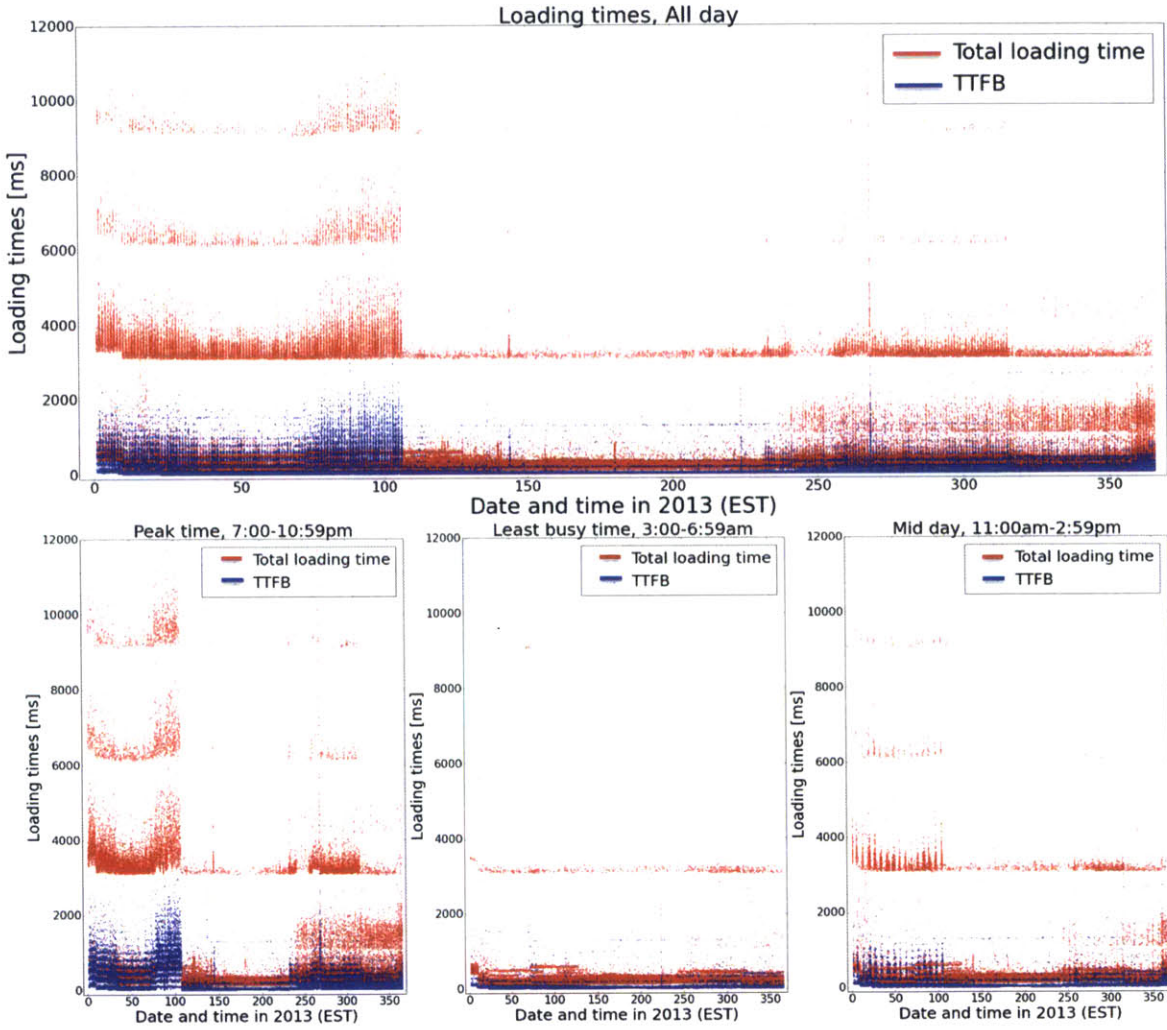


Figure 3-7: Yearlong time-series of loading times for Wikipedia, measured from units connected to the Internet through the Verizon network in the Northeast, with advertised speeds between 15 and 25 Mbps.

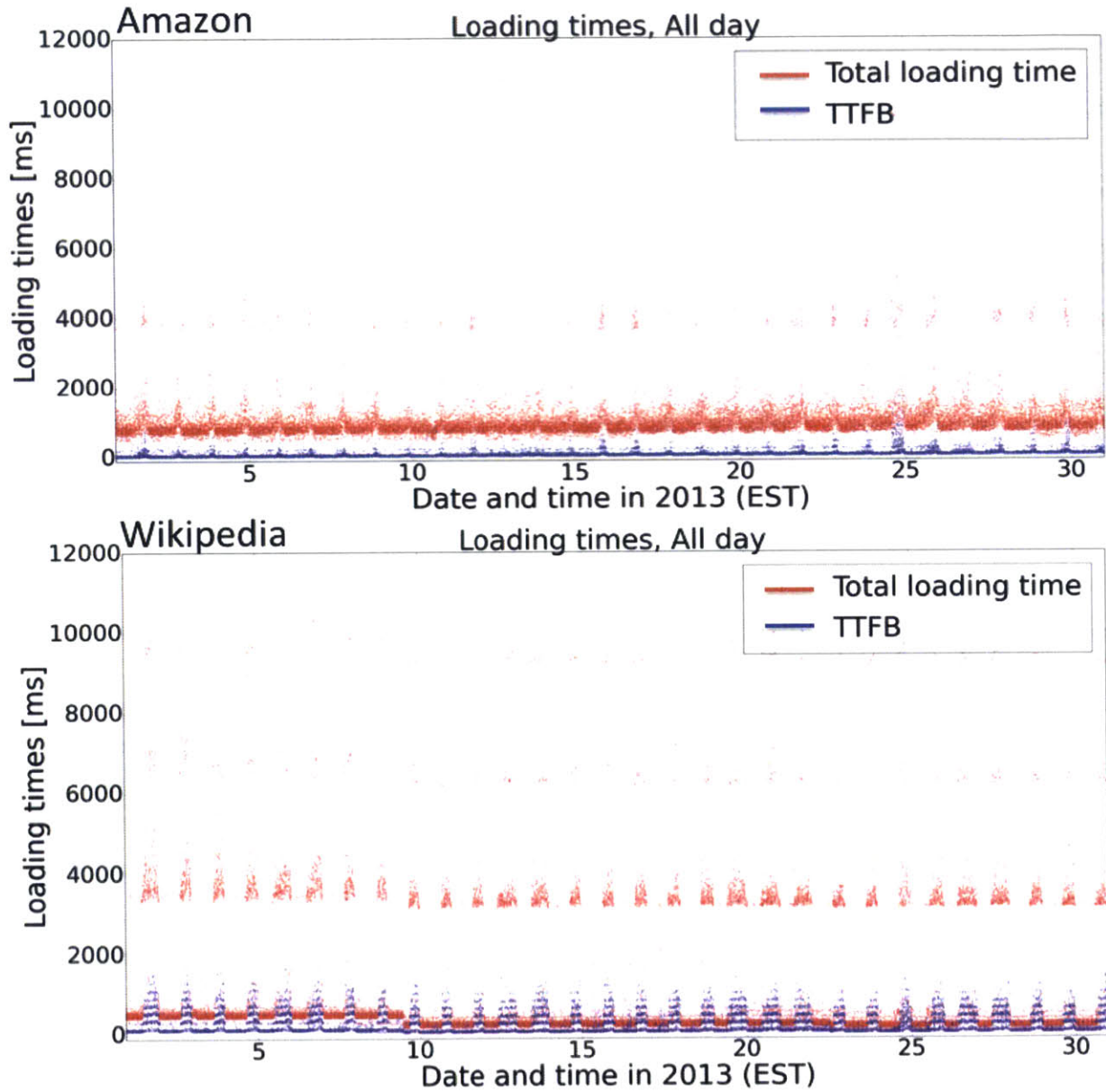


Figure 3-8: Time-series of loading times for Amazon and Wikipedia in January 2013, measured from units connected to the Internet through the Verizon network in the Northeast, with advertised speeds between 15 and 25 Mbps.

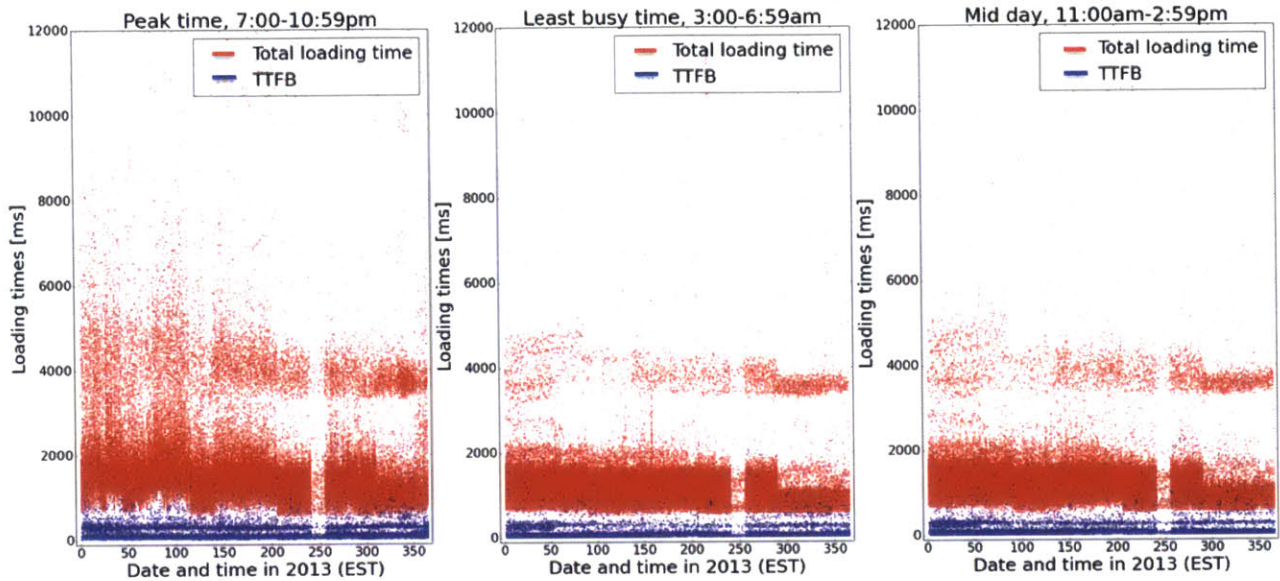


Figure 3-9: Yearlong time-series of loading times for CNN, measured from units connected to the Internet through the Comcast network in the Northeast, with an advertised speed of 20 Mbps.

this thesis’s supplemental materials¹³, diurnal patterns are present for all providers in the three U.S. regions studied.

3.2.5 Step Changes on Loading Times for a Specific Website

We find evidence of step changes in loading times for a single website when accessed from multiple providers. For example, *msn*, a content provider owned by Microsoft, shows similar patterns of change in total loading time during the second quarter of 2013, for ten of the twelve providers we analyzed. Data for four of these providers—Comcast in the Northeast, Time Warner Cable (TWC) in the Pacific and Cox and AT&T in the South—are shown on Figure 3-11.

Given the diversity in providers and geography for which this step change is present, it is unlikely a result of localized congestion in the providers’ network. This hypothesis is reinforced by the inexistence of a step change in loading times for other websites during that period.

Rather, evidence suggests there was a change in *msn*’s infrastructure. For example, there

¹³ Available at <http://dspace.mit.edu/> under the author’s name.

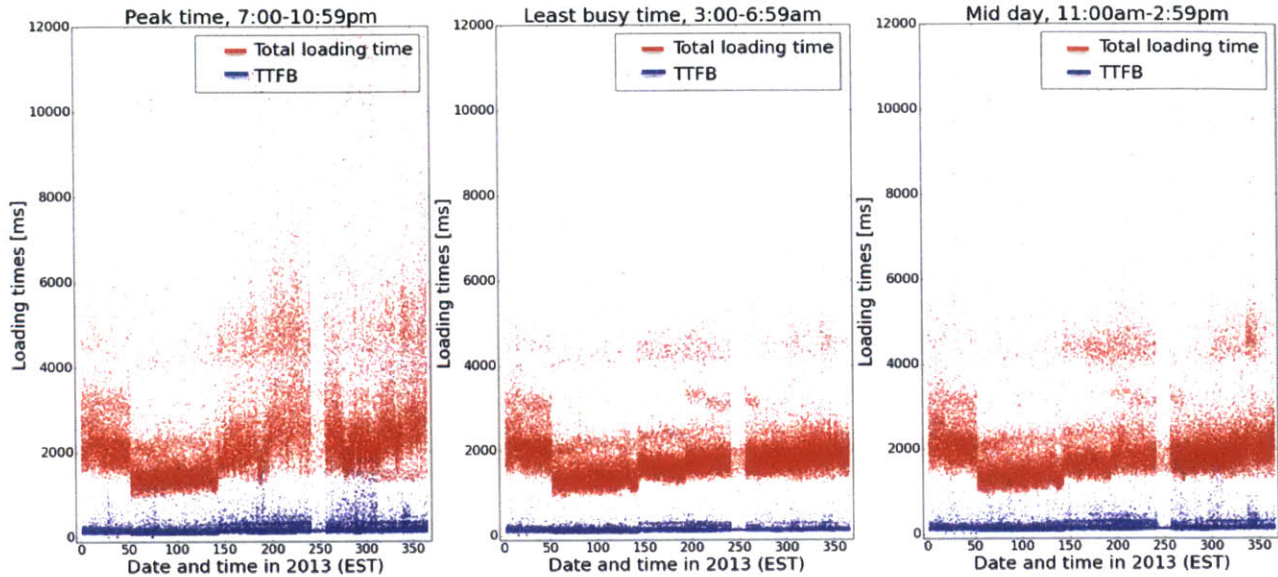


Figure 3-10: Yearlong time-series of loading times for Yahoo, measured from units connected to the Internet through the Time Warner Cable network in the South, with advertised speeds between 15 and 20 Mbps.

could have been a change in the way content is delivered to users, such as the use of a different CDN. Another possibility concerns the server-side infrastructure: msn could have been doing a major hosting upgrade during that period, resulting in increased response times to users' requests. Yet another possibility deals with the content of the msn website itself: changes in image compression can yield massive changes in loading times, since images are a sizable portion of most website's content (in terms of bytes transferred.) [21]

If the loading time changes described above are a consequence of msn's server-side modifications, IP addresses can provide an indication of such modifications. However, not all infrastructure modifications would result in IP address changes, given the use of multiplexing techniques such as Network Address Translation. IP addresses are only an imperfect metric for estimating size and configuration of website infrastructures.

For the case of msn, we observe IP address changes that correspond to the step changes on loading times, as can be seen in the period highlighted on Figure 3-12 for Cox in the South. Similar patterns to that observed for Cox are also present for all providers under study, underscoring the likelihood of a change on msn's server-side or delivery network. Importantly, the decrease in loading times—towards the end of the highlighted period—corresponds to the

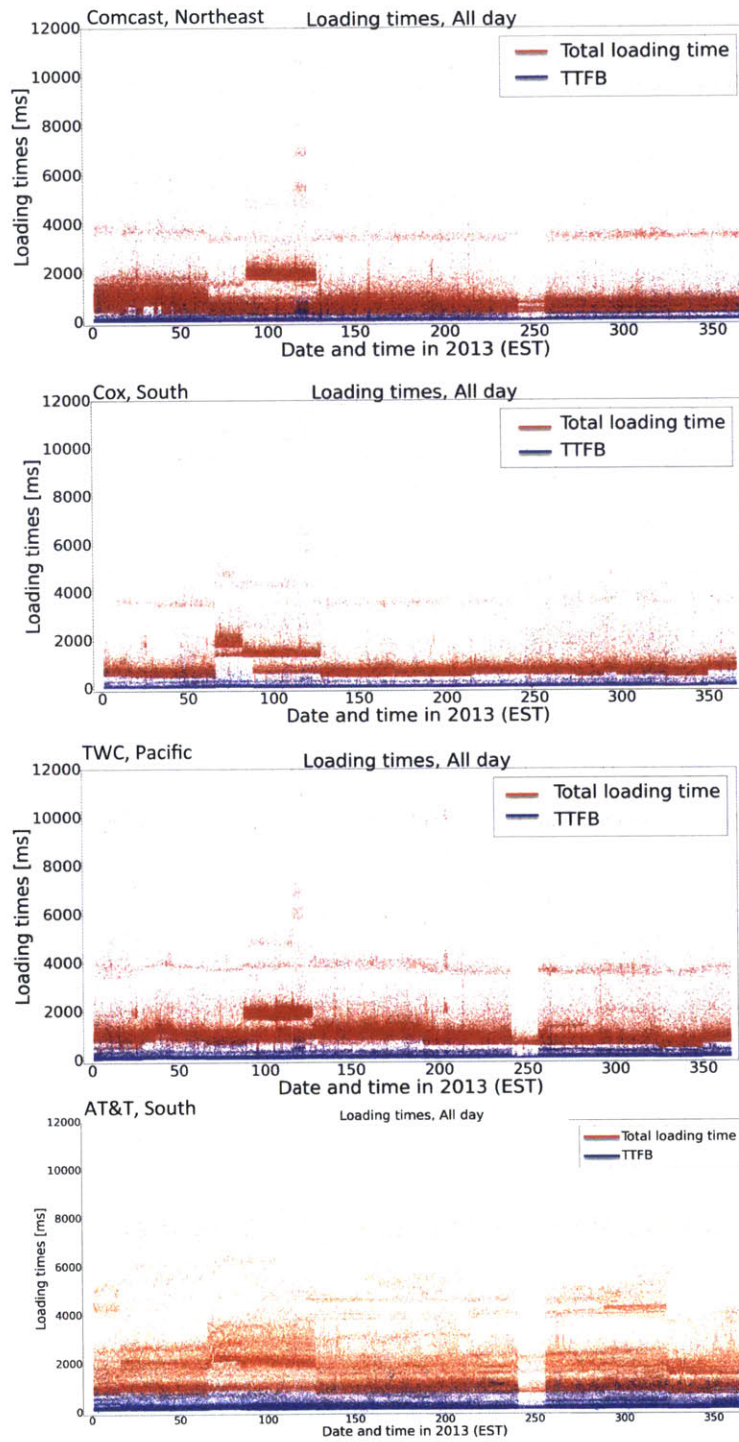


Figure 3-11: Yearlong time-series of loading times for msn, measured from four different providers in three regions. Note that although loading times are different in nominal terms, a similar increase can be seen during the second quarter (days 90-120).

use of a different IP address for delivering content for the requested URL.

Another example of server-end load-sharing, for the case of YouTube, can be seen on Figure 3-13. The two highlighted periods show a similar behavior to that described above for msn: changes in loading times correspond to IP address changes, suggesting load-sharing schemes on the server-end. These schemes are used to reduce loading times and presumably improve QoE, and can be seen for four target websites: Google, YouTube, Facebook and msn.

3.2.6 Drop in Facebook Loading Time and Bytes Transferred

Facebook's data shows a drastic drop in loading time and bytes transferred in early May, 2013, for all providers under study. Then, the number of bytes transferred falls to just 776—down from around 667 KB—and stays at this new value for the rest of the year. This factor of a thousand drop in bytes transferred is reflected on the total loading time shown on Figure 3-14, for data collected on the AT&T network on the West Coast.

Our hypothesis is that the 776 bytes correspond to an error message displayed by an anti-crawling mechanism deployed by Facebook. The social networking site's servers could be detecting that the whiteboxes are accessing the same (policies) website at around the same time every hour, a potential indication of crawling behavior.

3.3 Analysis of MIT-Collected Data

We set up a network probe¹⁴, located at MIT's Computer Science and Artificial Intelligence Lab, that measured total loading times to the nine websites studied by the F.C.C. We used a script provided to the public by SamKnows¹⁵. The test ran every five minutes from March 13th to June 20th, 2015. This section presents some of the results.

MIT-collected data was collected over a year after the most recent data included in the previous section. Results from this new data, then, cannot be compared in nominal terms

¹⁴The complete setup of the network probe, as well as some initial data analysis, was done by Steven Bauer at MIT.

¹⁵In a private communication between Steven Bauer (MIT) and Sam Crawford (SamKnows Limited), the latter confirmed that the test we are running is similar to that run by the whiteboxes.

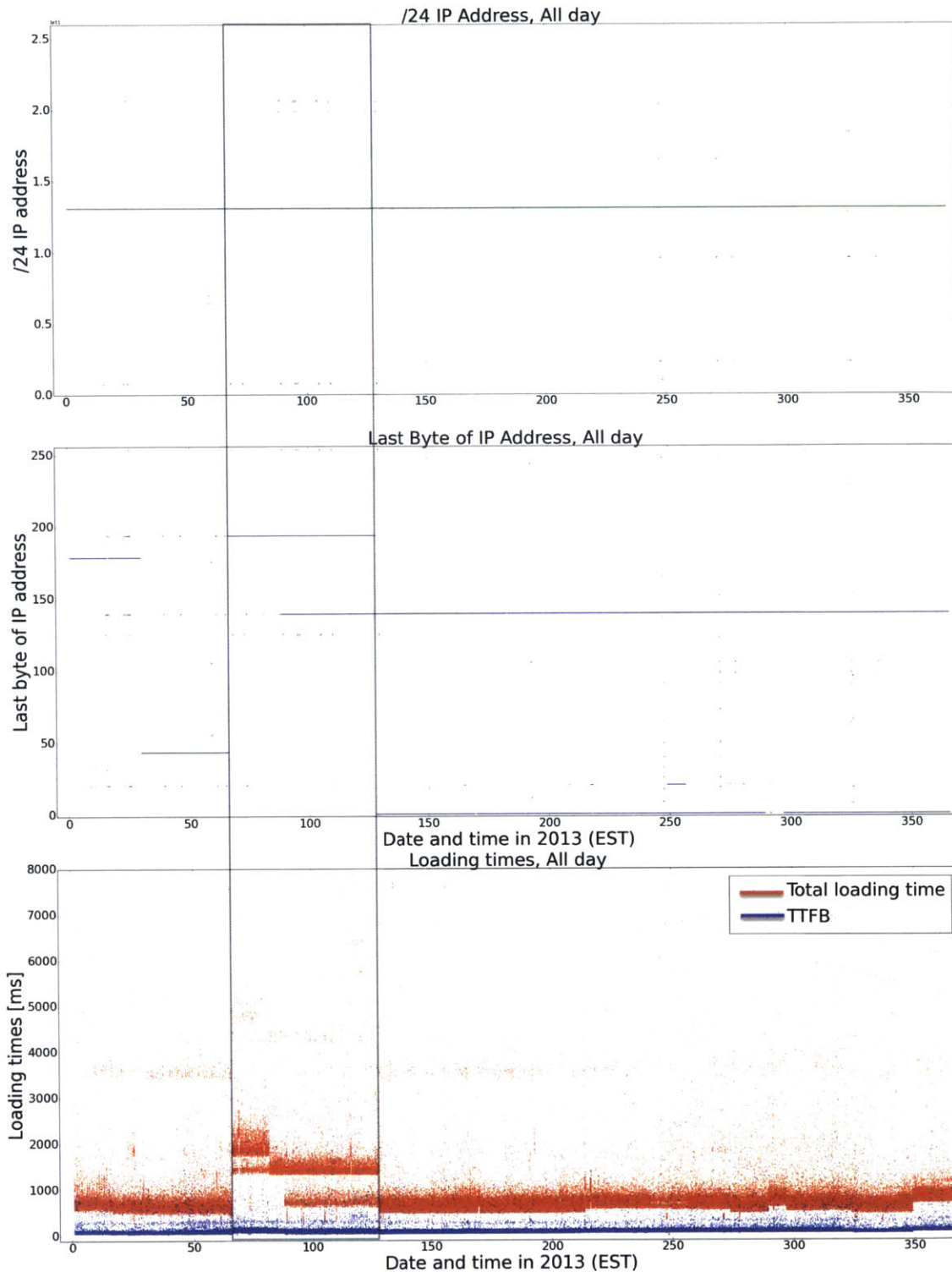


Figure 3-12: IP address and loading times for msn, measured from the Cox network in the South. The highlighted period shows variations in loading times that correspond to changes in the server's IP address. The last byte of the IP address is shown separately to highlight host changes.

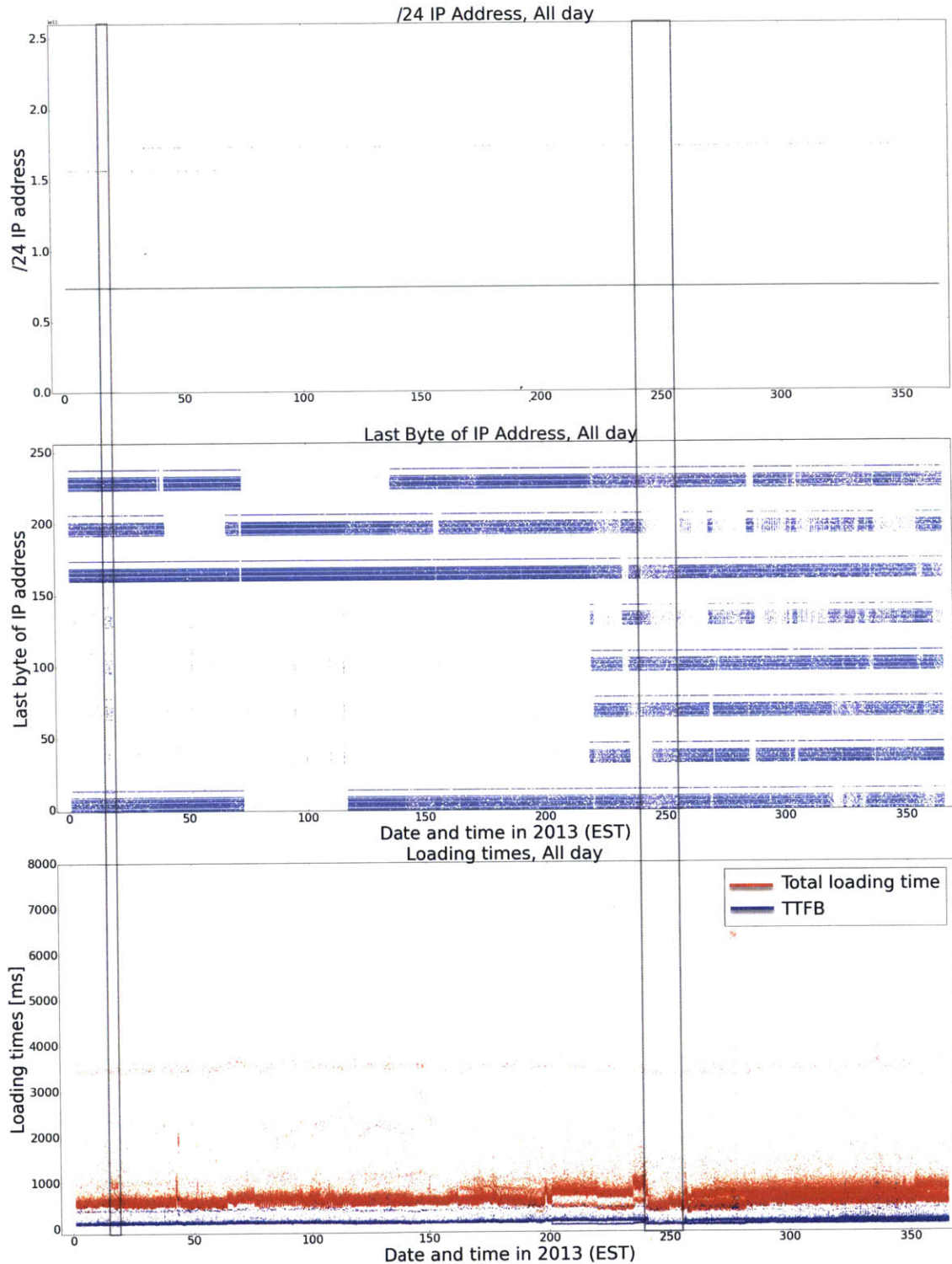


Figure 3-13: IP address and loading times for YouTube, measured from the Verizon network in the Pacific. The highlighted periods show variations in loading times that correspond to changes in the server’s IP address blocks. The last byte of the IP address is shown separately to highlight host changes.

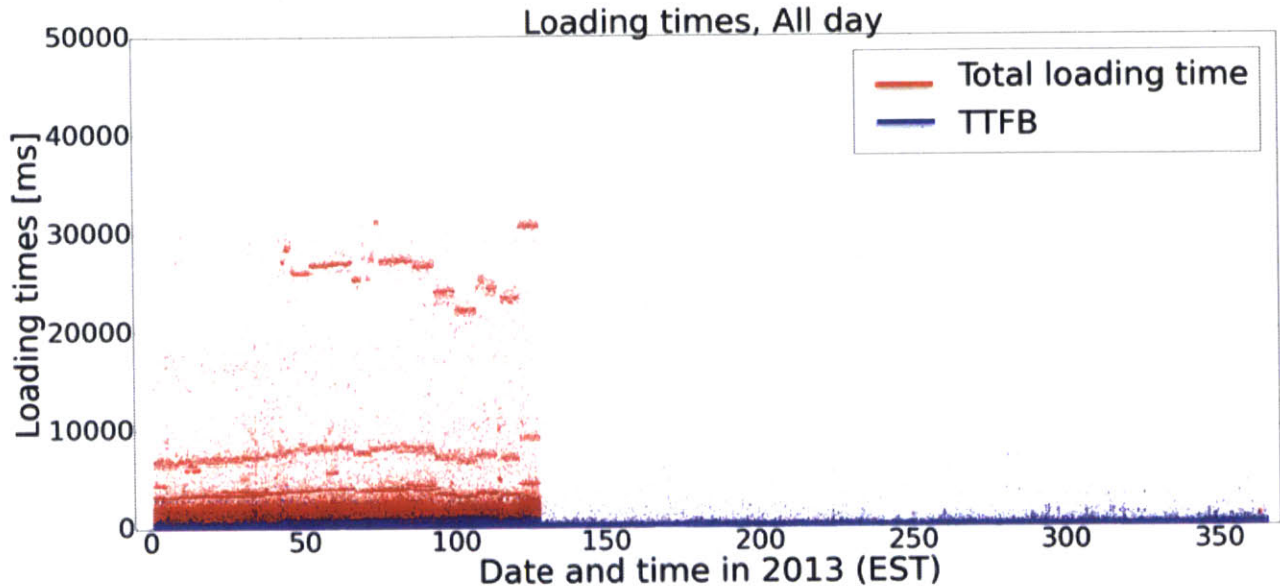


Figure 3-14: Total loading time for Facebook, measured from the AT&T network in the Pacific. The visible drop in loading times occurs on May 7th, 2013.

to those measured by the F.C.C.

Our analysis, in consequence, focuses on confirming two of the general inferences on web browsing. First, that the loading times depend heavily on infrastructure outside the broadband access provider; and second, that the presence of network congestion—including that inside the provider’s network—leads to diurnal variation in loading times.

Since MIT acts as an ISP for devices inside its network, and has network connectivity that is superior¹⁶ to all the units included in our analysis of the F.C.C. data, variation in total loading times are likely the result of systems outside the MIT’s network.

Figure 3-15 shows one instance of **step changes in loading time** for Facebook, when measured from MIT. As in Section 3.2.5, the highlighted periods on Figure 3-15 show increased loading times and a corresponding change in IP addresses on the server-side. Importantly, reductions in loading times correspond to content being served to our network probe from a new IP address block. Although not shown here, similar load-sharing patterns are seen in the MIT data for Google, YouTube and msn.

Differences in performance for multiple websites—when measured from the MIT network—

¹⁶In terms of nominal download and upload speeds.

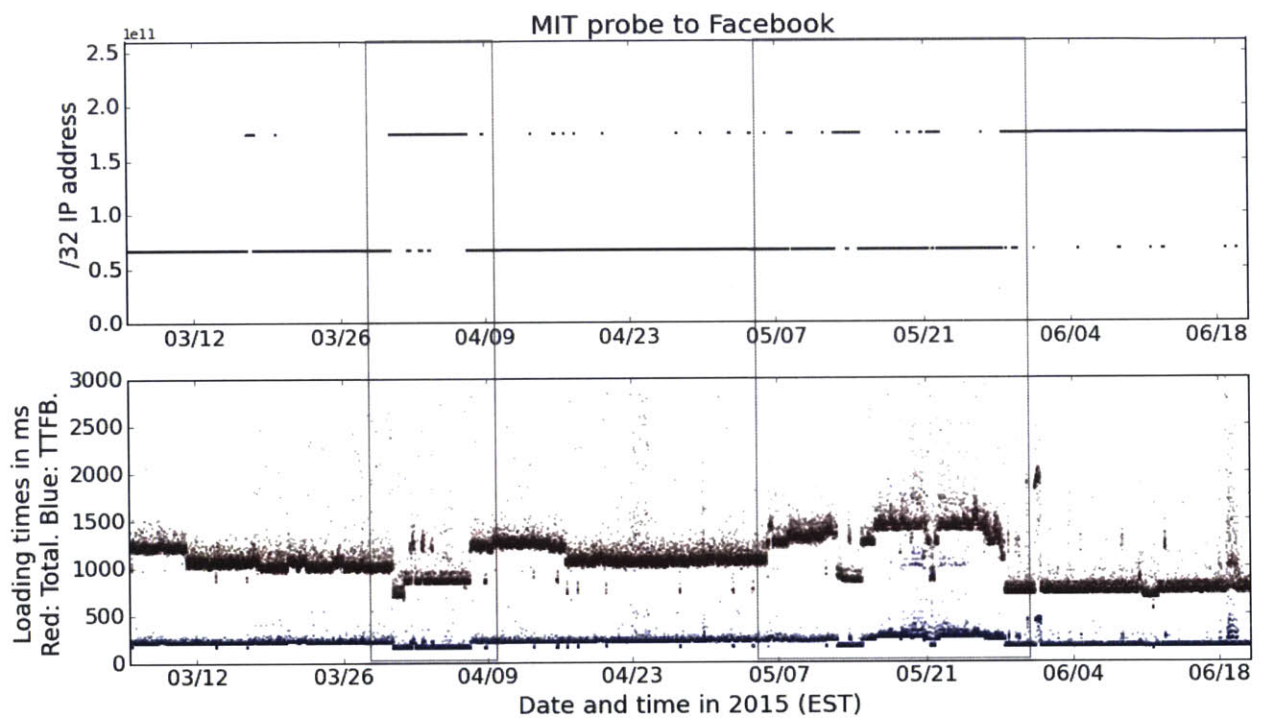


Figure 3-15: IP address and loading times for Facebook, measured from MIT. The two highlighted periods show variations in loading times that correspond to changes in the server's IP address blocks.

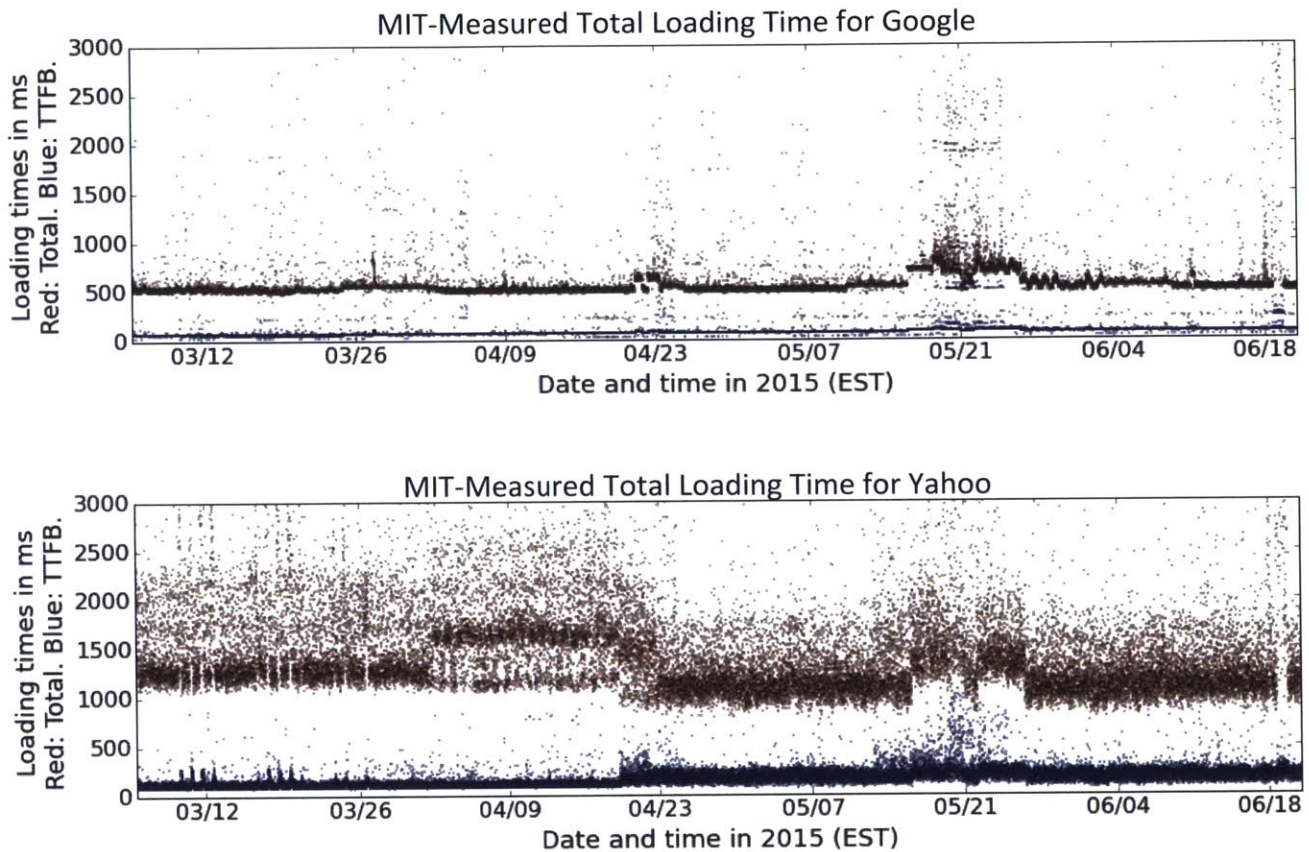


Figure 3-16: Total Loading Times for Google and Yahoo, measured from MIT.

can indicate differentials in target websites' hosting and delivery infrastructure. Lower loading times, as those seen for Google on Figure 3-16, are an indication of lower latency than that for Yahoo, which shows higher loading times on the same chart. Further, total bytes transferred for Google are more than twice those for Yahoo, so the differential in effective throughput is higher.

Finally, Figure 3-17 shows histograms of loading times for the nine websites studied by the F.C.C., measured from the MIT network. Google and msn show left-skewed, narrow distributions signaling better performance. This coincides with our analysis of the F.C.C. data on Section 3.2.2.

Wikipedia also shows good performance on Figure 3-17, but its very small page size (a median of 90 KB for the MIT data) is likely an important factor in loading times. Amazon

and Yahoo, on the other hand, exhibit longer tails in their distributions, signaling higher variability and increased loading times, both signals of worse performance.

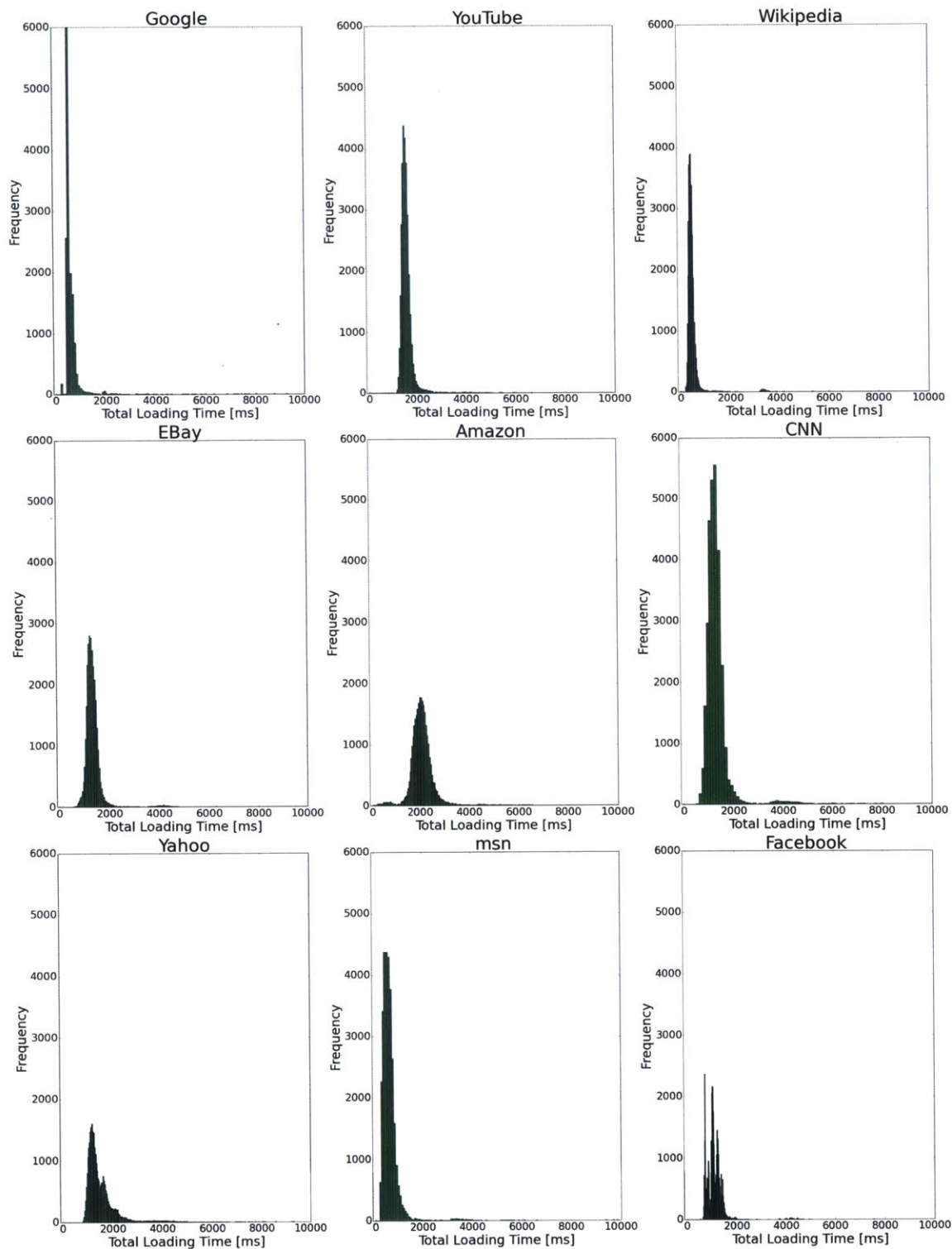


Figure 3-17: The vertical axis shows how many measurements of total loading time fell into each bin on the horizontal axis. Wider distributions are an indication of higher variability in loading time, whereas higher peaks on the left side indicate lower loading times.

Chapter 4

Implications for Quality of Experience Measurements and Public Policy

Start-ups, nonprofits and many other organizations—Wikipedia, for example—that use the Internet do not have the deep pockets of, say, Fox News or NBC to spend on enhanced access to their content.

Tim Wu in *Scientific American* [22], 2015.

Mr. Wu’s above statement discusses differences in performance for diverse websites, and is in some respects accurate: as we have discussed earlier, the infrastructure of the target website *does* greatly impact performance experienced by the users. This chapter will provide some recommendations for technical changes that can lead to better understanding of that differential in website performance. We begin by explaining the relative value of web browsing data when compared to throughput metrics, which have been the central focus of the F.C.C. reports to date. Then, we discuss why loading times measured by SamKnows are an underestimate of those experienced by the user. Further, we provide evidence of the impact of a target website’s resources, on performance, as measured by median web loading times and diurnal variation.

4.1 Relative Value of Web Browsing Data

Web browsing data may more adequately capture user experience than throughput measurements, especially for homes with high-speed connections. The F.C.C. dedicates only a marginal portion of its Measuring Broadband America reports to web browsing analysis, and instead focuses almost exclusively on throughput metrics for residential access networks. Throughput can only partially characterize broadband performance: as Sundaresan, Feamster et al. report on [8], throughput is the major factor determining page load times only for connections up to 16 Mbps. For higher speeds, latency is the leading performance determinant.

American consumers continue to migrate to faster broadband connections at home, with the average speed standing at 21.2 Mbps in 2013 [17]—the same year we analyzed—which makes the need to include extended web browsing metrics in the F.C.C. reports more acute. For a big and growing share of households, then, effective throughput is now less of a performance bottleneck, and the web browsing test could provide a more accurate picture of QoE.

4.2 Limitations and Policy Recommendations regarding Loading Times Reported by the F.C.C.

Current testing procedures by the F.C.C., reporting how long it takes to download initial content and associated resources, results in an underestimate of website loading times. When a real website is accessed on a browser, such initial download represents only a fraction of the time it takes to show the full page to the user. Other actions required to render the site include, for example, Javascript execution, often leading to additional (dynamic) objects being fetched from the original site or third party providers. Further, some content required might be specific to the user, including news and social media posts, which requires database queries at the server end. All of these actions result in additional delays.

To illustrate the difference, we ran a simple set of measurements using Google Chrome’s developer tools from a residential location connected to the Comcast network in Cambridge, Massachusetts, in February 2015. The results are presented on Table 4.1. Note that for

Target	Bytes Transferred	Load Time
Google	3.7 MB	3.14 s
Facebook	700 KB	1.86 s
Amazon	2.9 MB	3.42 s
EBay	2.1 MB	3.20 s
MSN	1.1 MB	1.47 s
Wikipedia	93.1 KB	1.01 s
Yahoo	1.2 MB	5.12 s
CNN	2.2 MB	6.12 s

Table 4.1: Total bytes transferred and loading times from Google Chrome’s developer tools.

most websites the loading time is of several seconds, and page sizes exceed one megabyte. This simple experiment’s results, although almost anecdotal, are in line with those reported on [23], a survey of major retail websites: “The median top 100 retail site takes 5.4 seconds to render primary content, and 10.7 seconds to fully load.” The disparity between real website loading times (what the users experience) and those measured by the F.C.C. becomes apparent when looking at our analysis in the previous chapter. Most website’s loading times reported for Comcast in the Northeastern United States, for example, are under two seconds¹.

If the web browsing test is to capture Quality of Experience, it needs to execute and render all the resources of the site to simulate actual loading times, so the reported results are in line with real world performance. The test already captures timings of all four actions executed by real browsers²; but, particularly for the "object download" phase, the F.C.C. reports an underestimate of loading times. When real users access a site, their browsers have to get all static and dynamic objects in order to show the full page—not just the initial HTML’s content and resources that whiteboxes currently fetch—and a new test should aim to capture the complete process.

Such modifications to the F.C.C.’s web browsing test might be, in fact, imminent: in

¹See Figure 3-1.

²A full description of all the actions that occur to load a website is included in Section 2.1.

a private email between Steven Bauer, a member of our lab at MIT, and SamKnows, the company suggests that a more comprehensive test may be rolled out in the near future.

In addition, it would greatly benefit the measurements research community if the F.C.C. collected and released trace route data for the web browsing test, showing the IP addresses of mid-point routers in the connection path between the website’s server and the whitebox. With such data, in addition to more fine-grained inferences on the source of performance impairments, researchers could explore interconnection paths between large broadband and content providers.

Data collected from networks other than residential broadband could complement Measuring Broadband America’s web browsing test, as we have shown in this thesis with data collected at M.I.T. This non-residential data has at least two advantages. First, smaller and relatively uncongested networks—such as those of research universities—are easier to characterize than residential access networks. Second, testing from locations with less internal diurnal variation (because of lower congestion and better link provision) can help better understand performance impairments *outside* access networks.

Finally, to the best of our knowledge there are no other analysis of F.C.C. web browsing data on this scale. One of the reasons for this lack of studies on MBA’s web browsing data—even though it has been around for multiple years—is arguably that the F.C.C. itself does not provide any funding for third-party analysis of the data it collects. Given the complexity of the technical questions addressed by the test, and the many man hours that need to be invested in understanding and working with F.C.C.’s custom data formats, more widespread research is unlikely to materialize unless public funding is made available to the research community.

4.3 Measuring the Impact of Content Delivery Networks

Our analysis shows that websites with larger infrastructures, signaled by a higher number of unique IP addresses to deliver content, generally have lower loading times; this, in turn, reflects on better QoE for users. Figure 4-1 shows this trend clearly for fiber and cable providers: msn, Google and YouTube, with superior infrastructure, also exhibit lower loading

times.

Importantly, other websites with more modest infrastructure, such as Wikipedia or EBay, exhibit higher loading times. It is possible that even smaller sites, for example that of a small credit union, could have worse performance. Although measuring to nine top websites, as the F.C.C. does, provides a benchmark for a large portion of users' interactions with the web, the typical user will also access other websites not included in this list.

Therefore, if the web browsing test is to capture the experience of real users, the list of target websites should be expanded to include a more diverse array of organizations: as it stands, the nine websites are either specialized content providers (CNN, Facebook, Google, msn, Wikipedia, Yahoo and YouTube) or large retailers (EBay and Amazon).

Most users will also access websites in other industries: education, healthcare and finance, to name a few; future versions of the F.C.C.'s web browsing test should include target websites in these and other relevant categories, such as sites of organizations in all levels of government.

4.4 Diurnal Variation for Each Target Website

As we showed on Section 3.2.4, there is evidence of diurnal variation for most target websites. However, each target is affected differently; some experience a greater increase in their loading times during peak hours, as is evident on Figure 4-2.

Once again, the sites with superior networking infrastructure—msn, YouTube and Google—deliver a better and, in this case, more consistent experience to their users. During peak hours, loading times for these three websites increases by less than 5%. For EBay that increase exceeds 20%.

Although the underlying measurements by the F.C.C. are an underestimate³ of actual website loading times, the data can still provide performance comparisons between different targets, as we have shown here; further, analysis of data from other periods can reflect improvements due to deployments of CDNs, caching techniques, and others.

We also computed diurnal variation metrics for the MIT-collected data⁴, shown on Figure

³As discussed previously in Section 4.2.

⁴Usage patterns from the MIT network are also different to those of residential users: in some instances,

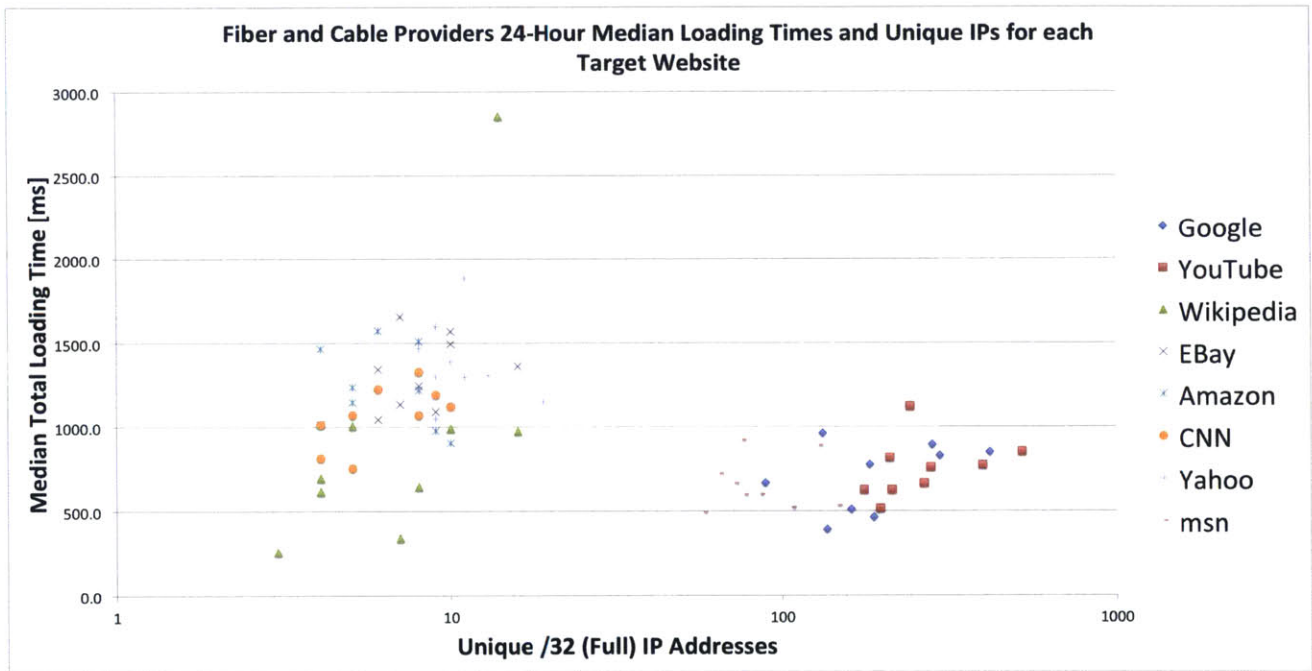


Figure 4-1: Median loading time and number of unique IP addresses for eight target websites, measured from fiber and cable providers (all but Frontier and AT&T.) Each data point represents the 24-hour loading time, for all of 2013, of a single provider in a geographic region, and the corresponding number of unique IP addresses for the same period. Note the logarithmic scale on the horizontal axis, and the exclusion of Facebook due to the data issues discussed in Chapter 3.

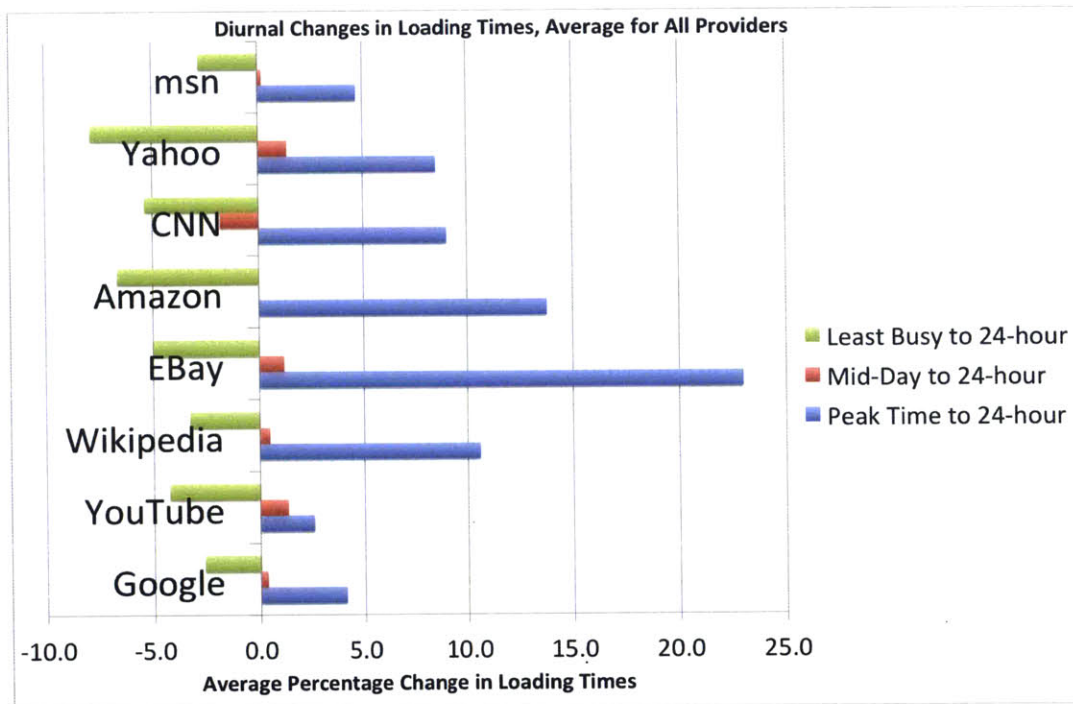


Figure 4-2: Average diurnal change in loading times for each target website, including data for all 12 providers in the three regions studied. Each period (3:00-6:59am, 11:00am-2:59pm and 7:00-11:59pm) is compared to the 24-hour median, obtained from the data for all of 2013. Note the exclusion of Facebook due to the issues discussed in Section 3.2.6.

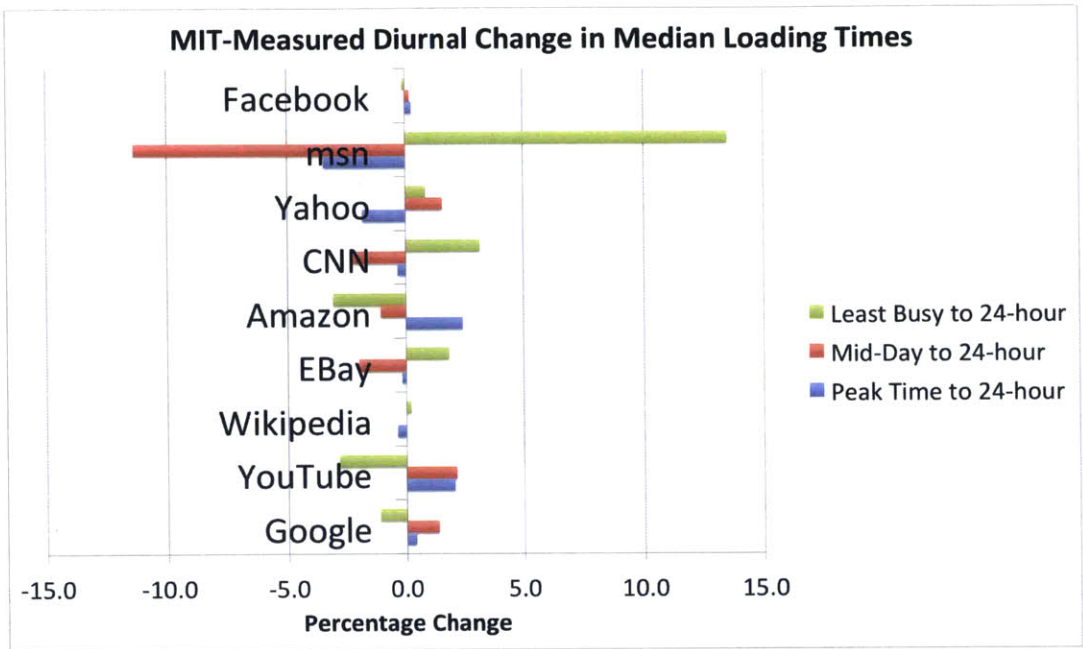


Figure 4-3: Change in loading times for each target website, measured at MIT from March 13th to June 20th, 2015. Each period’s median (3:00-6:59am, 11:00am-2:59pm and 7:00-11:59pm) is compared to the 24-hour median.

4-3. Note that all target websites except msn exhibit diurnal changes of less than 5%, and in all cases much lower differentials than those seen in Figure 4-2 for residential users.

There are at least three factors influencing the more consistent performance seen at MIT: first, latency in access networks—due to congestion or otherwise—has a large impact on website load time [8]. The relatively uncongested network at MIT likely improves performance for all websites.

Second, the MIT data is for 2015, whereas the F.C.C. data is for 2013. CDN deployments, as well as new or extended peering interconnection agreements, could all have reduced diurnal variation. Third, MIT is connected to Internet2, a consortium of research institutions with better interconnection to content providers than the typical residential access network.

median loading times at the "least busy" hours of 3:00-6:59am *increase* when compared to 24-hour values.

Chapter 5

Conclusion

It seems like we've been arguing about network neutrality for a long time.
What's taking so long?

Timothy Lee in *Vox* [6], 2015.

Debates on broadband performance—more specifically, on *who is to blame* for bad performance—have been present on the mainstream media for a few years now, and will likely continue to be so. The reason for such long public interest on a seemingly technical matter is simple: web browsing performance degradation can be difficult to attribute to a single actor. As we have shown in this thesis, this is because *all systems involved in web browsing clearly impact the quality of users' experience*. In this chapter we summarize our findings in that regard.

We present four main contributions: first, we find performance differentials across websites that are consistent across providers (both commercial ISPs and MIT) and regions, showing the impact of infrastructure of transit and content providers on loading times and Quality of Experience (QoE). Second, we find strong evidence of diurnal variation in loading times, highlighting the importance of network and server load on end-user QoE. Third, we show instances of localized congestion that severely impair the performance of some websites. Fourth, we find that web loading times correlate with the size of a website's infrastructure as estimated by the number of IP addresses observed in the data. We discuss how we reach each of these results in this chapter.

Using boxplots of loading times of nine major websites, measured from the networks of

eight providers in three geographic regions, we have shown that there is significant variation in web browsing performance for each website (See Section 3.2.2.) Further, each website shows similar loading times distributions when accessed from different providers, including M.I.T., highlighting the importance of target websites' infrastructure for loading times and consequently QoE. Of the websites studied, Google and msn deliver the best and most consistent performance, particularly for fiber and cable providers; this can be seen on their narrower loading time distributions. Our results coincide with what previous studies have found: target website and CDN infrastructure are critical for web loading times [8,24].

To our knowledge, we provide the first yearlong analysis of web browsing data collected by the F.C.C., resulting in two main inferences: the presence of diurnal variation of loading times, and instances of localized congestion. We discuss each finding separately in the following paragraphs.

Our temporal analysis, looking at loading times for each website by time of day, shows clear evidence of diurnal variation. This variation is both provider *and* website-dependent, as a single website can exhibit diurnal variation from one provider and not from another; and single providers show diurnal variation when connecting to some websites and not others (See Section 3.2.3.) In aggregate, Google, YouTube and msn show the lowest diurnal variation¹ on the SamKnows-collected data, with increases in median loading times lower than 5% on average. EBay, Amazon and Wikipedia, on the other hand, show the strongest diurnal patterns, with loading times increasing between 11 and 23%. The M.I.T. network, by contrast, shows changes in loading times of less than 5% for all target websites except EBay.

We also find evidence of localized² congestion for some provider and target website pairs. Determining the location of such congestion is nearly impossible with the data available. For large target websites, different providers can be served from also different hosts and networks, so it is possible that portions of the website's infrastructure is contributing to the performance degradation seen from an ISP.

¹Facebook was not included in this analysis, as data collected by the F.C.C. is faulty starting in May 2013; see Section 3.2.6 for a detailed discussion of the issue.

²We use the term "localized" congestion for three reasons: it does not affect all websites equally, nor is it visible on the data for all providers in a single region, and observed performance impairments dissipate after a few days or weeks.

We can, however, speculate on two possible explanations for localized congestion: popular and timely downloads, such as software updates, can create instances of congestion that dissipate after a few days³; or peering interconnection links between an ISP and a target website (or their hired CDN or transit provider) could be underprovisioned, leading to performance degradation that is only visible from the homes of the users that subscribe to the ISP involved *and* that access the affected website.

Finally, we show that target websites' infrastructure—as estimated by the number of IP addresses observed in the data—greatly influences website performance for all ISPs. We show this in two ways: by correlating step changes in loading times with changes in the server-end IP address that the whiteboxes most frequently connect to (Sections 3.2.5 and 3.3); and by comparing 24-hour median loading times for all of 2013 for each provider⁴–target website–region triplet with the size of the target websites infrastructure (Figure 4-1.)

In both of these analyses, Google, YouTube and Microsoft's msn—all three with superior delivery and hosting infrastructure—exhibit the highest flexibility to deal with instances of localized congestion. This flexibility results in better and more consistent performance seen by their users. Facebook exhibits similar load-sharing schemes when measured from M.I.T.

Despite the value of web browsing data—capable of producing all the insights outlined in this chapter—the F.C.C. dedicates a mere 150 words to web browsing in the latest 67-page Measuring Broadband America report [17]. In this regard, we provide three major policy recommendations.

First, we suggest a greater emphasis on web browsing inferences. As recognized by the F.C.C. and others [8, 17], throughput—the major focus of F.C.C. reports—only reduces web loading times significantly up to 15 Mbps. Second, we recommend two major changes to the web browsing test to make it more accurately reflect user experience: expand the list of target websites to organizations outside information technology and retail; and change the technical design so the whiteboxes download all static and dynamic objects of the websites (not just the initial content and resources as it does now.) Third, we recommend the provision of public funding to analyze F.C.C.-collected data; as well as support the collection of measurements

³One instance of event-driven congestion was privately discussed by the F.C.C. with the technical measurements community: Apple's iOS release in September 2013.

⁴Cable and fiber only.

from less congested, and better provisioned, non-residential networks such as those of research universities.

In this thesis, we have shown that web browsing performance—as experienced by residential users—is impacted by several systems in addition to the ISP’s network. Of particular importance is the infrastructure of the target website being accessed. We have also shown evidence of diurnal variation in loading times and localized congestion.

Further, we significantly improve the state of the art of residential web browsing performance analysis by: extending the timeframe to a full year; providing both temporal and statistical evidence of our inferences; analyzing data collected from over 800 units in 19 states; and validating our inferences with measurements collected at M.I.T., a non-residential access provider with superior network connectivity.

Appendix A

Loading Time Histograms

Figures A-1 and A-2 show loading time distributions for all websites¹ under study, measured from two providers in the Northeast: Comcast and Cablevision. In the Pacific and Southern U.S., the other two regions under study, similar patterns are seen on target website loading time histograms. Figures A-3 and A-4 present the histograms for loading times measured from the Verizon Pacific and Cox Southern networks, respectively.

¹Note that, due to a sharp and persistent drop in loading times for Facebook in May 2013 (discussed in Chapter 3), only data for January through April is included.

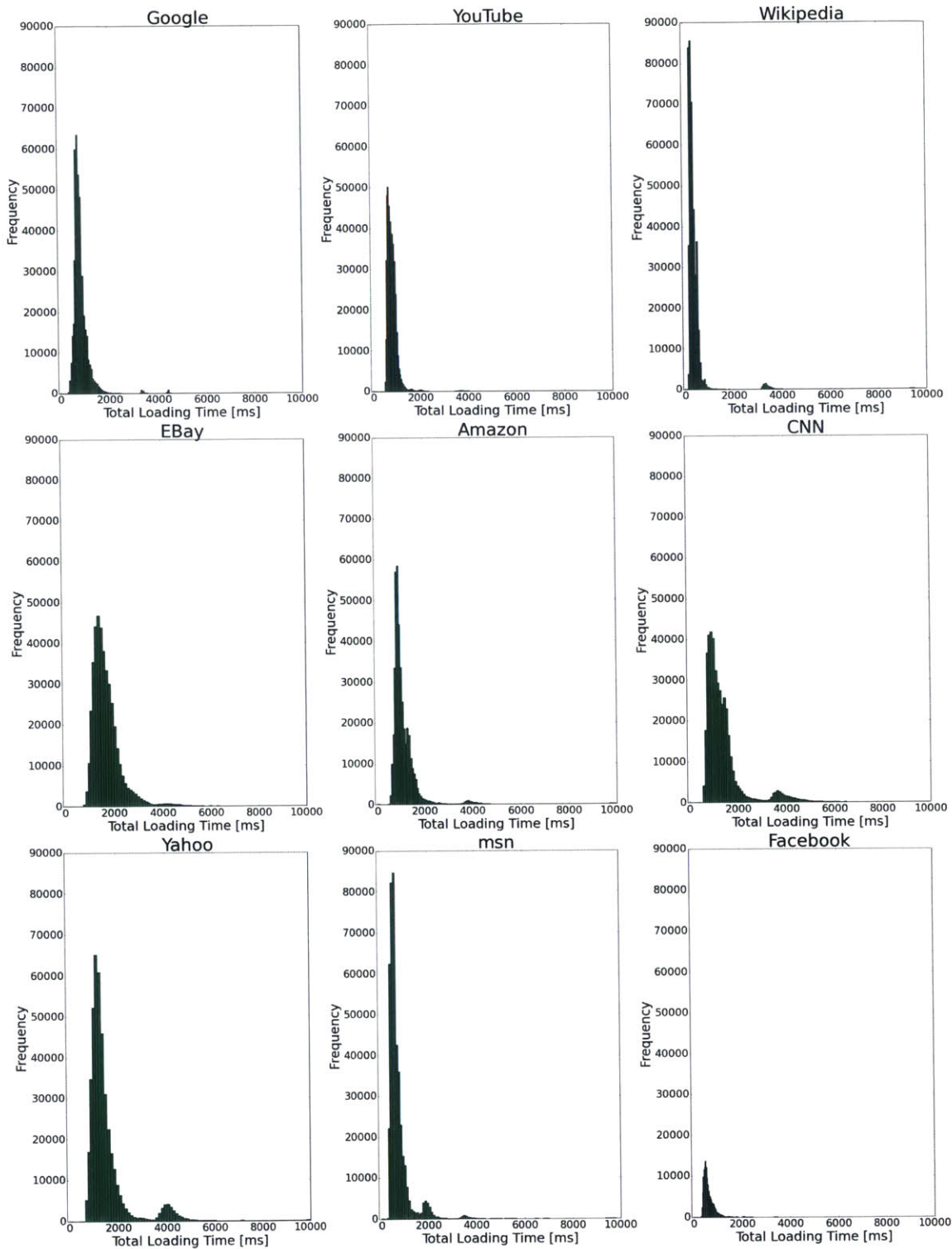


Figure A-1: Histograms of loading times of different websites for all of 2013, measured from the Comcast network in the Northeast, with an advertised speed of 20 Mbps. The vertical axis shows how many measurements of total loading time fell into each bin on the horizontal axis. Wider distributions are an indication of higher variability in loading time, whereas higher peaks on the left side indicate lower loading times. Facebook's data is for Jan-Apr.

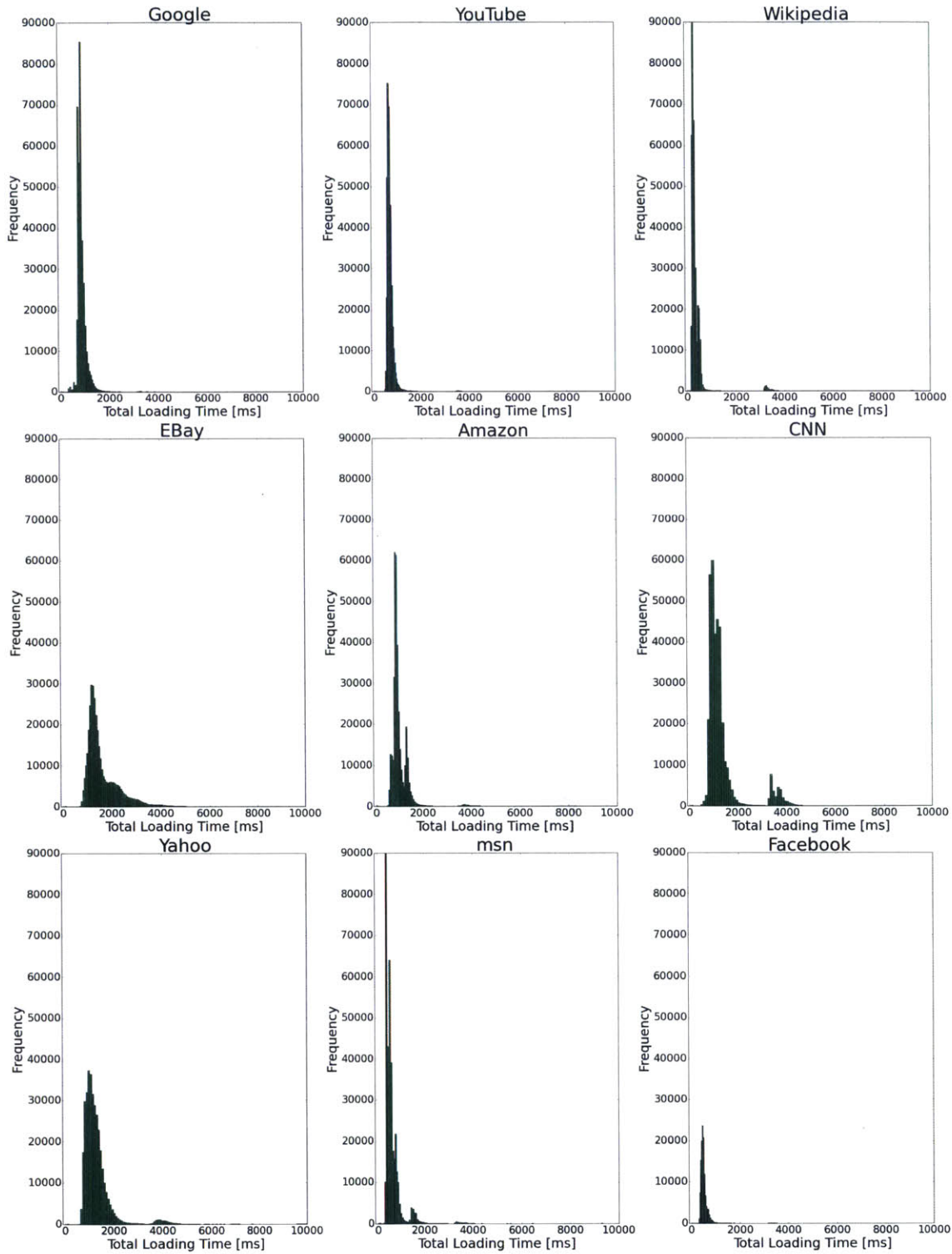


Figure A-2: Histograms of loading times of different websites for all of 2013, measured from the Cablevision network in the Northeast, with an advertised speed of 15 Mbps. The vertical axis shows how many measurements of total loading time fell into each bin on the horizontal axis. Wider distributions are an indication of higher variability in loading time, whereas higher peaks on the left side indicate lower loading times. Facebook’s data is for Jan-Apr.

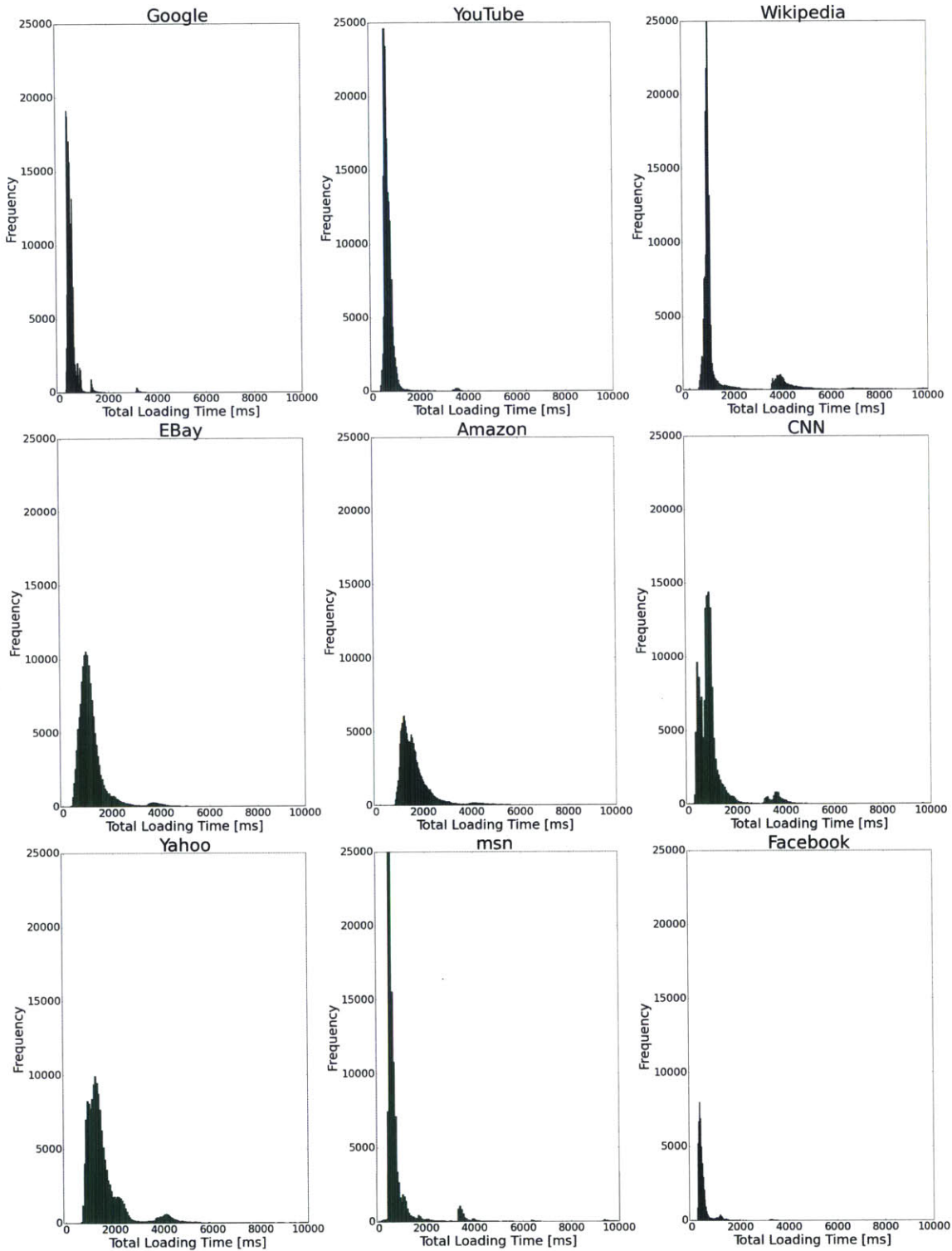


Figure A-3: Histograms of loading times of different websites for all of 2013, measured from the Verizon network in the Pacific, with advertised speeds between 25 and 50 Mbps. The vertical axis shows how many measurements of total loading time fell into each bin on the horizontal axis. Wider distributions are an indication of higher variability in loading time, whereas higher peaks on the left side indicate lower loading times. Facebook’s data is for Jan-Apr.

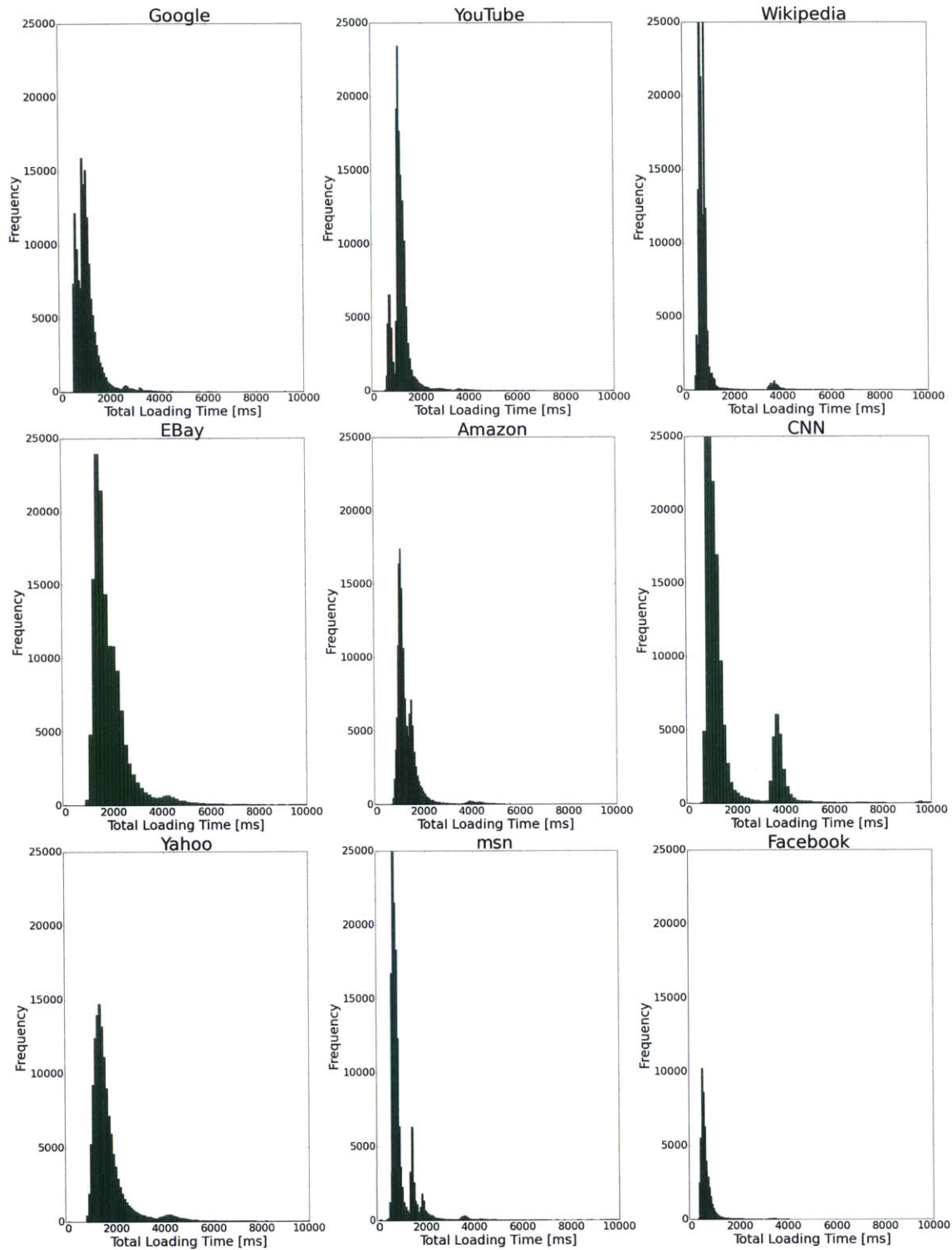


Figure A-4: Histograms of loading times of different websites for all of 2013, measured from the Cox network in the South, with advertised speeds between 12 and 25 Mbps. The vertical axis shows how many measurements of total loading time fell into each bin on the horizontal axis. Wider distributions are an indication of higher variability in loading time, whereas higher peaks on the left side indicate lower loading times. Facebook's data is for Jan-Apr.

Bibliography

- [1] “Statement by the President on Net Neutrality.” [Online]. Available: <https://www.whitehouse.gov/the-press-office/2014/11/10/statement-president-net-neutrality>
- [2] T. Wu, “Network Neutrality, Broadband Discrimination,” *2 J. on Telecomm. & High Tech. L.* 141, vol. 1, 2003.
- [3] L. Lessig, “Testimony Of Lawrence Lessig C. Wendell And Edith M. Carlsmith Professor Of Law Stanford Law School At The Hearing On “Network Neutrality”,” 2006. [Online]. Available: <https://cyberlaw.stanford.edu/files/publication/files/lessig-020706.pdf>
- [4] B. van Schewick and D. Farber, “Point/Counterpoint Network neutrality nuances,” *Communications of the ACM*, vol. 52, no. 2, p. 31, Feb. 2009. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1461928.1461942>
- [5] The Editorial Board, “Global Threats to Net Neutrality,” Apr. 2015. [Online]. Available: <http://www.nytimes.com/2015/04/10/opinion/global-threats-to-net-neutrality.html>
- [6] T. B. Lee, “9 questions about network neutrality you were too embarrassed to ask.” [Online]. Available: <http://www.vox.com/2014/11/10/7187281/9-questions-about-network-neutrality-you-were-too-embarrassed-to-ask>
- [7] B. Fung, “Your guide to net neutrality: Everything you need to know about today’s FCC vote,” Feb. 2015. [Online]. Available: <https://www.washingtonpost.com/blogs/the-switch/wp/2015/02/26/your-guide-to-net-neutrality-everything-you-need-to-know-about-todays-fcc-vote/>

- [8] S. Sundaresan, N. Feamster, R. Teixeira, and N. Magharei, “Measuring and mitigating web performance bottlenecks in broadband access networks,” *Proceedings of the 2013 conference on Internet measurement conference - IMC '13*, pp. 213–226, 2013. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2504730.2504741>
- [9] WiX, “Improving Your Site’s Loading Time | Support Center.” [Online]. Available: <https://www.wix.com/support/html5/technical-difficulties/live-site-issues/kb/improving-your-sites-loading-time>
- [10] F. Kuipers, R. Kooij, D. De Vleeschauwer, and K. Brunnstrom, “Techniques for Measuring Quality of Experience,” in *Wired/Wireless Internet Communications: 8th International Conference, WWIC*. Berlin Heidelberg: Springer-Verlag, 2010, pp. 216–227.
- [11] C. Hunt, “TCP/IP Network Administration,” Feb. 1999. [Online]. Available: http://www.diablotin.com/librairie/networking/tcpip/ch02_02.htm
- [12] D. Clark, S. Bauer, W. Lehr, B. Huffaker, M. Luckie, K. Claffy, and A. Dhamdhere, “Measurement and Analysis of Internet Interconnection and Congestion,” in *Telecommunications Policy Research Conference (TPRC)*, 2014.
- [13] S. Bauer, D. Clark, and W. Lehr, “The Evolution of Internet Congestion,” *37th Research Conference on Communication, Information and Internet Policy (TPRC)*, 2009.
- [14] D. Rayburn, “Netflix’s Network Congestion Message Rolling Out On All ISPs, Not Just Verizon,” 2014. [Online]. Available: <http://blog.streamingmedia.com/2014/06/netflixs-network-congestion-message-rolling-networks-just-verizon.html>
- [15] Microsoft Technet, “What is QoS?” [Online]. Available: <https://technet.microsoft.com/en-us/library/cc757120%28v=ws.10%29.aspx>
- [16] R. Buyya, P. Mukaddim, and A. Vakali, *Content Delivery Networks*. Berlin: Springer, 2008.
- [17] FCC’s Office of Engineering and Technology and And Consumer and Governmental Affairs, “2014 Measuring Broadband America Fixed Broadband Report,” Federal Communications Commission, Tech. Rep., 2014.

- [18] Federal Communications Commission, “2014 Measuring Broadband America Report - Technical Appendix,” Tech. Rep., 2014.
- [19] T. Everts, “How to create the illusion of faster web pages (while also creating actual happier users).” [Online]. Available: <http://www.webperformancetoday.com/2014/07/16/eight-tricks-improve-perceived-web-performance/>
- [20] J. Nielsen, “Response Times: The 3 Important Limits,” Jan. 1993. [Online]. Available: [ResponseTimes:The3ImportantLimits](http://www.useit.com/alertbox/930101/)
- [21] I. Grigorik, “Image Optimization.” [Online]. Available: <https://developers.google.com/web/fundamentals/performance/optimizing-content-efficiency/image-optimization?hl=en>
- [22] L. Greenemeir, “Net Neutrality Prevails in Contentious FCC Vote.” [Online]. Available: <http://www.scientificamerican.com/article/net-neutrality-prevails-in-contentious-fcc-vote/>
- [23] T. Everts, “Spring 2014 State of the Union: Ecommerce Page Speed & Web Performance [INFOGRAPHIC].” [Online]. Available: <http://www.webperformancetoday.com/2014/04/29/spring-2014-state-union-ecommerce-page-speed-web-performance-infographic/>
- [24] A.-J. Su, D. R. Choffnes, A. Kuzmanovic, and F. E. Bustamante, “Drafting behind Akamai (travelocity-based detouring),” *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4, p. 435, 2006.