TECHNIQUES FOR THE ANALYSIS OF MODERN WEB PAGE TRAFFIC USING ANONYMIZED TCP/IP HEADERS

Sean M. Sanders

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2017

Approved by:

Jasleen Kaur

Jay Aikat

Kevin Jeffay

Fabian Monrose

Raheem Beyah

ABSTRACT

Sean M. Sanders: Techniques for the Analysis of Modern Web Page
Traffic Using Anonymized TCP/IP Headers
(Under the direction of Jasleen Kaur)

Analysis of traces of network traffic is a methodology that has been widely adopted for studying the Web for several decades. However, due to recent privacy legislation and increasing adoption of traffic encryption, often only anonymized TCP/IP headers are accessible in traffic traces. For traffic traces to remain useful for analysis, techniques must be developed to glean insight using this limited header information. This dissertation evaluates approaches for classifying individual web page downloads — referred to as web page classification — when only anonymized TCP/IP headers are available.

The context in which web page classification is defined and evaluated in this dissertation is different from prior traffic classification methods in three ways. First, the impact of diversity in client platforms (browsers, operating systems, device type, and vantage point) on network traffic is explicitly considered. Second, the challenge of overlapping traffic from multiple web pages is explicitly considered and demultiplexing approaches are evaluated (web page segmentation). And lastly, unlike prior work on traffic classification, four orthogonal labeling schemes are considered (genre-based, device-based, navigation-based, and video streaming-based) — these are of value in several web-related applications, including privacy analysis, user behavior modeling, traffic forecasting, and potentially behavioral ad-targeting.

We conduct evaluations using large collections of both synthetically generated data, as well as browsing data from real users. Our analysis shows that the client platform choice has a statistically significant impact on web traffic. It also shows that change point detection methods, a new class of segmentation approach, outperform existing idle time-based methods. Overall, this work establishes that web page classification performance can be improved by: (i) incorporating client platform differences in the feature selection and training methodology, and (ii) utilizing better performing web page segmentation approaches.

This research increases the overall awareness on the challenges associated with the analysis of modern web traffic. It shows and advocates for considering real-world factors, such as client platform diversity and overlapping traffic from multiple streams, when developing and evaluating traffic analysis techniques.

To Candice

ACKNOWLEDGEMENTS

First, I would like to take the time to thank my advisor, Jasleen Kaur, for her advice, support, guidance, and patience during this process. I believe that I have greatly improved my research skills under her guidance which has helped make this dissertation possible. I would also like to thank Vladimir Jojic for helping me learn the fundamentals of machine learning and convex optimization. I am also very grateful to Kevin Jeffay, Fabian Monrose, Jay Aikat, and Raheem Beyah for being part of my dissertation committee. I really appreciate your valuable input and precious time that helped improve the quality of my work.

I would also like to thank the National Science Foundation and AT&T Labs for funding the vast majority of my graduate education. This external funding allowed me to have some freedom to pursue research topics that I am passionate about and will be successful in. I have been extremely blessed to have been supported by these organizations.

My friends and family are well aware of my desire of earning a PhD and have been very supportive of my graduate education. I thank them for their continued love and support. I would not have made it this far without you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Address record
ACK	Acknowledgement of TCP Segment
ADMM	Alternating Directions Method of Multipliers
AGL	Alexa Genre Labels
AJAX	Asynchronous Javascript
b	bit
B	Byte
CDF	Cumulative Distribution Function
CNAME	Canonical Name record
CSS	Cascading Style Sheets
CT	Classification Trees
DNS	Domain Name System
DPI	Deep Packet inspection
E-mail	Electronic Mail
EM	Expectation Maximization
FIN	TCP control flag indicating to close the TCP connection
FIN-ACK	Acknowledgement of FIN segment
FPR	False Positive Rate
FTP	File Transfer Protocol
Gb	Gigabit
GB	Gigabyte
GT	Ground Truth
HLS	HTTP Live Streaming
HMM	Hidden Markov Model
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority

iOS	iPhone Operating System
IP	Internet Protocol
IRB	Institutional Review Board
ISP	Internet Service Provider
KB	Kilobyte
KNN	K-Nearest Neighbors
KS test	Kolmogorov-Smirnov test
LDA	Linear Discriminant Analysis
Mb	Megabit
MB	Megabyte
ML	Machine Learning
ms	Millisecond
NAT	Network Address Translation
NB	Naive Bayes
NS	Network Simulator
P2P	Peer-to-Peer
PTR	Pointer record
RST	TCP control flag indicating to Reset or close the TCP connection
RTT	Round-trip time
s	Second
SSH	Secure Shell
SVM	Support Vector Machine
SYN	TCP control flag indicating to begin TCP connection establishment process
SYN-ACK	Acknowledgement of SYN segment
TCP	Transmission Control Protocol
TDL	Target Device Labels
TPR	True Positive Rate
TTL	Time to live
UDP	User Datagram Protocol

URI	Uniform Resource Identifier
URL	Uniform Resource Locator
U.S.	United States of America
VSL	Video Streaming Labels
WNL	Web page Navigation Labels
WWW	World Wide Web
XML	Extensible Markup Language

CHAPTER 1: INTRODUCTION

1.1 Motivation

1.1.1 Overview of the Web: The Most Popular Application on the Internet

The Web is an Internet application that links web pages, which are transferred via HTTP and viewed using browsers.[1] In its infancy, the Web consisted of mostly static web pages, which were used to display non-interactive content and supported only a limited set of applications. However, as browsers and other computing technology have advanced, web pages have become dynamic and are used for many *modern web services*, including video streaming, gaming, social networking, and email, which has vastly increased Web usage. In fact, Internetlivestats.com [2015] reports that the number of web sites has gone from 23,500 in 1995 to over 600 million in 2015, and Popa et al. [2010] reports that Web traffic accounts for over 80% of the Internet traffic mix. The Web is thus the most popular application on the Internet.

1.1.2 Applications of Traffic Trace Analysis

Given its popularity and scale, the Web must be monitored to ensure that it can meet the increasing demand. Content providers (e.g., Google and Facebook) and analytics companies (e.g., Quantcast and Alexa), who have access to servers and/or clients, own a vast amount of proprietary data that is used to study web usage. This proprietary data is not accessible to most Web researchers (enterprises, universities, and Internet service providers, or ISPs). Internet traffic traces, which capture all data transferred on a network link, are an accessible alternative to proprietary Web data [Xu et al., 2011]. Traffic trace analysis has been used for several decades by researchers and network administrators for many different applications, some of which are detailed below.

- *Network Monitoring, Forecasting, and Capacity Planning:* Understanding network traffic and its growth allows ISPs, universities, and enterprise network administrators to meet the increasing demand for Web services [Xu et al., 2011, Erman and Ramakrishnan, 2013]. In addition to monitoring

[1] Please refer to Chapter 2 for formal definition of web pages and other terms used in this chapter.

aggregate traffic volume, trace analysis is used to analyze application usage and performance trends to build more accurate forecasting models that predict future network use and identify potential capacity limitations so that they can be addressed before performance issues affect user experience [Ihm and Pai, 2011, Smith et al., 2001].

- *Traffic Modeling:* Traffic models are developed by studying the statistical properties of Internet traffic traces (e.g., packet sizes and inter-arrival times). The traffic generated with these models is used to drive simulation and testbed experiments that can be used for network forecasting and protocol testing before deployment [Weigle et al., 2006, Barford and Crovella, 1998].
- *Privacy and Security Analysis:* Traffic traces have been analyzed to examine packet payloads for signatures that correspond to known malware [White et al., 2013]. Traffic trace analysis techniques are routinely used by intrusion detection systems, specialized tools that improve network security [Paxson, 1999, Gu et al., 2008]. Security researchers also use traffic trace analysis to test data anonymization techniques [Yen et al., 2009].

Recent trends in the networking and Web communities may soon make traffic trace analysis more difficult. Some of these trends are discussed in the following section.

1.1.3 Motivation for Using Only Anonymized TCP/IP Headers

Deep-packet inspection of traffic traces has been used for applications such as botnet detection and traffic engineering [Tegeler et al., 2012, Sen and Wang, 2004], but it is viewed as an invasion of user privacy because it scans the content of network communications [Asghari et al., 2013]. Analysis of traffic traces that invade user privacy are becoming increasingly infeasible for the following reasons:

- *Privacy Legislation:* Recent privacy legislation in the United States explicitly forbids the analysis of content data [Law, Sicker et al., 2007] (i.e., HTTP header analysis or deep-packet inspection). In fact, according to the Federal Wiretap Act, analyzing the content of network communications without user consent or a provider protection exception [Sicker et al., 2007] is a felony. The Pen Register and Trap and Trace Act similarly limits the analysis of non-content of network communications, including TCP/IP headers.[2] These laws present problems for network traffic monitoring. First, many entities,

[2] Analysis of non-content of network communications was legal before revisions to The Pen Register and Trap and Trace Act that came bundled with the Patriot Act of 2001.

such as ISPs and universities, find it difficult to obtain consent from all users. Second, the provider protection exception limits the amount of data collected and the type of analysis that can be performed on the data. For instance, service providers that invoke this exception must specify the purpose of the network monitoring, collect the minimum amount of data necessary for analysis, and follow best practices for anonymizing the data. Arguably, the only non-content data that can legally be analyzed in the United States is anonymized TCP/IP header data.[3]

- *Traffic Obfuscation:* The increasing amount of obfuscated (encrypted or compressed) traffic makes deep-packet inspection infeasible, because obfuscation distorts the content. Approximately 86% of traffic is now obfuscated White et al. [2013]. New protocols—such as HTTP/2, which encrypts all traffic—exacerbate this problem [Register, Belshe et al., 2015, Jerome]. HTTP/2 is already being adopted; 2.9% of web sites currently support the protocol [W3Techs]. These changes suggest that it will soon be too arduous to obtain content data for traffic analysis.

As noted above, the challenges of privacy legislation and traffic obfuscation mean that, in some scenarios, *only anonymized TCP/IP headers are available for traffic trace analysis*, which offer only header and time-related (non-content) information.

1.2 Overview of Dissertation

For traffic trace analysis to remain useful as a network analysis methodology, techniques must be developed to glean insight from traffic traces using the limited information in anonymized TCP/IP headers. To that end, this dissertation evaluates approaches to *classifying* individual web page downloads when only anonymized TCP/IP header data is available; we refer to these approaches as *web page classification* approaches. Classifying individual web page downloads is a critical building block for numerous web measurement-related applications, including privacy analysis, user behavior modeling, traffic forecasting, and behavioral ad-targeting.

This dissertation also investigates two issues that are not usually considered in the traffic classification literature [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b, Dyer et al., 2012, Herrmann et al., 2009, Miller et al., 2014, Schatzmann et al., 2010]: classifying web page download traffic that is generated using

[3] In an anonymized TCP/IP header, the source and destination IP addresses cannot be directly resolved to a hostname, preserving user privacy.

multiple client platforms, and classifying traffic that is not separated according to individual web page downloads. Next, we present an overview of these issues.

- *Client platform choice may influence web page download traffic:* Real traffic traces consists of a mixture of traffic that is generated from multiple different client platforms, including different operating systems (e.g., Mac OSX, Windows, and Linux), browsers (e.g., Chrome, Firefox, Safari), and devices (e.g., laptop, smartphone, or tablet). Different client platforms may generate different web page traffic despite referencing the same source HTML. Researchers disagree about the degree of impact client platform choice has on traffic [Yen et al., 2009, Butkiewicz et al., 2011]. This dissertation collects a diverse traffic sample that includes traffic generated from different web pages and different client platforms in order to determine whether these factors impact the characteristics of modern web page traffic. If so, these factors should be considered in the design and evaluation of traffic analysis techniques.

- *Real web page download traffic is not well-separated:* Most methods that classify individual web page download traffic assume that the traffic data has been processed and that identifying the traffic for each individual web page download is trivial. However, real traffic traces consist of a mixture of different web page downloads that are difficult to demultiplex into individual page downloads. This type of mixed data requires an approach that can separate, or segment, web traffic into individual web page downloads. These approaches, referred to as web page segmentation approaches, have not been evaluated in over a decade, and better approaches to web page segmentation will likely improve the applicability of web page classification to real traffic traces.

Thesis Statement When only anonymized TCP/IP headers are available, web page classification performance is influenced by traffic from multiple client platforms and requires reliable segmentation of web page traffic. Its performance can be improved by using traffic features that are consistent across client platforms and using web page segmentation approaches that outperform the current state-of-the-art approaches.

Contributions This thesis details the results of three related studies which each support our high-level goal of making traffic trace analysis (specifically, traffic classification) more useful in practice. The first study investigates whether modern web page traffic is impacted by different client platforms (e.g., browsers, operating systems, etc); the second study investigates the web page classification problem; and the third study conducts a comprehensive evaluation of state-of-the-art web page segmentation techniques. These studies make meaningful contributions to the areas of Internet measurement and traffic analysis, which are described below.

1. This work conducts the *first comprehensive study* of modern web page traffic and of how client platform (i.e., browser, operating system, etc) affects it. This study, detailed in Chapter 3, includes measurements taken from HTTP headers, TCP/IP headers, DNS headers, and the HTML source. Results show that modern web page traffic differences across many different factors, including browser, operating system, device, and vantage point, are statistically significant. These results can help interpret aggregate web measurement data and encourage development of measurement analysis techniques (e.g., traffic classification) that are robust to these differences.

2. This work investigates whether well-separated web page downloads can be classified according to four orthogonal labeling schemes (namely, device-based labels, video streaming-based labels, navigation-based labels, and Alexa genre-based labels) using learning-based classification methods to analyze data from the anonymized TCP/IP headers of traffic that is generated by multiple different client platforms. The labeling schemes that we study are not prominently considered in the traffic classification literature [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b, Dyer et al., 2012, Herrmann et al., 2009, Miller et al., 2014, Schatzmann et al., 2010]. We argue that each of the considered labeling schemes is enabling for different application domains. This study also evaluates the performance of several popular supervised machine-learning methods (e.g., classification trees, Naive Bayes, etc.) and find that they perform better than random guessing. We also show that features that are relatively robust to client platform-specific differences perform better than those that are not, demonstrating the value of incorporating data generated from multiple client platforms into the feature-selection and training process. Lastly, we show that the web pages classified with our approach are statistically similar to the ground-truth labeled data. Thus, our approach can be used to label traffic that is similar to the training dataset.

3. This work conducts the *first comprehensive evaluation* of web page segmentation approaches in over a decade. Results show that idle time-based web page segmentation methods do not perform well on modern web page traffic. This work, which is provided in Chapter 5, proposes a new class of web

page segmentation techniques, which we call *change point detection methods*, and shows that these perform better than idle time-based methods. All tested web page segmentation methods had difficulty detecting web page downloads with small inter-arrival times (< 10s); however, some change point detection methods are able to approximate the statistical properties of real user browsing behavior, including the number of web pages downloaded and the average inter-arrival time of web page downloads. We also show that the choice of web page segmentation method impacts the performance of web page classification methods with both real user-generated browsing streams and synthetically generated browsing streams.

We provide a more detailed overview of the the related studies described above, including more details of the methodology used and a more complete summary of the results, in the rest of this introduction.

1.2.1 Study 1 - Determine whether modern web page traffic is impacted by different client platforms

Importance of Adequate Sampling To avoid bias, any sound web measurement study must ensure that the diversity of web traffic is represented in the sample traffic. Most web measurement studies address sampling issues by intentionally incorporating traffic diversity — for example, by increasing the number of web pages included in a sample [Xu et al., 2011, Barford and Crovella, 1998, Newton et al., 2013, Butkiewicz et al., 2011]. Because different web pages exhibit different traffic characteristics, larger traffic samples are less likely to be biased.

Problem Addressed: Examining Impact of Client Platforms on Web Page Traffic However, larger samples don't take into account the diversity of *client platforms* used. Modern client platforms consist of a wide variety of different browsers (e.g., Chrome, Safari, and Firefox), operating systems (e.g., Mac OS X, Linux, and Windows), devices (smartphone, tablet, laptop), and vantage points. There is some disagreement in the literature about whether the type of client platform used impacts web page download traffic [Yen et al., 2009, Butkiewicz et al., 2011], and most previous web measurement-related studies do not explicitly include traffic generated by different client platforms [Dyer et al., 2012, Butkiewicz et al., 2011, Xu et al., 2011, Neasbitt et al., 2014]; it is thus possible that many prior studies are biased towards a specific client platform and not generally applicable.

The goal of the first study presented in this dissertation is to determine whether modern web page traffic is impacted by different client platforms, a determination that has implications for the design and evaluation of measurement analysis techniques. For instance, if browser choice impacts web page traffic, then different browsers should be included in the methodology for evaluating traffic analysis techniques. This study aims to provide quantitative evidence about the impact of client-specific factors on web page traffic, which might affect research in several web-related application domains, including privacy and performance analysis [Butkiewicz et al., 2011, Neasbitt et al., 2014].

Basic Approach An overview of this study's basic methodology is provided below.

- *Diverse sample set:* Using data that is publicly available from Alexa [Inc.], a third-party Internet analytics provider, we compose a list of over 3600 web pages from the 250 most popular web sites in the world, which are responsible for the vast majority of the web traffic on the Internet [Callahan et al., 2013, Xu et al., 2011, Bump].

- *Multiple client platforms:* We use four different types of client platforms to download web pages: different browsers (e.g., Google Chrome, Internet Explorer, Firefox), operating systems (e.g., Windows 7 and Mac OSX), devices (e.g., laptops, smartphones, and tablets), and client locations. We also take repeated measurements of each web page download to determine whether the results are significantly impacted by the time of the measurement.

- *Quantitative feature extraction:* Each web page download is processed to extract quantitative features that summarize the traffic that was downloaded. We derive two different types of features, traffic-based and source HTML-based, from the web page downloads. Traffic-based features, which we derive from packets containing information related to the IP, TCP, HTTP, and DNS protocols, correspond to metrics for protocol performance. Source HTML-based features correspond to metrics for the frequency of HTML tags and unique terms present in the source HTML, and are useful in determining whether different client platforms reference different source HTML. Both traffic-based and source HTML features are useful for comparing the web page download traffic generated across client platforms. We consider 575 traffic-based and 127 source HTML-based features (702 web page features total). A complete list of these features is provided in Appendix 3.

- *Statistical testing:* We use a standard non-parametric statistical test, the Kruskal-Wallis test, to determine which traffic-based and source HTML-based features differ significantly across different client platforms.

- *Implications for application:* Lastly, we explicitly investigate the implications of any significant platform-specific differences that we observe. For instance, differences in traffic generated by different browsers may have implications for web performance, while differences in traffic generated by different vantage points may have implications for privacy.

Summary of Results The key findings of this study are provided below.

- We find that browser type has a significant impact (p-value < 0.05) on the majority of web page traffic features, including the number of bytes transferred ($p = 1.5 \times 10^{-198}$), the number of servers contacted ($p = 9.2 \times 10^{-45}$), the number of web objects transferred ($p = 4.8 \times 10^{-51}$), and the number of TCP connections established ($p = 1.4 \times 10^{-39}$). In total, we found that 534 of the 575 traffic-based features yielded $p < 0.05$.[4] We also found that many of these web page traffic features differ across browsers, even when the browsers reference the *same* source HTML. This research is the first to show such strong evidence that browsers have a significant impact on web page traffic features. Past work usually acknowledges only that browsers may manage TCP connections differently, rather than considering that the amount of data—the number of web objects transferred—is different [Butkiewicz et al., 2011, Yen et al., 2009]. For instance, we found that Chromium-based browsers (Chrome and Opera) can transmit more than *4 times* as many bytes than other browsers (Firefox and IE) when rendering a web page, because Chromium-based browsers request objects that other browsers do not. We also find that the majority of these objects have hostnames that generally correspond to servers that provide ads, tracking, and CDN services.

- We find that operating system has a minor, and mostly negligible, influence on web page traffic compared to browsers. The only statistically significant differences that we observe are the frequency with which the RESET ($p = 7.3 \times 10^{-121}$) and FIN ($p = 9.8 \times 10^{-123}$) flags are set in TCP segments. *Statistically significant* here implies only that there are differences in TCP implementation across operating systems, and it is well-known that TCP implementation differences can be used to fingerprint desktop operating systems [Nma]. We show that operating systems can be detected using a single coarse feature (number of segments with RESET flag set); other methods rely on more fine-grained (byte-level) features. Because we only observe significant differences at the TCP/IP level, not at the

[4] Most of these p-values were $< 10^{-10}$.

application level, our observations imply that web developers do not optimize web pages according to different operating systems; it is likely unnecessary to measure web pages across different desktop operating systems for different web-related applications. Prior web measurement studies simply assume this result, do not perform experiments to verify it, or do not consider the possibility that operating systems significantly impact web traffic [Butkiewicz et al., 2011, Maciá-Fernández et al., 2010, Miller et al., 2014, Newton et al., 2013, Dyer et al., 2012].

- We find that device type has a significant impact on web page download traffic: 541 out of 575 traffic-based features yielded p-values < 0.05. However, contrary to our browser-based analysis, the differences in web page traffic across devices are caused primarily by differences in the source HTML that is served when different types of devices request the same web page (servers redirect web requests by mobile devices to mobile-optimized versions of the web page). This difference in source HTML is evidenced by the observation that 66 out of 127 source HTML-based features yielded p-values $<$ 0.05.[5]When we ensure that the *same* source HTML is downloaded by different devices, we observe no statistically significant differences in the traffic across devices. Past work has shown that source HTML-based features are impacted by device type (i.e., mobile optimized web pages vs traditional web pages) and that these differences impact web page download traffic on desktops [Johnson and Seeling, 2014, Butkiewicz et al., 2011]. Our work builds on this by showing that the device itself does not impact traffic features when the source HTML is the same. We also build on this work, which considered only smartphones and desktops, by measuring the traffic generated by tablets; results show that tablets may download pages that are designed for smartphones (i.e., small screen), tablets (i.e., moderate screen size), or even laptops (i.e., large screen). This observation suggests that web designers currently do not agree on how to design web pages for tablets.

- As expected, we find that vantage point has a significant impact on temporal traffic features such as average RTT ($p = 6.1 \times 10^{-245}$). The differences in RTT observed across different vantage points are likely influenced by delays incurred by middle-boxes within a network, as well as by the distance between end-hosts. The average RTT observed in China is over 4 times the average RTT observed in Japan, despite requesting similar content and being in similar regions of the world. We observe

[5] For context, the number of source HTML-based features that were <0.05 for the browser-based and operating system-based analysis were eight and zero, respectively.

similar, yet less extreme, differences in RTT even within the United States. Prior studies have also shown that web page download performance is impacted by vantage point [Li et al., 2015b, Zaki et al., 2014, Ahmad et al., 2016, Bischof et al., 2015].

We also find that vantage point has a minor, and mostly negligible, impact on non-temporal traffic and source HTML-based features. Many of these minor differences are because hosts in certain countries download objects unique to that region. We also observe differences that are due to content personalization (i.e., downloading a different search result page for the same search query because the results are targeted to location) or to different web privacy laws in different countries (other countries' web privacy laws are generally more stringent than those in the United States [pri]). For example, the international web site of cnn.com displays a warning that their web site may track users, but their U.S. web site does not. These minor differences do not have a statistically significant impact on traffic-based features. The content-based differences (or similarities) observed across vantage points have been anecdotally known, but not widely studied. Investigating content-based differences in web pages—for example, price discrimination on web pages downloaded in different regions—is a growing research area that is related to the recent data transparency initiative [Laoutaris et al., Mikians et al., 2013].

1.2.2 Study 2 - Investigating the Web Page Classification Problem

Overview of Traffic Classification: One common technique for the analysis of anonymized TCP/IP headers (or other coarse information, such as NetFlow logs) is traffic classification. Traffic classification research has in the past focused primarily on two different types of classification problems: application layer protocol classification and web page identification.

1. *Application layer protocol classification:* Most traffic classification work using anonymized TCP/IP headers and flow-level data has focused on application protocol classification. Application categories that have been of particular interest in the past include HTTP, FTP, P2P, video streaming, and mail [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b]. While distinguishing between some types of applications can be useful, determining whether an application is Web (i.e., HTTP) or not is less informative; the Web itself includes a wide variety of diverse applications, including video streaming, gaming, social networking, mail, and file hosting. This diversity in web applications

is only expected to grow as the Web becomes the standard front-end for emerging services and as existing services continue to migrate to the Web [Labovitz et al., 2010]. In the near future, there will be little utility in classifying something as Web traffic, since most observed traffic will be Web traffic. Thus, it is important to ask whether we can classify HTTP traffic using more fine-grained labels.

2. *Web page identification:* The security and privacy research community have focused on another type of traffic classification problem, here referred to as *web page identification*. Prior studies have shown that it is possible to *identify the exact web page* that the traffic corresponds to with flow-level and/or TCP/IP packet-level data, using existing techniques such as similarity metrics (e.g., the Jaccard index) or learning-based classification (e.g., Naive Bayes) [Dyer et al., 2012, Herrmann et al., 2009]. Web page identification is thus possible, but "in the wild" it is useful for little besides showing that traffic analysis attacks are a threat; web page identification does not scale well with increasing numbers of web pages [Dyer et al., 2012]. There are too many web pages (on the order of millions) to measure, fingerprint, store, and reliably identify. Web page identification is typically used only when trying to identify a small set of targeted web pages.

Application layer protocol classification is too coarse a classification for modern traffic traces, and web page identification is too fine-grained. We propose investigating "web page classification," which we believe can serve as a middle ground between these two traffic classification frameworks.

Problem Addressed: Web Page Classification Web page classification gives traffic a label more fine-grained than just "web traffic," but avoids the scaling problems of web page identification by allowing multiple web pages to share the same label, reducing the number of classes of web pages that must be characterized and labeled. Labels for web applications are already commonly used for traffic trace analysis when available [Xu et al., 2011, Rao et al., 2011, Schneider et al., 2008, Butkiewicz et al., 2011]. The added information provided by labels on traffic gives additional insight to how the network is being used without compromising user privacy. The classification labels used depend on the intended purpose of the classification. Some specific examples of web page classification for different applications are provided below.

- *Profiling video streaming (application type) usage:* Video streaming is now reported to occupy nearly 50% of network bandwidth, and consumption is expected to grow [Sandvine, Bump]. The ability to distinguish between bandwidth-hungry video and non-video streams at critical traffic aggregation points can facilitate better network planning and control. For instance, a campus network manager may be able to prevent network abuse and/or rate-limit video streams destined for student dorms; researchers may want to build profiles of enterprise video traffic to facilitate traffic modeling and forecasting studies; and ISPs may want to limit resources per business interests [Brodkin].

- *Profiling mobile device usage:* By 2017, it is estimated that the average number of devices per Internet user will grow to 5 [Bort], most of them mobile devices. The ability to identify downloads of web pages targeted for mobile devices can help to build profiles of mobile web usage within an enterprise (for capacity planning, modeling, and forecasting purposes) and can be used to deliver personalized content and advertisements customized for the constrained displays, power, and connectivity of mobile devices.

- *Profiling web browsing navigation styles:* The way users navigate through web pages can be classified: they access a landing page (homepage), clickable content (non-landing pages), or a search result. This kind of navigation-based classification can be useful for identifying network misuse, such as web crawlers being misused for the purposes of web page scraping [Jacob et al., 2012]. Recent studies have shown that malicious bots can be identified by the pattern of web page navigation from a given end-point [Tegeler et al., 2012, Wang et al., 2013].

- *Profiling the content type of a web page:* The content of web pages can typically be categorized into genres: Games, Shopping, News, Education, Business, etc [Inc., Xu et al., 2011]. Knowing the genre of web pages downloaded by a given user may be used to gauge user interest, which is invaluable for delivering personalized content and targeted advertisements [Yan et al., 2009]. Service providers currently rely on deep-packet inspection to assess what content consumers are interested in [Corporation]. This classification will also be useful for some types of measurement studies, which perform content-based analysis of web traffic to better understand network use [Xu et al., 2011, Butkiewicz et al., 2011].

Basic Approach The basic approach used for this analysis is provided below.

- *Diverse sample:* As in the previous study, we collect a sample of web pages from the top 250 web sites in the world.

- *Ground-truth labels using multiple labeling schemes:* For classification methods to work, each web page must be labeled. We use four orthogonal labeling schemes: 1) Target device-based; 2) Video streaming-based; 3) Navigation-based; and 4) Genre-based. As previously noted, the labeling scheme used directly impacts the applicability of web page classification.

- *Multiple client platforms:* We apply the results from our previous study to this study's data collection and feature selection methodology. We explicitly consider traffic features that are relatively stable over time and consistent across client platforms, because selecting robust traffic features is likely to improve web page classification performance.

- *Evaluate classifications:* We empirically evaluate web page classification performance while using both parametric and non-parametric methods. We do this performance evaluation for each of the labeling schemes considered.

- *Investigate the applicability of web page classification:* Lastly, we study whether web page classification methods can be used for different web-related application domains. We conduct two application-specific case studies, one investigating web page classification's usefulness for applications in traffic forecasting and simulation modeling, and the other examining whether web page classification can be used to build and approximate user browsing profiles to gauge user interest.

Summary of Results The primary results and contributions of the study are given below.

- As previously mentioned, we classify web page download traffic according to four orthogonal labeling schemes, which are not usually considered in the traffic classification literature [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b, Dyer et al., 2012, Herrmann et al., 2009, Miller et al., 2014, Schatzmann et al., 2010].

- We find that non-parametric methods, such as K-Nearest Neighbors (KNN), outperform parametric methods, such as Linear Discriminant Analysis (LDA), on numerous metrics, including F-score, precision, recall, and accuracy. The performance difference between these methods ranges from 8% to 50% for all major metrics. We believe this is because theoretical distributions are not able to approximate the empirical distributions of the different traffic features. Prior traffic classification studies that use supervised machine-learning techniques also found that KNN and classification trees outperform other classification methods [Lim et al., 2010, Kim et al., 2008].[6]

- Features that are identified as being stable over time and consistent across client platform achieve higher classification performance (up to an increase of 12% in F-score). This analysis of whether different traffic features are effective across client platforms has not been explicitly considered in past work [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b, Dyer et al., 2012, Herrmann et al., 2009, Miller et al., 2014]. The study by Yen et al. [2009] is, to our knowledge, the only work in traffic classification that considers that traffic features may differ across client platforms. However, Yen et al. [2009] addresses this problem by filtering web page traffic according to browser before applying a web page identification technique. This browser-based filtering improves the performance of the classification technique, showing that considering client platform-specific differences in web traffic may improve classification performance.

- We find that the distributions of well-separated web page download traffic classified using our approach are *statistically indistinguishable* ($p > 0.05$) from distributions derived using ground-truth labels. This result indicates that web page classification can be used for web-related application domains such as traffic modeling and user profiling.[7] While prior work in traffic classification states that the results of classification can be used for traffic modeling, they do not provide results from a simulation study or conduct a statistical analysis to demonstrate this [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b].

1.2.3 Study 3 - Comprehensively Evaluating the State of the Art in Web Page Segmentation

Overview of Web Page Segmentation: The problem of grouping individual TCP/IP segments into the web pages that they collectively represent is called *web page segmentation*. Web pages must be segmented to perform any type of page-level analysis on traffic trace data, including general-purpose web traffic characterization and web page classification. These applications are described below.

[6] Supervised machine-learning methods are the current state of the art in traffic classification, outperforming unsupervised and heuristic methods for traffic classification. Please refer to Chapter 2 for a more complete overview of traffic classification methods.

[7] Our case studies also suggest this.

- *Web Page Characterization:* Measurement and characterization studies of web traffic have been conducted for a number of different applications, including traffic forecasting and web usage modeling [Hernández-Campos et al., 2003a, Choi and Limb, 1999, Butkiewicz et al., 2011, Newton et al., 2013, Mah, 1997, Ihm and Pai, 2011, Barford and Crovella, 1998]. Many of these studies attempt web page segmentation, but do not make it their primary focus and do not evaluate it. It is therefore not clear whether they work well for modern web traffic [Newton et al., 2013, Ihm and Pai, 2011], or even if these studies' conclusions might have been impacted in unexpected ways by their segmentation methods.

- *Web Page Classification:* While there have been significant recent advances in the design of classification and fingerprinting approaches, the vast majority of analysis techniques designed over the past two decades assume the availability of perfectly segmented web page traffic [Miller et al., 2014]. Although many recognize the critical need to understand how well web page segmentation can be performed in practice [Miller et al., 2014], few studies have tested actual web page segmentation in an applied setting.

While web page segmentation methods have been used for numerous traffic analyses in the past [Hernández-Campos et al., 2003a, Choi and Limb, 1999, Butkiewicz et al., 2011, Newton et al., 2013, Mah, 1997, Ihm and Pai, 2011, Barford and Crovella, 1998], it is unclear whether existing techniques are effective for traffic analysis on modern web page traffic, which introduces new challenges that make segmentation more difficult. Some of these challenges are described below.

- *Increase in Automatically Generated Traffic:* Modern web pages may use technology such as AJAX, which generates traffic without downloading a new page. For example, Facebook web pages load new content on the same page by scrolling, while Youtube web pages subsequently auto-play additional videos on the same page as the original video. This automatically generated traffic, which may be used to update the content of a web page, presents a problem for web page segmentation, which may treat it as a new page download.

- *Increase in Overlapping Traffic:* Modern browsers support multi-tab browsing, which allows multiple browser windows to be open simultaneously. This is a problem for web page segmentation approaches, because increasing the number of tabs increases the degree of traffic that overlaps [Miller et al., 2014]. Overlapping traffic (traffic that consists of multiple web pages) is tougher to demultiplex, or segment, than traffic that only includes a single web page. This problem is further exacerbated when we consider that modern web pages can automatically generate traffic in the background while the user is not directly consuming the content. There are also environments where multiple users share a single IP address (e.g., NATs) that also increase the degree of overlapping traffic [Guha and Francis, 2005, Tsuchiya and Eng, 1993].

Problem Addressed in this Work: Evaluation of Web Page Segmentation Approaches Previous literature has studied web page segmentation using TCP/IP headers as a timeseries analysis problem and has approached this problem using *idle time-based approaches*. These approaches estimate the beginning of a web page download by detecting whether the network activity level (e.g., the number of bytes observed) exceeds pre-defined thresholds after a certain amount of idle time [Maciá-Fernández et al., 2010, Newton et al., 2013, Ihm and Pai, 2011, Mah, 1997, Barford and Crovella, 1998]. Another class of segmentation approaches rely on *change point detection*, and these identify if/when a timeseries exhibits a substantial increase in network activity. Change point detection methods, such as fused lasso regression and hidden Markov models, have been applied in other fields, including computational biology and speech recognition [Rabiner, 1989, Tibshirani and Wang, 2008, Bleakley and Vert, 2011]. Neither idle time-based approaches nor change point detection approaches have been comprehensively evaluated on modern web page traffic, which produces noisier traffic and more difficult segmentation problems than older web traffic.

Through a comprehensive empirical evaluation of idle time-based and change point detection approaches using both synthetic and real browsing data, we will determine whether web page segmentation approaches can be used to analyze modern web page traffic using only anonymized TCP/IP headers. A successful web page segmentation approach should be able to approximate some of the statistical properties of real user browsing behavior, including the number of web page downloads a user requests and the average inter-arrival time between these downloads—metrics that can be used to model user behavior. A successful web page segmentation approach can also be used to enable and/or facilitate web page classification "in the wild," rather than in an isolated test environment.

Basic Approach An overview of the methodology used for this study is provided below:

- *Browsing stream generation:* Our data collection methodology considers both synthetically generated and real user browsing data. The synthetic data generation explicitly incorporates different inter-arrival time distributions, client platforms (browsers), and number of tabs used into the data collection methodology. The web pages browsed using the synthetic data is the same sample of web pages used in the previous two studies described. The real user browsing data was collected by recruiting 40 real users in an IRB-approved study [Sanders and Kaur, 2015c] in order to study personalized pages and those with user-interactive content, which may not be included in the synthetic data.

- *Web page segmentation:* We evaluate the performance of 2 types of web page segmentation methods: idle time-based methods and change point detection methods. We apply these web page segmentation methods to browsing streams derived from the number of bytes and the number of SYNs traffic features—two features that have been previously used to segment web pages using TCP/IP headers [Maciá-Fernández et al., 2010, Newton et al., 2013, Ihm and Pai, 2011, Mah, 1997].

- *Applicability of web page segmentation:* We study whether the web page segmentation method used has a measurable impact on different web traffic analysis domains. We do this by conducting case studies for the application domains of (i) user behavior modeling and (ii) web page classification.

Summary of Results A summary of the results of this empirical evaluation is provided below.

- For web page segmentation, the number of SYNs is a more robust and informative feature than the number of bytes. Using the number of SYNs rather than the number of bytes improves the true positive rate, or recall, of the best web page segmentation methods by approximately 5%, and some methods, such as the basic idle-time based method, improve by over 40%. A recent study by Newton et al. [2013] also found that the number of SYNs was more effective for web page segmentation than the number of bytes.

- We find that the best-performing change point detection methods tested (fused lasso and a heuristic method) yield F-scores that are 20-30% higher than idle time-based methods. Thus, we find that change point detection methods can more robustly segment modern web traffic than existing idle time-based approaches. This study is the first work to consider change point detection methods for web page segmentation, although in a recent study, Ihm and Pai [2011] showed that idle time-based approaches may perform poorly on modern web traffic. However, Ihm and Pai [2011] proposes that web pages should be segmented using content-based methods, which do not work on anonymized TCP/IP headers; we recommend change point detection methods, which do.

- We find that the inter-arrival time between web page downloads has a significant impact on web page segmentation performance. Our results show that web pages with small inter-arrival times ($< 5s$) are detected less than 30% of the time, while web pages with large inter-arrival time ($> 10s$) are detected approximately 85% of the time. We find that browser choice has only a small impact on segmentation performance, and that the number of tabs open in a browser begins to significantly impact segmentation performance only when the number of tabs increases from 4 to 8 (true positive rate and other metrics decrease by over 10%). These factors have not previously been considered in the literature on web page segmentation [Maciá-Fernández et al., 2010, Newton et al., 2013, Ihm and Pai, 2011, Mah, 1997, Xie et al., 2013, Neasbitt et al., 2014].
- We find that the performance of web page segmentation methods impacts the performance of different applications domains that leverage segmentation, such as user behavior modeling and web page classification. Higher-performing web page segmentation methods are beneficial for applications where web page segmentation is critical, with both synthetically generated and real user browsing data. This work is the first to evaluate the performance of multiple web page segmentation methods on traffic classification when only anonymized TCP/IP headers are available [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b, Dyer et al., 2012, Newton et al., 2013, Ihm and Pai, 2011].

1.3 Outline of Dissertation

The rest of this dissertation is organized as follows. Background and related work in the fields of Internet measurement, traffic classification, and web page segmentation is described in Chapter 2. The first contribution of this dissertation, a comprehensive measurement study that provides insight on the characteristics and diversity in modern web page traffic, is provided in Chapter 3. Web page classification and web page segmentation are investigated in Chapter 4 and Chapter 5, respectively. Concluding remarks and possible directions for future work are provided in Chapter 6.

CHAPTER 2: BACKGROUND AND RELATED WORK

The Web, the most popular application on the Internet, is a highly complex application because it uses multiple technologies — some of these technologies include web browsers (i.e., Google Chrome and Firefox), web pages (i.e., HTML and javascript), web servers (i.e., Apache), and communication protocols (i.e., HTTP, TCP/IP, and DNS). This complexity makes the discussion of the Web difficult because the term "Web" can be used to describe many related, yet different, technologies. Thus, it is important for any work that involves web technologies to clearly define the context for the aspect of "the Web" that is being considered. This dissertation focuses on the development and evaluation of techniques that analyze "the Web" using only anonymized TCP/IP headers — this focus touches upon multiple technologies and research areas that are each related to it. This chapter provides the context for the aspects of "the Web" that this dissertation considers and is divided into three parts:

- *Background:* The first part of this chapter provides a brief background on the Web. Here, we clearly define key terms such as the Web, a web page, and web page traffic. We also provide background on web technologies including Hypertext Markup Language (HTML), javascript, Hypertext Transfer Protocol (HTTP), and other protocols that are used by web technologies including Domain Name System (DNS), Transmission Control Protocol (TCP), and Internet Protocol (IP). Readers familiar with this background material may read ahead to Section 2.3.

- *State of the Art in Web Measurement Methods and Tools:* The second part of this chapter discusses the methods and tools used for Web measurement. This discussion provides additional background on how related studies collect data to study the Web.

- *Related Work:* The last part of this chapter discusses the literature on the research areas that is related to this dissertation. These include Web measurement studies, traffic classification methods, and web page segmentation methods.

2.1 Background

2.1.1 Background on the Web

What is the Web? The Web is an application that allow clients to obtain files from web servers [Kurose, 2005] over the Internet. A client is an end-host (i.e., device) that initiates the request for a file from a server, while a server is an end-host that responds to a client's request by sending the requested file to the client. HTTP is an application layer protocol that specifies the format for the communication between web clients and servers. Web browsers, such as Google Chrome, implement client-side components of HTTP, while web servers, such as Apache Tomcat, implement the server-side components of HTTP. Any communication between web clients and web servers that use HTTP as its application layer protocol is referred to as web traffic.

What is a Web Page? A *web page* is a file that uses the HTML markup language to specify how information should be displayed via a web browser.[1] The HTML language allows web pages to reference files that are needed to properly display the web page. These files are either located locally on the same end-host as the web browser, or externally on a web server, and are traditionally called web objects. Web objects are files that can be referenced with a Uniform Resource Locator (URL) — these include a number of different file types including HTML, image, audio, and video files to name a few. A URL is an identifier that specifies the location of a web object in terms of the hostname — a human-friendly identifier for a device — of the web server and the path where the web object can be retrieved on the web server. In the context of this dissertation, a web page is a term that refers to any HTML file that is referenced by a URL and is displayed/rendered on a web browser. Therefore, other file types (e.g., image, audio, and video files) and HTML files that are displayed on a mobile app are not considered web pages. Please note, that this definition also means that web pages with URLs that share the same hostname, say www.yahoo.com and www.yahoo.com/sports/article.html, are considered different web pages — instead, we use the term *web site* to refer to a set of web pages that share the same hostname.[2]

There are also many other features and technologies used by web pages including (i) hyperlinks which allow for navigation between web pages, (ii) javascript which allow for dynamic interaction with a web page,

[1] Web pages are stored as HTML files.

[2] We formally define terms such as hostname later.

and (iii) Cascading Style Sheets (CSS) which allow for externally modifying the style in which web pages are displayed. While these technologies are widely used for designing web pages and likely influence web page traffic (discussed next), we do not explicitly control for these in this dissertation. More specifically, we simply assume that these technologies exist and allow them to execute *without* controlling for, modifying, or disabling them. We do this primarily because this dissertation focuses on studying and analyzing *web page traffic*, but also because these technologies (particularly javascript) are difficult to manipulate and control for in a scientific manner without manually inspecting and modifying them for each web page studied. Explicitly studying the impact that these technologies have on web page traffic is valuable and is an avenue for future work.

What is Web Page Traffic? *Web page traffic* is a term that we use to refer to the data that is transmitted by web clients and servers that is used to display/render a web page via a web browser — this definition excludes data that is transmitted by non-browser web clients, such as mobile apps. In this dissertation, web page traffic includes the following:

- Data obtained from the different network protocols that comprise the Internet Protocol Stack [Kurose, 2005]. Data that is not transmitted over the network, such as data that is cached by a web browser, is not considered web page traffic.
- Data from all of the web objects (e.g., source HTML, image files, audio files, etc) that are transferred over the network that are explicitly referenced by a web page's source HTML. This includes traffic generated by web pages that are downloaded due to being referenced using the iframe HTML tag.
- Data that is automatically generated by web technologies, such as javascript, AJAX, or the web browser, that results from processing the source HTML. Please note that this includes traffic generated by modern web features such as instant search (updating the content of a page by typing), auto-play (automatically playing a video), and ever scroll (loading new content on a page by scrolling). This also includes traffic generated by DNS and web object prefetching features within browsers.

A *web page download* refers to the web page traffic that occurs when a browser downloads and renders a *single* web page. Thus, web page traffic consists of one or more web page downloads. This distinction between an individual web page download and web page traffic is important for some of the applications discussed in this dissertation (e.g., web page classification and web page segmentation) because it is the basic unit in which we make inferences on Web usage. Some details regarding the specifics of the definition of a web page download used in this dissertation, including when a web page download starts and ends, are provided below:

- A web page download starts when a web client (browser) sends the first packet to initiate the web page download process — we describe this process this detail later in this section.

- A web page download ends when the last packet is transmitted between the client and web servers involved in the web page download. This means that web page downloads that transmit traffic after the web page has displayed/rendered on a browser may continue for an indefinite amount of time. It is difficult to predict when a web page download ends unless the browser window in which the web page is rendered is closed. This is because many web pages leverage technologies that automatically generate traffic (e.g., AJAX technologies).

- Any interaction with a web page that initiates a new web page download or reloads a web page is considered the start of a new web page download. Thus, dynamic and automated web page interactions, such as instant search or video auto-play, that *update* an already rendered/displayed web page or simply download additional web objects are not categorized as a new web page download — the traffic generated by these interactions correspond to the web page download in which it originated. Some web page interactions which may initiate a new web page download include clicking on a hyperlink and submitting a search query in a search engine.

- It is possible for multiple web page downloads to overlap (i.e., occur during the same time interval on the same web client) — this is because many modern browsers support multi-tab and/or multi-window browsing. While the traffic from multiple web page downloads may overlap, we always assume that each packet observed originates from, or is considered part of, a single web page download.[3] However, it is difficult to demultiplex web page traffic (i.e., determine which traffic corresponds to which web page download) in practice — in fact, this problem is an active area of research [Xie et al., 2013, Newton et al., 2013, Neasbitt et al., 2014, Ihm and Pai, 2011, Barford and Crovella, 1998].

[3] Please recall that a web page download consists of traffic that is referenced by the source HTML of a single *web page* or any traffic generated by the browser that is due to processing this source HTML.

What is the Internet Protocol Stack? The Internet Protocol Stack is a collection of network protocols that are used for communication of packets over the Internet — the term packet is jargon for a formatted unit of data that is transferred over a network. Network researchers and designers organize these network protocols into *layers* to organize the structure and design of these network protocols [Kurose, 2005]. The Internet Protocol Stack consists of 5 layers — the application, transport, network, link, and physical layers [Kurose, 2005]. Each layer includes network protocols that implement services that are used by other layers. The Web requires the services from each layer of the Internet Protocol Stack. An overview of the services provided by each layer of the Internet Protocol Stack is provided below.

- *Application Layer:* The application layer includes protocols that specifies the format for the exchange of packets between applications on end-hosts [Kurose, 2005]. Packets at the application layer are referred to as *messages*. For example, HTTP is the application layer protocol that is used to request and transfer web objects, while DNS is the application layer protocol that translates between device identifiers that are understandable by humans and the network.

- *Transport Layer:* The transport layer provides services that support the delivery of application-layer messages. Packets at the transport layer are referred to as segments. TCP and UDP are two transport layer protocols used in the Internet. Some common services provided by transport protocols are provided below:

  - Error detection: Transport layer protocols use error detection mechanisms, such as a checksum, to determine whether the segment transferred between end-hosts have been corrupted. Corrupted segments are discarded and are not used by the application layer protocol.
  - Multiplexing: Transport layer protocols also support multiplexing. Multiplexing allows for the end-hosts to transfer segments for multiple network applications simultaneously.

  TCP also provides additional services such as reliable delivery, flow control, and congestion control. We provide additional details on some of these services in Section 2.1.2.

- *Network Layer:* The network layer provides services that support transferring network-layer packets, referred to as datagrams, from one host to another [Kurose, 2005]. IP is the network layer protocol that is used by the Internet. Two important services that are supported by the network layer are routing and forwarding. Forwarding refers to the action of transferring a datagram from an inbound link interface to an appropriate outbound link interface on a local host, while routing refers to the network-wide process that determines the end-to-end path that datagrams take between two end-hosts [Kurose, 2005].

- *Link and Physical Layers:* The communication medium that connects two adjacent hosts in a network are referred to as links. The link layer provides services that support the transfer of link-layer packets, referred to as frames, across the link between two adjacent hosts [Kurose, 2005]. Some examples of link layer services include medium access control and error detection [Kurose, 2005]. The physical layer transports individual bits within link layer frames across a link — please note that the notion of a packet does not exist at the physical layer because bits are the smallest unit of data for digital communication. There are physical layer protocols for that correspond to different mediums of links including wireless, coaxial cable, and fiber optic cable — in each case, a bit is transferred across a link differently [Kurose, 2005]. We do not discuss the link and physical layers in detail because this dissertation is concerned only with the information present at the application, transport, and network layers.

We present an overview of how some of network protocols, in particular HTTP, DNS, TCP, UDP, and IP, are used to download a web page in the next section.

2.1.2 Web Page Download Process - Web Traffic Background

High-level Overview of the Web Page Download Process Multiple steps are required in order for a web client to download and display a web page. These steps require the use of network protocols and web technologies.The high-level process of downloading and displaying a web page, with a focus on application layer protocols, is described below:

1. *Hostname resolution using DNS:* The first step to download a web page is for the web client to determine the IP address a web server given its hostname — this process is referred to as hostname resolution. An IP address is an identifier that is used by IP to route datagrams over the Internet. DNS is the protocol that is used resolve a hostname to an IP address. Figure 2.1 shows an illustration of this process. Here, the client makes a DNS request to a DNS server (i.e., a server that provides DNS services) to obtain the IP address for the hostname amazon.com. In this example, the DNS server responds to this request with the appropriate IP address. Please note that Figure 2.1 also illustrates

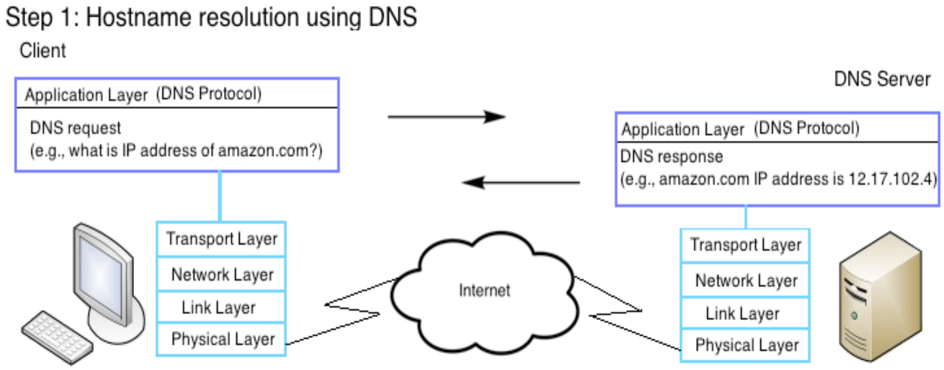

Step 1: Hostname resolution using DNS
Client
Application Layer (DNS Protocol)
DNS request
(e.g., what is IP address of amazon.com?)
Transport Layer
Network Layer
Link Layer
Physical Layer
Internet
DNS Server
Application Layer (DNS Protocol)
DNS response
(e.g., amazon.com IP address is 12.17.102.4)
Transport Layer
Network Layer
Link Layer
Physical Layer

Figure 2.1: Illustration of the hostname resolution process.

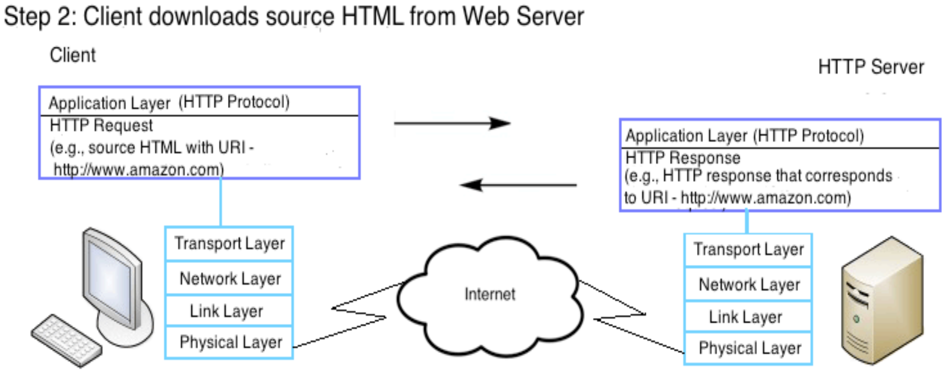

Step 2: Client downloads source HTML from Web Server
Client
Application Layer (HTTP Protocol)
HTTP Request
(e.g., source HTML with URI - http://www.amazon.com)
Transport Layer
Network Layer
Link Layer
Physical Layer
Internet
HTTP Server
Application Layer (HTTP Protocol)
HTTP Response
(e.g., HTTP response that corresponds to URI - http://www.amazon.com)
Transport Layer
Network Layer
Link Layer
Physical Layer

Figure 2.2: Illustration of a client downloading a source HTML file from a server.

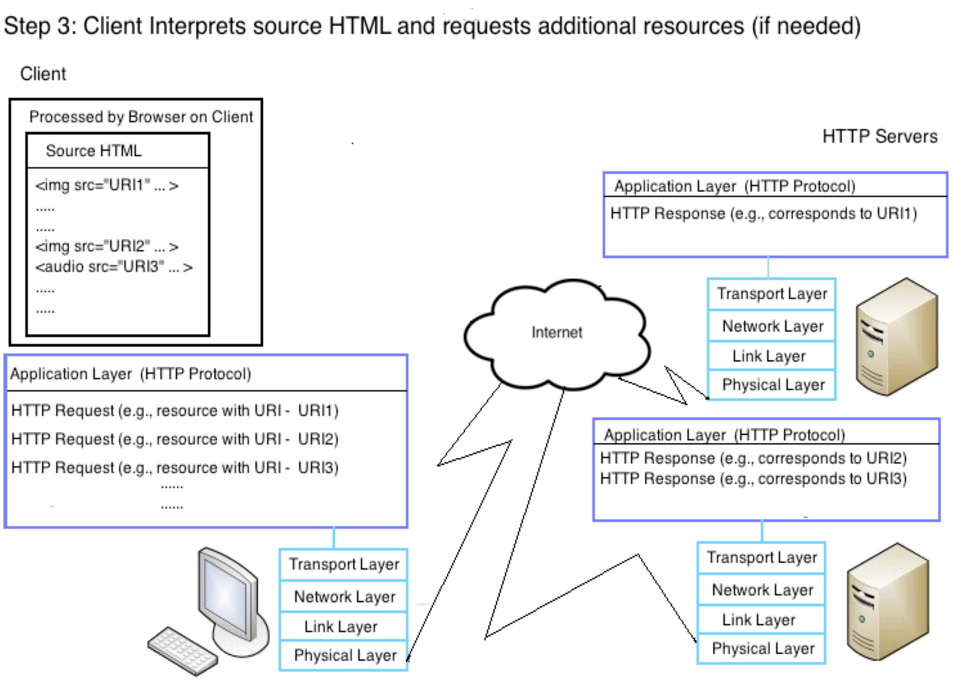
the existence of lower layer protocols on the end-hosts (i.e., Transport, Network, Link, and Physical layer protocols) — though, their functionality are not shown in detail for brevity. Application layer protocols, such as DNS and HTTP, are not aware of lower layer protocols and logically communicate directly over the Internet. This direct communication between the application layer protocol, DNS in this case, is shown with the black arrows in Figure 2.1.

2. *Client downloads source HTML from Web Server:* The second step to download a web page is for the client to download the source HTML from the web server. This step involves the following:

   - The client must first send an HTTP request for the HTML source file to the appropriate web server. This HTTP request is shown in Figure 2.2.
   - The web server then receives the HTTP request from the client. If the server can satisfy the request for the web object (in this case, the HTML source file), the server will send an HTTP response with the appropriate web object (HTML source file) to the client. This HTTP response is shown in Figure 2.2. If the server does not contain the web object it will send an HTTP response indicating that an error occurred.

3. *Client Browser interprets HTML source and Downloads Additional Web Objects if Needed:* For modern web pages, the HTML source will typically include references to other web objects that are located on other web servers. These web objects must be requested by the web browser using new HTTP requests. The process of requesting additional web objects continues until all the web objects referenced by the web page have been downloaded — though, the use of javascript and other technology allow for additional objects to be requested indefinitely. This process is illustrated in Figure 2.3. Please note that the additional web objects are not necessarily located on the same web servers — this concept is also illustrated in the Figure 2.3. Also note, that we do not include the black arrows that illustrate the logical communication between the application layers between end-hosts for simplicity.

These steps describe the high-level behavior of the application layer protocols, DNS and HTTP, that are used by the Web. We provide additional details for these protocols as well of other protocols and web technologies that are related to this dissertation next.



Step 3: Client Interprets source HTML and requests additional resources (if needed)
Client
Processed by Browser on Client
Source HTML
<img src="URI1" ... >
.....
.....
<img src="URI2" ... >
<audio src="URI3" ... >
.....
.....
Application Layer (HTTP Protocol)
HTTP Request (e.g., resource with URI - URI1)
HTTP Request (e.g., resource with URI - URI2)
HTTP Request (e.g., resource with URI - URI3)
......
......
Transport Layer
Network Layer
Link Layer
Physical Layer
Internet
HTTP Servers
Application Layer (HTTP Protocol)
HTTP Response (e.g., corresponds to URI1)
Transport Layer
Network Layer
Link Layer
Physical Layer
Application Layer (HTTP Protocol)
HTTP Response (e.g., corresponds to URI2)
HTTP Response (e.g., corresponds to URI3)
Transport Layer
Network Layer
Link Layer
Physical Layer

Figure 2.3: Illustration a client browser downloading additional web objects.

Hostname Resolution with DNS DNS is the protocol that is used by the Internet for hostname resolution. The hostname resolution process with DNS is described below:

- DNS request: The client sends a DNS request to its local DNS server — this request is either initiated by the browser or by the operating system on the browser's behalf. The DNS request includes the hostname that the client needs to resolve to an IP address.

- DNS response: The local DNS server responds to the DNS request with a DNS record. The DNS record includes the IP address of the requested hostname. DNS can also be used to obtain the hostname that corresponds to an IP address — this is known as a DNS reverse lookup.

Details of the HTTP Request and Response Process

- HTTP request: An HTTP request consists of an HTTP header that includes information regarding the web object that the client wants to download. *HTTP headers* contain many fields that organizes this information according to the HTTP protocol standard [Fielding et al., 1999]. Some of the most important fields include:

  - URL: An identifier that specifies the web object to download.

  - Host: The domain name of the server to contact to download the object specified by the URL. In many cases, browsers derive this information from the URL.
  - Method: Specifies the HTTP request type. The two most common types of HTTP requests (methods) are GET and POST. The GET method is typically used to specify that the client wants to retrieve whatever information that is identified by the URL. The POST method is used by the client to send information designated by the URL (i.e., HTTP content) to the server — please note that the actual action performed by the server when the POST method is received depends on the particular server that receives the POST method and the URL that is specified by the client. Other HTTP methods include TRACE, CONNECT, PUT, DELETE and HEAD [Fielding et al., 1999].
  - User-agent: Specifies details of the client that is making the HTTP request. These details may include the browser type, browser version, operating system, and even CPU-related information. The server may use the information provided in the User-agent field as a parameter to help target HTTP responses to different clients.
  - HTTP Version: Specifies the version of HTTP that is used during the communication between the client and server. Most web traffic that is transferred today uses HTTP/1.1.

- HTTP response: An HTTP response consists of an HTTP header that includes information about the web object that the server is sending to the client. HTTP responses may also include the web object that was requested by the client that make up the payload of the message. Some of the important fields in the HTTP header of an HTTP response include:

  - Status: The status field includes information about the attempt by the server to satisfy the HTTP request. There are many status codes, some of which are more popular than others. A status code of 2XX corresponds to a request that was successfully received and processed. Here, an "X" corresponds to any digit 0-9.[4] A status code that is not 2XX typically corresponds to an error that occurred while interpreting or satisfying the HTTP request. For example, status codes of 3XX correspond to redirections (i.e., the server sends the request to another server because it cannot satisfy it alone), while status codes of 4XX and 5XX correspond to client and server

[4] The most popular status code for a request that was processed successfully is 200.

errors respectively.

- Mime-type: The mime-type specifies metadata about the category of the web object that is sent by the server. The mime-type contains a type and a subtype of a web object. An example mime-type is text/html. Here, text is the type and html is the subtype. There are many different mime-types that specify the category of a web object including image/jpeg (i.e., an image of type jpeg), application/javascript (i.e, a javascript application), and video/mp4 (i.e., a video of type mp4). Freed et al. [2015] provides a more complete list of mime-types.
- Cache-control: This field specifies whether or not the web object sent should be cached. This information is accompanied with information regarding how long the object should be in the cache before expiring. An object that is in the cache (i.e., has been downloaded previously) can be reused as long as the cached object has not expired. Caching is a common strategy to improve network performance by reducing the number of redundant requests for data. The cache may be located on the client or on a proxy service within the Internet. The Pragma field in the HTTP header may also be used to inform clients or proxy services that the web object can be cached.
- HTTP Version: Specifies the version of HTTP that is used during the communication by the client.

HTML Formatting As noted before, the HTML source file contains information about how to display content. The information regarding how to display content is defined by HTML tags which are the building blocks of an HTML source file. For example, the <i> tag is used to denote that text should be italic, whereas the <table> tag is used to denote that information should be in the format of a table. Some HTML tags function as references to web objects (content external to the HTML source file) that should be included to display the web page. For example, the <img> tag can be used to reference an image that can be displayed on a web page. In this example, if the image is not located on the client it must be downloaded via an HTTP request. It is the responsibility of the web browser to (i) interpret the HTML source so that it can make all of the necessary HTTP requests for the referenced web objects and (ii) display the information and web objects as specified by the HTML source.

Underlying Transport Protocols (TCP and UDP) Application layer protocols, such as HTTP and DNS, require the services of transport layer protocols to support the delivery of messages over the Internet. TCP is a transport layer protocol that is used by HTTP, while UDP is used by DNS. Segments consists of both header information and payload information. Header information provides logic for how the segment should be treated by the network and end-host applications (depending on the layer of the protocol), while payload information is the actual application-layer message, including the application layer headers, that is transferred. The header information in transport protocols are used by end-systems and the Internet to successfully provide its services to application layer protocols.

TCP provides a number of services that support the transmission of segments over the Internet. Some of these services include:

- Reliability: A service ensures that segments are reconstructed at the receiver in order, despite any segment disordering, corruption, or loss that may have occurred during transmission.
- Congestion control: A service that controls the rate in which segments are transferred across a network without causing network congestion. Network congestion is not desirable because it can significantly reduce the rate in which segments are transferred between end-hosts.
- Flow control: A service that controls the rate in which the sender sends segments to the receiver such that the receiver can process the segments without being overwhelmed by the transmission rate of the sender.

TCP requires maintaining state about the status of segment transmissions to support these services. This state is established using a handshaking procedure known as the *three-way handshake* and is denoted as a *TCP connection*. A TCP connection is a sequence of segments between two end-hosts — TCP connections are also referred to as flows in this dissertation. Maintaining the state required for TCP connections may be costly in some scenarios. UDP, another transport protocol, is known as a *connectionless protocol* because it does not provide services such as reliability and congestion control that require maintaining state about the segments transferred. UDP is suitable for applications that are impacted by the overhead required of TCP (e.g., real-time applications).

A TCP header contains at least 20 bytes of information and includes many different fields that serve a specific purpose. A description of some of the most important fields in a TCP header is shown below, while an illustration of the structure of a TCP header is shown in Figure 2.4.

- Sequence Number: The sequence number is a value that is used by TCP to ensure that segments are received in the correct order.

- Acknowledgement Number: The acknowledgement number is a value that is used by TCP to verify that a segment with a corresponding sequence number was received.

- Source Port Number: A *port number* identifies the specific application process in which the segment can be forwarded to when it arrives to the receiver. The port number of the end-host that sends the segment is known as the source port number.

- Destination Port Number: The port number of the end-host that receives the segment.

- Flags: TCP headers include a 8 bit field for flags that act as control messages for TCP — this 8 bit field is shown in Figure 2.4 to the right of the Reserved field. Each bit of this 8 bit field corresponds to a different control message. The SYN flag is used to establish a TCP connection. The ACK flag is used to indicate that a segment was successfully transmitted and received. The FIN and the RST flags are used to close TCP connections, while the PSH flag is used to tell the receiving application to immediately load the segment (i.e., do not buffer the segment). Please note that more than one of these flags can be set at the same time. There are also other flags that are outside the scope of this work.

A UDP header, not shown, contains 8 bytes of data and also include fields that allow UDP to function properly. We do not provide any more information on UDP since it is not the focus on this dissertation.

Underlying Network Protocol (IP) IP is the network layer protocol that is used by both TCP and UDP. We focus our background discussion on IPv4 because it is the most widely used version of IP in the Internet today. An IP header contains at least 20 bytes of data and includes many different fields that serve a specific purpose. A description of some of the important header fields for IP is shown below, while an illustration of the structure of an IP header is shown in Figure 2.5:

- Protocol: A number indicating the transport layer protocol that is using IP. A value of 6 corresponds to UDP, while a value of 17 corresponds to TCP.

- Source IP address: The IP address of the end-host that sends the datagram.

- Destination IP address: The IP address of the end-host that receives the datagram.

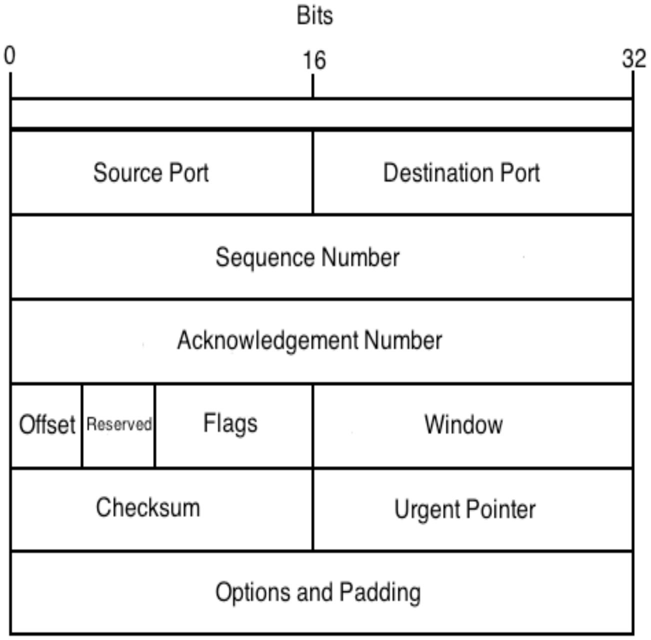


Figure 2.4: Illustration of the structure of a TCP header.

- Total Length: A 16 bit field that represents the number of bytes transmitted by the datagram.
- Options and Padding: The IP also contains an optional options and padding field. The options and padding fields are not commonly used, so most IP headers consists of 20 bytes of data.

We refer readers to the work by Kurose [2005] for more details about other protocols (i.e., link and physical layer protocols) that are also used by the Internet. We only provide an overview of the application, transport, and network layer protocols that are studied in this dissertation.

TCP/IP Headers and Anonymization A *TCP/IP header* is nomenclature for when analysts refer to both the TCP and IP headers of a datagram at the same time. This nomenclature is useful because TCP and IP headers are tightly coupled and many important concepts require analyzing both headers simultaneously. For example, datagrams from the same TCP connection have the same source port, destination port, source IP address, destination IP address, and transport layer protocol. At a high-level, an *anonymized TCP/IP header* is TCP/IP header where the source and destination IP addresses have been anonymized or obfuscated in a manner in which they cannot be easily used to obtain the hostname of the servers they reference — recall that a reverse DNS lookup can be used to determine the hostname of a server given its IP address. Anonymizing IP addresses is a technique that is used to increase the privacy of users by making it more difficult for an

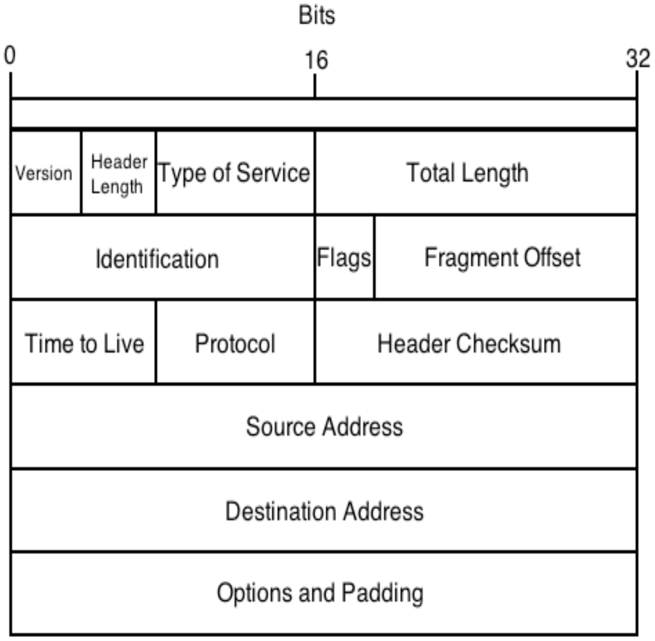


Figure 2.5: Illustration of the structure of an IP header.

analyst to determine which servers were communicating with each other. This process is done because hostnames alone may contain information about the type of content that a user is consuming. For example, a user that contacts a server owned by webmd.com is likely consuming health-related content.

A simple approach to anonymizing a TCP/IP packet trace is to map each *real* IP address observed in the trace to another *random*, or otherwise *obfuscated*, address. Usually, the only requirement is that the mapping from each real IP address to each random address be one-to-one [Fan et al., 2004b] — the anonymization of the TCP/IP trace is achieved by keeping the mapping between these addresses secret [Fan et al., 2004b]. There are two basic methods to performing this simple anonymization: 1) Table-based methods; and 2) Cryptography-based methods. These are described below:

- *Table-based* anonymization methods create a one-to-one mapping between a *real* IP address and a random address by simply generating a new random IP address for each real IP address observed in the trace — some tools that anonymize traces in this manner include TCPdpriv [Minshall, 1996] and IPsumdump [Kohler]. One drawback of table-based methods is that the mapping between real IP addresses to random IP addresses is *random* — that is, a different mapping will usually be generated each time a table-based method is applied to a traffic trace. This is a potential problem because the one-to-one mapping between IP addresses will not be preserved across multiple traces which may include traffic from hosts that have the same IP address — this drawback also means that table-based methods cannot be easily used in a parallel/distributed computing environment.

- *Cryptography-based* anonymization methods create a one-to-one mapping between IP addresses by leveraging concepts in key-based encryption [Daemen and Rijmen, 1999, Fan et al., 2004b]. Essentially, given a key and an encryption algorithm, the one-to-one mapping is achieved by *encrypting*, or hashing, a *real* IP address to an obfuscated IP address — the anonymization is achieved by keeping the key secret.[5] Also, contrary to table-based methods, cryptography-based methods are able to create one-to-one mappings that are consistent across multiple traces provided that the key used for the anonymization is the same. The most well-known cryptography-based anonymization tool is perhaps Crypto-PAn [Fan et al., 2004a] — many other network measurement tools incorporate or extend Crypto-PAn [Armitage and But, 2007, Kirstoff, 2005, Moore et al., 2001].

By default, the one-to-one mapping of IP addresses does not preserve the bit-order of the original IP addresses. Indeed, the bit order of an IP address reveal information about the relative location of a device/host and the structure of the network — such information is useful for some applications such as routing performance analysis [Fan et al., 2004b]. Anonymization algorithms are referred to as *prefix-preserving* if the bit-order of the IP addresses in the trace are preserved such that the network structure is maintained. Most of the anonymization tools that we discussed (both table-based and crytography-based) support prefix-preserving anonymization [Minshall, 1996, Fan et al., 2004a].

It is important to note that while anonymization methods make it more difficult to determine which hosts were communicating on a network, they may still leak information that can be used to determine the hostname of users. For example, an analyst may combine frequency-related information obtained from an anonymized TCP/IP trace, say a list of the most frequently observed anonymized IP addresses, with information that is obtained from third party analytics services, say a list of the most popular hostnames, to *guess* the hostnames that correspond to some of the anonymized IP addresses. This risk of information leakage is a legitimate concern that must be considered when using anonymization methods in general [Fan et al., 2004b]. Please note that prefix-preserving anonymization methods leak potentially sensitive information because they intentionally preserve the network structure. Thus, prefix-preserving methods should not be used unless knowledge of the structure of the network is *needed* for the intended application/study.

[5] In the field of cryptography, the encryption algorithm is usually assumed to be known [Stallings, 2006].

In this dissertation, when we refer to an anonymized TCP/IP traffic trace we assume that:

- All application layer header and payload information was removed — this includes DNS and HTTP headers. Thus, application layer information such as hostname, MIME-type, and HTTP status codes, are not included in the dataset. While methods exist for anonymizing application layer and payload information [Koukis et al., 2006, Pang and Paxson, 2003], we do not leverage these because we assume that traffic is encrypted.[6]

- Simple anonymization was applied to the TCP/IP headers such that the mapping between *real* IP addresses and *obfuscated* addresses is one-to-one. We do not assume that the anonymization is prefix-preserving since, in this dissertation, we do not explicitly consider applications where understanding network structure is critical. Also, we do not make *strong* assumptions about the frequency or structure of the IP addresses observed in the trace which may weaken the anonymization itself — more specifically, we do not use the IP address to resolve or link the hostnames of different clients and servers in a non-trivial way.

- All other information present in the TCP and IP headers, such as the port numbers, length fields, and TCP flags, remain unchanged.[7] Please refer to Table 2.4 and Table 2.5 for a visualization of these fields for the TCP and IP headers, respectively.

Please note that we refer to the information present in an anonymized TCP/IP traffic trace as *limited* because some information in the original traffic trace was either removed (application layer headers) or obfuscated (IP address).

Limitations of Using Different Types of Web Page Traffic The Web is studied by analyzing the functionality of these different types of technologies and protocols that are used by the Web. Some aspects of the Web can be more easily analyzed using certain types of data than others. For example, HTML source data is useful for understanding the semantics of a web page but it cannot be used to directly understand the network implications of a web page download. Similarly, DNS data is useful for understanding the hostnames and IP addresses of the servers that are referenced by a web page but it cannot be used to *exactly* understand which

[6] Hence, simply removing the application layer header and payload information is reasonable for our scenario.

[7] It is important to note that these fields may be obfuscated to improve user privacy [Koukis et al., 2006]. However, most popular anonymization tools do not support/perform such obfuscation [Minshall, 1996, Fan et al., 2004a].

web objects were requested via HTTP — although, methods exist that are able to determine which DNS and HTTP messages that are *likely* associated [Krishnan and Monrose, 2011]. The many different types of data available (e.g., HTML source, HTTP header, TCP header, IP header, DNS header, and message payloads) allows for the analysis of the Web from many different perspectives. Generally speaking, HTML source, HTTP header, and message payload data are most useful for analyzing Web usage in terms of the types of pages that are downloaded and the patterns in which users navigate web pages, while TCP/IP headers are most useful for analyzing the performance of the network. *The challenge addressed in this dissertation research is inferring web usage (i.e., the types of web page downloaded and the patterns in which users download pages) using only the limited information available in anonymized TCP/IP headers.*

2.2 State of the Art - Web Measurement Methods and Tools

2.2.1 Web Measurement Methods

In the previous section, we present a broad overview of the web page download process and the many technologies that are needed in order for it to complete successfully. Web pages can be studied by collecting and analyzing web page traffic. The type of analysis that can be performed using web page traffic depends on the data collection methodology. Web page traffic can be collect either (i) on the server, (ii) on an aggregate link on the network, and/or (iii) on the client. In this section, we describe some of the common approaches to measuring web page traffic and the corresponding types of analysis that can be performed.

Measuring web page traffic on the server Servers host the web objects that clients download. A traffic monitor can be placed on the server to collect and analyze the traffic that it sends and receives. Some common metrics that are collected by servers include the timestamp of the HTTP request, the IP address of the client, the requested web object URL, and the status code — many of these metrics are derived from analyzing TCP/IP and HTTP headers. It is not uncommon for servers to maintain state about how users access its web objects — cookies exchanged in HTTP request-responses or login-related information are commonly used for this purpose [Spiliopoulou et al., 2003, Jansen and Spink, 2006]. This type of information is typically recorded in a server log. Server logs have great value to content providers and are routinely processed for a number of purposes including security analysis, performance analysis, and to improve the functionality of their services. It is because of this value that server logs are considered proprietary information and are not commonly made available to the public. Content providers are able to leverage both the provider protection and informed consent exceptions in the law that allow them to analyze server logs without limits — this monitoring is allowed because one party of the communication (i.e., the server) consented to the communication [Ohm et al., 2007].[8]

While collecting and analyzing web page traffic on the server is important for understanding how a content provider's web service is used by users, it only provides data about a single server [Jansen and Spink, 2006]. Monitoring a single server results in a limited view of the web page traffic that is observed on other servers. This limitation is particularly an issue because modern web pages may refer to objects that are located on dozens of servers. Monitoring many servers from a diverse sample of content providers is an alternative to monitoring a single server. However, this alternative will require a significant effort to place monitors at the different servers which may be owned by other content providers to be useful [Jansen and Spink, 2006].

Measuring Web page traffic on an aggregate Link ISPs and network administrators typically use a Data Acquisition and Generation (DAG) card to monitor web page traffic that flows on their network on an aggregate link — these devices/monitors are placed within the Internet [Benevenuto et al., 2009, Smith et al., 2001, Newton et al., 2013, Lim et al., 2010, Gu et al., 2008, Ihm and Pai, 2011, Shafiq et al., 2012, Xu et al., 2011, White et al., 2013, Erman and Ramakrishnan, 2013, Schneider et al., 2008]. These monitors capture packets and store them for post-analysis — some tools, such as NetFlow [Claise, 2004], record statistics of the network properties instead of capturing and storing each individual packet.[9] This web page traffic can be archived for an indefinite amount of time.[10] This approach results in a rich dataset that includes the communication between many clients and servers, where the scope of the measurement depends on where the monitor is placed. For example, an ISP will have access to more web page traffic than a University because its network is larger and reaches more users.

HTTP headers, message payloads, and TCP/IP headers are captured when collecting web page traffic on an aggregate link.[11] However, due to privacy laws, HTTP headers, message payloads, and IP addresses cannot be analyzed by these entities without a provider protection exception or consent [Ohm et al., 2007].

[8] Please note that in some regions in the U.S., both parties of the communication must consent to being monitored. In either case, the provider protection exception can easily be applied.

[9] Makes copies of the packets.

[10] On the order of days to years. The archival time usually depends on the amount of disk space available for archival.

[11] Please note that HTML source information can be extracted from payload of messages.

Contrary to the analysis of data collected on the server, consent is difficult to obtain because network monitors are eavesdropping on the communication between clients and servers. In other words, the client and the server are not aware that they are being monitored. Thus, obtaining consent to analyze client and server communications is not easy. The provider protection exception is difficult to obtain as well because one must justify the need to collect content-level information for managing a network. Thus, most aggregate measurement collection efforts focus primarily on improving the performance and efficiency of the network instead of analyzing how users use web services. This limitation places significant barriers on potential applications of aggregate trace data such as behavioral ad targeting and user behavioral modeling [Smith et al., 2001, Maciá-Fernández et al., 2010, Sanders and Kaur, 2015b].

Collecting web page traffic on the client Clients download web objects from servers. Thus, monitoring the client is another approach to collecting web page traffic [Gavaletz et al., 2012, Drago et al., 2012, Butkiewicz et al., 2011, Sanders and Kaur, 2014b, 2015b, Sun et al., 2002, Liberatore and Levine, 2006]. In this scenario, however, all the web objects that are needed to load a web page is present because all of the web objects requested by the client will be transferred to the client by all the servers contacted. In addition, HTTP headers, TCP/IP headers, and message payloads are legally available because the client consented to the communication with the servers. A limitation of collecting web page traffic on the client is that it provides limited diversity of client behaviors that can be observed on the Web. A web crawler is a common tool that is used to collect a large and diverse sample of web page traffic to overcome this limitation [Heydon and Najork, 1999, Boldi et al., 2004].

Some entities may install tools or plug-ins on client browsers that send messages to servers on how users are browsing the Web [Liu et al., 2010]. One method to collect client usage statistics is to reference a web object that needs to be requested within a web page — except in this case, the corresponding HTTP request serves as a beacon for client usage. This type of technology serves as the foundation of modern analytics and tracking services. Many users are unaware of the prevalence of such tracking technology [Chen and Stallaert, 2014, Li et al., 2015a].

2.2.2 Web Measurement Tools

In addition to the many ways that web page traffic can be collected, there are many tools available to analyze it. Most of the tools that are widely available and used are specialized for a particular type of web page traffic such as HTML source data or TCP/IP and HTTP headers. Some of these tools are described next.

Tools for the Analysis of HTML Source Data HTML source data is typically stored as a simple text file. Tools that analyze HTML source data must be able to parse the HTML syntax to extract information about the tags and attributes that describe how the web page should be displayed and the objects that it references. Popular tools that process HTML source files include JSoup, a java-based tool, and Beautiful Soup, a python-based tool [Hedley, 2010, Richardson, 2015].

Tools for the Analysis of Header Data Web page traffic that is transferred on the network is commonly stored using a packet capture tool. Tcpdump and windump are packet capture and analysis tools that use the libpcap library [Jacobson et al., 1989, Deogioanni et al., 2000, Jacobsen et al., 2005]. The libpcap library contains the details of the link technology used for the network communication and is operating system dependent [Paxson, 1999] — Tcpdump is used for MacOSX and Linux operating systems, while windump is used for Windows operating systems. Wireshark is another tool that can capture and analyze web page traffic [Combs et al., 2007]. Wireshark, however, is a GUI-based tool while tcpdump and windump and command-line based. These packet capture tools store data in a .pcap file. These packet capture tools can also analyze web page traffic and output log files that organize protocol header-related information (e.g., IP, TCP, UDP, or DNS headers) in a format that can be easily analyzed using scripts. NetFlow is a tool that processes captured packets and output summary statistics on the flows that correspond to the observed packets [Claise, 2004]. NetFlow is a popular tool that is used by routers to record TCP connection information of the observed packets and is a light-weight alternative to a packet capture tool.[12] Packet capture, NetFlow, and similar tools have been used primarily to study the performance of the network and the communication patterns between the end-hosts [Smith et al., 2001, Gu et al., 2008].

The analysis of HTTP headers using the logs generated by packet capture tools is more difficult than the analysis of other protocols because HTTP header fields are not extracted by these tools by default. Pcap2har is a tool that extracts HTTP header fields from .pcap files and outputs a .har file [pcap2har]. A .har file is a log file that includes HTTP header fields in the, easy to process, json format. There are also browser plugins that can capture and analyze HTTP header such as HTTP Watch and Firebug [Watch, Mozilla] —

[12] Packet capture tools record information about each packet, while NetFlow only record summaries about the flows that the packets correspond to. Though, some information is lost in the process of summarizing packet transfers as TCP connections.

these tools also store the HTTP headers as a .har file. These browser-based tools only capture and analyze HTTP headers and do not analyze TCP or other protocols. These browser-based tools are typically browser-specific so it is difficult to use the same tool for multiple browsers — Firebug, for example, is only available for Firefox. HTTP header analysis tends to be more content-based than say TCP/IP header analysis which tends to be more performance based — though some metrics on network performance can be inferred from HTTP headers as well.

Some browsers, such as Safari and Chrome, include a suite of development tools that are used to analyze and inspect web pages in real-time [Chr, Saf]. These tools include features that are able to perform a number of web page monitoring and analysis tasks which include analyzing source HTML (including javascript), logging messages to the console (both source HTML and HTTP messages), and analyzing network performance (say, request and response times for HTTP requests) [Chr, Saf]. These tools are typically used by developers for debugging and optimizing web pages, and are not commonly used by researchers for Internet measurement studies.

Specialized deep packet inspection tools Packet capture tools can be tuned to record payload/content data — in this context payload data corresponds the payload of TCP segments. However, it takes a significant amount of effort to analyze payload data. The process of analyzing payload data is known as deep packet inspection — HTTP headers are considered payload data here because it can contain sensitive information about the content of the data (i.e., URLs) [Ohm et al., 2007]. Specialized deep packet inspection tools have been developed that perform packet filtering, traffic classification, malware detection, and other network management tasks. nDPI, Libprotoident, and L7-filter are examples of deep packet inspection tools [Deri et al., 2014, Levandoski et al., 2008, Alcock and Nelson, 2012]. Deep packet inspection tools are ineffective on encrypted or obfuscated traffic [White et al., 2013]. Though, recent methods have been developed that can infer the content of payloads using the statistical properties of the HTTP and TCP/IP headers transferred over the network [Barford et al., 2002, Lakhina et al., 2005, Soule et al., 2005, Brauckhoff et al., 2006, White et al., 2011]. We discuss some of these methods in Section 2.3.2.3.

2.3 Related Work

We next discuss the literature that is relevant to the goals of this dissertation research. An overview of the work in the areas of Internet measurement, traffic classification, and web page segmentation is provided below:

- *Internet Measurement:* The first part of this section discusses the literature on Internet measurement. This prior work consists primarily of Internet measurement studies that focus on studying web usage and investigating methods to improve web performance. The literature in the Internet measurement area is diverse because there are many different types of data that can be used for Web measurement and analysis. For example, Web usage has been studied using HTML, HTTP, TCP/IP, and DNS data. While these different types of data all contain information about web usage, each type of data reveals different information about web usage and/or performance. We discuss Web measurement studies that use these different types of data in detail in this section.

- *Traffic Classification:* The second part of this section presents literature on traffic classification. The traffic classification literature consists of three primary types of methods: 1) Methods that rely on deep packet inspection; 2) Methods that focus on classifying broad categories of traffic (e.g., classifying traffic as email or not); and 3) Methods that focus on identifying a predefined set of web pages. An overview of each of these general methods to traffic classification with specific examples of each is provided in this section.

- *Web Page Segmentation:* The last part of this section discusses prior work that is related to web page segmentation. We discuss both methods that examine HTTP headers and methods that work when using only anonymized TCP/IP headers.

2.3.1 Internet Measurement

Prior literature that analyze web page traffic, in particular those that analyze source HTML, HTTP, TCP/IP, and DNS traffic, are discussed next.

2.3.1.1 HTML-based studies

Text and link-based analysis One of the first problems associated with analyzing web page traffic is being able to collect data. The process of collecting web page traffic using a client, particularly HTML-based data, is known as web crawling or simply crawling. Web crawlers are client-side tools that measure the Web by downloading many web pages. Cho et al. [1998] noted that it is infeasible to download all web pages in a reasonable amount of time due simply to the sheer number of web pages that exist and the amount of memory required to store those pages. Thus, Cho et al. [1998] addressed the problem of how to efficiently crawl such that one can crawl and store web pages that are important to a particular query or goal.

One of the first approaches to studying web pages was to use text mining metrics and techniques to analyze the source HTML. For example, Cho et al. [1998] used word count features and inverse term document frequency similarity metrics to analyze the textual content of web pages to determine whether the page was relevant to the input query. Other work leverages some of the well-defined structure of HTML data to help with text mining. Smith and Chang [1997], for example, performs a text-based analysis but also focused on analyzing the visual content referenced by HTML data (e.g., images and video).

The work by Spertus [1997] proposed using link-based information, not just hyperlinks, to study the Web. This study notes that links for HTML data can correspond to hyperlinks between web pages, relationships between the directory structure of web pages on the same web site, relationships between the domain names of web pages, and links related to the taxonomy of web pages [Spertus, 1997]. This work, however, only discusses using this different link information for studying the Web at a high-level and did not perform a large scale study of the relationship between web pages using this link information. Page and Brin [1998] leveraged the hyperlink structure present in HTML data to understand how web pages were organized. Page and Brin [1998] constructed a hyperlink-based graph of the Web where web pages (i.e., HTML document) were represented as vertices. Here, directed edges between vertices were formed from vertex i to vertex j if a hyperlink in web page i exists that referenced web page j. By applying graph theoretical techniques to this hyperlink-based web graph, Page and Brin [1998] showed that important web pages tend to be referenced by many other web pages. The PageRank score, a metric derived from this technique, is used to rank web pages according to their popularity as measured by the number of hyperlinks that reference them. The PageRank score has been used by web crawlers to help make their crawl more efficient by crawling pages with high importance. In fact, Najork and Wiener [2001] showed that the breadth first search crawl strategy is an efficient method to crawling web pages because the web pages crawled tend to have high PageRank. The analysis of source HTML is relevant to this dissertation because source HTML reference the web objects that are transferred over the Internet. Thus, any observation in web page traffic is linked to the source HTML. In Chapter 3, we use these techniques to help quantify the impact that source HTML has on other types of web page traffic including HTTP headers and TCP/IP headers.

Challenges of Crawling Web Pages Much of the previously mentioned prior work highlight some key limitations of crawling web pages. The most important of these issues include the limited amount of disk space available for storage and the amount of time required to perform a crawl.[13] One factor that makes this issue worse is that web pages may change over time. Cho and Garcia-Molina [1999] performed a study that focused on studying how often web pages change. This type of study is important for web crawling efforts because it has implications on how often web crawlers need to download the same web pages.[14] To address this issue, Cho and Garcia-Molina [1999] crawled 720,000 web pages everyday for over 4 months. A key result from this study was the observation that over 50% of all web pages and over 80% of web pages with the .com domain will change within 4 months — Douglis et al. [1997] observed similar results. These results imply that web crawlers need to incrementally crawl the Web to obtain "fresh" versions of a web page. One limitation of this study was that the metric used only identified whether a page changed and not how much a page change. This limitation is a problem because web pages that do not change much may not need to be crawled again. Fetterly et al. [2003] performed a similar study and this limitation by using a text mining technique that can quantify the degree of change between web pages. One of the key findings that Fetterly et al. [2003] report is that the degree of change between web pages varies a lot, and that larger web pages change more often and to a larger degree than smaller web pages.

These observations have been used when performing large web crawling efforts. For example, the Internet Archive has been using incremental crawls to repeatedly download web pages for over a decade [Notess, 2002]. Though, with the continued rapid change and growth of the Web and the diversity of content, large web crawling efforts (e.g., web archival and search engines) still have difficulty collecting a vast amount of fresh web pages [Costa and Silva, 2012]. This issue is exacerbated as web pages become increasingly personalized and depend more on factors such as location [Mikians et al., 2013] and client platform [Johnson and Seeling, 2014]. Understanding the limitations of web crawlers it is important for this dissertation research because we must use one to collect the data used for our analysis. Any significant issues with web crawling, such as not crawling a diverse set of web pages, will limit the scope and application of this work.

Common Specialized Applications of HTML-based data While designing search engines and monitoring the evolution of web pages are examples of analysis-driven applications of HTML-based data, there has

[13] The amount of time to do a representative crawl can take months.

[14] The logic here being that web pages only need to be crawled again if they change.

been much work in other applications as well. For example, Qi and Davison [2009] and Choi and Yao [2005] describe many different techniques for classifying web pages using HTML-based data. These techniques leverage a wide variety of information represented by HTML content to classify web pages including link/graphical properties, textual properties (word counts), and information derived from HTML tags. Shen et al. [2004] show that web page classification performance can be improved if summaries of the text data are taken before applying the classifier. This document summarization procedure improves classification performance by reducing the amount of noise in the feature set. Document summarization for HTML-based data is commonly used in the Web analysis literature [Pera et al., 2010, Delort et al., 2003].

An alternative to doing web page classification using HTML-based data is to do web page topic modeling [Lin et al., 2014, Zhang et al., 2013]. Web page classification and topic modeling are different because web page classification methods use supervised approaches to categorize web pages while web page topic modeling methods use unsupervised approaches to cluster web pages. Supervised approaches require a training dataset that has been labeled/categorized while unsupervised techniques does not. Unsupervised techniques can be preferred over supervised techniques because obtaining labeled data can be prohibitively difficult. Past work on web page topic modeling typically use widely accepted text mining approaches such as Probabilistic latent semantic analysis and Latent Dirichlet allocation [Lin et al., 2014, Zhang et al., 2013].

Pera et al. [2010] and de Boer et al. [2014] use sentiment analysis techniques for analyzing online customer reviews. These techniques are used by content providers to understand and better address the needs of their customers. HTML-based data can even be used for security analysis [Canali et al., 2011, Likarish et al., 2009, Seifert et al., 2008]. These studies use features derived from HTML-based data including the number of tags, features derived from javascript code, and URL information to help predict whether a web page is malicious or not. These approaches are considered to be lightweight because they use textual information (i.e., HTML files) instead of network or other systems-based information [Canali et al., 2011]. These approaches are useful because they can be used as filters to more reliable and computationally expensive systems-based tools for identifying malicious pages [Wang et al., 2006].

This past research, which is summarized in Table 2.1, is complementary to the goals of this dissertation research because it involves developing learning-based techniques to help analyze web pages. The work presented in this dissertation, however, focuses on developing techniques that can analyze web page traffic using only anonymized TCP/IP headers. Results on this study on web page classification using anonymized TCP/IP headers is provided in Chapter 4.

TABLE 2.1: Summary of Prior Work on Internet Measurement (HTML-based studies)

Author(s)	Purpose of Measurement	Type of Data	Scale of study	Analysis Type	Key Insight/Notes
Cho et al. [1998]	Evaluation of Web Crawling Methods	Source HTML	179,000 web pages	Performance Analysis	PageRank performs the best according to numerous metrics for web page importance
Page and Brin [1998]	Evaluation of Information Retrieval System	Source HTML	24 million web pages	Performance Analysis	Modeling the web as a graph is useful for ranking pages
Najork and Wiener [2001]	Evaluation of Web Crawling Methods	Source HTML	325 million web pages	Performance Analysis	Breadth-first crawling strategy is yields pages with high PageRank.
Cho and Garcia-Molina [1999]	Evaluation of Web Crawling Methods	Source HTML	720,000 web pages (4 month span)	Exploratory Analysis	Web crawlers must be selective in which pages to crawl because some web page update more often than others
Douglis et al. [1997]	Evaluation of Web Cache Effectiveness	Web objects	474,000 web objects	Exploratory Analysis	Half of the web pages measured updated within 2 weeks
Fetterly et al. [2003]	Evaluation of Factors Correlation with Web Page Change	Source HTML	150,836,209 web pages (11 week span)	Exploratory Analysis	Rate of change in HTML is correlated with factors such as domain name, number of words, and file size
Mikians et al. [2013]	Investigating Price Discrimination across different Locations	Source HTML	1,500 requests (6 month span)	Exploratory Analysis	Price discrimination exists across different locations
Zhang et al. [2013]	Evaluation of Topic Modeling methods	Source HTML	6.6 million wikipedia pages	Topic Modeling	Online dictionary learning algorithms outperform traditional topic modeling methods
Lin et al. [2014]	Evaluation of Topic Modeling methods	Source HTML	1,119,264 Twitter pages	Topic Modeling	Sparse Topic Modeling outperforms prior topic models
Pera et al. [2010]	Evaluation of Summarization for Sentiment Analysis Approaches	Source HTML	1,000 reviews web sites	Classification	Summarization approaches impact sentiment analysis performance
Shen et al. [2004]	Evaluation of Web page classification	Source HTML	153,019 web pages	Classification	Using summarization for feature selection improves classification
Canali et al. [2011]	Evaluation of Features for Classification of Web Pages	Source HTML	205,073 web pages	Classification	No single class of feature (URL, javascript, etc) is sufficient for classification
Smith and Chang [1997]	Evaluation of Information Retrieval System	Source HTML + Images	16,733 web pages	Classification	It is possible to classify web pages based on the type/subject of the web page
Likarish et al. [2009]	Evaluation of classification methods for detecting malicious javascript	Javascript source	63 million scripts	Classification	One of the first studies that evaluated script-level features for classifying malicious scripts
Seifert et al. [2008]	Evaluation of classification methods for detecting malicious web pages	Source HTML + HTTP responses	561,000 web pages	Classification	Simple rules-based classification can detect malicious web pages with high accuracy

2.3.1.2 HTTP-based studies

HTTP headers provide information about the type of content (e.g., image, audio, etc) that are being transferred over the Internet. The focus of Web measurement studies conducted in the late 1990s and early 2000s, however, were motivated primarily by improving the network performance of the Internet. Thus, early web traffic measurement studies opted to use TCP/IP headers instead of HTTP headers even though they were studying the Web [Crovella and Bestavros, 1997, Barford and Crovella, 1998]. HTTP headers were not used for web traffic measurement until the focus of Internet analysis shifted from being performance-oriented to content-oriented. This shift occurred mainly because the type of content transferred over the Internet has implications on network performance, user experience, and user privacy [Xu et al., 2011, Rao et al., 2011]. This shift also occurred because of an industry-wide trend to add HTTP support for interactive and dynamic applications such as video streaming and file sharing [Popa et al., 2010]. While HTTP headers, TCP/IP headers, and even DNS headers can be analyzed in isolation of one another, it is best to analyze the behavior of all of these network protocols because, their behavior and performance are tightly coupled for Web-based applications. Though, most measurement studies focus on a subset of these protocols to reduce and narrow the scope of the study. We focus on studies that *use* HTTP header information for traffic analysis in this section and discuss studies that use TCP/IP headers and DNS headers in Section 2.3.1.3 and Section 2.3.1.4 respectively.

Aggregate Web Measurement Studies Callahan et al. [2010] conducted a measurement study on the evolution of web traffic at a small institute over a 3 year period of time. This study showed the benefit of analyzing web traffic over an extended period of time. Callahan et al. [2010] shows that web traffic has been growing over the years and even highlighted the prevalence and growth of certain types of technologies including content distribution networks. Other types of analysis including understanding which hostnames were frequently (and infrequently) contacted by users, and investigating the cause and frequency of errors were also performed. This study also showed that the behavior of AJAX applications, such as gmail, can be identified using HTTP headers — in particular, the authors tracked the adoption of gmail by monitoring the number of POST requests over time. Schneider et al. [2008] performed a measurement study that focused exclusively on understanding the network characteristics of AJAX applications. AJAX applications are essentially web pages that automatically transfer traffic without the need for a user to request a new web page. The methodology adopted in this study was to use HTTP headers to filter web traffic depending on whether it was an AJAX application or a traditional web application. AJAX applications were identified by performing regular expressions on the URL field of the HTTP header — the regular expressions used by Schneider et al. [2008] targeted four AJAX applications. Schneider et al. [2008] found that AJAX applications tend to be make many more HTTP requests and consume more bandwidth than traditional web applications because they tend to aggressively prefetch web objects to improve user experience by reducing network latency.

Ihm and Pai [2011] conducted a longitudinal study of web traffic, similar to the study conducted by Callahan et al. [2010]. Ihm and Pai [2011] however presented more results about the trends of web traffic. Some of these results include observing an increase in CSS and javascript web objects that is correlated to the increased adoption of AJAX-based technology. Other key results from this study is that URLs of web objects tend to become more popular despite an increase in the number of unique and unpopular URLs. This result has an implication on web object caching because caching popular URLs will likely help with caching while caching unpopular URLs will likely hurt caching. Ihm and Pai [2011] also showed that client location is correlated with a number of factors including (i) the browser preference among the user population and (ii) the types of web services that are used by the user population. These prior studies are related to this dissertation because they show that web page traffic has evolved over time and provide insight on the properties of modern web page traffic. The results of these studies, which are included in Table 2.2, also suggests that prior traffic analysis methods, such as traffic classification, may be impacted when considering the characteristics of modern web traffic.

Measurement Studies for Emerging Web Applications The Web was widely considered to be the most popular application on the Internet in the late 1990s and early 2000s [Crovella and Bestavros, 1997, Christiansen et al., 2000]. However, the Web's popularity began to decline in the mid 2000s as video streaming, audio streaming, file sharing, email, and most notably P2P applications became more popular [Popa et al., 2010]. In particular, this shift in the popularity of Internet applications resulted in a number of measurement studies that focus on studying the traffic characteristics of P2P applications and being able to compare and distinguish it from web applications [Sen and Wang, 2004, Basher et al., 2008, Erman et al., 2007b, Choffnes and Bustamante, 2008]. Recent studies show, however, that this trend has reversed and web traffic is regaining its traffic share from other applications [Popa et al., 2010]. In fact, an increasing number of applications including video streaming (Youtube and NetFlix) [Rao et al., 2011, Reed and Aikat, 2013], social networking (Facebook) [Benevenuto et al., 2009], and file sharing (Dropbox) [Drago et al., 2012] are using HTTP. Indeed, HTTP has become the defacto standard of communication over the Internet [Popa et al., 2010].

HTTP headers have been used to analyze the traffic generated by these new web applications. However, studying how these new applications are used using HTTP headers is not trivial and depends on the application that is being studied. Rao et al. [2011] and Reed and Aikat [2013] analyzed Web-based video streaming traffic. The web traffic for these studies were collected using a client side measurement methodology — so heuristics were not needed to filter video streaming traffic from non-video streaming traffic. These studies found that the web traffic observed for streaming technologies depends on many factors including the type of web technology used (say Flash or HTML5), browser choice (say Firefox or Internet Explorer) and the provider of the video streaming service (say Youtube or NetFlix)[Reed and Aikat, 2013, Rao et al., 2011]. Rao et al. [2011] explicitly showed that web traffic collected using a client-side measurement methodology can be used to help understand aggregate traffic. In particular, this work used their client-side traffic measurements to develop a model for aggregate video streaming traffic. This model was used to show that tuning certain video streaming parameters can produce smoother video streaming traffic. Reed and Aikat [2013] made similar observations and modeling efforts for video streaming applications except it focused more on simulation applications.

Benevenuto et al. [2009] performed a study about understanding how users use social networking services. This work initially performed client-side measurements to understand the URL structure present in HTTP headers so that it can be used to (i) identify social networking traffic on an aggregate link and (ii) identify the way that users were interacting with different social networking application features (e.g., viewing a profile, uploading a photo, etc). Benevenuto et al. [2009] showed that understanding web traffic in a controlled environment can be used to study web traffic on an aggregate link — though the authors highlight that their approach is error prone. This methodology was used in this study to compare and contrast the web traffic of different social networking services. Drago et al. [2012] use a similar methodology to study the traffic characteristics of Dropbox, an online storage system. Drago et al. [2012] found that users may prefer to access some Web-based services (e.g., storing a file on the cloud) by interfacing with a specialized application that is installed on the client instead of interfacing with a general purpose web browser.

There is a current trend for developers to develop specialized applications for interacting with the Web using mobile devices. These specialized applications are referred to as mobile apps. Xu et al. [2011]

performed a measurement study that focused on characterizing the web traffic observed on mobile apps. Xu et al. [2011] found that the web traffic that corresponds to mobile apps can be easily filtered in aggregate traces using the User-agent field in HTTP headers. This User-agent field uniquely identifies the mobile app that generated the traffic. The biggest contribution of this study is showing that mobile app traffic already make up a non-negligible portion of aggregate web traffic and that the type of device a user has impacts their web browsing experience. Efforts have already been taken to improve the performance [Huang et al., 2010], and user-friendliness of the Web for mobile devices [Johnson and Seeling, 2014].

These studies on emerging applications,which are included in Table 2.2, provide additional insight on the diversity of modern web page traffic. We use this insight in this dissertation to collect a diverse dataset that includes these emerging applications. This diverse dataset is used throughout this dissertation research.

Web Page Performance Bouch et al. [2000] conducted a study to determine if network performance metrics (i.e., QoS metrics) are correlated with a user's subjective perceptions of web performance quality. The results of this study showed that a user's subjective perceptions of web performance quality are, in fact, related to network performance. Galletta et al. [2004] performed a similar study that showed that a user's attitude towards web performance quality diminished as the web page load time increases. Delays in loading web pages can have serious economic consequences. In fact, several recent studies have found that even slight increases in page load delays, say 100ms - 1s, can cause a noticeable drop in web application use and can result in significant revenue losses for major web sites [Schroter, 2011, Group, 2011].

While most web performance studies are done using TCP/IP headers, there has been some recent work to analyze web performance using HTTP headers. Browser-based tools, such as YSlow and PageSpeed, have been developed in industry that use the information present in HTTP, and even DNS, headers to make recommendations on how to improve web page download performance [Yahoo, Google]. However, these browser specific tools do not make performance recommendations based on web page load time — web page load time is not used because it cannot be reliably and consistently measured despite using the same browser and under similar network conditions [Gavaletz et al., 2012]. Instead, these tools provide recommendations on how to improve web page performance based on rules-of-thumb — example rules include avoiding HTTP errors and minimizing the number of HTTP requests. Butkiewicz et al. [2011] conducted a client-side study to specifically study the relationships between HTTP-related features and web page load time. Some of the HTTP-related features that this study highlighted that impact web page download performance include the number of HTTP requests, the number of javascript objects (HTTP responses with MIME type text/javascript or application/javascript) and the number of hosts contacted. These results support many of the rules-of-thumb that are used by industry tools. Ihm and Pai [2011] also studied web page load performance but focused primarily on identifying opportunities for caching web objects — this study was described before.

There have also been some recent studies that study how SPDY [Jerome], a fairly new networking protocol developed primarily at Google, performs on modern web traffic. HTTP/2.0, the most recent version of HTTP that is currently being deployed, adopts many of the ideas and features developed in SPDY [W3Techs]. Wang et al. [2014] performed a client side measurement study to evaluate the performance of SPDY and found that SPDY improves the efficiency of web objects transferred in a single TCP connection. However, Wang et al. [2014] also notes that SPDY does not necessarily improve web page load times. The authors attributed this observation to the fact that most web pages reference objects that are distributed across multiple servers, and the dependency between these web objects is a performance bottleneck that is not addressed by SPDY. The performance of SPDY is also significantly impacted by other network characteristics such as high RTT and high loss rates [Wang et al., 2014]. Erman et al. [2013] performed a similar study on SPDY performance that confirmed that networks with high RTT and loss rates, such as cellular networks, do not benefit from the performance enhancements implemented in SPDY. Erman et al. [2013] recommends using a cross-layer approach to improve the performance of the Web, because web object transfers are dependent on multiple protocol layers.

These prior studies on web performance, which are summarized in Table 2.2, is related to this dissertation because we consider web performance analysis as an application domain. Specifically, we identify some of the limitations of the tools used for web page performance analysis and (ii) show that web page classification can be used to supplement traditional web page performance analysis. These contributions are made in Chapters 3 and 4, respectively.

User modeling using server-side logs Web page traffic can also be analyzed using server-side logs. While these server-side logs are not necessarily considered a HTTP header, they are usually derived from HTTP requests and responses between a client and server.[15] The most unique type of analysis that is possible using

[15] Please note that these server-side logs can also include other types of web traffic. We consider server-log data in the HTTP header analysis portion of this discussion because it is primarily derived from HTTP headers.

TABLE 2.2: Summary of Prior Work on Internet Measurement (HTTP-based studies)

Author(s)	Purpose of Measurement	Type of Data	Scale of study	Analysis Type	Key Insight/Notes
Schneider et al. [2008]	Study Characteristics of AJAX	HTTP, TCP, and IP Headers	≈ 30 million HTTP requests	Exploratory Analysis	AJAX applications send more HTTP requests and consume more bandwidth than normal
Ihm and Pai [2011]	Evaluation of Web Cache Performance	HTTP, TCP, and IP Headers	>138 million requests across 187 countries (5 year span)	Exploratory Analysis	AJAX traffic has been steadily increasing over time and caching behavior has also improved
Rao et al. [2011]	Study Characteristics of Video	HTTP, TCP, and IP Headers	≈ 15,000 videos	Exploratory Analysis	Different streaming technologies have different traffic characteristics
Benevenuto et al. [2009]	Study Characteristics of Social Media	HTTP, TCP, and IP Headers	802,574 HTTP requests (12 day span)	Exploratory Analysis	First to show access patterns of users can be used to improve social network sites
Drago et al. [2012]	Study Characteristics of Cloud Storage	HTTP, TCP, and IP Headers	≈ 4.2 million flows (42 day span)	Exploratory Analysis	First to find that storage applications are highly used and some inefficiencies in traffic management
Johnson and Seeling [2014]	Comparison between mobile vs desktop pages	HTTP Headers	1 million downloads (2 year span)	Exploratory Analysis	Mobile web pages reference fewer objects than desktop pages
Xu et al. [2011]	Study Characteristics of Smartphones	HTTP, TCP, and IP Headers	≈ 22,000 apps (1 week span)	Exploratory Analysis	First to perform large-scale study on smartphone app usage
Callahan et al. [2010]	Longitudinal Study of Web Traffic	HTTP, TCP, and IP Headers	≈ 29 million connections (3.5 year span)	Exploratory Analysis	Some aspects of web traffic have been fairly static over time while others have changed
Liu et al. [2010]	Study Characteristics of User Dwell Time	HTTP headers	205,873 web pages with 10,000+ visits each	Exploratory Analysis	Web page dwell time follows a weibull distribution moreso than an exponential distribution
Bouch et al. [2000]	Identify Network Metrics that are correlated with User performance	HTTP, TCP, and IP Headers	User Study of 22 web pages and 30 participants	Exploratory Analysis	Metrics such as delay is correlated with user-perceived web performance
Butkiewicz et al. [2011]	Study Characteristics of Web Traffic	HTTP Headers	1,700 web pages at 4 locations (7 week span)	Exploratory Analysis	Web page category seems to impact traffic. Also, number of objects and servers impact page load times
Wang et al. [2014]	Evaluation of Web Performance of HTTP vs SPDY	HTTP, TCP, and IP Headers	200 web pages	Performance Analysis	The performance gains of using SPDY over HTTP is usually marginal
Huang et al. [2010]	Evaluation of Web Performance Differences Across Smartphones	HTTP, TCP, IP, and UDP Headers	≈ 20 web pages (4 different smartphones)	Performance Analysis	Factors such as device type and network technology (WiFi and 3G) impact web performance
Gavaletz et al. [2012]	Evaluation of In-browser measurement tools	HTTP Headers	1,000 HTTP responses (across 5 browsers)	Performance Analysis	In-browser measurement tools should not be used because measurements vary wildly
Erman et al. [2013]	Evaluation of Web Performance of HTTP vs SPDY	HTTP, TCP, and IP Headers	20 web pages (over LTE and WiFi networks)	Performance Analysis	HTTP and SPDY perform similarly

this type of server-side logs is user-browsing modeling. One of the first steps to analyzing server-side logs is to reconstruct user sessions. A user session is defined here as the ordered set of HTTP requests and responses that correspond to a single user. Sessions are often limited in duration, so a user that accesses a web service on two different occasions is considered two different sessions. Reconstructing a web session using server-side logs is difficult because HTTP headers do not necessarily uniquely identify users nor provide a feature that informs a server when a user is done browsing. Spiliopoulou et al. [2003] discusses a number of methods and heuristics to overcome these issues using HTTP headers. One common method to identifying unique users is to use cookie field in HTTP headers.[16] The primary limitation of using the cookie field for identifying users is that all users may not enable cookies within their browser. Content providers can get around this issue altogether by requiring users to login before using their service. This login information, which is unique to a user, can be merged with HTTP headers to uniquely identify users. Login and cookie information are not enough to identify the beginning and end of sessions for a number of reasons including the fact that users may not be actively browsing a web page despite being logged in. Timeout heuristics, where a session ends after a user does not interact with a service after a predefined amount of time, are typically used to approximate the end of a session. Common timeout values used in this domain are 10 minutes, 15 minutes, and 30 minutes [Jansen and Spink, 2006, Spiliopoulou et al., 2003].

A number of studies have used server-side logs and the concept of user sessions to model user browsing behavior. Early studies of user web browsing behavior found that user navigation patterns can be approximated using a markov model [Chen et al., 2005, Cadez et al., 2003].[17] These results were not surprising given that web pages can be navigated using hyperlinks that are present on a page. Please recall that the hyperlink structure between web pages is the theoretical foundation for many web analysis methods including PageRank [Page and Brin, 1998]. Chen et al. [2005] and Cadez et al. [2003] showed that user browsing patterns that are approximated using markov models can be used to cluster users according to their browsing patterns. Content providers use these markov models to understand how users browse to improve user experience and the utility of their web sites [Borges and Levene, 2004]. A recent study by Chierichetti et al. [2012] has challenged the notion that Web users are markovian. Indeed, features like the “back” button and multi-window browsing, introduce “memory” into a web browsing session. Chierichetti et al. [2012] found

[16] IP addresses from other data sources, say TCP/IP headers, can be used in a similar way. Though, this approach has limitations because multiple users can share the same IP address.

[17] The next page visited by a user only depends on the current page the user is viewing and not on the entire browsing history of the user.

that user browsing behavior can depend on as many as the past 7 web pages visited by a user, which is much more than only considering the current page that is being viewed. These results show that user browsing behavior may change as web sites and/or client browsers evolve and provide different features and/or services to users.

Often times, content providers are interested in how users utilize their services so that they can improve them. For example, a search engine provider may look at search engine user logs to determine how a user searches for content. Jansen and Spink [2006] performed an analysis of search engine transaction logs from multiple search engine providers. The key observation of this study is that different search engines are used in a similar manner by users especially if they have similar features — similar observations were made in the social networking study by Benevenuto et al. [2009]. Another key observation is that users tend to select the first result of the search result page and will rather make a new search query than click on multiple pages. This observation is important because it shows that the first search result is critical for engaging a user and highlights the importance of ranking web search results. Liu et al. [2010] studied user behavior by investigating the amount of time users spent on a web page. Liu et al. [2010] found that the amount of time a user spends on a web page, referred to as dwell time, follows two extremes. Dwell times are either really short, under 10s, or really long, on the order of dozens of seconds. This observation is useful because it shows that users determine whether they will continue reading a web page fairly quickly. The cause of this result, whether it was that users left web pages prematurely because of performance issues (slow loading web page) or because the user found the web page boring, is unknown.

Prior work that used server-side logs to analyze web page traffic is relevant to this dissertation because the techniques used for user behavioral modeling are related to the problem of web page segmentation which this dissertation addresses. Some of the methods for modeling users that were developed for server-side logs, in particular user session timeouts, might be useful when using anonymized TCP/IP headers. These related techniques are considered and evaluated in Chapter 5.

2.3.1.3 TCP/IP-based studies

Most studies that use TCP/IP headers to study the Web are focused on web monitoring, modeling, and performance evaluation applications. We provide an overview of the most related prior work that primarily use TCP/IP data next.

TCP/IP for Monitoring and Characterizing Web Usage Crovella and Bestavros [1997] performed one of the earliest studies on the workload characteristics of web traffic using primarily TCP/IP headers. Crovella and Bestavros [1997] found that web workloads exhibit many of the properties of self-similar processes — similar observations were already made about Internet traffic as a whole [Willinger et al., 1997, Leland et al., 1994]. Self-similar processes follow heavy-tailed distributions and are autocorrelated. Crovella and Bestavros [1997] noted that the fact that users share the same links and access similar web pages are some of the primary factors that explain the self-similar characteristics that web traffic exhibits. Other factors include the underlying distributions of file sizes and the notion that users pause, or think, before requesting new web pages. Many of these factors still influence web traffic properties today [Ihm and Pai, 2011].

Smith et al. [2001] also conducted a measurement study on web traffic using TCP/IP headers. Smith et al. [2001] show that TCP/IP headers can be used to make inferences about how the Web is used. Some examples that were highlighted in their study include being able to comment on the increased adoption of banner ads by content providers and the increased adoption of Web-based email services using only TCP/IP headers. There are also studies that tracked the evolution of the Web using only TCP/IP headers [Hernández-Campos et al., 2003a, Newton et al., 2013]. The goals and key results of these studies are similar to the HTTP header based studies previously discussed [Ihm and Pai, 2011, Callahan et al., 2010]. Paxson [1999] proposed a real-time monitor that can be used for security and network performance analysis. The monitor proposed by [Paxson, 1999], called Bro, processes TCP/IP headers in real-time and generates logs that correspond to events that the network administrator may want to investigate further. Moore et al. [2001] and Roesch et al. [1999] proposed tools, called CoralReef and Snort respectively, that perform similar tasks. These tools can perform deep packet inspection to obtain more information about how the network is being used that can be used detect when a network has been compromised. Gu et al. [2008] also proposed a similar tool that analyzes TCP/IP headers to determine whether a network is being used in a malicious manner. Gu et al. [2008] used TCP/IP headers for this analysis because most applications on the Internet use TCP and the system is intended to be application layer protocol independent. Gu et al. [2008] showed that host communication patterns derived from TCP/IP headers could be used to detect malicious activity without application layer protocol headers. These studies show that application layer headers are not always needed to analyze web traffic.

One of the most popular applications of web traffic monitoring is web traffic modeling and traffic generation. Barford and Crovella [1998] proposed one of the first tools, called SURGE, that generates realistic web traffic. SURGE generates traffic using statistical models that approximate the distributions of network features that influence the network properties that are observed on a real link. Examples of network features that are modeled include file size and the duration of active/inactive periods in traffic. Barford and Crovella [1998] showed that theoretical distributions, such as weibull, pareto, and lognormal, can be used to model these network features in a manner that approximates real web traffic. However, theoretical distributions are not enough to model web traffic as it becomes more complex. Hernandez-Campos [2006] acknowledged this issue and proposed models for replaying web traffic at the level of HTTP requests and responses.[18] Weigle et al. [2006] used the ideas behind the models by Hernandez-Campos [2006] to develop a web traffic generator, called TMIX, that can replay web traffic measured "in the wild" in a manner that approximates the properties of real web traffic.

This past research, which is included in Table 2.3, is relevant to this dissertation because it shows that TCP/IP headers can be used to study the properties of web traffic. In this dissertation, we build upon this past research by expanding the scope in which TCP/IP headers can be used to study the modern Web to include web page traffic classification. We also use many of the analysis methods and tools, including TMIX and theoretical distributions, to study and model web page traffic.

Evaluating Web Performance using TCP/IP It is inevitable to use TCP/IP headers when evaluating the performance impact that protocol enhancements at the network and transport layers have on the Internet. Nielsen et al. [1997] designed experiments to determine whether the adoption of new technology will impact web performance. In particular, this study determined that HTTP/1.1 outperformed HTTP/1.0. This study also showed that the widespread use of CSS style sheets and the use of a more compact PNG image representation improved web performance. These results were obtained by analyzing only TCP/IP header data via a carefully designed experimental methodology. More or less, the authors ran multiple experiments, where one enabled an existing feature, say HTTP/1.0, and another enabled a different feature, say HTTP/1.1. Conclusions were drawn from such experiments by simply comparing the TCP/IP data generated using the experiments. This simple approach is heavily used for TCP/IP based analysis of web traffic [Le et al., 2007, Wang et al., 2014, Christiansen et al., 2000].

Le et al. [2007] and Christiansen et al. [2000] conducted studies similar to Nielsen et al. [1997] except their focus was to analyze the impact that different active queue management methods have on web

[18] This method is more generally referred to as replaying traffic at the *source* level

performance. Active queue management (AQM) refers to a class of approaches that attempt to improve the performance of the Web at the network layer by managing the amount of datagram queuing at routers. These methods typically achieve this goal by having routers either (i) drop datagrams[19] or (ii) sending explicit notifications to end-hosts to reduce their sending rates. Christiansen et al. [2000] found that actively managing router queues by simply dropping datagrams does not significantly improve web performance alone. Le et al. [2007] found that active queue management approaches that send explicit notifications to end hosts in addition to dropping datagrams can noticeably improve web performance in many scenarios. Le et al. [2007] noted that performance improvements were less significant for TCP connections that have a high variance in RTTs. Recent performance evaluations that use TCP/IP data investigate whether SPDY improves web performance over HTTP/1.1 [Wang et al., 2014, Erman et al., 2013]. We discussed these the prior section since they also leverage HTTP headers.

This body of work, which is included in Table 2.3, is related to this dissertation because it highlights that differences at other layers of the Internet Protocol Stack, say at the network and application layers, can be studied using TCP/IP headers. In this dissertation, we focus on developing techniques that can classify such differences using anonymized TCP/IP headers.

Comments on Anonymization of Web Traffic TCP/IP headers include IP address information which can be used to determine sensitive hostname information that can have implications on user privacy. The studies by Hernández-Campos et al. [2003a], Newton et al. [2013], and Smith et al. [2001] make extra efforts to preserve user privacy by anonymizing the IP addresses so hostnames cannot be easily obtained. However, the details of the anonymization procedures used in Web studies are rarely described [Sicker et al., 2007]. This lack of detail in the anonymization process is a serious issue because data that is weakly anonymized can still be a privacy concern. There have been a number of instances where a content provider releases “anonymized” user data only to have the data be analyzed by others to obtain the private information that the content providers were trying to protect via anonymization [He and Naughton, 2009]. Such cases have made the owners of web traffic data (e.g., researchers, Content providers, ISPs, etc) more reluctant to share/release it. Content providers are also increasingly encrypting network content to address privacy issues [White et al., 2013]. There have been a number of efforts to improve anonymization approaches [Fan

[19] TCP will reduce the sending rates of end-hosts when datagrams are dropped/lost. Dropping a datagram is an indirect and aggressive approach of reducing send rates.

et al., 2004b, Koukis et al., 2006, Le Blond et al., 2013, Schneier, 2013, Chen et al., 2013]. Despite these efforts to improve user privacy, it is still possible to deanonymize portions of anonymized web traffic and to infer the content of encrypted communications. We discuss these approaches when we discuss the related work on traffic classification.

2.3.1.4 DNS-based studies

Resolving domain names is the first step in requesting and rendering web pages. While DNS is essential to the functionality of the Web, DNS data is not typically used to study the Web because it does not provide much additional information for how the Web is used outside of an obvious impact on web performance. Jung et al. [2002] performed one of the first large scale studies of DNS functionality and performance. This study investigated different avenues for improving web performance by understanding the cause of DNS query failures and the impact of the TTL value on DNS cache performance. A recent study by Callahan et al. [2013] found that there is still room for web performance improvement by addressing client-side caching issues with DNS. Krishnan and Monrose [2011] showed that, at times, DNS-related web performance improvements (i.e., DNS prefetching) can increase privacy risks. This finding is important because DNS data is not encrypted. In fact, there are no proposed enhancement to DNS, even DNSSEC, that utilize encryption. Not encrypting DNS traffic provides an avenue for malicious users to eavesdrop on network communications [Paxson et al., 2013]. Paxson et al. [2013] found that malicious users leverage DNS to send covert messages over the Internet despite not being encrypted — DNS is used for this purpose because malicious users know that DNS has not been monitored and analyzed in the past [Paxson et al., 2013]. Most recent DNS-based studies focus on analyzing the malicious aspects of DNS use [Antonakakis et al., 2011]. These DNS studies, which are included in Table 2.4, show that (i) DNS performance can impact the temporal properties of HTTP headers, and (ii) adding features to web technologies (e.g., browser prefetching) can be reflected in DNS traffic. These observations are relevant to this dissertation because it shows that the behavior of multiple protocols and technologies are dependent on each other and that observations using one protocol can be reflected in others — we made similar observations previously when we discussed HTTP-based studies.

TABLE 2.3: Summary of Prior Work on Internet Measurement (TCP/IP-based studies)

Author(s)	Purpose of Measurement	Type of Data	Scale of study	Analysis Type	Key Insight/Notes
Leland et al. [1994]	Study of Statistical Properties of Internet Traffic	TCP and IP Headers	Over 1 billion packets measured (1-2 day span)	Exploratory Analysis	First study to show that Internet traffic is exhibits properties of a self-similar process
Willinger et al. [1997]	Study of Statistical Properties of Internet Traffic	TCP and IP Headers	11025 source-destination pairs (27 hour span)	Exploratory Analysis	Found that self-similarity in network traffic occurs on a flow-level
Crovella and Bestavros [1997]	Study of Statistical Properties of Web Traffic	HTTP, TCP, and IP Headers	575,775 URLs requested (6 week span)	Exploratory Analysis	First to find that Web traffic is self-similar and discussed possible causes for it
Smith et al. [2001]	Study of Characteristics of Web Traffic	TCP and IP Headers	$\approx$ 2.7 billion packets over numerous 1 hour periods (2 year span)	Exploratory Analysis	First to show that TCP/IP headers can be used to make inferences about web usage
Hernández-Campos et al. [2003a]	Longitudinal Study of Web Traffic	TCP and IP Headers	$\approx$ 200 million objects over numerous 1 hour periods (4 year span)	Exploratory Analysis	The size of web objects and the number of web objects has been steadily increasing
Newton et al. [2013]	Longitudinal Study of Web Traffic	TCP and IP Headers	$\approx$ 1.5 billion packets over numerous 1 hour periods (13 year span)	Exploratory Analysis	TCP Connection duration length and the number of servers contacted per web page download has increased
Barford and Crovella [1998]	Comparison of Web Workload Generators	TCP and IP Headers	Synthetically Generated Workloads	Performance Analysis	Discuss challenges associated with generating realistic web workloads
Hernandez-Campos [2006]	Comparison of Models for Generating Traffic Workloads	TCP and IP Headers	Numerous 1 hour capture periods [Smith et al., 2001]	Performance Analysis	Proposed sequential and concurrent a-b-t models for replaying web traffic and showed that it approximates the properties of real traffic
Weigle et al. [2006]	Proposed traffic generation tool for ns2	TCP and IP Headers	Collected over numerous 1 hour capture periods [Smith et al., 2001]	Performance Analysis	Proposed tool which uses a-b-t model generates traffic that approximates the characteristics of real traffic
Nielsen et al. [1997]	Evaluation of HTTP/1.1 with and without Pipelining	TCP and IP Headers	Single “Test” web site	Performance Analysis	HTTP/1.1 outperformed HTTP/1.0 for all tests done
Christiansen et al. [2000]	Evaluation of performance impact of RED on web traffic	TCP and IP Headers	Laboratory Experiment [Crovella and Bestavros, 1997]	Performance Analysis	RED provides little gain in web performance
Le et al. [2007]	Evaluation of performance impact of ECN on AQM	TCP and IP Headers	Laboratory Experiment [Smith et al., 2001]	Performance Analysis	ECN does not always improve web performance

TABLE 2.4: Summary of Prior Work on Internet Measurement (DNS-based studies)

Author(s)	Purpose of Measurement	Type of Data	Scale of study	Analysis Type	Key Insight/Notes
Jung et al. [2002]	Evaluation of Cache Performance of DNS	TCP, IP, and UDP Headers + DNS queries	≈ 14 million TCP connections and 10 million DNS lookups (2 day span)	Exploratory Analysis	There a significant number of inefficiencies in DNS
Callahan et al. [2013]	Study Characteristics of Modern DNS Traffic	TCP, IP, and UDP Headers + DNS queries	Residential Network of 85 clients — 200 million DNS queries (14 month span)	Exploratory Analysis	Observed DNS Prefetching and Caching Failures
Krishnan and Monrose [2011]	Evaluation of Effectiveness of DNS Prefetching	HTTP, TCP, IP, and UDP headers + DNS queries	Campus Network — ≈ 40-65 million queries per day (5-7 month span)	Exploratory Analysis	DNS prefetching leaks private information about web usage and provides minimal performance benefits
Paxson et al. [2013]	Study Information Content in DNS Payloads	DNS queries	Campus and Enterprise Networks — ≈ 230 billion queries	Exploratory Analysis	Developed approach to measure the amount of content that is transmitted in a covert manner over DNS
Antonakakis et al. [2011]	Evaluate Techniques for Detecting Malware with DNS	DNS queries	Domain Name Registers — 100000 domain names (8 month span)	Classification (Malicious or not)	Malicious domain names can be detected by analyzing query resolution patterns

2.3.1.5 Open Relevant Issues in Internet Measurement

The past work on Internet measurement that we discussed has given tremendous insight into the state-of-the-art of web traffic analysis and how to address different types of Web-related problems. However, there have been few studies that understand the Web in terms of *the web page*. While, there has been much interest in understanding web page traffic using an aggregate measurement methodology, most prior work either cannot reconstruct traffic into web pages or use error-prone heuristics for doing so. The work that is most related to understanding web page traffic is the work by Butkiewicz et al. [2011]. This work analyzes HTTP traffic that correspond to individual web page downloads using the Firefox browser. There are little issues in teasing out traffic into individual web page downloads since the traffic is generated and collected in a controlled environment. This study shows that there is tremendous diversity in web page traffic. However, the conclusions that can be drawn from it is limited due to the following:

- *Analysis of only HTTP headers:* The traffic analysis is restricted to HTTP headers. This leaves many research areas including the characterization of DNS and TCP/IP traffic for performance and privacy implications.

- *Impact of client platform:* The web page traffic was collected on a single platform using the Firefox browser. However, modern users have a choice of different client platforms including browser (e.g., Chrome, Internet Explorer, Opera), operating system (e.g., Windows, Linux, Mac OSX), and device (e.g., Smartphone, Laptop, and Tablet). It is possible that web page traffic will differ significantly across different client platforms. Client access point and cache settings may also impact web page traffic [Joumblatt et al., 2012, Callahan et al., 2013]. It is imperative to study the impact that different client platforms, and other factors, have on web page traffic so that it will be easier to understand the sources of variation in observed web page traffic (e.g., the web page itself or the client platform). HTML source data can also be used to help determine whether any differences in traffic is due to implementation differences on client platforms or due to client platforms processing different HTML source.

- *Limited diversity in web page type:* This study focuses entirely on characterizing landing pages. There are other types of web pages that users browse including news articles, video pages, and search engines that make up a substantial share of traffic [Bump, Gill et al., Zhou et al., 2010, Cheng et al., 2010]. Restricting a study to a particular type of web page may yield misleading conclusions on the diverse characteristics of web page traffic.

The above factors are important to address in order to adequately characterize and understand the diversity of web page traffic. *This understanding can then be used to develop techniques that can identify individual web pages from aggregate traffic traces (i.e., web page segmentation) and perform web page traffic inference (i.e., web page classification).*

2.3.2 Traffic Classification

The broad problem of attaching a label to traffic is called *traffic classification*. This dissertation research directly contributes to the field of traffic classification. Traffic classification problems vary according to the type of traffic analyzed (e.g., anonymized TCP/IP headers, HTTP headers, etc) and the type of labels used (e.g., application protocol – HTTP, P2P, etc). Background on the state-of-the-art methods in traffic classification that are related to this dissertation is presented in this section.

2.3.2.1 Classification via Commonly Used Port Numbers

The simplest approach to traffic classification is to label traffic according to commonly used port numbers. The Internet Assigned Numbers Authority (IANA) defines designated ports that should be used by certain applications [Touch et al., 2013]. For example, TCP segments with a port number of 80 correspond to applications that use HTTP, or web applications, while TCP segments with a port number of 22 correspond to SSH applications — please refer to the work by Touch et al. [2013] for a complete list of commonly used port numbers in the Internet. While these port numbers are designated for certain applications, there are no strict requirements for content providers to use them. In fact, many content providers, especially those hosting undesirable or illegal services (i.e., some P2P and file sharing applications), may intentionally avoid using IANA designated port numbers in order to avoid being detected by a firewall. To avoid being detected by a firewall, a content provider may "tunnel" their application through another application or simply use a different port number [Borges and Levene, 2004]. Deri et al. [2014] notes that many popular applications that used to use other designated ports, including video streaming, audio streaming, and email, are now using port 80 (i.e., the port designated for HTTP).[20] With HTTP rapidly becoming the application layer

[20] These applications may also use HTTPS (port 443).

protocol of choice to transmit a wide variety of services (e.g., audio streaming, social networking, etc), the effectiveness of port number-based classification decreases. Thus, much research has been done to develop different approaches to classify applications [Deri et al., 2014, Schatzmann et al., 2010, Kim et al., 2008]. The limitations of port number-based classification is one of the primary motivations of this dissertation research.

2.3.2.2 Classification via Deep Packet Inspection

Moore and Papagiannaki [2005] proposed a number of approaches and heuristics for classifying traffic without necessarily relying on port numbers. Some of these approaches and heuristics include (i) classifying simplex flows[21] as malicious and (ii) classifying flows by comparing the first few transport-layer segments or bytes of a flow, also known as signatures, with known flows from applications that have matching signatures. The process of classifying flows via matching the signatures of unknown flows in traffic with the flows of applications with known signatures is called signature detection. Moore and Papagiannaki [2005] also uses IP addresses and the traffic pattern history of hosts to help classify flows. Sen et al. [2004] used the signature-based detection approach that Moore and Papagiannaki [2005] used to classify traffic. Sen et al. [2004] focuses on identifying P2P applications (e.g., Kazaa, Bittorent, Gnutella, eDonkey) that were particularly undesirable to ISPs because they consumed a substantial amount of bandwidth and the content of the communication were widely viewed as being illegal.[22] The signature-based detection method used by Sen et al. [2004] is costly because signatures must be manually generated by analyzing the behavior of the traffic that is being classified. Haffner et al. [2005] showed that machine learning approaches can be used to automatically generate traffic signatures.

There has been some effort to make additional optimizations to deep packet inspection methods to improve the classification process by making them faster. Aceto et al. [2010] showed that deep packet inspection can be effective when only considering the first 32 bytes of application-layer message payloads. Alcock and Nelson [2012] expanded upon this idea and showed that the first 4 bytes of application-layer message payload could be used to classify traffic. Using less data not only increases the performance of the of deep packet inspection, but also makes the analysis less privacy sensitive.

Many of the ideas for classifying applications using deep packet inspection are used by commercial deep

[21] Flows that transmit traffic in only one direction.

[22] The P2P applications were used to transfer music, video, and other multimedia files.

packet inspection tools that are used today [Shen and Huang, 2012, Deri et al., 2014, Levandoski et al., 2008, Alcock and Nelson, 2012]. Carela-Español et al. [2014] conducted a study and showed the performance of these deep packet inspection tools can be quite different. In particular, Carela-Español et al. [2014] found that the commercial tool, PACE [Shen and Huang, 2012], consistently outperformed open-source tools such as nDPI[23] and Libprotoident [Deri et al., 2014, Alcock and Nelson, 2012]. These performance differences are likely due to extra features included in a commercial tool (PACE), such as an increased number of application signatures and the use of flow statistics to supplement performance when signature information is not enough for classification alone.

While deep packet inspection is widely used, it does not work on obfuscated traffic. In fact, a recent study shows that an increasing share of traffic is obfuscated which implies that deep packet inspection may not be a viable option for traffic classification in the future [White et al., 2013]. White et al. [2013] proposed using a hypothesis testing approach to filter traffic that cannot be analyzed by DPI tools. White et al. [2013] found that their filtering approach free up valuable computational resources[24] to allow a DPI tool to process more application-layer messages and detected more events in their data that otherwise would have not been detected. Uceda et al. [2015] proposed a method that uses regular expression matching to detect ASCII characters in application-layer message payloads. The assumption of this work is that non-ASCII data, likely encrypted or otherwise obfuscated, cannot be analyzed by DPI engines. Thus, the authors designed a tool that can be used to filter this undesirable traffic. While these approaches are used to filter obfuscated traffic, additional methods are still needed to classify it. One goal of this dissertation research is to develop techniques that can perform such classification.

2.3.2.3 Classification using TCP/IP Headers

Application Protocol Classification Since deep packet inspection approaches do not work on encrypted traffic and raise privacy concerns, approaches that use only TCP/IP headers need to be developed to classify traffic. Karagiannis et al. [2004] proposed a method for classifying traffic as being either a P2P application or not. Karagiannis et al. [2004] leveraged a common behavior exhibited by P2P applications where they use both UDP and TCP during the same time interval to distinguish it from other applications. Karagiannis et al. [2004] also used other heuristics such as segment size to help classify P2P applications. Karagiannis

[23] Formerly known as OpenDPI.

[24] Deep packet inspection is an CPU and Memory intensive process.

et al. [2005] extended the ideas of this work by using additional heuristics to identify other applications in addition to P2P including HTTP, chat, or even an attack. For example, Karagiannis et al. [2005] classified flows that do not send transport-layer segment payloads as attacks. Karagiannis et al. [2005] also proposed using segment transmission patterns to identify a collection of hosts that frequently communicate — this collection of hosts are referred to as a community. The authors used these communities to identify gaming applications. Dewes et al. [2003] conducted a study that was focused on identifying chat traffic. Dewes et al. [2003] found that segment interarrival time, TCP connection duration, and small segment sizes were particularly informative for distinguishing chat traffic from other types of traffic.

These prior work used a wide variety of features to classify applications. However, these methods rely on heuristics to perform the classification. There are a number of techniques with a strong theoretical foundation that can be used for traffic classification. Crotti et al. [2007] proposed classifying traffic by using probability distributions of applications and determine whether unknown traffic matches the stored traffic distribution within a certain predefined threshold. Similar to prior approaches, Crotti et al. [2007] classified traffic that corresponds to application layer protocols including POP3 and HTTP. Moore and Zuev [2005] used the Naive Bayes classifier, a probabilistic machine learning approach, for traffic classification. Roughan et al. [2004] also proposed using machine learning approaches, particularly Linear Discriminant Analysis and K-Nearest Neighbors, for traffic classification. The study by Roughan et al. [2004] is unique from most prior classification studies in that it was interested in classifying traffic for traffic engineering applications. Traffic engineering is a problem where ISPs strategically allocate network resources for traffic with different performance or Quality of Service (QoS) requirements. Thus, the categories of traffic that were targeted were not based entirely on application protocols. Some of the classes that Roughan et al. [2004] targeted were bulk (large file transfers), interactive (realtime applications that require user input such as a remote login), streaming (real-time applications such as video), and transactional (applications that transfer a small number of requests). Schatzmann et al. [2010] used the Support Vector Machines (SVM) machine learning approach to classify traffic as either being mail or non-mail. Schatzmann et al. [2010] found that some temporal features (e.g., duration of a TCP connection and TCP connection interarrival times) can be reliably used to differentiate mail from non-mail traffic.

Kim et al. [2008] conducted a study to compare the performance of the many different methods used in the traffic classification literature. Kim et al. [2008] found that the machine learning approaches outperforms heuristic methods [Karagiannis et al., 2005] and port-based approaches [Moore et al., 2001, Touch et al., 2013]. Kim et al. [2008] also highlighted that machine learning methods require a large amount of training data in order to achieve high performance. Lim et al. [2010] conducted a study that investigated which machine learning approaches performed the best. Lim et al. [2010] concluded that machine learning methods can achieve similar performance if appropriate, and sometimes essential, preprocessing and data transformation methods are used on the traffic features. Though, Lim et al. [2010] also found that some techniques such as classification trees and K-Nearest neighbors performed well without any additional preprocessing. We note that the authors of this study did not tune each machine learning method during their evaluation. Thus, it is unclear whether other methods, such as Support Vector Machines or Naive Bayes, can perform well without preprocessing as well.

The machine learning methods that have been discussed are supervised machine learning techniques. Recall, supervised machine learning techniques require labeled training data to work. There have also been traffic classification methods that use unsupervised machine learning techniques which do not require labeled data. Unsupervised machine learning techniques are commonly referred to as clustering algorithms. While clustering algorithms do not require training data, additional effort, and most likely the assistance from a domain expert, is needed to label and interpret the clusters that are output from them. McGregor et al. [2004] investigated the feasibility of using the expectation maximization (EM) algorithm (a clustering method) for traffic classification. McGregor et al. [2004] found that the resulting clusters included many different types of applications. Ideally, different applications should appear in different clusters. Erman et al. [2006] also investigated whether clustering algorithms can be used for traffic classification. Erman et al. [2006] evaluated the DBSCAN, K-means, and Autoclass clustering approaches and found that approximately 150 clusters were needed for the majority of each cluster to include a single type of traffic. 150 clusters is a large number considering that Erman et al. [2006] were trying to distinguish less than 10 classes of traffic. This result implies that traffic behavior is diverse and a single cluster cannot be used to classify an application. Hernández-Campos et al. [2003b] used a different class of clustering approach, hierarchical clustering, for distinguishing traffic. Hierarchical clustering groups instances in a hierarchy and is more robust than traditional clustering methods at clustering diverse datasets. Hernández-Campos et al. [2003b] found that the results from hierarchical clustering were able to distinguish applications such as P2P and Web. This body of work, which is included in Table 2.5, is related to this dissertation because it consists of the state-of-the-art traffic classification methods that use TCP/IP headers.[25] However, the related work that we discussed focuses primarily on the problem of application classification. The work presented in Chapter 4 uses similar techniques, particularly learning-based classification methods, to advance the state-of-the-art in traffic classification by focusing on the problem of web page classification.

Anomaly detection is an area of traffic analysis that is related to traffic classification. The goal of anomaly detection is to identify instances in network traffic that deviate from "normal" behavior. Thus, anomaly detection problems at their simplest level consider two classes of traffic, normal and abnormal/anomalous. Most anomaly detection methods in the literature are not focused on understanding and classifying web applications nor other application layer protocols. These methods are instead more focused on the network layer information and identifying security or network management-related issues (e.g., port scanning, flash crowds, attacks, routing errors, etc) [Barford et al., 2002, Lakhina et al., 2005, Soule et al., 2005, Brauckhoff et al., 2006, Ringberg et al., 2007, John and Tafvelin, 2007, Nychis et al., 2008, Milling et al., 2012, Silveira et al., 2010, Yan et al., 2012].

2.3.2.4 Web Page Identification

The security and privacy research community have focused on a problem similar to traffic classification that is best referred to here as *web page identification*. Web page identification is the problem of *identifying the exact web page* given an encrypted traffic trace. In other words, successfully solving the web page identification problem means that fine-grained information such as the web page visited by a user can still be inferred from traffic despite being encrypted. The presence of encrypted traffic means that only TCP/IP headers are available for web page identification.

Sun et al. [2002] showed that exact web pages can be identified using information derived from network traffic. One primary metric used for this problem is the size of a web object. Here, the size of a web object, or any type of web traffic (say transport-layer segments), is the number of bytes it contains. Sun et al. [2002] first derived web object sizes using TCP/IP headers, and then used the jaccard similarity classifier to compare traffic signatures that are known with those present in the traffic. The primary assumption of this approach is that the sizes of web objects that are referenced by different web pages are different. Sun et al. [2002] went on to propose mechanisms to help make their method less effective — hence, improving network security. These mechanisms include padding transport-layer segments (increasing size of web objects to be larger

[25] Erman et al. [2006] also published another similar study that only considered the K-means method Erman et al. [2007b].

TABLE 2.5: Summary of Prior Work on Traffic Classification

Author(s)	Classification Labels	Features Used	Methods
Moore and Papagiannaki [2005]	Application Type (Mail, Web, P2P, Games, Multimedia, Services, Interactive, Bulk, Database, Malicious)	Packet payloads + Coarse flow-level features	Signature Detection + Heuristics
Sen et al. [2004]	Application Type (P2P vs Non-P2P)	TCP Payloads (string matching on payloads)	Signature Detection
Aceto et al. [2010]	Application Type (P2P, Web,Unknown, Services, Encryption, Network management, Mail, Multimedia, Tunneling, Filesystem, Bulk, Games, Interactive)	Packet Payloads (restricted to first 32 bytes of payload) + TCP Protocol Type	Signature Detection
Alcock and Nelson [2012]	Application Type (ESP over UDP, Web, Bittorent, Razor, Garena, Skype, RTMP, Xbox Live)	Packet Payload (First 4 bytes) + IP Address + Port number	Signature Detection
Karagiannis et al. [2004]	Application Type (P2P vs Non-P2P)	TCP, UDP, and IP Headers (non-temporal features such as Packet Size and IP address)	Heuristics
Karagiannis et al. [2005]	Application Type (P2P, Web, Mail, Chat, FTP, Network management, Games)	TCP, UDP, and IP Headers (non-temporal features such as Packet Size and IP address)	Heuristics
Dewes et al. [2003]	Application Type (Chat vs Non-Chat)	Packet Payloads + TCP, IP, UDP, and HTTP Headers (specifically, temporal features)	Heuristics
Crotti et al. [2007]	Application Type (Web, SMTP, POP3, Other)	TCP and IP Headers (e.g., port numbers, temporal features, and segment size)	Heuristics
Moore and Zuev [2005]	Application Type (Bulk, Database, Interactive, Mail, Services, Web, P2P, Attack, Games, Multimedia)	TCP/IP headers (e.g., port numbers, temporal features, packet size, and TCP flags)	Supervised Machine Learning (NB)
Roughan et al. [2004]	Application Type (Domain, FTP, HTTP/Web, P2P, Telnet, HTTPS)	TCP and IP Headers (e.g., port numbers, temporal features, packet size, and TCP flags)	Supervised Machine Learning (KNN and LDA)
Schatzmann et al. [2010]	Application Type (Webmail vs Non-webmail)	Coarse flow-level features (specifically, temporal features)	Supervised Machine Learning (SVM)
Kim et al. [2008]	Application Type (Web, DNS, Mail, Chat, FTP, P2P, Streaming, Game)	TCP, UDP, and IP Headers (e.g., port numbers, packet size, and TCP flags)	Supervised Machine Learning (NB, CT, KNN, SVM)
Lim et al. [2010]	Application Type (Web, P2P, Attack, FTP, DNS, Mail, Streaming, Network Operation, Games, Encryption, Chat, Unknown)	TCP, UDP, and IP Headers (non-temporal features such as port numbers, packet size, and TCP flags)	Supervised Machine Learning (NB, LDA, CT, KNN)
McGregor et al. [2004]	Application Type (Web, ICMP, SMTP, IMAP, NTP, FTP)	TCP, UDP, and IP Headers (e.g., port numbers, packet size, temporal features, and TCP flags)	Clustering (Expectation Maximization)
Erman et al. [2006]	Application Type (Web, POP3, Database, P2P, Other, FTP, limewire)	TCP, UDP, and IP Headers (e.g., port numbers, packet size, and TCP flags)	Clustering (K-means, AutoClass, DBScan)
Hernández-Campos et al. [2003b]	Application Type (Web, HTTPS, POP3, Gnutella, Telnet, POP, FTP,SMTP, NNTP, Database)	TCP, UDP, and IP Headers (e.g., port numbers, packet size, and TCP flags)	Clustering (Hierarchical Clustering)

than expected) and adding extra background traffic (increasing the number of web objects referenced by a web page to be higher than expected).

Liberatore and Levine [2006] performed a similar study except they focused on using transport-layer segment size as the primary feature instead of web object sizes — the direction of the communication was also used as a feature for classification. Liberatore and Levine [2006] used segment sizes because some encryption technologies hide the notion of a web object in a manner that cannot be recovered using TCP/IP headers alone. Liberatore and Levine [2006] showed that web page identification is still possible using segment size as the primary feature. Liberatore and Levine [2006] also showed that the Naive Bayes classifier performed better than the Jaccard Similarity classifier primarily because the Jaccard similarity does not account for some traffic features such as the number of segments transferred. Herrmann et al. [2009] and Panchenko et al. [2011] conducted similar studies except Herrmann et al. [2009] used a Multinomial Naive Bayes Model while Panchenko et al. [2011] used an SVM model for web page identification. The method proposed by Panchenko et al. [2011] is different from most prior methods in web page identification in that it uses coarse traffic features (i.e., features derived from multiple segments), such as TCP connection duration and bandwidth, in addition to fine-grained features (i.e., features that use each segment) such as the distribution of segment sizes observed in traffic.

Dyer et al. [2012] compared many of these web page identification approaches to determine which one works best. Dyer et al. [2012] found that the classification method used did not matter as much as the types of features used for the classification. In particular, Dyer et al. [2012] highlighted that the coarse features used by Panchenko et al. [2011] were particularly informative for web page identification. Coarse features are especially useful for identifying web pages when different countermeasures are used because they are more robust to the noise that these countermeasures add to the traffic. Cai et al. [2012] conducted a similar study that showed that web page identification can still work without the use of the segment size feature and despite the use of countermeasures. Yen et al. [2009] and Coull et al. [2007] also use coarse features for web page identification. Though, these studies only consider coarse features while the study by Panchenko et al. [2011] considers both coarse and fine-grained features. Both studies show that coarse features can be used to effectively identify web pages without fine-grained features. A key result from the study by Yen et al. [2009] is that it shows that browsers have an impact on web page identification and being able to detect browsers can help in building specialized classifiers for each browser that work better than a single classifier.

Miller et al. [2014] conducted a study that investigated whether browser-specific features such as browser caching can impact web page identification performance. The results of this study show that the browser cache has an impact on web page traffic and can consequently impacts web page identification performance. Miller et al. [2014] also found that a hidden Markov model can be used to provide navigation information to help identify the web page that a user visited. The assumption of this approach is that the navigation patterns of a user can be used to predict the domain of the web page (i.e., web site) that was visited — thus reducing the set of possible web pages that a user visited. Cai et al. [2012] and Danezis also showed that the navigation patterns of users can be used to supplement web page identification techniques.

The majority of the work in web page identification assume a closed world model [Cai et al., 2012, Yen et al., 2009, Dyer et al., 2012, Liberatore and Levine, 2006, Herrmann et al., 2009, Miller et al., 2014]. The closed world model assumes that the web page traffic trace includes only the traffic resulting from downloading a finite set of *k* known web pages. Indeed, in a real-world scenario one will observe web page traffic that includes download traffic from unknown web pages. The open world model considers such a scenario. Panchenko et al. [2011], Coull et al. [2007], and Sun et al. [2002] conducted some of the few studies that consider the open world model.

A summary of the prior work described in this section is provided in Table 2.6. The body of work presented in this section is related to this dissertation because (i) it is related to the general problem of traffic classification or (ii) it involves classifying web page traffic using TCP/IP headers.

2.3.2.5 Open Relevant Issues in Traffic Classification

More and more applications including video streaming, social media, mail, and file sharing are using web technologies. Simply classifying network traffic as Web is no longer enough information to provide useful insight about the content of traffic because the vast majority of traffic on the Internet is Web. Thus, a method to further classify web traffic into meaningful categories is desirable. One well-researched problem for further classifying web traffic is web page identification. However, web page identification is a problem that does not scale in the wild because there are too many web pages to measure, fingerprint, store, and reliably identify. Web page classification, the process of classifying *web page traffic* into a web page category, is an alternative to web page identification. Web page classification is different from web page identification because multiple web pages may have the same classification label — thus, web page classification will likely scale better than web page identification because there will be a substantially smaller number of classes of web pages to characterize and subsequently label. Web page classification is thoroughly investi-

TABLE 2.6: Summary of Prior Work on Web Page Identification

Author(s)	Fundamental Data Type	Features Used	Methods	Open/Closed World
Sun et al. [2002]	Web Object	Web object size	Similarity-based (Jaccard Similarity)	Open
Liberatore and Levine [2006]	Single Flow (SSH Tunnel)	TCP, UDP, and IP Headers — TCP segment size and direction	Supervised machine learning (Naive Bayes with Gaussian kernel). Also, compared performance with Jaccard similarity.	Closed
Herrmann et al. [2009]	Single Flow (SSH Tunnel)	TCP, UDP, and IP Headers — TCP segment size and direction	Supervised machine learning (Naive Bayes with Multinomial kernel). Also, compared performance with Jaccard similarity and Naive Bayes with Gaussian kernel.	Closed
Panchenko et al. [2011]	Single Flow (Tor browser)	TCP, UDP, and IP Headers — TCP segment size and direction + Coarse Flow-level features	Supervised machine learning (Support Vector Machines)	Considered both Open and Closed
Dyer et al. [2012]	Single Flow (SSH Tunnel)	TCP, UDP, and IP Headers — TCP segment size and direction + Coarse Flow-level features	Supervised machine learning (Naive Bayes)	Closed
Cai et al. [2012]	Single Flow (Tor browser)	TCP, UDP, and IP Headers — TCP segment size and direction + Coarse Flow-level features + User navigation patterns	Supervised machine learning (Support Vector Machines with edit distance-based kernel)	Closed
Yen et al. [2009]	Multiple Flows	Coarse Flow-level features	Supervised machine learning (Support Vector Machines)	Closed
Coull et al. [2007]	Multiple Flows	Coarse Flow-level features	Kernel Density Estimation and Bayes Belief Networks	Open
Miller et al. [2014]	Multiple Flows	TCP, UDP, and IP Headers — TCP segment size and flows + User Navigation patterns	Supervised Machine Learning (Logistic Regression)	Closed
Danezis	Multiple Flows	TCP, UDP, and IP Headers— TCP segment size and flows + User Navigation patterns	Supervised Machine Learning (Hidden Markov Model)	Closed

gated in this dissertation.

2.3.3 Web Page Segmentation

Web page segmentation approaches are used to identify when an individual web page download begins. Web page segmentation is a problem that is directly addressed by this dissertation research. Much of the past work in Internet measurement and traffic classification described in this Chapter make use or assume the availability of a web page segmentation approach as a critical step in their analysis [Yen et al., 2009, Maciá-Fernández et al., 2010, Choi and Limb, 1999, Hernández-Campos et al., 2003a]. Web page segmentation approaches fall into two categories: 1) Content-based approaches and 2) Idle time-based approaches. These prior approaches are described in this section — please refer to Table 2.7 for a summary of these approaches.

2.3.3.1 Content-based Approaches

Content-based approaches leverage HTTP headers to segment web pages from web page traffic. Choi and Limb [1999] proposed a method that assumes that each HTML object (a web object with MIME type text/html) corresponds to a new web page while non-HTML objects do not. Indeed, modern web pages may be reference of multiple HTML objects, making this method outdated. Ihm and Pai [2011] proposed a method, called StreamStructure, that leverages the fact that some web pages reference easily identifiable web objects to web analytics companies once they finish loading (e.g., Google Analytics). StreamStructure uses these web objects, also referred to as analytics beacons, to identify when the traffic for a particular web page download ends. StreamStructure also uses a number of heuristics used in prior methods to determine the beginning of a web page. These include using HTML objects to identify the "start" of a page and using a timeout heuristic that treats multiple HTML objects that occur in quick succession as a single web page.

Ihm and Pai [2011] presented results that showed that only a relatively small fraction of web pages, approximately 24%, reference analytics beacons. Thus, StreamStructure is not an approach that can be generally applied "in the wild". Xie et al. [2013] proposed a method, called Resurf, that leverages the referer field in HTTP headers to segment web pages. Similar to StreamStructure, Resurf also uses timing heuristics to assist in the web page segmentation process. Xie et al. [2013] showed that Resurf can perform web page segmentation more effectively than StreamStructure. Neasbitt et al. [2014] proposed using a method, called Clickminer, that uses the assistance of a browser to help segment web pages. At a high-level, Clickminer uses HTTP headers as input to a browser emulator to approximate and model the possible user click behavior that resulted in the observed network traffic [Neasbitt et al., 2014]. Neasbitt et al. [2015] also developed a similar tool, Webcapsule, which is designed to run within a browser, in real-time. One advantage that Webcapsule has over Clickminer is that it is able to use additional information that is collected by the browser to help with web page segmentation — this additional information consists of details related to user-browser interactions (e.g., scrolling on page, clicking a link, etc), messages related to web-rendering API events, and more [Neasbitt et al., 2015].

Nelms et al. [2015] developed a web security tool, called Webwitness, which classifies web malware download paths (or a sequence of web page downloads) such that administrators can more effectively determine *how* malware is downloaded on a host. Some categories of the paths considered in this work include social engineering and drive-by downloads. However, in order to classify web page download paths Webwitness first determine the web pages that should be included in each web page download path. This is done by first identifying different types of heuristics which likely associate two web page downloads. Some of these include the determining whether the referer field in HTTP header match other web page downloads, determining whether domain names between downloads match, and determining whether different downloads include similar content types (e.g., .swf, .pdf, and .jar). These are used to build a web page download graph which models the likelihood that web page downloads follow the same path — additional heuristics are applied to this graph to determine which web page downloads belong to the same path [Nelms et al., 2015].

There are also related approaches that attempt to identify web page sessions using server logs — many of these approaches were previous mentioned when we discussed the study by Spiliopoulou et al. [2003]. However, identifying web page sessions and segmenting web pages are different problems. Though, some of the ideas used for these different problems, such as using idle time-based heuristics, are applicable in both problem domains [Ihm and Pai, 2011, Spiliopoulou et al., 2003]. Specific idle time-based heuristics that have been used for web page segmentation are described next.

2.3.3.2 Idle time-based Approaches

While there are many content-based web page segmentation methods, they require the use of HTTP headers and are ineffective against encrypted traffic. Idle time-based approaches are used when HTTP headers are not available due to privacy concerns or encryption. Idle time-based approaches assume that new web page downloads begin when new web traffic is observed after a predefined idle-time period [Barford and

Crovella, 1998, Hernández-Campos et al., 2003a] — in this case, new web page traffic corresponds to any additional segments that are observed. This idea was originally proposed by Barford and Crovella [1998] and was later identified to be effective by others [Smith et al., 2001]. This assumption was reasonable when web pages were static and referenced few web objects. However, modern web pages, which uses dynamic technology, such as Ajax, and references many web objects, will generate web traffic that will reduce the effectiveness of idle time-based approached. Newton et al. [2013] identified this issue and proposed using an additional predefined threshold for the amount of web page traffic features (in particular the number of segments with the SYN flag set) for detecting new web page downloads. The addition of a threshold for web page traffic is intended to reduce the chances of falsely identifying automatically generated traffic as a new web page download. However, this method was not extensively evaluated so it is unclear whether it works for modern web page traffic. This dissertation research performs such an evaluation.

2.3.3.3 Open Relevant Issues in Web Page Segmentation

Content-based approaches (i.e., methods that use HTTP headers) are ineffective against encrypted traffic and raise privacy concerns. Thus, idle time-based approaches are the only existing methods that are applicable to anonymized TCP/IP header traces. However, idle time-based approaches have not been extensively tested on modern web page traffic. An extensive evaluation of idle time-based approaches, and other appropriate approaches, on modern web page traffic is conducted in this dissertation to determine whether modern web pages can be segmented using only anonymized TCP/IP headers.

TABLE 2.7: Summary of Prior Work on Web Page Segmentation

Author(s)	Type	Content/Idle time-based	Features Used	Scope of Evaluation	Key Insight/Notes
Barford and Crovella [1998]	Web page	Idle time-based	TCP/IP headers	Did not explicitly evaluate segmentation method	Idle time in traffic is correlated to new web page downloads — used 1s as threshold for idle time — please note that Hernández-Campos et al. [2003a], Smith et al. [2001], and Maciá-Fernández et al. [2010] use the same approach as Barford and Crovella [1998].
Newton et al. [2013]	Web page	Idle time-based + Threshold	TCP/IP headers	Did not thoroughly evaluate method	Proposed using an additional threshold for network activity to reduce false web page detection rate. Used SYN+ACKs over bytes as a feature. The parameterization that Newton et al. [2013] found performed best was a threshold of network activity of 2 SYN+ACKS and an idle time of 2.5s.
Choi and Limb [1999]	Web page	Content-based	HTTP headers	Did not explicitly evaluate the proposed segmentation method.	MIME-type field in HTTP response headers are key for determining boundaries of web pages. Specifically, response headers with MIME-type ".html", ".cgi", or ".asp".
Ihm and Pai [2011]	Web page	Content-based	HTTP headers	The evaluation included comparison with methods by Barford and Crovella [1998] and Choi and Limb [1999]	The referer field of HTTP headers are useful for segmentation. Requests from known analytics beacons are useful as well.
Xie et al. [2013]	Web page	Content-based	HTTP headers	Extensive evaluation against the method by Ihm and Pai [2011].	The referer field in HTTP headers and time-related information can be used to reconstruct web pages.
Nelms et al. [2015]	Web page download paths	Content-based	HTTP headers	Evaluation did not include comparison with other methods.	Used the location, hostname, and referer fields in HTTP headers as features.[26]
Neasbitt et al. [2014]	Web page	Content-based + Browser-assisted	Browser and HTTP headers	Adequate evaluation with the work by Xie et al. [2013].	The browser can be used to replay captured traffic to determine whether certain actions (web-interactions) caused the traffic observed
Neasbitt et al. [2015]	Web page	Content-based + Browser-assisted	Browser and HTTP headers	Evaluation did not include comparison with other methods.	Information obtained from browser is useful for segmentation (e.g., rendering API events)
Spiliopoulou et al. [2003]	Web sessions	Content-based + Idle time-based	Server Logs, HTTP headers, and TCP/IP headers	Extensive evaluation which compared many techniques was performed	Web browsing cookies, login information, and idle time features are useful for reconstructing web sessions

CHAPTER 3: CHARACTERIZATION OF MODERN WEB PAGE TRAFFIC[1]

Any sound web measurement study must appropriately sample traffic such that the diversity in web traffic is considered. The results of web measurement studies that ignore this are at an increased risk of being biased because the dataset analyzed only represents a targeted subset of traffic. Such biased results are undesirable because they only apply to the subset of traffic considered in the study and may not be generally applicable "in the wild". Many web measurement-related studies — including traffic characterization, traffic classification, and privacy analysis — attempt to avoid this bias by ensuring that diversity in web page traffic is included in their traffic sample by studying a large volume of different *web pages* — this is typically done by considering either (i) a large volume of web traffic or (ii) a set of popular web pages which account for a large fraction of traffic [Dyer et al., 2012, Herrmann et al., 2009, Butkiewicz et al., 2011, Xu et al., 2011, Neasbitt et al., 2014, Rao et al., 2011].

However, the diversity in web page traffic may also be influenced by the different client platforms that are available to users when downloading web pages, including browser types, operating systems, the type of device (e.g., laptops, smartphones, and tablets), and vantage points. Buckler acknowledges this diversity by reporting usage statistics for a plethora of client platforms including 5 desktop browsers, 6 mobile browsers, and 9 operating systems (including mobile, desktop, and video game consoles). This diversity in client platform usage, and its impact on traffic, is only expected to grow as the types of devices that are web-enabled (e.g., smartphones, video game consoles, smartwatches, etc) and their adoption by users increases; in fact, the average number of devices per user is estimated to grow from 3 in 2014 to 5 by 2017 [Bort]. While there is a large number of client platforms available to users, it is unclear whether these impact web traffic measurements differently; in fact, there is some disagreement on this issue in the literature [Yen et al., 2009, Butkiewicz et al., 2011]. Despite this possible issue, most prior web measurement-related studies do not explicitly consider the potential impact that client platforms have on their study, nor on related applications [Dyer et al., 2012, Herrmann et al., 2009, Butkiewicz et al., 2011, Xu et al., 2011, Neasbitt et al., 2014]. The results and implications of such studies are at risk of being biased towards a specific client

[1] The contents of this chapter includes material from articles published in two conferences [Sanders and Kaur, 2014b, 2015a]

platform and hence not generally applicable.

To address this problem, we perform the first comprehensive measurement study that investigates the question — Do different client platforms generate different web page traffic despite processing the same source HTML? This study can be used to assist in the development of traffic analysis tools and techniques that are robust to client platform-specific traffic differences. More specific to this dissertation, we use the results of this study to help design the web page classification and segmentation techniques considered in Chapters 4 and 5, respectively. We provide details of the methodology, results, and implications of this measurement study in the rest of this chapter.

3.1 Data Collection Methodology

To answer the question considered in this study, it is critical to collect a dataset where the web page traffic has been (i) filtered on a per web page basis, (ii) labeled according to the client platform that was used to load the page, and (iii) labeled according to the web page that was downloaded. These properties will allow us to definitively state what type of traffic is generated by an individual web page and determine whether it is influenced by client platforms. There are two data collection methodologies that are available to us for web page traffic measurement. These are the widely used aggregate measurement methodology and the less common client-side measurement methodology— details on these measurement methodologies are provided in Chapter 2. While the aggregate measurement methodology is more popular and has the advantage of easily measuring more web traffic than the client-side measurement methodology [Smith et al., 2001, Hernández-Campos et al., 2003a, Schneider et al., 2008], little is known about the web pages and the client platforms that generated the actual traffic that is measured using this methodology. This is due to the following:

1. *Difficulty in identifying web page download events:* Modern web pages consist of many individual objects that are distributed across multiple sources. Given a traffic trace that is collected on an aggregate link, it is non-trivial to group these individual objects into individual web page units—this process is important because information about the properties of individual web pages is difficult to understand when the individual web objects that make up the individual page are not clearly identifiable. Although several methods have been recently proposed to solve this important problem, such methods are error-prone [Xie et al., 2013, Neasbitt et al., 2014, Ihm and Pai, 2011].

2. *Difficulty in identifying client platform:* Privacy concerns, recent legislation, and the increase in encrypted and compressed traffic dictate that only relatively coarse data such as anonymized TCP/IP headers and NetFlow logs are available for analysis. Information about browsers, operating systems, device type, and client location is not obtainable from this data.

Due to these issues, we instead use the less common client-side methodology where the web page requested and the client platform is known and can be controlled for. Using a client-side measurement methodology means that we must be sure to explicitly study a large and diverse sample of web pages and client platforms in order for our results to be generalizable to traffic observed "in the wild". To cater to different types of client platforms available today, our measurement methodology is split into: (i) desktop-based measurement, (ii) mobile device-based measurement, and (iii) planetLab-based measurement. We describe the details of our client-side measurement methodology next.

3.1.1 Desktop-based Measurement Methodology

Our desktop-based measurement methodology addresses several issues, including diversity in web pages studied, using multiple operating systems and browsers, and repeated measurements. These are discussed below:

- *Diversity in web pages studied:* The Web consists of over 600 million web sites [Internetlivestats.com, 2015]. Since downloading web pages from each web site is infeasible, we focus primarily on the most popular web sites. In particular, we study web pages from the top 250 web sites in the world [Inc.] — recent studies show that the majority of web traffic (over 90% of bytes) originate from these 250 web sites [Callahan et al., 2013, Xu et al., 2011, Bump]. A list of the URLs for the landing page of these web sites is provided in Appendix 1. However, landing pages alone do not account for the diversity of web applications that is present on the modern Web. Such web applications include search engines, video streaming, and audio streaming. To account for this diversity in our study, we manually browse each of these 250 web sites to collect a list of URLs that correspond to search result pages and other non-landing pages (e.g., Facebook user profiles, news articles on CNN.com, videos on Youtube.com, etc). This list was collected by (i) clicking on 4 random links on each landing page and (ii) using the search box to make 4 different search queries that were arbitrarily chosen.[2] We repeat this process for

[2] The 4 search terms that we used were "golden", "ray lewis", "stone mountain", and "mcdonalds".

TABLE 3.1: Desktop-based Clients (8 clients - 2 devices)

Operating System	Chrome	Opera	IE	Firefox	Safari
Mac OSX 10.9.4	v38.0	v25.0	N/A	v33.0	v537.77.4
Windows 7	v38.0	v25.0	v11	v33.0	N/A

web sites that have mobile versions. Thus, for each web site, we download 18 web pages if we find a mobile optimized web page and 9 web pages if we do not — we download fewer web pages if the web site does not include a search service. In total, we analyze 3614 diverse web pages from these 250 web sites. A list of the URLs for these web pages is provided in Appendix 2.

- *Operating Systems and Browsers Studied:* A goal of this study is to determine whether client platforms impact web page download traffic. To achieve this goal, we must download these web pages using multiple browsers and operating systems. The 5 browsers that we use are Google Chrome, Mozilla Firefox, Safari, Internet Explorer, and Opera. We focus on these browsers because they are the most popular and generate over 98% of the traffic share in the world [Sta, b]. The operating systems that we use are Windows 7 and Mac 10.9.4. These operating systems were selected because they are currently supported by Mozilla Firefox, Opera, and Google Chrome — Safari and Internet Explorer, however, are not supported on Windows and Mac OS X platforms respectively. We only study operating systems that support the *same* browser type and version to determine whether operating systems influence web page traffic when the same browser is used — differences in browser type and version across operating systems may introduce noise in our analysis which may make our conclusions unreliable. Please note that Linux-based operating systems are not considered because (i) the most popular browsers are out-of-date as compared to their Mac OS X and Windows 7 counterparts and (ii) the other browsers that are supported by Linux-based operating systems, such as Konquerer, Chromium, and Web, are not supported by Mac OS X and Windows 7 [Sta, a]. Table 3.1 provides details on the versions of the different browsers and operating systems used. These clients are located on a University campus that access the Internet via 1Gbps Ethernet NICs.

We use tcpdump/windump to capture all web page traffic that is generated [Jacobsen et al., 2005]. The data collection procedure is *automated* as:

1. Start packet capture tool

TABLE 3.2: Mobile Device-based Clients (6 clients- 5 devices).

Device Type and Operating System	Browser
Mac OSX 10.9.4 (Laptop)	Chrome v38.0
Mac OSX 10.9.4 (Laptop)	Safari v537.77.4
Android 4.4.2 Samsung Galaxy (Tablet)	Chrome v35.0
iOS7 iPad 3 (Tablet)	Safari v9537.53 Mobile/11A501
iOS8 iPod Touch (Smartphone)	Safari v600.1.4 Mobile/12A405
Android 4.4.4 Motorola X (Smartphone)	Chrome v37.0

2. Start a browser with a web page URL as an argument

3. Close the browser after 60 seconds

4. Start a browser with a web page URL as an argument to capture traffic with a full cache

5. Clear the local DNS resolver cache

6. Clear the browser cache

7. Go to Step 1 using a new URL

Using these steps, we perform multiple web page measurements to minimize the effects that outliers may have on our analysis. Our data is collected over a period of 16 weeks between December 12, 2014 and March 20, 2015. It takes approximately 4 weeks to download 3614 web pages for 4 browsers for two cache states (empty cache and full cache). We also stagger the time of day at which each iteration begins by 6 hours each, as a modest attempt to account for the impact of peak traffic hours. Overall, our dataset includes 4 repeated measurements for each browser, operating system, and cache state.

3.1.2 Mobile device-based Measurement Methodology

The key differences between the above measurement methodology for mobile devices as opposed to desktops are:

- *Devices and Browsers Studied:* We take measurements using 5 devices — these include 2 tablets, 2 smartphones, and a laptop. Table 3.2 shows details on the versions of the operating systems and browsers used for each device. We only study mobile devices (i.e., tablets and smartphones) that use either the Android and iOS operating systems since they are the most popular — these two operating systems account for over 95% of the mobile operating system market [IDC]. We study Google Chrome on Android platforms and Safari on iOS platforms because they are the most popular and are exclusive to these operating systems. The laptop is included in this study as a baseline for comparison against the mobile devices — the browsers used for the laptop, Safari and Chrome, are intended to match the browsers used by the different mobile devices.

- *Web Traffic Trace Collection:* The web traffic trace collection for mobile devices is different from desktop machines because it is not straightforward to run scripts that (i) open web pages and (ii) start tcpdump. We instead tether our mobile devices to a desktop machine running Mac OS X 10.9.4 using a bluetooth Internet connection — the Internet sharing utility on the desktop is used for this purpose.[3] This desktop machine connects to the Internet using a 1Gbps Ethernet link and captures all of the traffic that is transferred on this interface, including the web traffic generated by the mobile devices, using tcpdump. We develop a web page whose primary purpose is to automatically open and close web pages on mobile devices. This process is done using a simple javascript command (i.e., window.open("url")) that allow us to open web pages as a popup on any browser, irrespective of device type.[4] We also install Zend Server v 3.0, a web server, on the desktop machine [Zen]. Zend Server is used by the desktop to synchronize the web traffic generated by the mobile devices that are tethered to it. This synchronization process is done by transferring AJAX commands that notify a single mobile device that it may request a new web page. The other mobile devices are notified to load a "dummy" web page that is hosted on the desktop machine — this dummy page is used to end any background web traffic generated by the browser and will not be captured by the wired interface on the desktop machine. However, other apps may still generate traffic in the background and may be captured — we only allow a minimum number of apps that are essential to the functionality of the mobile devices to remain active (e.g., settings app).

We run into two issues when measuring web pages using mobile devices. First, bluetooth connections are unstable on mobile devices and may get disabled intermittently, which halts data collection until the procedure has been manually rebooted. Second, Android devices do not allow for the sleep feature to be disabled, which also halts our data collection procedure. We use the StayAwake and NoLock apps that

[3] We tether bluetooth connections instead of WiFi connections to comply with campus policies.

[4] The only requirement is that the browser allows for pop-up blocking to be disabled — else the requested web pages will be blocked. Our methodology does not work for Windows Phone because this option is not currently available for the mobile version of Internet Explorer.

TABLE 3.3: Hostnames of planetLab nodes.

Hostname	Geographical Region
pl2.eng.monash.edu.au	Austraila
planetlab1.pop-mg.rnp.br	Brazil
cs-planetlab4.cs.surrey.sfu.ca	Surrey, Canada
planetlab2.cs.ubc.ca	Vancouver, Canada
planetlab2.cqupt.edu.cn	China
planetlab4.goto.info.waseda.ac.jp	Japan
pluto.cs.brown.edu	Rhode Island, USA
planetlab2.csee.usf.edu	Florida, USA
planet-lab1.cs.ucr.edu	California, USA
planetlab1.csuohio.edu	Ohio, USA
planetlab4.mini.pw.edu.pl	Poland
planetlab3.cs.uoregon.edu	Oregon, USA
pl3.cs.unm.edu	New Mexico, USA

are designed to avoid sleep events on Android devices but it only partly addresses the issue [StayAwake, NoLock]. Despite using these apps, our data collection procedure for mobile devices is significantly slower than for desktops. To account for this slowdown, we reduce the number of web pages studied for mobile devices from 3614 web pages to 1010 — else, the data collection for mobile devices will take a prohibitively long amount of time (i.e., over 1 year). These 1010 web pages correspond to the first 1010 URLs provided in Appendix 2. We still, however, collect 4 repeated measurements for each web page download.

3.1.3 PlanetLab-based Measurement Methodology

We use 13 planetLab nodes to understand the impact that vantage points have on web page download traffic. These nodes are located in Australia (1 node), China (1 node), Japan (1 node), Brazil (1 node), Canada (2 nodes), Poland (1 node), and the United States (6 nodes – Oregon, Rhode Island, California, Florida, New Mexico, and Ohio) — a list of these nodes is provided in Table 3.3. Each planetLab node runs a single browser, Firefox v17.1.[5] Otherwise, the traffic trace collection procedure for the planetLab data follows the same procedure as the desktop-based measurement methodology.

3.1.4 Potential Issues Related to Data Collection

There are two potential caveats associated with the collection and analysis of web page download traffic that are worth mentioning — these may impact the reproducibility of the results presented in this work. The

[5] The PlanetLab nodes all run Linux that is based on Fedora Core 8 and have a fairly old version of Firefox installed (v 17.1).

first issue is the frequency in which web pages are updated by content providers. Web pages that constantly update may generate different web page traffic — thus, any observed differences in web page traffic may either be due to the client platform or due to the web page itself. Indeed, it is important to check whether web pages change over time before drawing conclusions — this issue is investigated in Section 3.7.1. The second issue is the possibility that browsers may auto-update during our data collection period [Bott]. In our dataset, we find that the Internet Explorer, Safari, and Opera browsers did not update, while the Chrome and Firefox browsers updated multiple times during our data collection period— each update that occurred during this period, along with the date of the update, is provided in Appendix 6.[6]

The fact that these updates occur is important to note because an update in browser version may impact web page download traffic [DUp]. While it would be ideal to study the traffic generated by a single browser version, some browsers auto-update so frequently that it is difficult to obtain multiple web page samples for a single browser version — for instance, Chrome updated 9 times during our data collection period which includes 4 repeated measurements. Thus, our data collection procedure does not completely control for the impact that browser version may have on web page traffic. Despite this limitation, we find that the key results across each repeated measurement for each browser type during our data collection procedure are similar. An overview of the results for these repeated measurements, which include differences in source HTML and browser version, is provided in Section 3.7.2 — additional/supplemental figures which show the similarities across the repeated measurements are provided in Appendix 7.

3.2 Data Analysis Methodology

3.2.1 Feature Extraction

A traffic feature is a measurable property of web page traffic that can be derived from our tcpdump traffic traces. These traffic traces include the first 60 seconds of web page download traffic and consist of HTTP, TCP/IP, and DNS headers, and source HTML files. Some details of the traffic features that we derive are provided below.

DNS-based Features: We developed scripts to extract DNS-based traffic features from tcpdump logs. We use port-numbers (destination port 53) to identify DNS messages. From this, we are able to compute traffic features such as the number of DNS requests, the number of DNS responses, and the TTL of DNS records.

[6] Please note, however, that the Firefox browser on the PlanetLab nodes did not update.

We are also able to compute temporal metrics such as the response time of DNS requests and the inter-arrival times of DNS requests.

HTTP-based Features: HTTP(S) messages are also identified using a port-based approach (80 and 443). We use pcap2har to convert .pcap files (i.e., tcpdump traffic traces) to .har logs (i.e., JSON formatted HTTP archive logs) and developed scripts to extract traffic features from these [pcap2har]. We process the .har log files to extract fields from HTTP headers, including the MIME type, hostname, and length of the web object. We are unable to process encrypted HTTP messages (i.e., HTTPS).[7]

TCP/IP-based Features: We developed scripts to extract TCP/IP-based traffic features from tcpdump logs. We use the 5 tuple heuristic to identify TCP connections and classify it as a TCP connection if we observe the three-way handshake during the web traffic trace. We use TCP header fields to compute the number of TCP flags observed (i.e., PUSH, RESET, FIN, SYN, and ACK), and the number of bytes transferred within each segment. In addition to these statistics, we compute temporal metrics such as the duration of TCP connections and round-trip-time (RTT). RTT is defined as the time between the initial SYN segment sent by the client and the time it receives a SYN-ACK response segment from the server. We also measure the *inter-connection arrival time*—this metric is defined as the amount of time between the start of consecutive TCP connections, where the sending of the initial SYN segment is the event that is taken to be the start of a TCP connection. The last temporal traffic feature that we measure is TCP connection duration — this metric is defined as the amount of time between the first and last segments transmitted during a connection.

Heuristic for web page download times: Web page download time is approximated by taking the amount of time between the first DNS request sent by the client and the last payload byte sent by the server [Huang et al., 2010]. To compute robust statistics of this metric, we also compute several measures of the time until P% (50%, 90%) of the bytes in a given web page are downloaded. Although this metric does not completely correspond to user perceived load times of web pages, it is correlated with user perceived web page load times [Yahoo, Huang et al., 2010].

HTML-based Features: We use the BeautifulSoup python library to extract features from the source HTML of the web pages [Richardson, 2015]. Count statistics that are derived from tags that represent (i) hyperlink-level information (e.g., “a” and “link” tags) and (ii) the *extensions* of embedded objects that are referenced by a page (e.g., .jpeg, .gif, and .png extensions for embedded image objects) are used for *object-*

[7] 18% of our observed TCP connections transfer HTTPS traffic.

based features and analysis— these have commonly been used in other HTML-based analysis [HTM, Canali et al., 2011]. We also derive a feature to analyze the textual-related differences between HTML documents. We use a simple bag-of-words model to count the frequency of all the words that are present in a document — bag-of-words models are commonly used in natural language processing, machine learning, and computer vision [Weinberger et al., 2009]. A *word* in this model is defined as any sequence of characters that is present in an HTML document that is delimited by >, <, ", newline, or whitespace characters. This model allows us to derive features that can measure the overall text-related differences between two documents. To compactly represent these text-related differences, we derive the number of different words feature which is computed as the number of words that are different between two documents (that is, a baseline document and a test document). We use this feature simply as a measure to flag significant differences in text for further analysis.

Please note that our study is focused primarily on the analysis of web page *traffic*. We analyze source HTML primarily to assist in understanding the traffic differences that we observe. In particular, we want to understand if any difference we are observing in web page traffic across client platforms is the result of differences in the source HTML, or is due to the choice of client platform.

3.2.2 Feature Selection Procedure

In total, we derive 26 primary traffic features, 549 secondary traffic features, and 127 HTML-based features (702 features total). Here, *primary traffic features* are traffic features that represent coarse traffic information, such as the number of bytes transferred, number of PUSH flags, number of objects, and number of TCP connections. *Secondary traffic features* are more fine-grained traffic features, such as the number of javascript objects or the number of HTTP responses with status code 404, and derived multi-flow features such as the average number of bytes sent per TCP connection or the maximum object size observed. Lastly, *HTML-based features* are features that are derived from the source HTML. A complete list of the features used in this study are provided in Appendix 3.

We also leverage statistical tests for our analysis that focuses on determining the influence that client platforms have on web page download traffic. In particular, we use a standard non-parametric statistical test, the Kruskal-Wallis test, to determine which traffic features differ significantly across different client platforms. The Kruskal-Wallis test yields p-values that represent the statistical significance of each feature for different client platforms. Here, lower p-values correspond to results that have greater statistical significance. We deem results *statistically significant* if the p-value is less than 0.05. We then use these results to dig deeper into our dataset to determine the source of any statistically significant difference.

3.3 Impact of Browser on Web Page Traffic Features

In this section, we examine whether browser choice has an impact on traffic. Browser choice may impact traffic for different reasons. For example, browsers may generate different traffic because servers may send different browsers different source HTML despite referencing the same URL — indeed, different source HTML can impact web page traffic even when the browser considered is the same. It is also possible that differences in traffic can occur even when the source HTML is the *same* across browsers. For instance, different browsers may process the same source HTML differently because they may implement browser-specific features which influence traffic (e.g., object prefetching) [Wu and Kshemkalyani, 2004, Vieira]. Even if browsers implement the same features in a similar manner, browsers may encounter scripts that have conditional statements that depend on browser type (that is, different browsers may execute different paths in the same source HTML).

We perform experiments to understand the differences between browsers with respect to both traffic and source HTML. The results of these are used determine whether any significant differences in traffic are likely due to differences choice of browser. We do this by comparing the statistical properties (e.g., cumulative distributions) of the features derived from source HTML with the features derived from traffic — we only attribute differences to be due to choice of browser if the statistical properties for the traffic is substantially different even when the source HTML is the same. This general approach to the analysis of our dataset is used for each type of client platform considered.

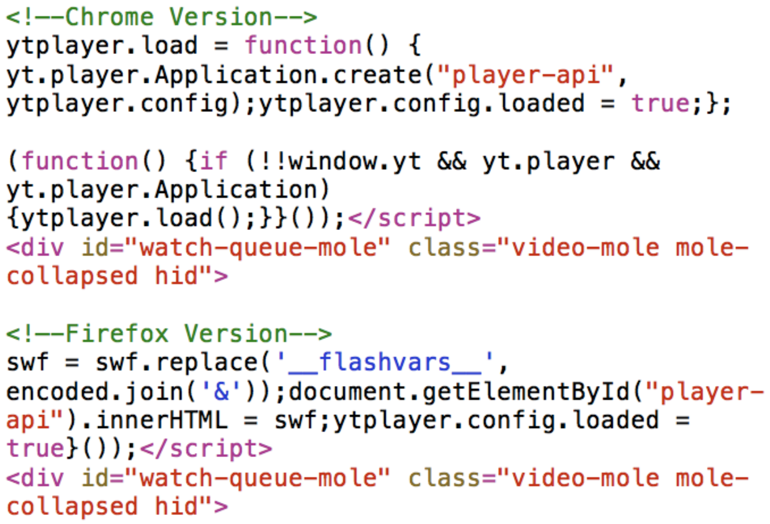
3.3.1 Does the source HTML differ, when the same URL is downloaded by different browser type?

We compare the latest versions of the Internet Explorer, Chrome, Firefox, and Opera browsers for the Windows 7 operating system. The Kruskal-Wallis test for the HTML-based features yields 8 features that have p-value $< .05$ across browser platforms — in fact, these p-values are generally less than 10^{-3}. These 8 statistically significant features are: the number of “label” tags, the number of “tr” tags, the number of “table” tags, the number of “td” tags, the number of “style” tags, the number of “legend” tags, javascript length (i.e., the number of characters present between script tags), and the number of different words. Upon further analysis of our data, we find that these statistically significant features correspond to the following:

- *Differences in javascript:* We find that many content providers such as soundcloud.com and bing.com (particularly image search results) use different javascript code that is suited for different browsers — these javascript related differences were identified using the number of different words feature. We find that some javascript methods are implemented differently across browser platforms and/or have conditional statements that branch for different client browser platforms. For example, soundcloud.com uses conditional statements that takes the client platform into account during javascript execution to determine whether HLS (HTTP Live Streaming) is supported by the client platform. Alternatively, Figure 3.1 shows an example where a Youtube.com page has javascript that is browser-specific — here the Chrome javascript for loading a video appears to be HTML5-based while the Firefox javascript appears to be flash-based (this flash content is identified by the "swf" references in Figure 3.1). It is known that if different client platforms are not taken into account, rendering differences across browsers can occur when the same source HTML is processed — for example, target.com has differences in rendered tables across browsers despite referencing the *same source code* that renders that portion of the page.

- *Ads:* We also observe "ads" that attempt to get a user to download a particular browser or app that is browser dependent. For example, Yahoo.com recommends that users update to the latest version of Firefox for client platforms that are not Firefox, whereas target.com recommends that Chrome users download the Target mobile app. These ads seem to be attempts to get users to utilize software that is fully supported by the content provider.

Figure 3.2 plots the cumulative distribution of the number of different words between the Chrome and the IE, Firefox, and Opera browsers. Please note that web pages that are the *same* across browser will have a number of different words value of 0 — hence, this feature for the baseline curve (Chrome vs Chrome) is always 0. This plot shows that over 50% of source HTML across different browsers do not have any differences (i.e., a value of 0 for the number of different words feature). In fact, approximately 75% of web pages differ by fewer than 50 words. This result shows that while there are some statistically significant features across browser platforms, many web pages are not influenced by client platforms at all.



```
<!--Chrome Version-->
ytplayer.load = function() {
yt.player.Application.create("player-api",
ytplayer.config);ytplayer.config.loaded = true;};

(function() {if (!!window.yt && yt.player &&
yt.player.Application)
{ytplayer.load();}}());</script>
<div id="watch-queue-mole" class="video-mole mole-
collapsed hid">

<!--Firefox Version-->
swf = swf.replace('__flashvars__',
encoded.join('&'));document.getElementById("player-
api").innerHTML = swf;ytplayer.config.loaded =
true}());</script>
<div id="watch-queue-mole" class="video-mole mole-
collapsed hid">
```

Figure 3.1: Example where javascript is different for different browsers (Chrome vs Firefox).

3.3.2 Does the source HTML differ, when the same URL is downloaded by different browser version?

We also briefly consider the impact that browser *version* may have on source HTML pages. Our data collection methodology, which downloads a web page using a browser, must be modified for this browser version analysis because it is difficult to use multiple versions of the *same* browser on a single device. We instead use the urllib2 python library to make HTTP requests for each web page listed in Appendix 3 and extract features from the corresponding HTTP responses (i.e., the source HTML) [Documentation]. Please note that the urllib2 library is used because the User-Agent field in the HTTP request header can be modified such that the server will believe that the request originated from the client platform of interest (i.e., operating system, browser type, browser version, etc). This approach, however, cannot be used to approximate the actual web page traffic downloaded when a browser processes source HTML —this is because different browsers have different source code and may generate different web page traffic despite referring to the same HTML.[8] Hence, the only way to measure web page traffic generated by a browser is to download the page using said browser. Thus, we do not consider web page traffic features in this analysis.

We consider four pairs of browsers on Windows 7; 1) Firefox v 33.0 and Firefox v 17.0; 2) Internet Explorer v 11.0 and Internet Explorer v 9.0; 3) Chrome v 38.0.2125.122 and Chrome v 33.0.1750.154; and 4) Opera v 25.0.1614.68 and Opera v 12.16. When comparing two samples of web pages, a sample

[8] Understanding whether browsers indeed generate different traffic is part of the goal of this study.

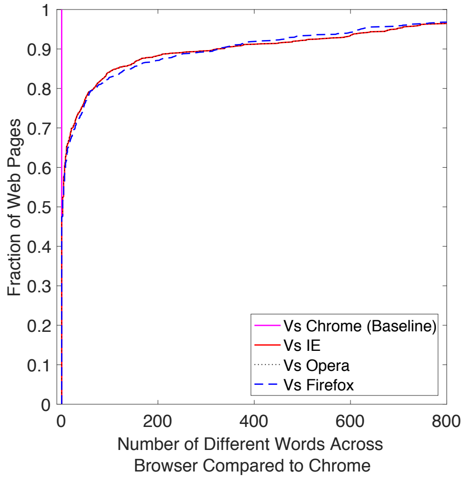


Figure 3.2: Cumulative distribution of number of different words across browser as compared against Chrome.

comprised of the four outdated browsers vs a sample of four up-to-date browsers, our statistical test yields 13 statistically significant features. The most notable features that are not also influenced by browser platform, say Internet Explorer vs Firefox, are the number of script tags and the number of HTML5 tags. With respect to the number of script tags, we observe similar differences in scripting behavior as we did with the differences in browser platform. With respect to the number of HTML5 tags, we observe that there tends to be more HTML5-related tags for the latest browser versions as compared to the older versions — we believe this occurs because HTML5 is the newest and current version of HTML that is likely only supported on newer browsers.

We also observe cases where content providers treat outdated or unsupported browsers in the following ways. First, the content provider can respond to the HTTP request, but provide a warning to the user that their browser needs to be updated (e.g., zillow.com and soundcloud.com) — this may also result in failed HTTP requests. Second, the content provider can respond to the HTTP request by sending source HTML that is compatible with the user's browser. This is explained next.

We find multiple instances when browser version has an impact on some of the HTML tag-based features. Figure 3.3 shows Google search results that were requested and rendered using an outdated Opera browser, while Figure 3.4 shows Google search results using an up-to-date Chromium-based Opera browser — these web pages are displayed differently for these different versions of Opera. These observed differences are almost purely stylistic with respect to image size and the visibility of URLs on images — though there are some images present in the new version that are not present in the original. We next show a different

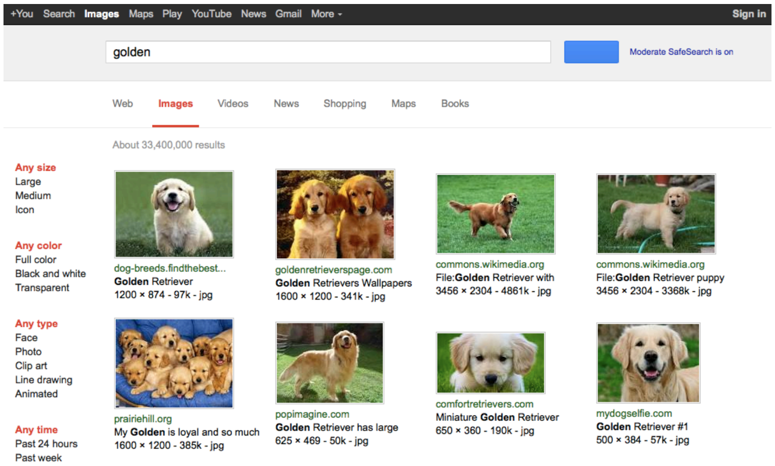


Figure 3.3: Search result page for old version of Opera

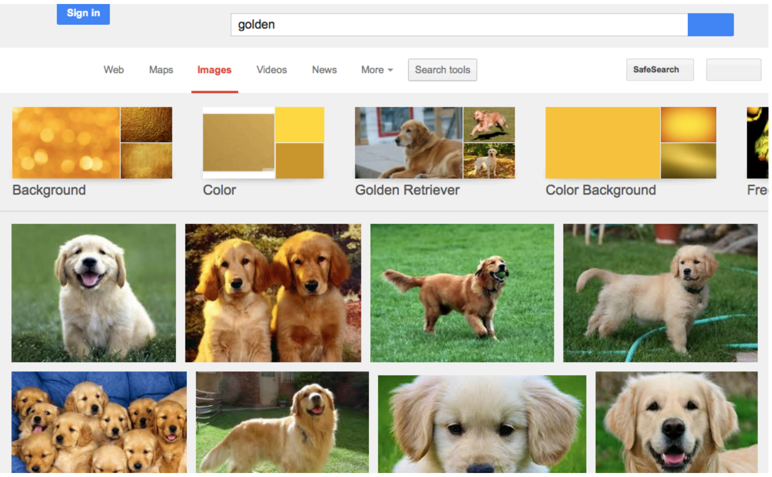
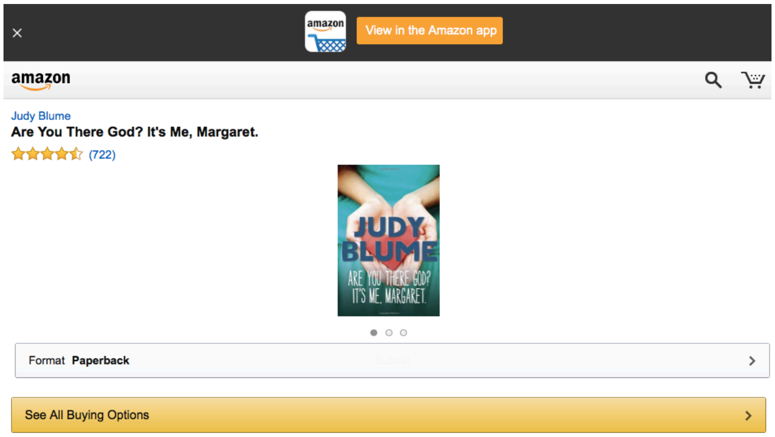


Figure 3.4: Search result page for new version of Opera.

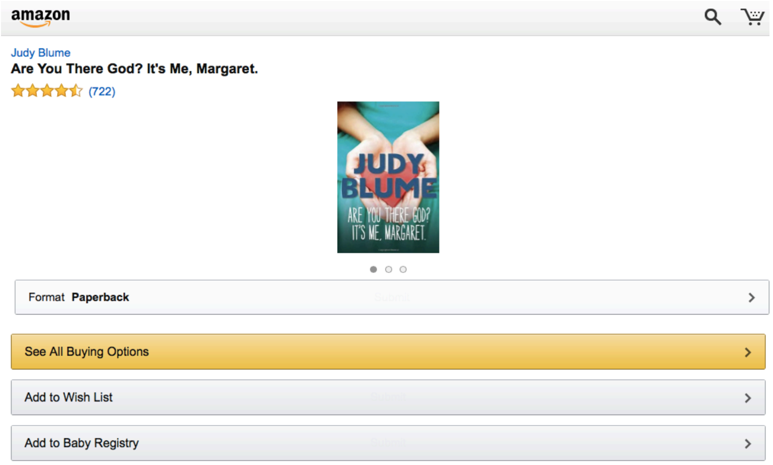
example of when a web server responds with a web page for an outdated browser. Here, the HTTP request is for a mobile web page of a product on Amazon.com. Figure 3.5 shows that when a mobile web page is requested using an *up-to-date mobile device and browser* (an iPhone in particular), the request is satisfied as expected. When we make the same request for a mobile web page using an *outdated Firefox browser on a laptop* we also get the *same* mobile web page — though we do not observe an ad for downloading an app. This web page is shown in Figure 3.6. Figure 3.7 shows that when the same request is made to Amazon.com using an *up-to-date Firefox browser on a laptop* we get a *different* mobile web page that is clearly representing the same product shown in Figure 3.5. It is clear that these downloaded web pages are both (i) mobile-optimized web pages and (ii) different, where the version of the page shown in Figure 3.7 appears to be an older mobile web page design than the page shown in Figure 3.5. We conclude two things from these



Judy Blume
Are You There God? It's Me, Margaret.
(722)

Format Paperback

Figure 3.5: Product web page for Amazon site as rendered using Safari browser on iPhone.



Judy Blume
Are You There God? It's Me, Margaret.
(722)

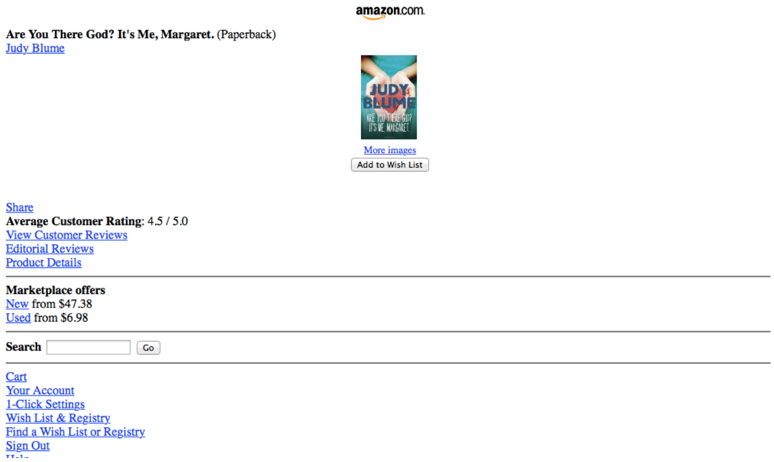
Format Paperback

Figure 3.6: Product web page for Amazon site as rendered using an old Firefox Browser.

observations: 1) mobile web pages may sometimes be used to fulfill HTTP requests to outdated browsers (we observe similar behavior for yahoo.com and att.com — please refer to Figure 3.8 for an example of a mobile page being returned for an outdated Firefox browser and Figure 3.9 for the normal up-to-date version); and 2) interesting and unexpected quirks exist for some HTTP requests that are influenced by browser choice.[9] The impact that browser version has on web page downloads is important for web crawling tools because (i) web crawlers may be used for years without receiving any significant upgrades and (ii) content providers may respond to known web-crawler User-Agents in a manner that results in errors or downloading data that is limited (in a manner similar to mobile web pages) [Notess, 2002].

Implications of HTML-based differences: We found that source HTML can be influenced by browser

[9] Please note that the significant differences discussed here are primarily true for browser version analysis for Opera and Firefox because we have the largest range in release dates for these two browsers.



Are You There God? It's Me, Margaret. (Paperback)
Judy Blume

More images
Add to Wish List

Share
Average Customer Rating: 4.5 / 5.0
View Customer Reviews
Editorial Reviews
Product Details

Marketplace offers
New from $47.38
Used from $6.98

Search Go

Cart
Your Account
1-Click Settings
Wish List & Registry
Find a Wish List or Registry
Sign Out

Figure 3.7: Product web page for Amazon site as rendered using a new Firefox Browser.

type and version. A summary of the implications of our analysis is below:

- Our results show that it is definitely possible that web pages can differ across client platform. Thus, any measurement study or web-related application that relies on source HTML, such as web page archival, should verify that the web pages that they visit is not influenced by the browser that is used — else, it is possible that the downloaded source HTML is biased for a particular browser which will limit the scope of the analysis.

- We find that most of the differences observed across browser platforms correspond to compatibility issues across browser type and version or browser-specific ads. In fact, most of the significant differences that we observe involve HTML tags that impact the way the page is displayed. These features do not tend to influence the number of web objects that are referenced by the page. Thus, it is not expected that these differences will have a dramatic impact on the TCP/IP and HTTP headers that correspond to these web pages.

- We find that approximately 50% of web pages differ, to some degree, across browser platform. While this is a large fraction of web pages, we are aware that time is a possible factor that may bias our results — we investigate this factor in Section 3.7. Nevertheless, this observation is important in the context of this dissertation because it provides a baseline for determining whether source HTML is the only cause for any possible differences in traffic observed across browsers. Specifically, if more than 50% (i.e., the approximate percentage of web pages that have different source HTML) of web page traffic differs across browsers we can confidently state that there are cases where choice of browser

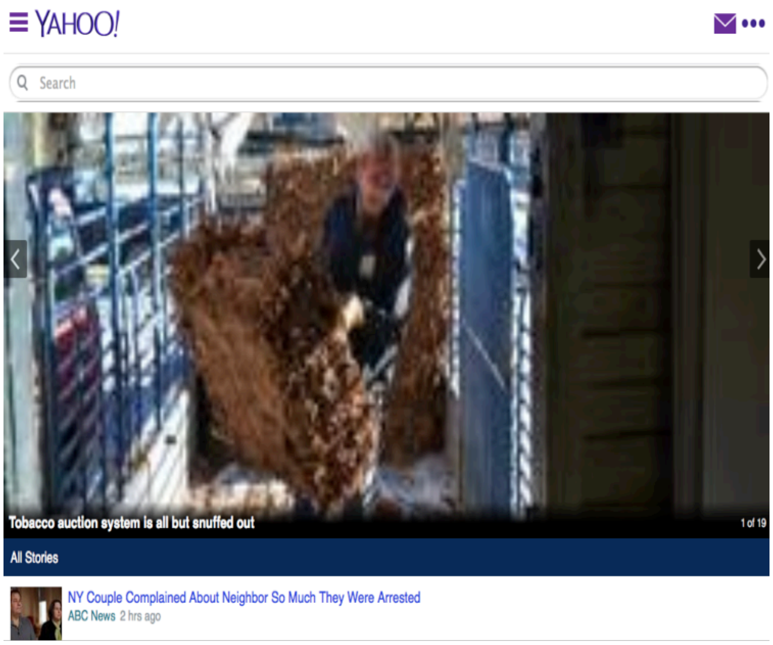

YAHOO!
Search
Tobacco auction system is all but snuffed out
1 of 19
All Stories
NY Couple Complained About Neighbor So Much They Were Arrested
ABC News 2 hrs ago

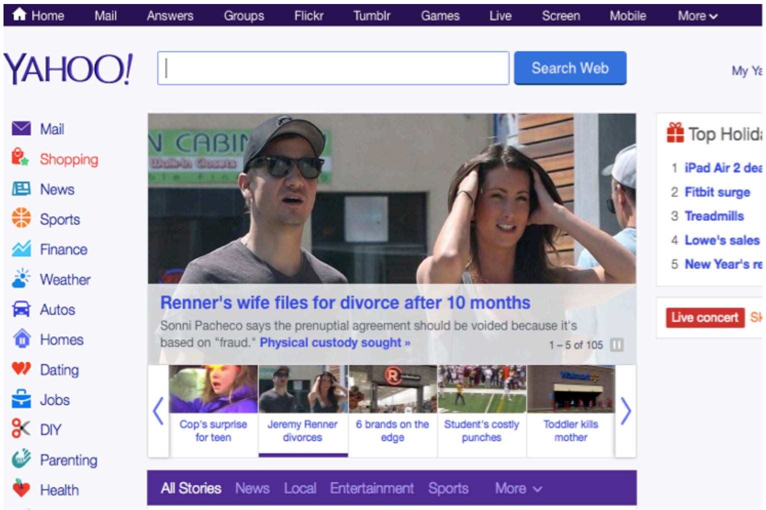
Figure 3.8: Example where mobile web page is returned for an old desktop Firefox browser.

influence web page traffic when the source HTML is the same.

3.3.3 Does traffic differ across different browsers?

We next study whether web page download traffic differs across client platforms and investigate whether such differences can occur when the source HTML is the same. We start by analyzing the differences in traffic as observed due to different browsers, but on the same operating system. We focus on the Windows 7 OS since it is the most popular. Using our statistical testing methodology, we find that a *vast majority of the traffic features* (534 out of the 575 traffic features) have p-values less than .05 in our dataset — in fact, most p-values are $< 10^{-10}$. Most of these features are secondary traffic features which are derived from the 26 primary traffic features that we defined in Section 3.2.2.[10] Thus, differences in the primary traffic features are likely the root cause of the significant differences that we observe. In this section, we summarize the significant primary traffic features when secondary traffic features are also statistically significant to avoid redundancy in our discussion — we also present secondary traffic features when primary traffic features are not statistically significant.

[10] Recall that primary and secondary traffic features are the only two types of traffic features that we consider. Please refer to Appendix 3 for a list of the primary traffic features



Renner's wife files for divorce after 10 months

Sonni Pacheco says the prenuptial agreement should be voided because it's based on "fraud." Physical custody sought »

Figure 3.9: Example where traditional web page is returned for an up-to-date desktop Firefox browser.

3.3.3.1 Differences in the Number of Bytes Transferred

We analyze the total number of application layer bytes in traffic observed when a web page is downloaded. This traffic feature yielded a highly significant p-value that is less than 10^{-20}. Figure 3.10 plots the cumulative distribution of this feature — please note that the x-axis is shown on a log-scale. Figure 3.10 shows that there are two groups of browsers that transmit a similar number of bytes — Opera and Chrome vs. IE and Firefox. Chrome and Opera are both based on the Chromium browser [CCh, OPC] (referred to as Chromium browsers), so it is likely that the Chrome and Opera browsers share much of the same source code. The IE and Firefox browsers are proprietary and are not based on the Chromium browser (referred to as non-Chromium browsers). So, while IE and Firefox transfer a similar number of bytes and do not share the same source code, they also transfer a consistently different number of bytes than Chrome and Opera. This observation suggests that Chromium-based browsers include unique features which cause them to behave differently than the other browsers.

Figure 3.10 shows that Chrome and Opera consistently send more bytes than IE and Firefox. For each web page in our dataset, we compute the ratio of the number of bytes transferred by Chrome, a Chromium-based browser, and IE, a non-Chromium browser. Figure 3.11(a) plots the cumulative distribution of this ratio. This figure shows that the median ratio of bytes transferred by Chrome and IE is 2. Figure 3.11(a) also shows that Chrome sends over 4 times more bytes than IE for 20% of web pages downloaded. Similar results for the ratios between Chrome and Firefox, Opera and IE, and Opera and Firefox, and are shown in

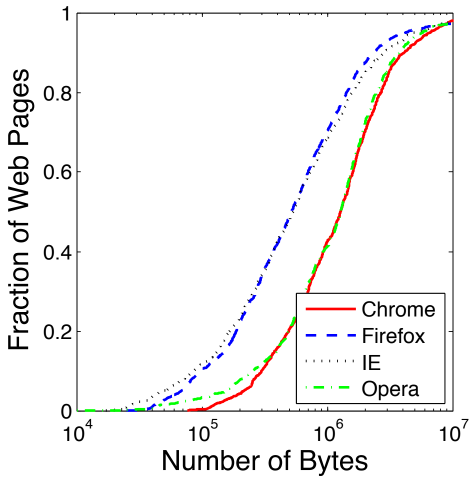


Figure 3.10: Network impact of browser selection (Raw cumulative bytes).

Figures 3.11(b), (c), and (d) respectively.

Implication - Network Impact Object prefetching, a browser-specific feature that uses heuristics to preemptively download objects, is a possible explanation for the large difference in the number of bytes feature we observe across browsers [Wu and Kshemkalyani, 2004, Vieira]. This large difference in bytes transferred may result in unwanted and unnecessary congestion and capacity issues, as well as unwanted charges for network usage. Please note that our dataset includes only *popular* web pages. Thus, such differences in the number of bytes downloaded across browser platforms are likely observable on aggregate links in the Internet today.

3.3.3.2 Differences in Number of Web Objects Transmitted

The source HTML may differ across browser because content providers may use the User-Agent field in HTTP requests to serve web pages that are targeted for specific browser platforms. Thus, we next consider whether the primary source of the differences in the number of bytes downloaded across browser is due to differences in the source HTML downloaded across browser. We found that over 50% of the source HTML of the web pages downloaded did not show any difference across browsers (i.e., our number of different words feature that was derived from our bag-of-words model was 0 for over 50% of pages). We also observe similar results, with respect to the number of bytes transferred, when considering only web pages that have the same source HTML — this is shown in Figure 3.12. These results mean that choice of browser *must be causing some* of the differences that we consistently observe across browser since traffic

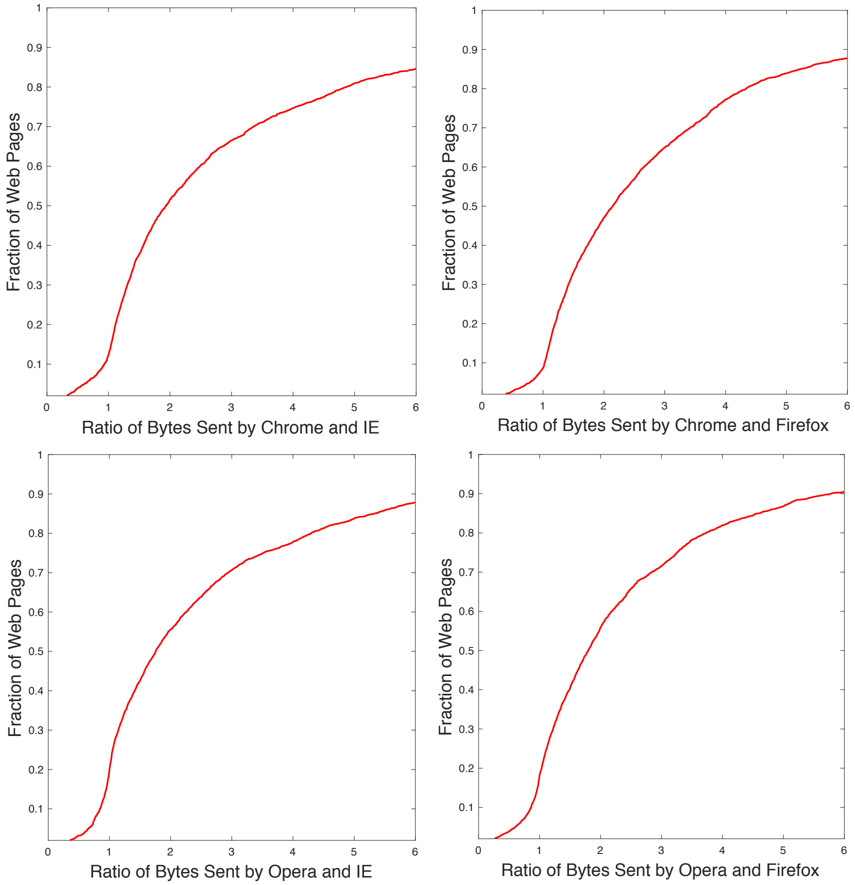


Figure 3.11: Network impact of browser selection (Ratio of bytes across browsers)

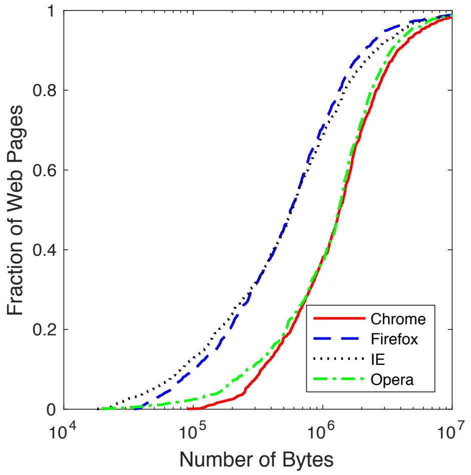


Figure 3.12: Network impact of browser selection when the source HTML is the same (Raw cumulative bytes).

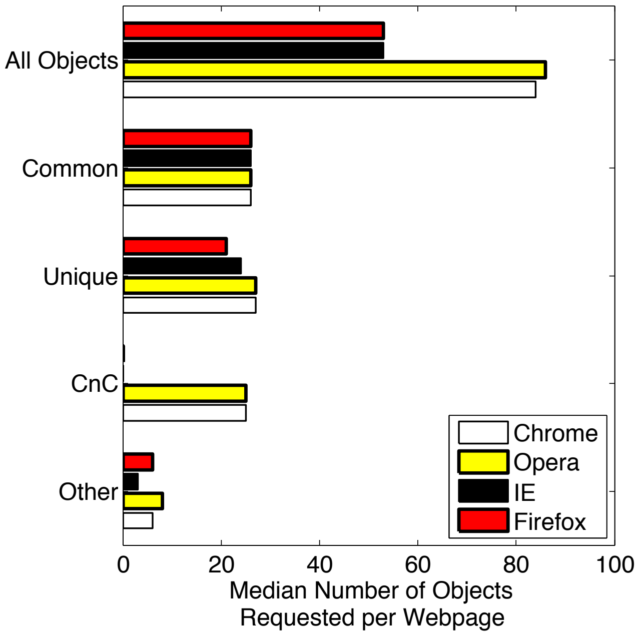


Figure 3.13: Breakdown of the average number of objects transmitted for each web page download for different browsers.

differs for more than 50% of the web pages downloaded — the observation that Chromium-based browsers behave similarly also supports this claim.[11] We assume that the large differences in the number of bytes downloaded are the result of browsers downloading a different number of objects to load the same web page — prefetching mechanisms within browsers may influence such behavior [Wu and Kshemkalyani, 2004, Vieira]. We next investigate the objects that are downloaded by each browser to gain a better understanding of the byte-related differences that we observe.

We measure the total number of objects as the number of HTTP responses observed in traffic when a web page is downloaded. The median number of total objects transmitted by each browser to download a web page is shown in the top-most bars in Figure 3.13 (labeled "All Objects"). Similar to the number of bytes feature, Chromium-based browsers transmit more objects than IE and Firefox. This result shows that it is likely that the number of bytes transmitted is different because the number of objects transmitted is different — the logic here being that more objects transmitted result in more bytes being transmitted. In fact, Figure 3.13 shows that Chromium-based browsers transmit *approximately 33 more objects* than Firefox and IE on average. Figure 3.14(a) plots the cumulative distribution of the number of objects feature

[11] We do not know whether browser of choice are causing the significant differences for the 50% of web pages with *different* source HTML.

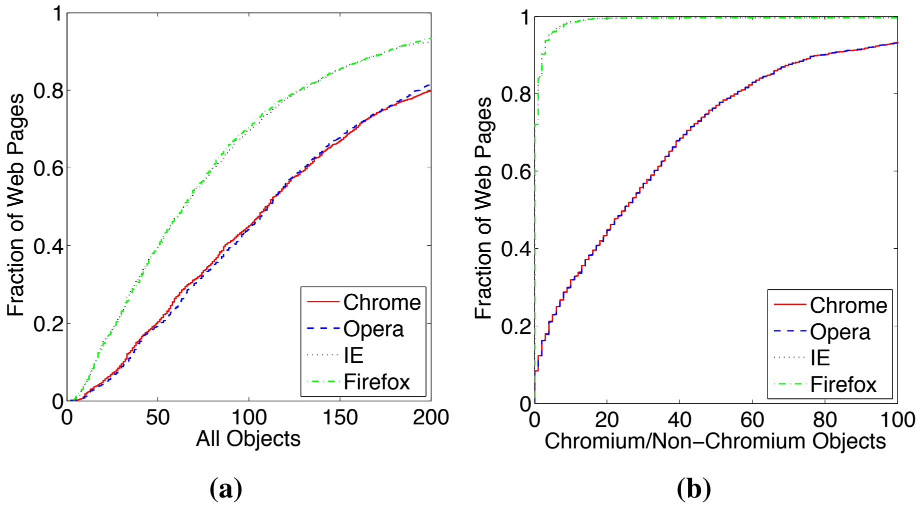


Figure 3.14: Cumulative distribution of Number of objects observed across browsers.

for each browser. Figure 3.15(a) shows a similar plot where only web pages with the same source HTML is considered. These plots confirm that the difference in the total number of objects transmitted between Chromium and non-Chromium objects is indeed consistent and is not skewed by the tails of the distribution. We next categorize different classes of objects to further analyze the difference in the total number of objects observed across browser. We find that:

- *Common Objects:* We define a common object as an object that is downloaded at least once for *every* browser per web page downloaded — here an object is identified using the string referenced in the URL field of the HTTP response header.[12] Figure 3.13 shows that the median number of common objects are similar across browser. Thus, common objects do not account for a significant amount of the difference in the total number of objects observed across browsers. Figure 3.13 also shows that common objects account for 30-50% of the total number of objects downloaded per web page.[13] This result shows that less than half of the objects downloaded across browsers per web page are similar.

- *Unique Objects:* Unique objects are objects that are observed *exclusively* on one of the four browsers per web page download. Figure 3.13 shows that unique objects account for approximately 30-45% of the total objects downloaded per web page — please note that the percentage of objects downloaded depends on the browser used. We find that unique objects account for a difference of approximately

[12] All browsers may download the *same* object multiple times when loading *same* web page. Downloading multiple copies of the same object *occurs for 0.7% of objects* and do not significantly impact our results.

[13] The percentage of common objects downloaded per web page load depends on the browser used.

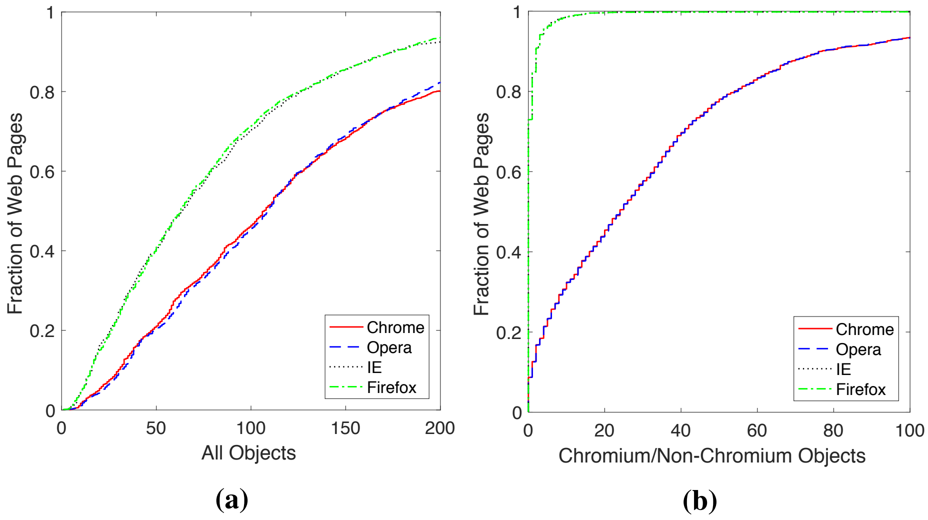


Figure 3.15: Cumulative distribution of Number of objects observed across browsers when the source HTML is the same.

7 out of the 33 total objects that differ across Chromium/Non-Chromium browsers (approximately 21% of the difference in total objects). While this difference is non-negligible, it does not account for the majority of the difference in the total number of objects transmitted by different browsers. We provide an overview of the hostnames that send the unique objects to clients in Section 3.3.3.3 to further analyze possible sources of the differences in the number of objects observed across browser.

- *CnC Objects:* We next study objects that exclusively occur on Chromium or non-Chromium browsers. Chromium objects are objects that are present exclusively on *both* Opera and Chrome, while non-Chromium objects are present exclusively on both Firefox and IE — Chromium and non-Chromium objects are collectively referred to as CnC objects. Figure 3.13 shows that the median number of CnC objects are substantially different across browsers. In particular, CnC objects account for less than 2% of the objects downloaded for non-Chromium browsers, while they account for approximately 30% of the objects downloaded for Chromium-based browsers. There is a difference of approximately 25 CnC objects between the Chromium-based and non-Chromium browsers. These objects correspond to approximately 75% of the difference in the median number of objects observed across these browsers (25 CnC objects out of the 33 total number of different objects). Figure 3.14(b) plots the cumulative distribution for the number of CnC objects for each browser. This plot shows that the differences in the number CnC objects transmitted between Chromium-based and non-Chromium browsers is consistent across the whole distribution — not just the median. Please note that similar observations are made even when we consider web pages with the *same* source HTML — this result is shown in Figure 3.15(b). We provide an overview of the hostnames of the servers that send the CnC objects to clients in Section 3.3.3.4 to further analyze possible sources of the differences in the number of objects observed across browser.

- *"Other" Objects:* The last type of object that we consider are "other" objects — these objects correspond to the objects that are *not* categorized as common, unique, nor CnC. Figure 3.13 shows that "other" objects account for a small amount of the differences (i.e., approximately 2-4 objects) in the median number of the total objects observed across browsers. While the difference in the number of "other" objects is small compared to the differences observed for unique and CnC objects, it makes up a larger fraction of the differences than common objects.

In summary, the categories of objects that differ the most across browsers are unique and CnC. We next investigate whether there are any clear patterns in the frequency in which different servers, as identified using their hostname, transmit unique and CnC objects. We perform this analysis because it may provide insight into the hostnames of the servers (i.e., which content providers), and likely the types of web-related services (i.e., ads, tracking, or CDNs), that are frequently contacted by different browsers — please note that we study the frequency of the hostnames of the objects instead of the full URL of the object because full URLs are much more difficult to interpret.

3.3.3.3 Hostnames for Unique Objects

We first investigate the frequency of hostnames of the unique objects in our dataset. 6976 different hostnames are observed in our dataset. We observe 5318 hostnames that correspond to objects that are unique to a particular browser. Despite this large number of hostnames that correspond to servers that host unique objects, 400 hostnames corresponds to approximately 75% of the unique objects observed in our dataset. We observe that the most popular hostnames in our dataset are associated with ads, tracking, and analytics services.[14] Table 3.4 shows 30 hostnames which correspond to servers that send the most unique objects that we observe (35% of the total unique objects) in our dataset (please note that these observations

[14] The classification we use here is obtained using details from the URL and services such as Crunchbase and search engines. URLs, hostnames, and descriptions from different services (e.g., search results and Crunchbase) that include the keywords "ad", "track", "trk", "analytics", "trc", "static", "google", and "beacon" are included in the ads, tracking, and analytics group.

were repeatable across multiple iterations).[15] While many of these unique objects yield a similar number of objects across browsers, there are some exceptions that are worth mentioning. First, gstatic.com, a hostname that is managed by Google, is primarily only contacted as an unique object for the Chrome browser. Other observations include that the servers with the hostname bid.g.doubleclick.net are visited significantly more for the Chrome browser than others, and that servers with the hostnames googleads4.g.doubleclick.net and s0.2mdn.net (both owned by Google) are visited more on Chromium-based browsers than others. We believe that it is by design that these Google-based services are more prevalent on the Chrome and/or Chromium-based browsers.

We also observe that some browsers may contact servers with different hostnames than other browsers for similar services. For example, apx.moatads.com and v4.moatads.com both serve approximately 1000 more objects for the Opera, Firefox, and IE browsers than the Chrome browser — this difference equates to approximately 2000 more objects that are transmitted to a moatads.com server for browsers other than Chrome. However, it seems that Chrome makes up for this 2000 object difference by requesting approximately 2000 objects from afs.moatads.com, whereas the Opera, Firefox, and Internet Explorer browsers do not request any unique objects from this server. Clearly, there is some type of browser dependence for the hosts that are contacted. This result may be due to load balancing efforts by moatads.com — multiple sources state that Chrome accounts for approximately half of the browser usage in the U.S [W3C, Sta, b].

There are many other examples of where the prevalence of the differences in the number unique objects for a hostname differs across browser, but the explanations are similar. We present only the top differences in Table 3.4 for brevity.

3.3.3.4 Hostnames for CnC Objects

We next discuss the frequency of hostnames that correspond to CnC objects. We observe 1571 hostnames that correspond to CnC objects, where the top 200 hostnames account for 75% of these objects. Table 3.5 and Table 3.6 together show a list of approximately 30 hostnames that correspond to these objects (40% of the total CnC objects). CnC objects consists of both ad and tracking services (e.g., pagead2.googlesyndication.com and b.scorecardresearch.com), shown in Table 3.5, and CDN services that serve images or other content (e.g., g-ecx.images-amazon.com, img.s-msn.com, and foodnetwork.sndimg.com),

[15] Please note that our dataset downloads 3614 web pages for each iteration. So, 3614 is a good reference point for understanding how often an object from a server with a particular hostname is requested per web page download.

TABLE 3.4: Frequency in Hostnames for Objects that are Uniquely Observed Across All Browsers (List Consists Primarily of Ads and Tracking Services).

Hostname	Chrome	Opera	IE	Firefox
b.scorecardresearch.com	3005	2466	2755	2581
pubads.g.doubleclick.net	2759	2507	2151	2065
pagead2.googlesyndication.com	2250	2466	1596	1913
v4.moatads.com	1116	2196	2149	2190
www.google-analytics.com	1623	1653	1614	1646
ad.doubleclick.net	1712	1601	1342	1390
googleads.g.doubleclick.net	1420	1305	1347	1288
www.google.com	1282	1303	1250	1225
ping.chartbeat.net	1168	1024	1009	1046
data.t.bleacherreport.com	459	811	1237	1449
apx.moatads.com	191	1143	1082	1142
www.gstatic.com	1007	76	68	42
l.betrad.com	882	722	667	775
s0.2mdn.net	1095	959	428	405
c.betrad.com	953	817	454	431
stats.g.doubleclick.net	543	592	439	622
secure-us.imrworldwide.com	701	459	498	442
pixel.quantserve.com	592	442	512	486
afs.moatads.com	2028	0	0	0
ib.adnxs.com	530	487	476	470
beacon.krxd.net	470	504	447	492
dt.adsafeprotected.com	590	331	365	481
trk.vidible.tv	80	116	1006	534
cm.g.doubleclick.net	482	398	457	372
optimized-by.rubiconproject.com	329	357	435	387
crl.microsoft.com	217	305	772	178
ads.yahoo.com	348	332	403	329
bid.g.doubleclick.net	499	263	299	289
trc.taboola.com	343	343	317	314
googleads4.g.doubleclick.net	585	526	20	20
at.atwola.com	397	196	211	218
beacon.walmart.com	249	252	241	263

TABLE 3.5: Frequency in Hostnames for Objects that are Chromium Only or Non-Chromium Only Across All Browsers (Ads and Tracking Services List).

Hostname	Chrome	Opera	IE	Firefox
pagead2.googlesyndication.com	1090	1126	56	56
b.scorecardresearch.com	955	955	0	0
global.fncstatic.com	819	819	0	0
tpc.googlesyndication.com	703	708	0	0
www.google.com	706	704	1	1
c.betrad.com	658	661	56	49
partner.googleadservices.com	604	604	1	1
static.huluim.com	566	566	12	12
cbsnews2.cbsistatic.com	540	540	0	0
cbsnews1.cbsistatic.com	508	508	0	0
static.ak.facebook.com	488	488	0	0
fonts.gstatic.com	440	479	8	16

shown in Table 3.6. The observation that ads and tracking services account for a large fraction of CnC objects shows that these services are influenced by choice of browser [Research]. Also, the observation that many of the CnC objects originate from CDN services means that it is likely that many of these objects have a MIME type of image [Butkiewicz et al., 2011]. The transmission of such objects is a possible explanation for the significant difference in the number of bytes observed between the different types of browsers in Figure 3.10 because image objects usually consist of a large fraction of the bytes in web page traffic [Butkiewicz et al., 2011].

Implication - Privacy Concerns We observe that many of the objects that make up the differences in the number of objects transmitted when loading a web page across browsers correspond to ad, tracking, and analytics services. This observation raises privacy concerns for users because the more objects that are transmitted from such services to a browser increases the chance and the extent that a user's browsing behavior is being tracked. Users may not be aware that browsers contact different servers, particularly ad and tracking servers, when browsing the *same* web pages. If users are aware of such differences they may choose to use a browser that tend to request a fewer number of web objects from ads, tracking, and/or analytics services (i.e., Non-Chromium browsers) [Chaabane et al., 2014, Li et al., 2015a].

TABLE 3.6: Frequency in Hostnames for Objects that are Chromium Only or Non-Chromium Only Across All Browsers (CDN or Other Services List).

Hostname	Chrome	Opera	IE	Firefox
www.gannett-cdn.com	5192	5190	7	7
img.s-msn.com	2350	2350	88	88
z-ecx.images-amazon.com	981	981	0	0
foodnetwork.sndimg.com	825	825	0	0
tags.tiqcdn.com	764	764	12	12
i.dailymail.co.uk	737	737	0	0
l.yimg.com	707	706	10	10
g-ecx.images-amazon.com	697	697	3	3
data.t.bleacherreport.com	645	633	442	437
s.huffpost.com	577	577	0	0
o.aolcdn.com	568	569	7	7
www.usmagazine.com	540	540	0	0
cdn.krxd.net	490	490	0	0
assets.macys.com	474	484	4	4
www.nbc.com	453	453	0	0
www.bing.com	447	447	13	13
ak1.ostkcdn.com	428	428	8	8
www.concast.com	401	401	0	0
img.webmd.com	390	390	0	0
platform.twitter.com	396	382	0	0
i2.cdn.turner.com	382	382	1	1

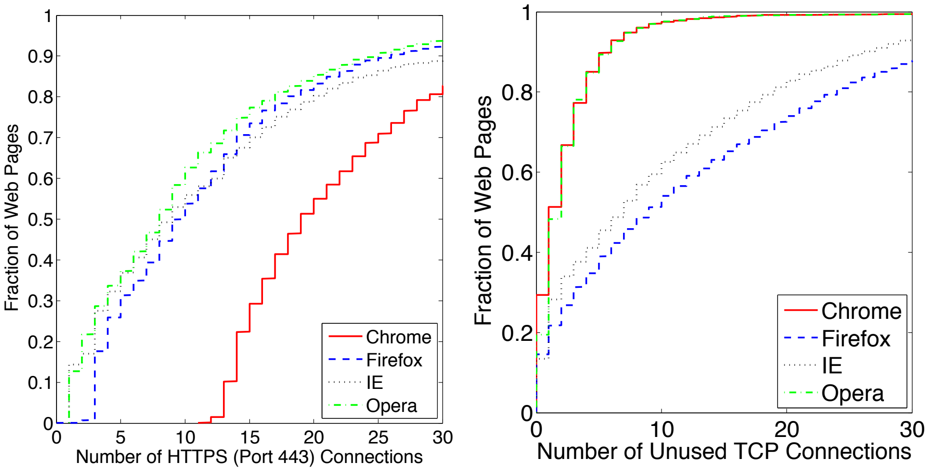
3.3.3.5 Differences in other TCP Connection-related features

We also discuss other primary traffic features that vary significantly across browser that are not directly related to the differences in web page traffic that we discussed earlier (i.e., the number of bytes and objects transmitted). We first discuss the difference between the number of secure TCP connections established across browser. Secure TCP connections are TCP connections where the destination port number is 443. Figure 3.16(a) shows a cumulative distribution plot of this feature. We find that the number of secure TCP connections established for the Chrome browser is consistently higher than other browsers. We refer to DNS traces to understand the hostnames of the servers that many of these connections are contacting.[16] We find that different browsers contact servers with unique hostnames for almost every web page download: (i) Opera contacts sitecheck2.opera.com and update.geo.opera.com, (ii) Firefox contacts ocsp.digicert.com, safebrowsing.google.com, and dtex4kvbppovt. cloudfront.net, (iii) Internet Explorer (IE) contacts go.microsoft.com and iecvlist.microsoft.com, and (iv) Chrome contacts accounts.youtube.com, mtalk.google.com, clients4.google.com, www.googleapis.com, accounts.google.com, ssl.gstatic.com, www.google.com, translate.googleapis.com, and clients4.google.com. Clearly, many of these requests are contacting servers that are associated with the developer of these respective browsers. In particular, the Chrome browser consistently contacts 9 hostnames while the other browsers tend to contact 2-3 hostnames — this difference accounts for most of the consistent differences across browser that we observe in Figure 3.16(a) (i.e., the shifts in the cumulative distribution curves in Figure 3.16(a)). We believe that the hostnames of these servers correspond to authentication, location-based, and tracking services. The other hostnames that account for the remainder of the difference between the curves (approximately 1-2 secure TCP connections on average) in Figure 3.16(a) do not consistently appear for each web page download — these secure TCP connections instead depend more on the particular web page that is downloaded. Please note that we observe similar results when considering only web pages that have the same source HTML — this is shown in Figure 3.17(a).

Another traffic feature that we discuss is the number of unused TCP connections. Unused TCP connections are TCP connections that do not transmit bytes that correspond to application layer data. Unused TCP connections are possibly due to the TCP pre-connect feature in browsers which uses heuristics to establish TCP connections preemptively to possibly reduce web page load times [Newton et al., 2013, Souders]. Fig-

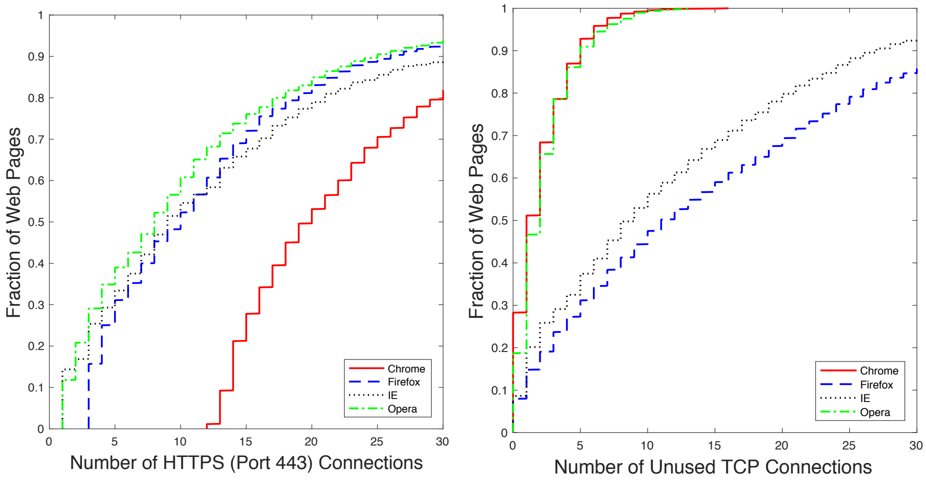
[16] Please note that we cannot observe the details of the objects transferred in secure TCP connections

ure 3.16(b) plots the cumulative distribution of this feature. One observation from this plot is that (i) the curves for non-Chromium and Chromium browsers differ to a large degree and (ii) non-Chromium browsers tend to establish more unused TCP connections than Chromium-based browsers. Similar observations can be made when the source HTML is the same — this is shown in Figure 3.17(b). Clearly, different browsers rely on these features to different extents.



(a) **(b)**

Figure 3.16: Differences in number of TCP connections established across browsers.



(a) **(b)**

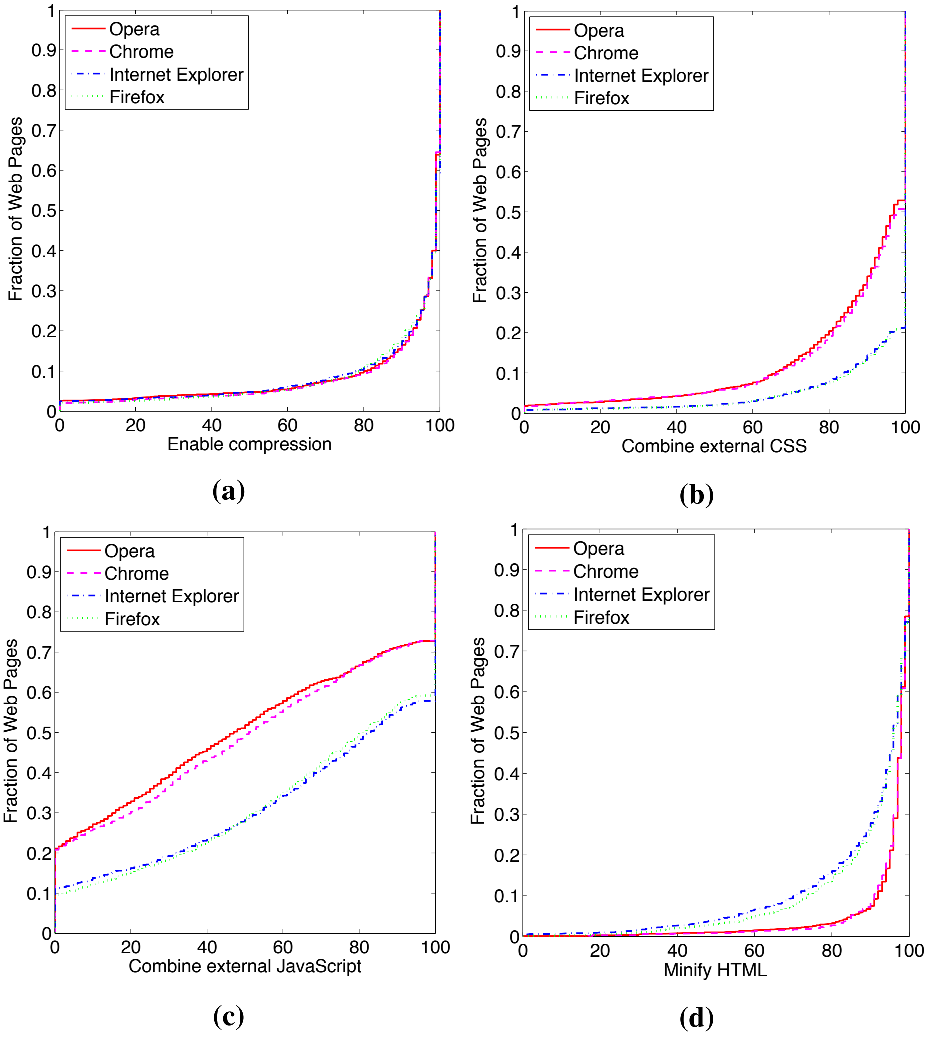
Figure 3.17: Differences in number of TCP connections established across browsers when the source HTML is the same.

3.3.3.6 Implications on Performance Analysis

Tools and metrics have been developed/used to evaluate the performance of web pages. One popular tool is Pagespeed [Google], a trace-driven tool for performance analysis developed by Google. Pagespeed uses a set of rules to determine whether web pages are designed using best-practices — some of these rules include "combine external CSS", "enable compression", "combine external Javascript" , "minify Javascript", and "minimize redirects". For each rule, Pagespeed outputs a score that ranges from 0 to 100, where higher scores are better, to denote how well a web page is designed according to that rule. Please note that Pagespeed does not use time-based metrics, such as RTT, to evaluate web page performance. Thus, Pagespeed scores are robust to the influence that vantage point and/or network conditions (e.g., congestion) may have on web page performance.

Our observation that the number of objects and bytes downloaded is different across browsers leads us to believe that Pagespeed will output results that may vary according to the type of browser that is used.[17] We use Pagespeed to investigate whether browser choice will impact the scores for some rules output from this tool for different web pages. Figure 3.18 plots the cumulative distribution for several scores that correspond to the rules that Pagespeed uses to evaluate web pages. Figure 3.18(a) shows that the score for the "enable compression" rule is similar across all browsers, whereas Figure 3.18(b) shows that the score for the "combine external CSS" rule follows different patterns for Chromium/Non-Chromium browsers that we observed before. In particular, Figure 3.18(b) shows that there are 30% more web pages that have a score of 100 for non-Chromium browsers than Chromium-based browsers. Figure 3.18(c) shows a similar result for the score of the "combine external Javascript" rule except that the difference in the percentage of web pages with a score of 100 is approximately 10% in this case. Figure 3.18(d) shows a cumulative distribution plot of the score for the "minify Javascript" rule. Contrary to other examples, Figure 3.18(d) shows that the of Pagespeed scores for a rule that generally performs worse for IE and Firefox browsers than Chromium-based browsers (i.e., the curves for the Chromium-based browsers are consistently below the curves for the Firefox and IE browsers). These results imply that browsers can influence the scores for multiple Pagespeed rules for different web pages to a differing degree. These results also highlight the importance of testing web pages on multiple browsers to validate that the web page behavior is intended and/or acceptable. We believe that it may be better to have Pagespeed, or a similar tool, evaluate the performance of a page by using only

[17] Browsers may impact the temporal properties of traffic.



(a)

(b)

(c)

(d)

Figure 3.18: Cumulative distribution plots for Pagespeed scores across browsers.

the objects that are "common" across browsers instead of including objects that are exclusively present for some browsers to facilitate a more apples-to-apples web page performance comparison across browsers.

3.3.4 Summary of observations

Browser type has a statistically significant impact on the majority of web page traffic features. The differences in the number of objects and bytes transmitted features are large and have implications on multiple domains including network utilization, privacy analysis, and web performance analysis. While we only focus on a few of the statistically significant traffic features that differ across browser in this section, we observe many other features that are different across browser (e.g., number of TCP RESET flags) — many of these differences are likely due to the fact that there is a different amount of data/objects that are transferred

TABLE 3.7: P-values for Notable Traffic Features.

Traffic Feature	P-value
Number of RESET segments	6.6×10^{-38}
Number of PUSH segments	3.2×10^{-51}
Number of FIN segments	3.3×10^{-218}
Number of Packets	1.2×10^{-162}
Number of servers contacted	9.2×10^{-45}
Number of TCP connections	7.9×10^{-39}
Number of Port 80 TCP connections	1.1×10^{-2}
Number of DNS requests	2.4×10^{-155}
Average RTT	7.4×10^{-28}
Average TCP connection interarrival time	9.19×10^{-57}
Average TCP connection duration	3.7×10^{-28}

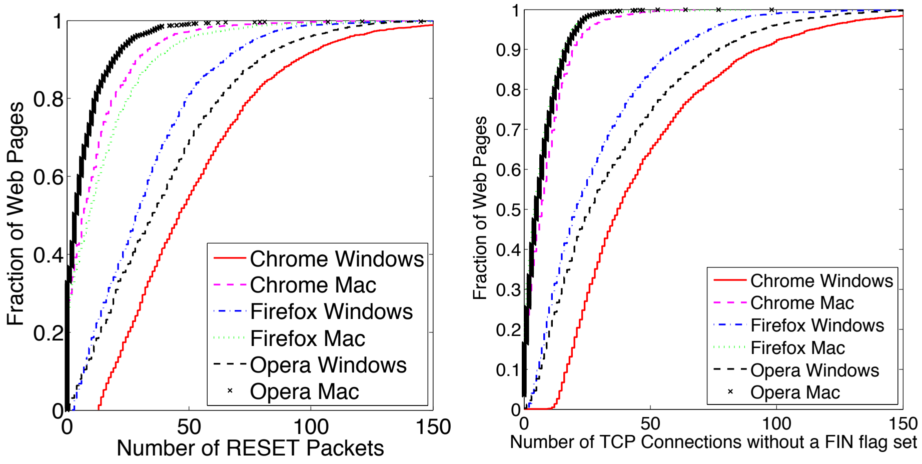
across browsers. For example, requesting a different number of objects that are from different servers will naturally lead to a different number of servers being contacted and a different number of TCP connections being established when loading a web page. We do not discuss such results in detail because it does not add to the value of our discussion — statistically significant p-values for the other primary traffic features are provided in Table 3.7. Our results show that differences in source HTML is not the only possible cause of the differences in traffic that we observe across browser platforms. *In fact, our results suggest that a large number of the differences observed across browsers are due to browser choice (e.g., aggressive prefetching, browser specific ads, TCP pre-connect, etc).* Please note that we cannot verify this claim without the source code of each browser.[18]

3.4 Impact of operating system on traffic

3.4.1 Does web page traffic differ across different operating systems?

We next investigate the impact that operating system may have on web page traffic features. We focus only on the Opera, Chrome, and Firefox browsers that are currently supported on both MacOSX 10.9.4 and Windows 7. We find that only two primary traffic features, the number of packets with the RESET flag set per web page and the number of packets with the FIN flag set per web page, differ significantly across operating systems. Figure 3.19(a) plots the cumulative distribution for the number of packets with the RESET flag feature. This plot shows that the RESET flag is set more often on Windows 7 platforms than

[18] While portions of Chromium-based browsers source code is available, we do not have access to the source code of proprietary browsers such as IE and Firefox.



(a) **(b)**

Figure 3.19: Cumulative distributions of differences observed across operating system.

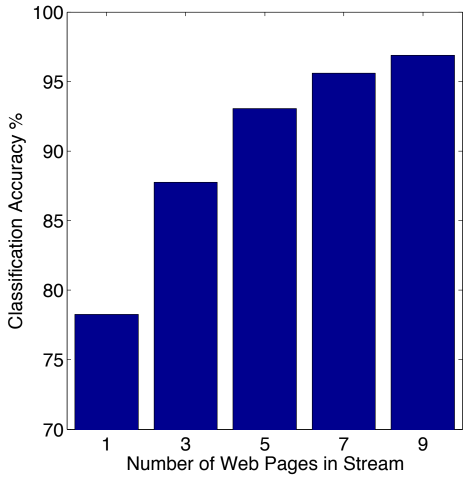
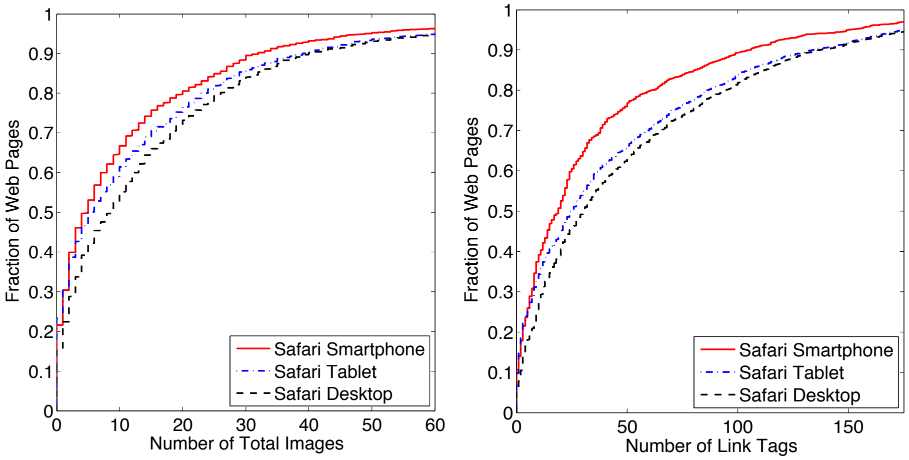


Figure 3.20: Operating system detection performance.

on MacOSX 10.9.4. Figure 3.19(b) plots the cumulative distribution for the number of TCP connections that do not include a single FIN flag per web page. This plot shows that this feature follows a similar pattern as the number of RESET flags feature. Given these observations, it is likely that MacOSX 10.9.4 closes TCP connections with the RESET flag at a higher rate than Windows 7. We do not find any statistically significant HTML source features that occur for the *same* browser across *different* operating systems. So, we are confident that these differences in traffic are, in fact, due to an operating system specific change.

Implication: OS Classification The number of RESET flags feature differs so much across operating systems (Windows 7 and MacOSX 10.9.4) that it may be used as a feature to identify operating systems via traffic analysis. We developed a simple classifier using a classification trees model to determine the efficacy of such an approach. We train our model on half of our dataset using the number of RESET flags per web page download as the only traffic feature. We test our model on the other half of the dataset using a streaming-based approach. This steaming-based approach randomly selects a stream of N web pages from the test set from a given operating system (Windows or MacOSX) and classifies each of the N pages using the trained model. The final label assigned for the stream of N web pages is the mode of the individual classifications of the N web pages. We believe that this classification approach is reasonable because users are likely to request more than one page for a single operating system. This streaming-based approach simply leverages the fact that multiple web pages can be used to help classify a host's operating system. The results of this classification is provided in Figure 3.20. Figure 3.20 shows that the operating system classification performance using a single web page sample can achieve an accuracy of 78%, whereas a 9 web page sample can achieve an accuracy of 97%. This result can be used in network security applications where the detection of malicious and/or vulnerable hosts commonly depends on operating system or browser [Yen et al., 2009].



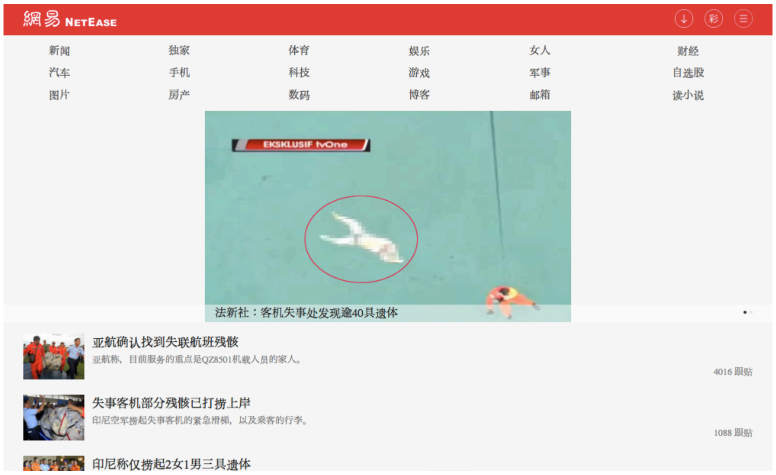
(a) **(b)**

Figure 3.21: Device type has a significant impact on HTML features.

3.5 Impact of Device on Web Page Features

3.5.1 Does source HTML differ across different devices?

We next study the impact that different devices have on web page downloads. We start by comparing the iOS 7 iPhone smartphone, iOS 7 iPad tablet, and the MacOSX 10.9.4 laptop where each device runs a



法新社：客机失事处发现逾40具遗体

亚航确认找到失联航班残骸

亚航称，目前服务的重点是QZ8501机载人员的家人。

4016 跟贴

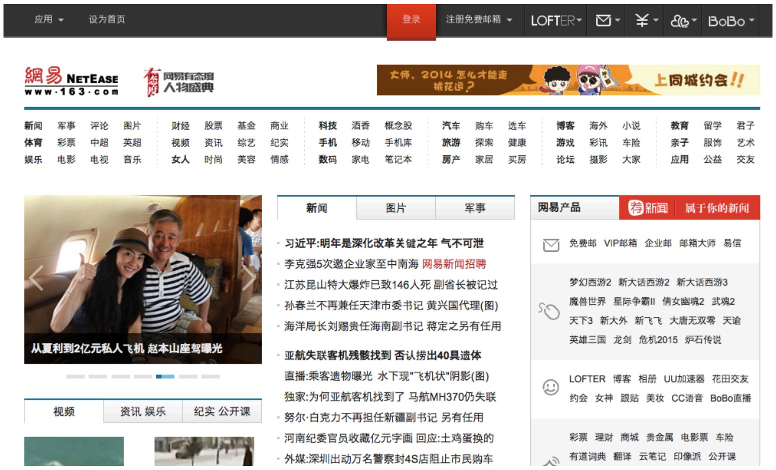
失事客机部分残骸已打捞上岸

印尼空军捞起失事客机的紧急滑梯，以及乘客的行李。

1088 跟贴

印尼称仅捞起2女1男三具遗体

Figure 3.22: Example mobile web page for 163.com.



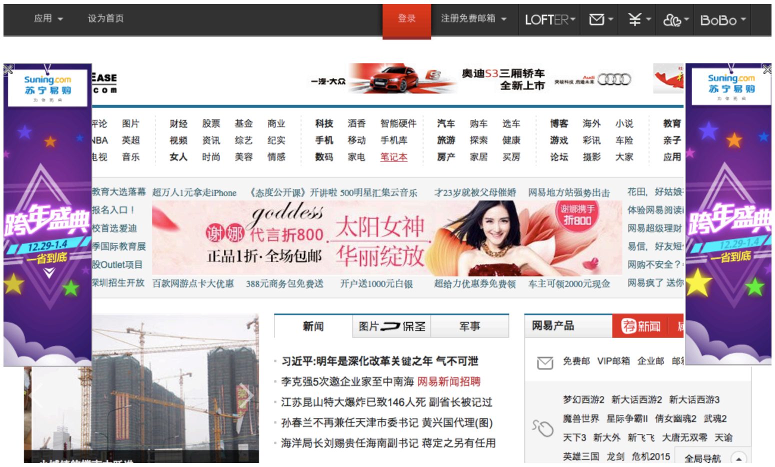
从夏利到2亿元私人飞机 赵本山座驾曝光

- 习近平:明年是深化改革关键之年 气不可泄
- 李克强5次邀企业家至中南海 网易新闻招聘
- 江苏昆山特大爆炸已致146人死 副省长被记过
- 孙春兰不再兼任天津市委书记 黄兴国代理(图)
- 海洋局长刘赐贵任海南副书记 蒋定之另有任用
- 亚航失联客机残骸找到 否认捞出40具遗体
- 直播:乘客遗物曝光 水下现"飞机状"阴影(图)
- 独家:为何亚航客机找到了 马航MH370仍失联
- 努尔·白克力不再担任新疆副书记 另有任用
- 河南纪委官员收藏亿元字画 回应:土鸡蛋换的
- 外媒:深圳出动万名警察封4S店阻止市民购车

Figure 3.23: Example tablet web page for 163.com.

version of Safari. The key results and implications of this analysis are summarized below:

1. *Device type has a statistically significant impact on web pages:* Pages designed for the small screens of mobile devices are likely to have simpler and smaller content. Thus, as can be expected, devices have a statistically significant impact on many features (66 total) by design intent — please refer to Appendix 4 for a full list of these features. The most prominent features that differ across phones, tablets, and laptops are embedded object-related features such as the number of images, scripts, and CSS references found in an HTML source, content-related features such as the total number of words present on a page, and the total number of links — all of these features have p-values that are on the order of 10^{-10} or less.



平论 图片 财经 股票 基金 商业 科技 酒香 智能硬件 汽车 购车 选车 博客 海外 小说 教育
NBA 英超 视频 资讯 综艺 纪实 手机 移动 手机库 旅游 探索 健康 游戏 彩讯 车险 亲子
电视 音乐 女人 时尚 美容 情感 数码 家电 笔记本 房产 家居 买房 论坛 摄影 大家 应用

教育大选落幕 超万人1元拿走iPhone 《态度公开课》开讲啦 500明星汇集云音乐 才23岁就被父母催婚 网易地方站强势出击 花田，好姑娘
报名入口！
校首选爱迪
季国际教育展
股Outlet项目
体验网易阅读
网易超级理财
易信，好友短
网购不安全？
深圳招生开放 百款网游点卡大优惠 388元商务包免费送 开户送1000元白银 超给力优惠券免费领 车主可领2000元现金 网易疯了 送你

- 习近平:明年是深化改革关键之年 气不可泄
- 李克强5次邀企业家至中南海 网易新闻招聘
- 江苏昆山特大爆炸已致146人死 副省长被记过
- 孙春兰不再兼任天津市委书记 黄兴国代理(图)
- 海洋局长刘赐贵任海南副书记 蒋定之另有任用

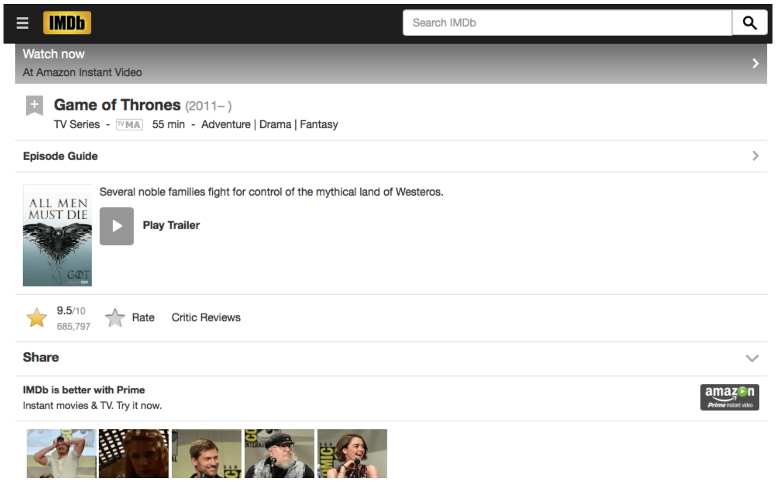
免费邮 VIP邮箱 企业邮 邮
梦幻西游2 新大话西游2 新大话西游3
魔兽世界 星际争霸II 倩女幽魂2 武魂2
天下3 新大外 新飞飞 大唐无双零 天谕
英雄三国 龙剑 危机2015

Figure 3.24: Example traditional web page for 163.com.

2. *Lack of consensus on the design of tablet-specific web pages:* Figure 3.21(a) shows the cumulative distributions of the number of images and Figure 3.21(b) shows the number of link tags stratified by device type. The cumulative distributions in these figures show that smartphone devices tend to exhibit the fewest number of images and link tags, while laptop devices tend to exhibit the largest. The cumulative distributions for the tablet devices tend to occur between the smartphone and laptop device distributions. In fact, the cumulative distributions for the tablet devices overlaps with the cumulative distributions of both the smartphone and laptop cumulative distributions at different ranges. This behavior of tablet devices is attributed to the lack of a consensus among content providers on the design of web pages for tablets. Content providers tend to either have (i) a unique web page design for each device type (e.g., 163.com shown in Figure 3.22, Figure 3.23, and Figure 3.24), (ii) similar web page designs for both laptop and tablet devices (e.g., imdb.com shown in Figure 3.25, Figure 3.26, and Figure 3.27), or (iii) similar web page designs for both tablet and smartphone devices (e.g., twitter.com shown in Figure 3.28, Figure 3.29, and Figure 3.30). We also find that different tablet manufacturers may receive different web pages. For example, android devices may receive ads to download android apps while iOS devices will receive ads to downloads apps on the Apple store. More interestingly, we find that the Amazon Fire Tablet will receive a smartphone version of a web page (espn.com) while the iPad Tablet will receive the desktop version of the page — this occurrence suggests that screen size can be a more important factor in determining which page is downloaded than simply referring to the device as a tablet or smartphone.

3. *Inconsistent redirect behavior that is based on device type across content providers:* We also find that there is a lack of consistency in the device-triggered redirect behavior across content providers. For example, some content providers will redirect mobile web page requests made by laptop clients to its corresponding laptop-based web page, while other content providers will not redirect requests in such a manner. This observed redirect behavior for devices is similar to the redirect behavior we observed for browser versions. This behavior can be problematic for some web-related applications. For instance, web crawlers may be redirected from the mobile view of a web page to the laptop view of a web page (in an undesired manner). This result has an impact for web page archival because undesirable, or less informative views of a page (mobile or desktop), may be archived instead of the desired page. This observation also raises concerns for information sharing across social media (e.g., search engines and social networking) because users can be referring to *different* views of information, or, at times, entirely different information altogether, via the *same* hyperlink. For example, if a user shares a link on a social media site, say Facebook.com, and a friend uses a different client platform to view it, the two users could be observing different content (especially comments and recommendations listed on a page). This can be particularly difficult if one user is referring to a particular comment or review on a page that is not immediately viewable by another user.

4. *Different search result sets for web search queries:* Device type is taken into account by web search engines such as bing.com and google.com when returning search results. We find that generally, smartphones tend to have more mobile optimized web pages included in a search result set than tablet and laptop devices — search providers take into account the mobile-friendliness of a web page when providing search results [Mob, b]. We also find that the search result set may have different semantics on different devices — this difference is mainly because search engines are increasingly providing web content to users instead of simply links to pages. For example, the search result set for the “nba standings” search query yields a different order of the basketball team rankings for a smartphone and a laptop (division rankings vs conference standings). This result further underscores the impact that device type can have on information sharing and other applications because a user may refer to portions of a page, say the rank of a basketball team, where a friend does not immediately see the same ranking that is being referenced.



IMDb Search IMDb

Watch now
At Amazon Instant Video

Game of Thrones (2011–)
TV Series - TV MA 55 min - Adventure | Drama | Fantasy

Episode Guide

Several noble families fight for control of the mythical land of Westeros.

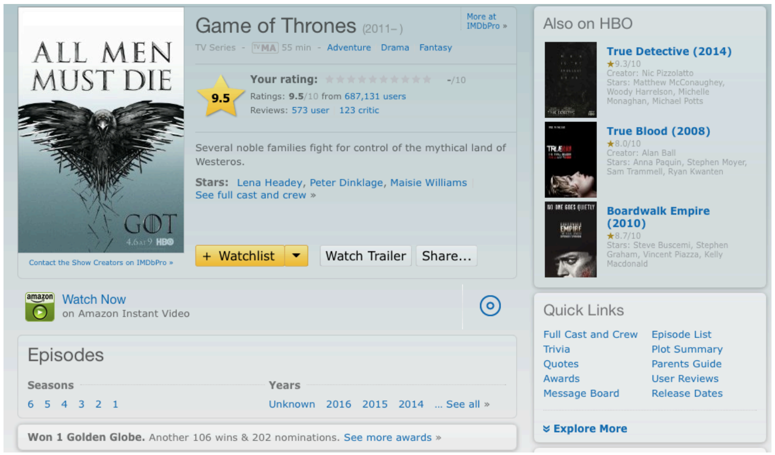
Play Trailer

9.5/10
685,797
Rate Critic Reviews

Share

IMDb is better with Prime
Instant movies & TV. Try it now.

Figure 3.25: Example mobile web page for imdb.com.



Game of Thrones (2011–)
TV Series - TV MA 55 min - Adventure Drama Fantasy

More at IMDbPro »

Your rating: -/10
9.5 Ratings: 9.5/10 from 687,131 users
Reviews: 573 user 123 critic

Several noble families fight for control of the mythical land of Westeros.

Stars: Lena Headey, Peter Dinklage, Maisie Williams
See full cast and crew »

Contact the Show Creators on IMDbPro »

+ Watchlist Watch Trailer Share...

Watch Now
on Amazon Instant Video

Episodes

Seasons
6 5 4 3 2 1

Years
Unknown 2016 2015 2014 ... See all »

Won 1 Golden Globe. Another 106 wins & 202 nominations. See more awards »

Also on HBO

True Detective (2014)
★9.3/10
Creator: Nic Pizzolatto
Stars: Matthew McConaughey, Woody Harrelson, Michelle Monaghan, Michael Potts

True Blood (2008)
★8.0/10
Creator: Alan Ball
Stars: Anna Paquin, Stephen Moyer, Sam Trammell, Ryan Kwanten

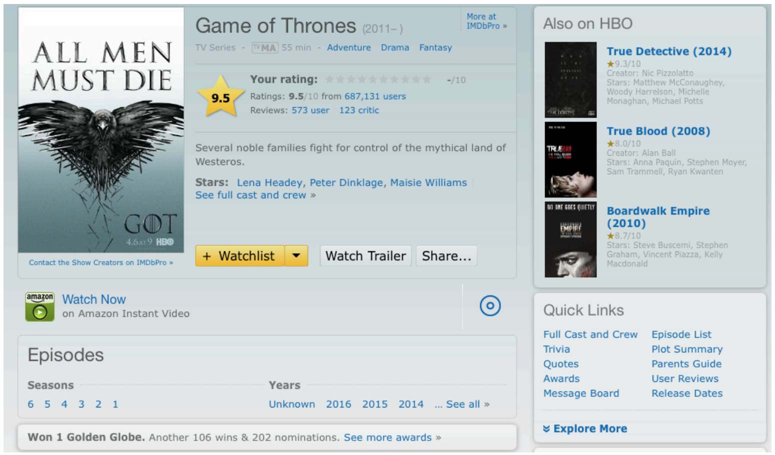
Boardwalk Empire (2010)
★8.7/10
Stars: Steve Buscemi, Stephen Graham, Vincent Piazza, Kelly Macdonald

Quick Links

Full Cast and Crew
Trivia
Quotes
Awards
Message Board
Episode List
Plot Summary
Parents Guide
User Reviews
Release Dates

Explore More

Figure 3.26: Example tablet web page for imdb.com.



Game of Thrones (2011–)
TV Series - TV MA 55 min - Adventure Drama Fantasy

More at IMDbPro »

Your rating: -/10
9.5 Ratings: 9.5/10 from 687,131 users
Reviews: 573 user 123 critic

Several noble families fight for control of the mythical land of Westeros.

Stars: Lena Headey, Peter Dinklage, Maisie Williams
See full cast and crew »

Contact the Show Creators on IMDbPro »

+ Watchlist Watch Trailer Share...

Watch Now
on Amazon Instant Video

Episodes

Seasons
6 5 4 3 2 1

Years
Unknown 2016 2015 2014 ... See all »

Won 1 Golden Globe. Another 106 wins & 202 nominations. See more awards »

Also on HBO

True Detective (2014)
★9.3/10
Creator: Nic Pizzolatto
Stars: Matthew McConaughey, Woody Harrelson, Michelle Monaghan, Michael Potts

True Blood (2008)
★8.0/10
Creator: Alan Ball
Stars: Anna Paquin, Stephen Moyer, Sam Trammell, Ryan Kwanten

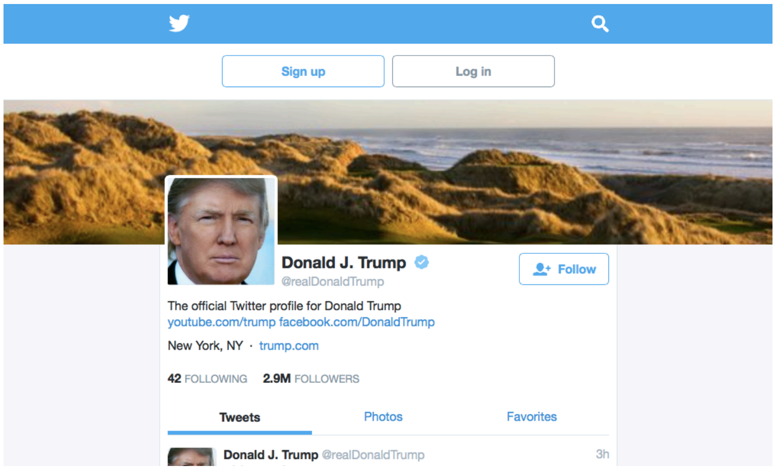
Boardwalk Empire (2010)
★8.7/10
Stars: Steve Buscemi, Stephen Graham, Vincent Piazza, Kelly Macdonald

Quick Links

Full Cast and Crew
Trivia
Quotes
Awards
Message Board
Episode List
Plot Summary
Parents Guide
User Reviews
Release Dates

Explore More

Figure 3.27: Example traditional web page for imdb.com.



Donald J. Trump
@realDonaldTrump

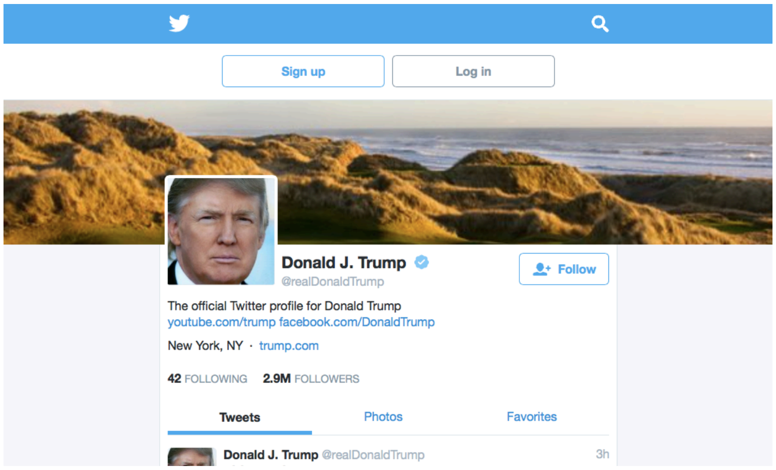
The official Twitter profile for Donald Trump
youtube.com/trump facebook.com/DonaldTrump

New York, NY · trump.com

42 FOLLOWING 2.9M FOLLOWERS

Donald J. Trump @realDonaldTrump 3h

Figure 3.28: Example mobile web page for twitter.com.



Donald J. Trump
@realDonaldTrump

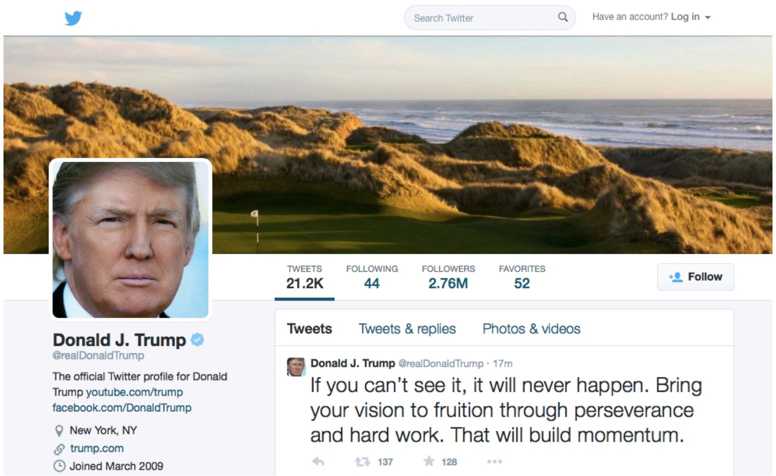
The official Twitter profile for Donald Trump
youtube.com/trump facebook.com/DonaldTrump

New York, NY · trump.com

42 FOLLOWING 2.9M FOLLOWERS

Donald J. Trump @realDonaldTrump 3h

Figure 3.29: Example tablet web page for twitter.com.



TWEETS 21.2K FOLLOWING 44 FOLLOWERS 2.76M FAVORITES 52

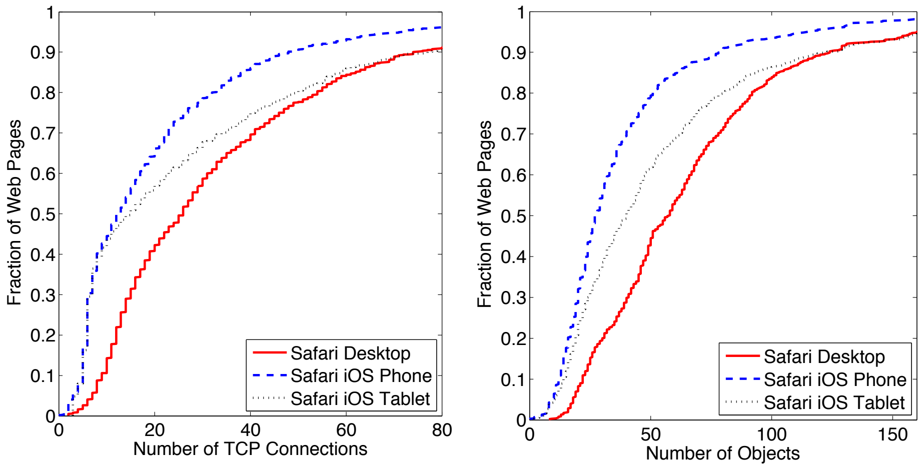
Donald J. Trump
@realDonaldTrump

The official Twitter profile for Donald Trump youtube.com/trump facebook.com/DonaldTrump

New York, NY
trump.com
Joined March 2009

Donald J. Trump @realDonaldTrump · 17m
If you can't see it, it will never happen. Bring your vision to fruition through perseverance and hard work. That will build momentum.
137 128

Figure 3.30: Example traditional web page for twitter.com.



(a) **(b)**

Figure 3.31: Cumulative distributions of differences observed across device (Safari browsers).

3.5.2 Do other traffic features differ across devices despite referencing the same source HTML?

We next focus our analysis on investigating whether device type has an impact on other traffic features. Our statistical analysis reveals that, similar to the browser analysis, most of the traffic features are significant across devices (541 out of 575 traffic features). Figure 3.31(a) plots a cumulative distribution of the number of TCP connections established per web page, while Figure 3.31(b) plots the number of objects observed per web page across different devices. These plots show that these traffic features differ in a largely expected manner. The mobile and desktop versions of Safari tend to exhibit the fewest and largest number of TCP connections and objects respectively — while the tablet version of Safari behaves in the middle, where it is similar to a mobile browser for some web pages, and is similar to the desktop browser for others. This result shows that there is a current lack of consensus among web designers on the properties of a web page that is designed for a tablet. We observe similar results for the different Chrome browsers measured on the different devices — these are shown in Figure 3.32.[19]

We next investigate whether these observed, and largely expected, differences in traffic features across devices are influenced by differences in the source HTML or not. Our statistical analysis reveals that most of the HTML-based features (66 of the 127 HTML-based features) are statistically significant — please note that for the browser analysis this difference was a modest 8 features. This result suggests that the root cause of the traffic differences we observe across devices is likely due to the differences in the source HTML.

[19] Though, the values are different because of browser-related differences that we previously discussed.

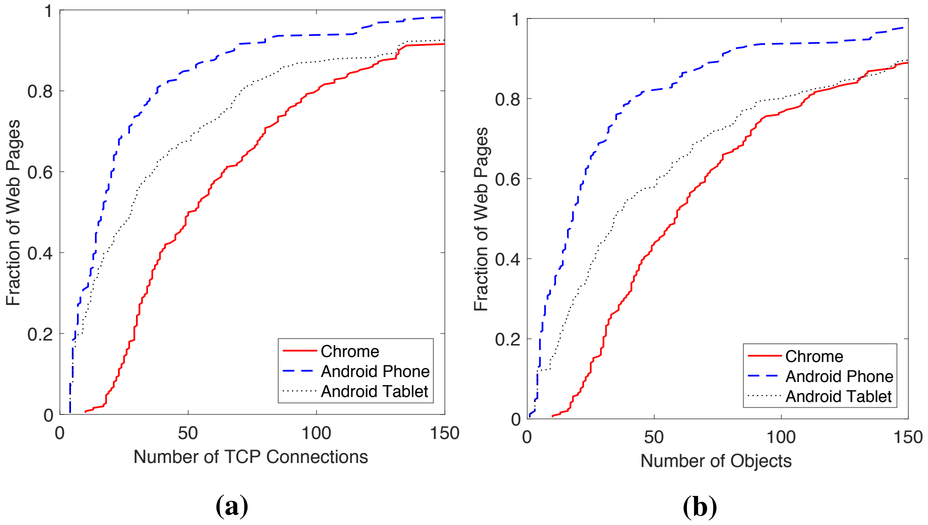
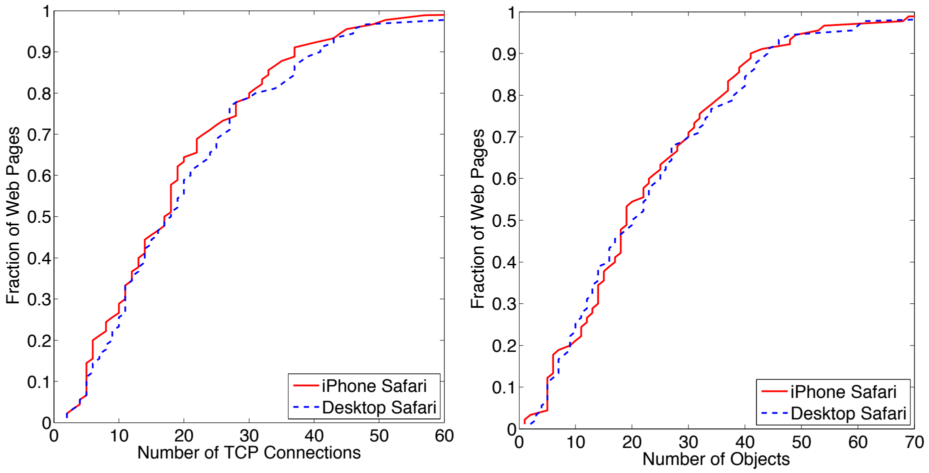


Figure 3.32: Cumulative distributions of differences observed across device (Chrome browsers).

The source HTML is different across devices because content providers typically redirect web page requests such that they are targeted for certain devices (e.g., laptops, smartphones, and tablets). We must compare the traffic generated by devices when the web pages are not redirected to confirm whether the root cause of the differences in traffic observed across different devices is due to source HTML differences.

Typically, servers will respond to requests by mobile devices for the full/desktop version of a web page by redirecting the request to a mobile optimized web page — this behavior does not work for our purposes. However, servers do not redirect requests by desktops for the mobile-optimized version of web pages to the full/desktop version of a page. Figure 3.33(a) plots the cumulative distribution for the number of TCP connections established per web page, while Figure 3.33(b) plots the cumulative distribution for the number of objects established be web page of when the same source HTML is used by different devices. This plot was generated by loading mobile optimized web pages using both the iOS iPhone Safari and the desktop Safari browsers — the 332 mobile optimized web pages used for this analysis are listed in Appendix 5. The number of TCP connections, shown in Figure 3.33(a), and the number of objects downloaded, shown in Figure 3.33(b), are similar across devices — our statistical analysis shows that these two features are statistically similar. All of the other traffic features are also statistically equivalent across devices in this scenario. This result shows that, while the traffic generated across devices are statistically significant by default (that is, without ensuring the same source HTML is downloaded), the *root cause of these statistically significant differences is the source HTML when considering mobile optimized web pages*. Though, we acknowledge that it is indeed possible that devices can influence traffic in a statistically significant manner



(a) **(b)**

Figure 3.33: Cumulative distributions of similarity of mobile web page traffic across different devices.

when the HTML is the same or even different when considering other types of web pages (say full/desktop version of a page) — we cannot verify this with our current methodology because mobile browsers tend to get redirected when requesting the full/desktop version of a web page.

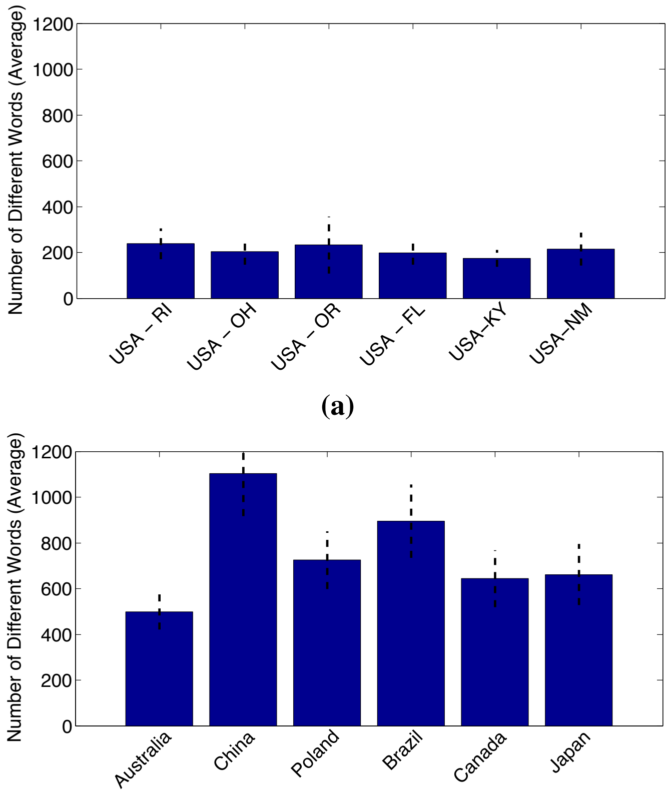
3.5.3 Summary and Implications of Results

The results in this section show that mobile devices have a significant impact on source HTML and web page traffic. However, we did not find any strong evidence that device-specific enhancements are the root cause of the differences in traffic that we observe. These results suggest that web page designers assume the majority of the responsibility for designing efficient mobile web pages. Thus, there is plenty of room for research into performing device or browser-based optimizations for mobile platforms. These results also show that web measurement studies must specifically account for mobile web pages because they exhibit traffic characteristics that is substantially different from traditional web pages — we explicitly do this in Chapters 4 and 5.

3.6 Impact of Location on Web Page Features

3.6.1 Does source HTML differ across vantage point?

We next discuss the impact of vantage point on base HTML pages. Our data includes web pages requested from 13 vantage points using the planetLab system across different continents including Asia, North America, South America, Europe, and Australia — please refer to Table 3.3 for a full list of the vantage points studied in this Section [Chun et al., 2003]. The web pages considered for this analysis are the 3614 web pages that were used in Sections 3.3 and 3.4, which are provided in Appendix 2. These web pages were downloaded using Firefox v 17.1 on Linux systems that are based on Fedora Core 8 [Chun et al., 2003] — these systems are outdated as compared to the systems used in the previous section, which limits our ability to compare our vantage point results across other sections. We perform Kruskal-Wallis tests on each feature where each group in this test corresponds to a single vantage point. Thus, there are 13 groups for the Kruskal-Wallis tests used in this analysis. We find that:



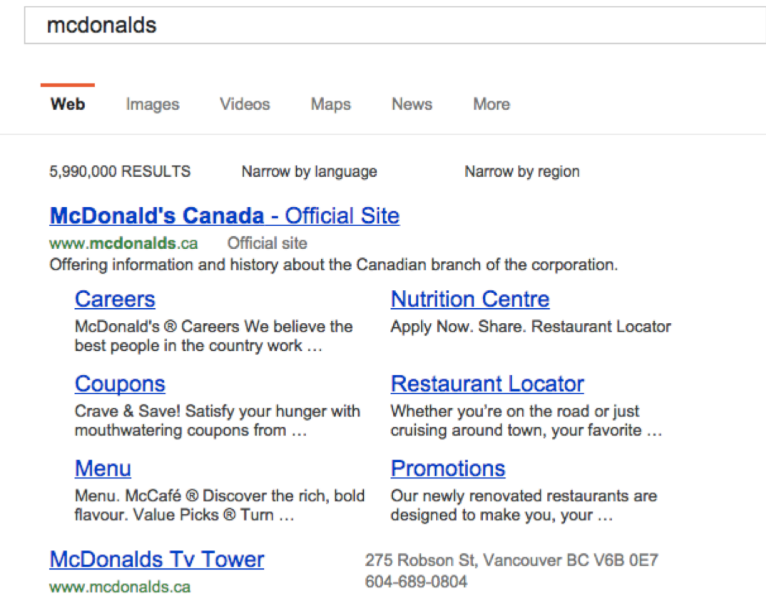
(a)

(b)

Figure 3.34: Impact of vantage point on number of different words (U.S.(a) and World(b)). The baseline for comparison located in California.

1. Our statistical analysis shows that *none* of the HTML tag-based features are significantly impacted by vantage point. This result shows that web page design and formatting is not significantly influenced by location. This result is surprising given cultural preferences in content layout and appearance.

2. Vantage point has a significant impact on the number of different words feature. This feature is com-



mcdonalds

5,990,000 RESULTS

McDonald's Canada - Official Site
www.mcdonalds.ca Official site
Offering information and history about the Canadian branch of the corporation.

Careers
McDonald's ® Careers We believe the best people in the country work …

Nutrition Centre
Apply Now. Share. Restaurant Locator

Coupons
Crave & Save! Satisfy your hunger with mouthwatering coupons from …

Restaurant Locator
Whether you're on the road or just cruising around town, your favorite …

Menu
Menu. McCafé ® Discover the rich, bold flavour. Value Picks ® Turn …

Promotions
Our newly renovated restaurants are designed to make you, your …

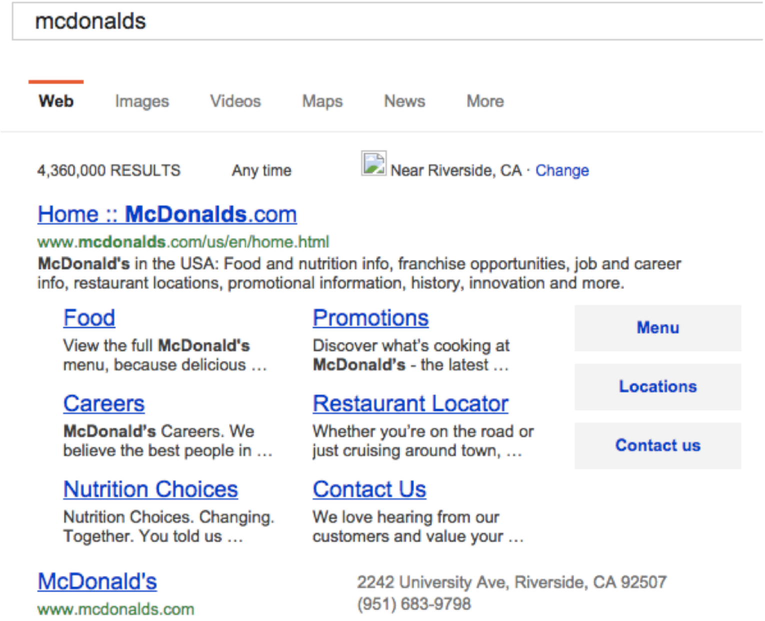
McDonalds Tv Tower
www.mcdonalds.ca
275 Robson St, Vancouver BC V6B 0E7
604-689-0804

Figure 3.35: Example search result of a bing.com web page that was targeted for an audience in Vancouver, Canada.

puted as the number of words that are different between two documents (that is, a baseline document and a test document). For this analysis, the baseline documents are web pages that were downloaded from the planetLab node in California (planet-lab1.cs.ucr.edu) while the test documents are web pages from the other nodes. Each pair of baseline and test documents correspond to web pages that were requested using the same URL. The average number of different words metric for a node is the mean of the number of different words computed for all test documents (or web pages) for that node. Figure 3.34 (a) shows that the average number of *different* words for each vantage point in the U.S is roughly 200-250 words, while Figure 3.34 (b) shows that the average number of different words for each vantage point that are outside of the U.S is over 500 — Figure 3.34 (a) and Figure 3.34 (b) both include 95% confidence interval bars around the average.[20] Many of these differences across all vantage points correspond to differences in search result sets that depend on location, while a few of these differences correspond to specialized versions of web pages that occur outside the United States. We discuss these below:

- Differences in search results: Bing search results, whether it is Web, news, or image search, may yield different links, ads, and images across different vantage points — please note that

[20] Please note that while we study the top 250 web sites in the world, many of these sites are served by content providers that are in the U.S.



mcdonalds

4,360,000 RESULTS Any time Near Riverside, CA · Change

Home :: McDonalds.com
www.mcdonalds.com/us/en/home.html
McDonald's in the USA: Food and nutrition info, franchise opportunities, job and career info, restaurant locations, promotional information, history, innovation and more.

Food
View the full McDonald's menu, because delicious …

Promotions
Discover what's cooking at McDonald's - the latest …

Careers
McDonald's Careers. We believe the best people in …

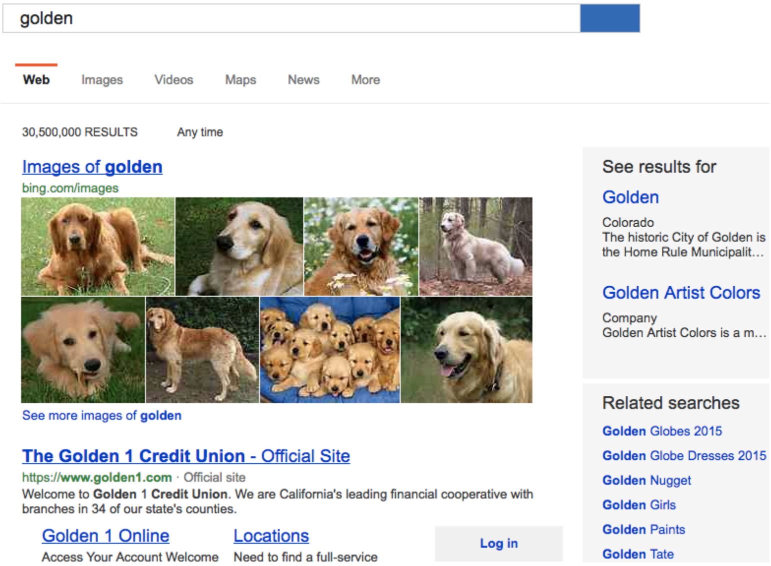
Restaurant Locator
Whether you're on the road or just cruising around town, …

Nutrition Choices
Nutrition Choices. Changing. Together. You told us …

Contact Us
We love hearing from our customers and value your …

McDonald's
www.mcdonalds.com
2242 University Ave, Riverside, CA 92507
(951) 683-9798

Figure 3.36: Example search result of a bing.com web page that was targeted for an audience in California, USA.



golden

30,500,000 RESULTS Any time

Images of golden
bing.com/images
See more images of golden

The Golden 1 Credit Union - Official Site
https://www.golden1.com · Official site
Welcome to Golden 1 Credit Union. We are California's leading financial cooperative with branches in 34 of our state's counties.

Golden 1 Online
Access Your Account Welcome

Locations
Need to find a full-service

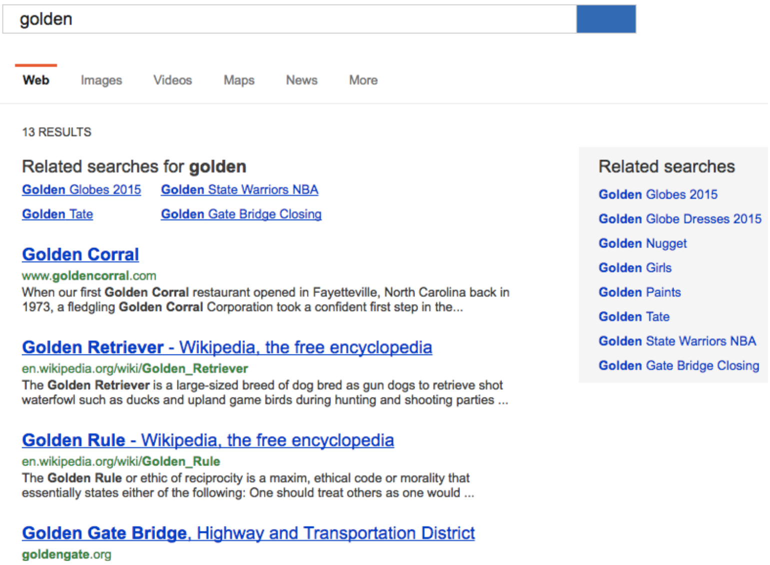
See results for

Golden
Colorado
The historic City of Golden is the Home Rule Municipalit…

Golden Artist Colors
Company
Golden Artist Colors is a m…

Related searches
Golden Globes 2015
Golden Globe Dresses 2015
Golden Nugget
Golden Girls
Golden Paints
Golden Tate

Figure 3.37: Example search result of a bing.com web page that was targeted for an audience in the CA-USA.



13 RESULTS

Related searches for **golden**

Golden Globes 2015 | **Golden** State Warriors NBA
Golden Tate | **Golden** Gate Bridge Closing

Golden Corral
www.**goldencorral**.com
When our first **Golden Corral** restaurant opened in Fayetteville, North Carolina back in 1973, a fledgling **Golden Corral** Corporation took a confident first step in the...

Golden Retriever - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Golden_Retriever**
The **Golden Retriever** is a large-sized breed of dog bred as gun dogs to retrieve shot waterfowl such as ducks and upland game birds during hunting and shooting parties ...

Golden Rule - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Golden_Rule**
The **Golden Rule** or ethic of reciprocity is a maxim, ethical code or morality that essentially states either of the following: One should treat others as one would ...

Golden Gate Bridge, Highway and Transportation District
goldengate.org

Related searches
Golden Globes 2015
Golden Globe Dresses 2015
Golden Nugget
Golden Girls
Golden Paints
Golden Tate
Golden State Warriors NBA
Golden Gate Bridge Closing

Figure 3.38: Example search result of a bing.com web page that was targeted for an audience in RI-USA.

we verified that this is not primarily a consequence of time.[21] Some of these differences are obvious due to location-based searches, say when a user is searching for McDonalds and the search engine returns the address of the nearest McDonalds — most of these differences occur even when the search result is made in the same country. Examples of such search results for a "McDonalds" query made in Vancouver, Canada and California, USA are provided in Figure 3.35 and Figure 3.36 respectively. Other differences are more complex, such as when more generic and random search queries such as "a hello berry" and "golden" yield different search results — example search results for the query "golden" for the vantage point in California and Rhode Island are shown in Figure 3.37 and Figure 3.38 respectively. The impact of vantage point on search results is important to note because search engines are a primary tool for various applications including web page scraping [Jacob et al., 2012] and web security [Leontiadis et al., 2011]. Vantage point driven search results also impact users because location can be misleading for users who access the Web via 3G or 4G services — thus, the wrong location can be used to target search results.

- International versions of web pages: We observe two cases where the landing page of a web site is different when requested outside of the United States. We believe that the larger difference

[21] We discuss results pertaining to bing.com because other search engines such as google.com are blocked in some countries.

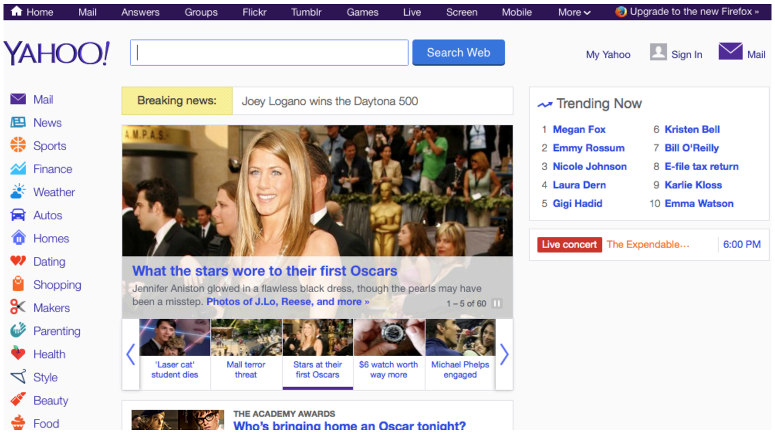
between the vantage points around the world and the United States is mainly due to content providers having international versions of content that is likely to be of interest to the local population. Figure 3.39 and Figure 3.40 show examples of yahoo.com web pages as rendered in different countries — the United States and Australia in this case. A similar example for cnn.com where the United States version of the web page is shown in Figure 3.41 and the international version, as observed from China, is shown in Figure 3.42. However, in this example we find that the United States version of cnn.com, shown in Figure 3.41, does not explicitly notify the user that their browsing behavior is being tracked while the international version, shown in Figure 3.42, does — the international version even informs the user about (i) how they are being tracked, (ii) why they are being tracked, and (iii) explicitly asks users to agree to being tracked. These differences are likely due to differences in privacy laws across different countries [Pri]. In particular, privacy laws in the United States do not require content providers to explicitly notify users that they are being tracked while privacy laws in many countries outside of the United States do [Pri]. We do find other web sites that have additional privacy-related information when accessed outside of the United States including foxnews.com and twitter.com.[22] This observation suggests that content providers will not be transparent with users on how their browsing behavior is being tracked and used unless privacy laws are in place that require them to do so.

3.6.2 Do other traffic features differ across vantage point?

We next investigate the impact that vantage point has on web page traffic. We find that temporal features, including RTT, TCP connection interarrival time, and TCP connection duration, are heavily influenced by vantage point, while only one non-temporal traffic feature, the number of DNS requests, is influenced by vantage point in a statistically significant manner. We discuss these in detail below:

- Temporal features are significantly influenced by vantage point: Vantage point has the biggest impact on the temporal properties web page traffic — these include the RTT, TCP connection arrival time, and TCP connection duration features discussed in Section 3.2. The average RTT metric for a given web page corresponds to the mean of the RTTs measured for each TCP connection established during each web page download. Figure 3.43 shows that the average RTT across different vantage points is

[22] We do not have an exact number of how often this behavior occurs in our dataset, due to the difficulty in extracting such information from source HTML.



Breaking news: Joey Logano wins the Daytona 500

What the stars wore to their first Oscars

Jennifer Aniston glowed in a flawless black dress, though the pearls may have been a misstep. Photos of J.Lo, Reese, and more »

1 – 5 of 60

'Laser cat' student dies | Mall terror threat | Stars at their first Oscars | $6 watch worth way more | Michael Phelps engaged

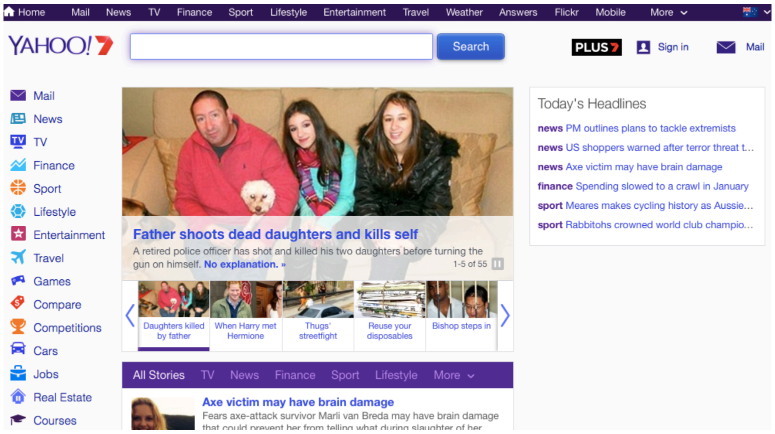
THE ACADEMY AWARDS
Who's bringing home an Oscar tonight?

Trending Now

1	Megan Fox	6	Kristen Bell
2	Emmy Rossum	7	Bill O'Reilly
3	Nicole Johnson	8	E-file tax return
4	Laura Dern	9	Karlie Kloss
5	Gigi Hadid	10	Emma Watson

Live concert The Expendable... 6:00 PM

Figure 3.39: Example of a yahoo.com web page that was targeted for an audience in the United States.



Father shoots dead daughters and kills self

A retired police officer has shot and killed his two daughters before turning the gun on himself. No explanation. »

1-5 of 55

Daughters killed by father | When Harry met Hermione | Thugs' streetfight | Reuse your disposables | Bishop steps in

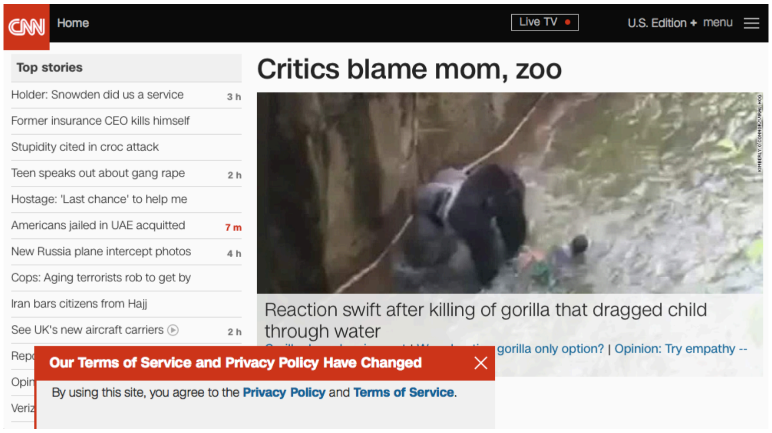
All Stories TV News Finance Sport Lifestyle More

Axe victim may have brain damage
Fears axe-attack survivor Marli van Breda may have brain damage that could prevent her from telling what during slaughter of her

Today's Headlines

news PM outlines plans to tackle extremists
news US shoppers warned after terror threat t...
news Axe victim may have brain damage
finance Spending slowed to a crawl in January
sport Meares makes cycling history as Aussie...
sport Rabbitohs crowned world club champio...

Figure 3.40: Example of a yahoo.com web page that was targeted for an audience in Australia.



Top stories

Holder: Snowden did us a service	3 h
Former insurance CEO kills himself	
Stupidity cited in croc attack	
Teen speaks out about gang rape	2 h
Hostage: 'Last chance' to help me	
Americans jailed in UAE acquitted	7 m
New Russia plane intercept photos	4 h
Cops: Aging terrorists rob to get by	
Iran bars citizens from Hajj	
See UK's new aircraft carriers	2 h
Rep	
Opin	
Veriz	

Critics blame mom, zoo

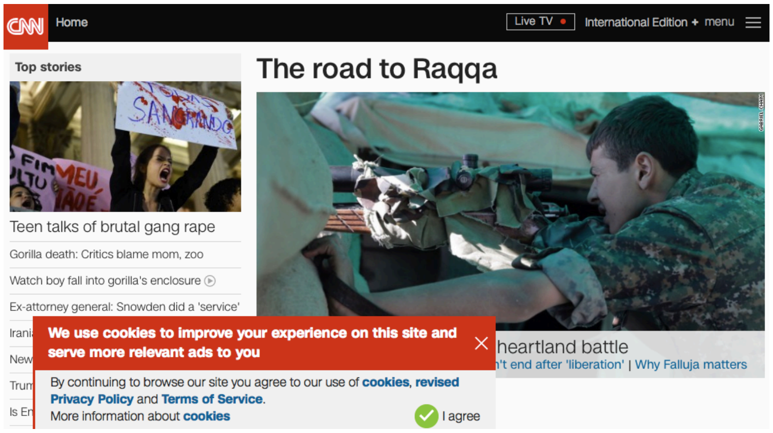
Reaction swift after killing of gorilla that dragged child through water

gorilla only option? | Opinion: Try empathy --

Our Terms of Service and Privacy Policy Have Changed

By using this site, you agree to the Privacy Policy and Terms of Service.

Figure 3.41: Example of a cnn.com web page that was targeted for an audience in the United States.



CNN Home

Live TV

International Edition + menu

Top stories

Teen talks of brutal gang rape

Gorilla death: Critics blame mom, zoo

Watch boy fall into gorilla's enclosure

Ex-attorney general: Snowden did a 'service'

Irani

New

Trun

Is En

The road to Raqqa

heartland battle

n't end after 'liberation' | Why Falluja matters

We use cookies to improve your experience on this site and serve more relevant ads to you

By continuing to browse our site you agree to our use of cookies, revised Privacy Policy and Terms of Service.

More information about cookies

I agree

Figure 3.42: Example of a cnn.com web page that was targeted for an international audience.

significantly impacted by vantage point — this figure also shows 5%-95% percentile bars for each vantage point. This impact is most significant for vantage points that are located outside of North America including Australia, China, Brazil, Japan, and to a lesser extent Poland. We believe that the average RTT is higher for vantage points that are outside of the United States because the most popular web sites in the world are largely hosted in North America [Inc.].

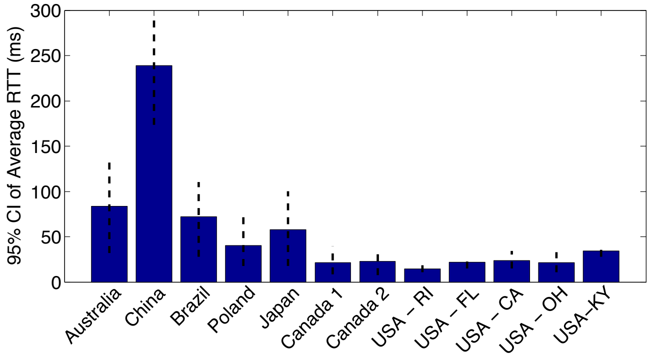


Figure 3.43: Average RTT observed across vantage point with 5%-95% percentile bars.

Figure 3.43 shows that there is also an *unusually* high RTT for the vantage point in China. More specifically, we find that while China and Japan are both in a similar region in the world, the average RTT of China is over 4x higher than the RTT in Japan. We also observe that many of the hostnames that are observed at other vantage points are not observed in China — we discuss more details about this later. These hostnames, including google.com and nytimes.com, are missing in China because network services are used in China that block web content from certain servers [Chi]. We believe that these network services within China perform extra processing that make the RTT appear to be larger than it would be otherwise. We observe similar differences in the average RTT measurements, though to a lesser extreme, across vantage points within the United States as well. Overall, our results show that the average RTT observed across different vantage points are likely influenced by delays incurred by middle-boxes within a network in addition to the distance between end-hosts. This observation is important because it shows that moving servers closer to end-users may not improve performance as much as expected because other factors may serve as a bottleneck in network latency.

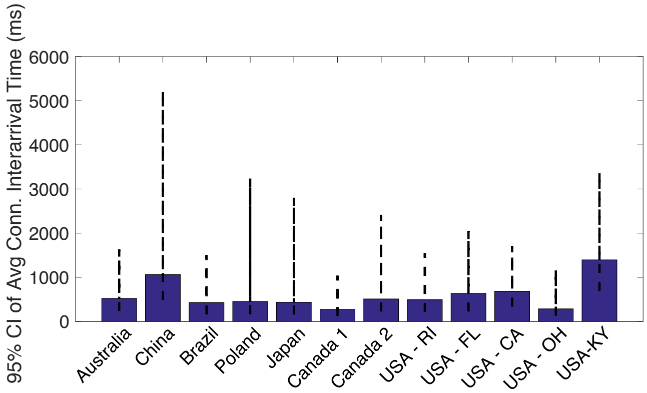


Figure 3.44: Average TCP connection arrival time observed across vantage point with 5%-95% percentile bars.

We also compute metrics for the average TCP connection interarrival time and TCP connection duration per web page download. The average TCP connection interarrival time metric for a given web page corresponds to the mean of the TCP connection interarrival times measured during each web page download, while the TCP connection duration is computed in a similar manner. These metrics are shown in Figure 3.44 and Figure 3.45 respectively. However, the results shown in these figures are difficult to explain. For instance, in Figure 3.44, we find that the average TCP connection interarrival time is moderately higher for the China and Kentucky vantage points as compared to the others. In Figure 3.45, we find that the average duration of TCP connections tends to be slightly lower for vantage points that are outside the United States. While we know that these features are influenced by RTT and that networking conditions may differ across vantage points [Zaki et al., 2014], we do not have a more specific explanation for the behavior that we observe in these figures.

- Most non-temporal features are mildly influenced by vantage point: Our statistical analysis shows that all but one of the non-temporal traffic features are *not* impacted by vantage point in a statistically significant manner. Figure 3.46 shows a bar plot of the average number of objects transmitted with

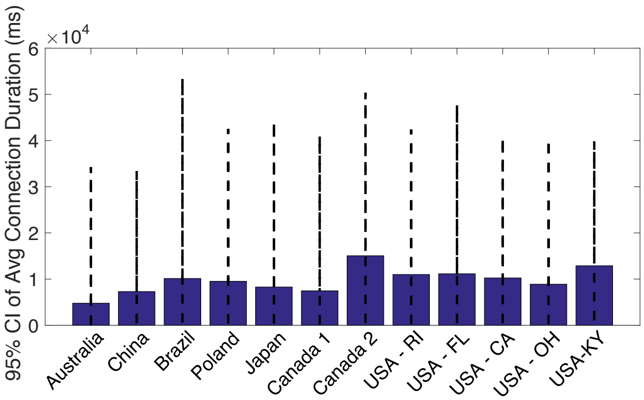


Figure 3.45: Average TCP connection duration observed across vantage point with 5%-95% percentile bars.

5%-95% bars when loading a web page across different vantage points. Our statistical analysis and Figure 3.46 shows that the number of objects transmitted when loading a web page do not vary in a statistically significant manner across vantage points.[23] While the number objects transmitted is not statistically significant, there are some differences in the objects that are downloaded across vantage point that are worth explaining. For instance, the vantage point located in China has fewer objects as compared to other sites — this result occurs because many of these hostnames are blocked in China including google.com, nytimes.com, and phcdn.com [Chi].

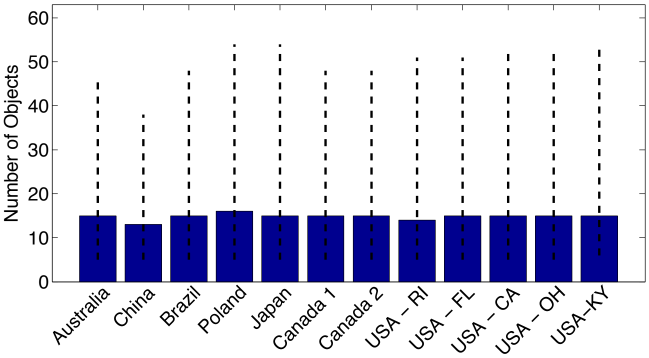


Figure 3.46: Average number of objects observed across vantage point with 5%-95% percentile bars.

We also observe objects that are unique to a vantage point. For example, some hostnames that are only present at the China vantage point include cn.bing.com, static.googleadsserving.cn, admaster.com.cn, and image.skype.gmw.cn, whereas some hostnames that are only observed in Japan in-

[23] Though the number of objects observed is lower than in prior plots. PlanetLab nodes have an older version of Firefox installed on nodes. We find that browser version impacts the number of objects downloaded because of (i) outdated javascript engines and (ii) some hosts such as yahoo.com will serve a mobile-optimized page to clients with outdated clients instead of a desktop page

clude cs.genieessp.jp, overseas.weibo.com, and mmtest2.wechatos.net. Similar results occur for other vantage points including Australia, Brazil, and Canada. In general, the unique hostnames that we observe for individual vantage points only account for a dozen or so objects in our entire dataset — so unique objects are rare. There are also instances where we visit a web page and only a select few countries make certain requests. For example, when we visit a product page on ebay.com, we observe objects from ads.rubiconproject.com only for the clients in Japan and China. However, hostnames that generally occur outside the United States (i.e., objects that are exclusively present in multiple countries outside the USA) are much more common — we observe 448 such hostnames in our dataset. Some of these hostnames include edition.cnn.com (the international version of CNN), adtech.de, ins.traffic.com, extended.dmtracker.com, and user.lucidmedia.com.

- Number of DNS requests varies greatly across vantage point: Despite the fact that the number of objects requested on each vantage point are statistically similar, the number of DNS requests made are not. In fact, the number of DNS requests made across vantage point is the only statistically significant non-temporal traffic feature observed in our vantage point analysis (p-value = 2.5^{-15}). Figure 3.47 shows a bar plot of the average number of DNS requests made when loading a web page for each vantage point with 5%-95% percentile bars. We investigate the DNS data to determine whether different DNS requests are being made at each vantage point. We find that hosts in different planetLab nodes, not necessarily different countries, send unique DNS requests to resolve hostnames within their respective local network. These DNS requests to local servers may be duplicated to a different degree at different planetLab nodes — for example, the host in Canada (Canada 1 in Figure 3.47) sends 4 DNS requests to resolve cs-planetlab4.cs.surrey.sfu.ca, while the host in Florida sends a single DNS request to resolve planetlab2.acis.ufl.edu (USA-FL in Figure 3.47). When excluding these duplicate DNS requests, all other DNS requests made are largely the same across vantage points— please note the mean difference of 3 DNS requests shown in Figure 3.47 for this example. We do not find any obvious explanation for this duplication of DNS requests to resolve the hostnames of local servers given that (i) the responses to requests are received before duplicate requests are made, (ii) the responses provide the same information, and (iii) the TTL is set to large enough values such that duplicate requests are not warranted (greater than 1 hour). Thus, we suspect that a planetLab-related quirk is causing these duplicate DNS requests and are unsure that this result is generalizable to non-PlanetLab vantage points.

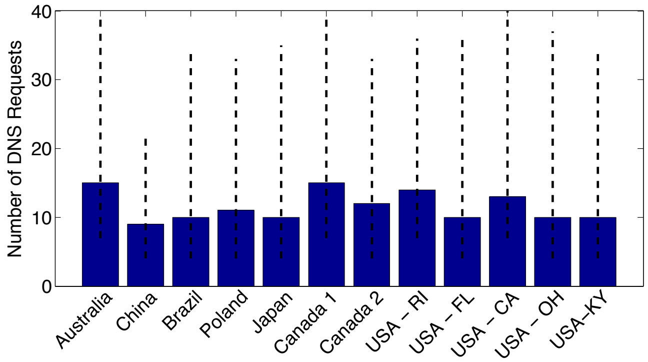


Figure 3.47: Average number of DNS requests observed across vantage point with 5%-95% percentile bars.

3.6.3 Summary of Observations

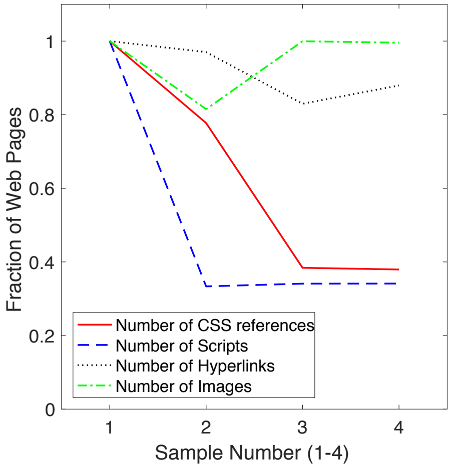
The key findings of our vantage point analysis are the following:

- *Location impacts the content of source HTML:* We observe that many web pages, particularly search results and landing pages, have content that is customized for different locations. However, tag-based source HTML features are not impacted by vantage point in a statistically significant manner. Thus, while content is customized to different user populations, the general format of web pages are the *same* irrespective of location. This result shows that there is plenty of room in the market for content providers to tailor web page templates to the different preferences that different user populations may have.

- *Location impacts temporal traffic features:* As expected, location has a substantial impact on the temporal features of traffic. This result emphasizes that temporal traffic features should be avoided when developing measurement analysis techniques such as traffic classification [Lim et al., 2010] to maximize the chance of scaling to multiple environments. There may also be some quirks in DNS requests across planetLab hosts on different networks that can also impact the performance of measurement analysis techniques that use DNS data.

- *Object requests are largely similar for the same browsers:* The majority of the objects that are requested for a given browser is largely the same when the location is different. Though, in rare cases the location of the client can influence a small number of unique objects that are requested.

3.7 Impact of Time on Web Page Traffic Features

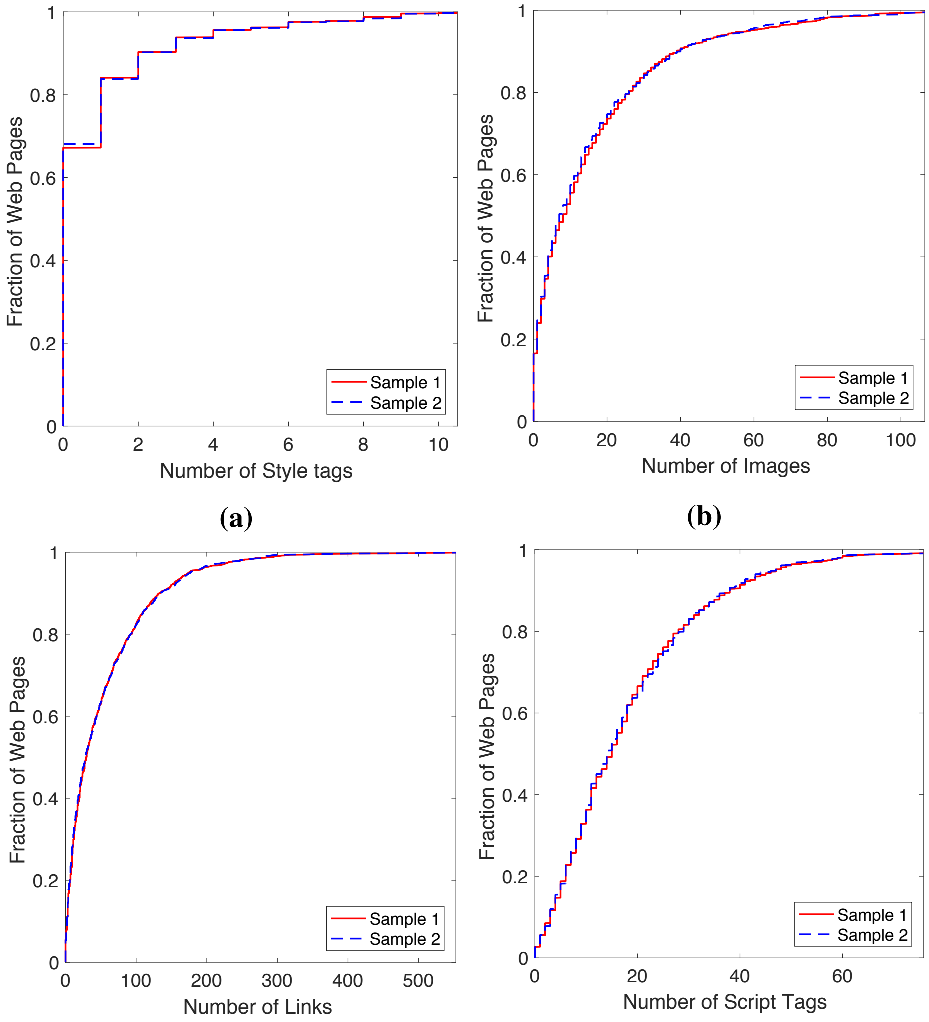


(a)

Figure 3.48: P-values for numerous HTML-based features between 4 samples which are ≈ 1 month apart each (comparison across time).

3.7.1 Does time influence our HTML-based results?

We investigate the impact that time has on base HTML source files. We perform many univariate Kruskal-Wallis tests between our first measurement (i.e., baseline measurement) and four subsequent measurements that was taken at 1 month increments. Each sample/measurement includes a single web page measurement for each client platform measured at that instance in time — that is, 4 browsers × 2 operating systems × 3614 web pages + 13 vantage points × 3614 + 6 devices/browsers × 1010. A plot which shows the differences observed across time for some of the HTML tag-based features is provided in Figure 3.48. Here the y-axis corresponds to the p-value of the difference between the baseline measurement (Sample 1) and the subsequent sample measurements, while the x-axis corresponds to the sample/measurement number — please note that the p-value for Sample 1 is one since it is a comparison to itself. Figure 3.48 shows that all of the HTML tag-based features shown have p-values that are higher than .05 for each sample measurement— in fact, we observe this for all HTML tag-based features. Thus, time does not have an impact on the HTML tag-based differences that we discussed earlier (specifically, across browser and device). To illustrate this, cumulative distribution plots for some of the HTML-based features measured across client platforms across the first two samples is provided in Figure 3.49. This figure shows that the cumulative distribution for many of the HTML-based features, including the number of images, number of script tags,



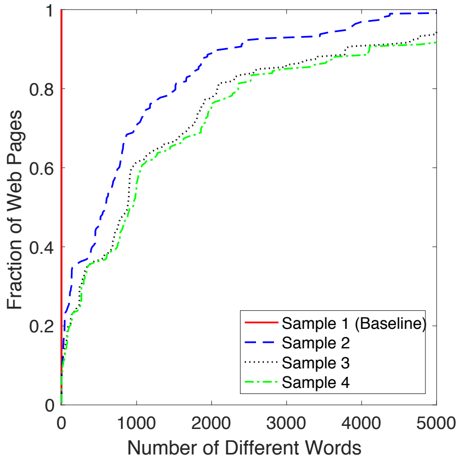
(a) (b)

(c) (d)

Figure 3.49: Cumulative distributions showing that time does not significantly impact many HTML-based features.

number of style tags, and the number of hyperlinks, overlap — that is, the distribution of these HTML-based features do not differ significantly over time.

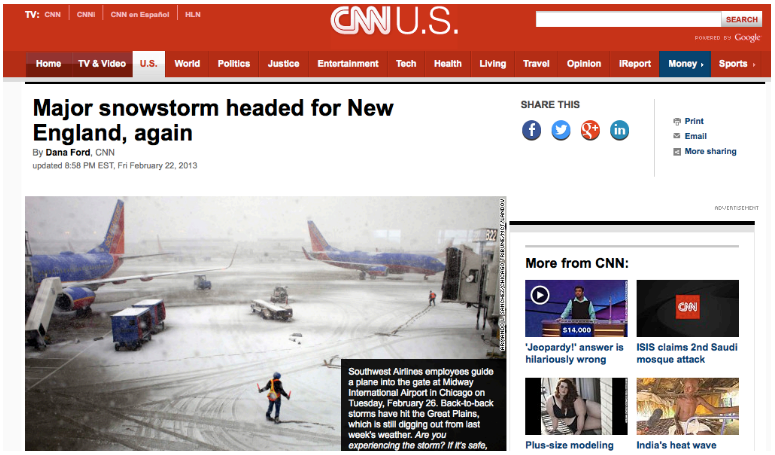
We also consider the influence that time may have on the number of different words feature that we report. Our statistical analysis shows that time has a statistically significant impact on this feature. Figure 3.50 shows the cumulative distribution of this feature as computed when comparing web pages measured which were measured 4 times for all client platforms. The measurements or samples shown are separated by approximately 1 month — that is, the time difference between Sample 1 and Sample 2 is one month, while the time difference for Sample 1 and Sample 3 is two months. This plot shows that only approximately 12% of web pages do not change at all after 4 months, whereas the rest do change to an increasing degree as time increases. This observation means that the implications and results presented relating to the number of different words feature could have been influenced by time.



(a)

Figure 3.50: Cumulative distribution plot of number of different words feature between 4 samples which are ≈ 1 month apart each.

We do, however, find rare cases where web page design has changed over time. For example, Figure 3.51 shows that the format for CNN web pages changed during our data collection procedure. We observe the new format for the CNN page (Figure 3.52) for all browsers and devices and conclude that CNN made this format change in order to serve a single web page that adapts to various screen sizes instead of serving multiple web pages to different device types. We also find that Overstock.com will display different versions of a page, one that includes product recommendations and another that does not, at different points in time (we find similar results for zillow.com with respect to content recommendations and imdb.com with respect to ads that completely change the layout of a page). We observe these differences over several browsers and believe that product recommendations are missing at certain instances in time for performance reasons — dynamically generating pages with up-to-date recommendations or ads may be costly. It is important to note these dynamic changes in web page design because it will impact the effectiveness of web page parsing tools that are optimized for a particular page design. These changes may also impact web crawling procedures because some pages may have links to related/recommended pages while others do not.



Major snowstorm headed for New England, again

By Dana Ford, CNN

updated 8:58 PM EST, Fri February 22, 2013

Southwest Airlines employees guide a plane into the gate at Midway International Airport in Chicago on Tuesday, February 26. Back-to-back storms have hit the Great Plains, which is still digging out from last week's weather. *Are you experiencing the storm? If it's safe,*

More from CNN:

'Jeopardy!' answer is hilariously wrong

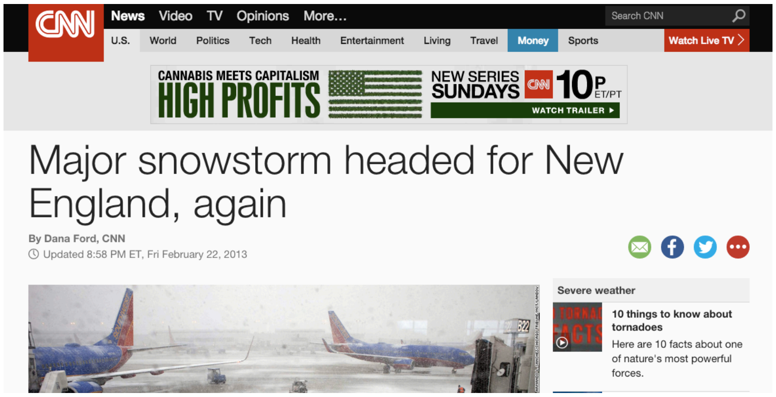
ISIS claims 2nd Saudi mosque attack

Plus-size modeling

India's heat wave

(a)

Figure 3.51: Web page layout and design can be changed over time (initial layout)



Major snowstorm headed for New England, again

By Dana Ford, CNN

Updated 8:58 PM ET, Fri February 22, 2013

Severe weather

10 things to know about tornadoes

Here are 10 facts about one of nature's most powerful forces.

(a)

Figure 3.52: Web page layout and design can be changed over time (updated layout).

3.7.2 Does time influence our traffic-based results?

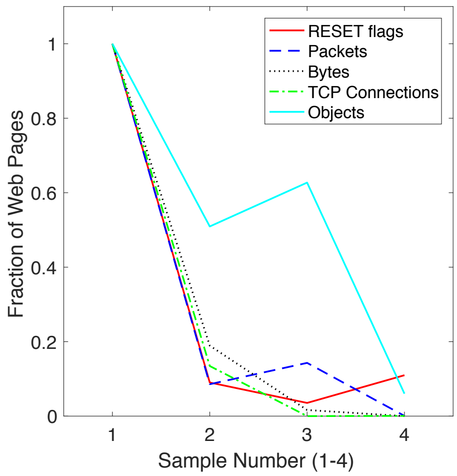
We also investigate whether the traffic features we consider vary significantly over time. Some p-values for prominent features which we identified to vary significantly for each of the 4 measurements taken are shown in Figure 3.53. It is important to note that while some of the p-values observed are less than .05 none of them are less than .001 — we observe similar results for all non-temporal features. For instance, please recall the p-value for the number of bytes differences across browser or RESET flags across operating system were on the order of $10^{?100}$. Thus, the differences observed across time are small relative to the p-values of the significant differences we discussed in this chapter.

Figure 3.54(a) and Figure 3.54(b) also visually shows this result for the number of TCP connections and number of bytes traffic features for each web page measurement across browsers, operating systems, and vantage points (i.e., each sample includes measurements from (i) web pages generated from the Firefox, Chrome, and Opera browsers on Windows 7 and Mac OSX and (ii) the web pages requested using the 13 different vantage points considered in this study). Similar observations are made for other traffic features such as the number of objects and servers — these are shown in Figure 3.55. Thus, we conclude that time is not a primary factor that influenced the non-temporal differences that we observed across browser, operating system, and vantage point in previous sections — cumulative distribution plots for the most significant traffic-based differences across client platform for each repeated measurement are provided in Appendix 7 to show that the non-temporal differences highlighted in this chapter are repeatable. We also find that the temporal traffic features including average RTT, TCP connection inter-arrival time, and TCP connection duration differs in a statistically significant manner over time across browser, operating system, and vantage point. In fact, each of the p-values for these temporal features are less than 10^{-5}. Thus, these differences in temporal traffic features may be due to randomness and/or client platforms.

3.8 Contributions and Concluding Remarks

In this chapter, we attempt to understand how modern web page traffic is impacted by different client platforms. We generate traffic in a controlled manner that includes samples from different (i) browsers, (ii) operating systems, (iii) devices, and (iv) vantage points. We find that:

- Browser type has a significant impact on the majority of the web page traffic features including the number of objects requested, number of servers contacted, and the number of TCP connections established. We find cases where the web page traffic across different browsers differ in a statistically significant manner despite the browsers referencing the *same* source HTML — we attribute these differences to choice of browser [Sanadhya et al., 2012]. One key result is that Chromium-based



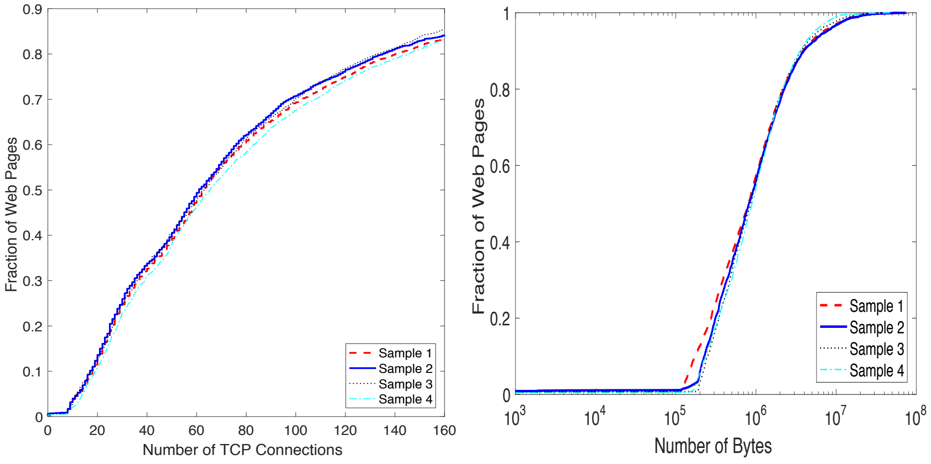
(a)

Figure 3.53: P-values for numerous traffic-based features between 4 samples which are $\approx$ 1 month apart each (comparison across time).

browsers (Chrome and Opera), at times, can transmit over *4 times* more bytes when rendering a web page than other browsers (Firefox and IE). We find that this large difference in the number of bytes transferred is due to the fact that Chromium-based browsers tend to request objects that other browsers do not. We also find that the majority of these objects have hostnames that generally correspond to servers that provide ads, tracking, and CDN services. Some of the implications of these results are:

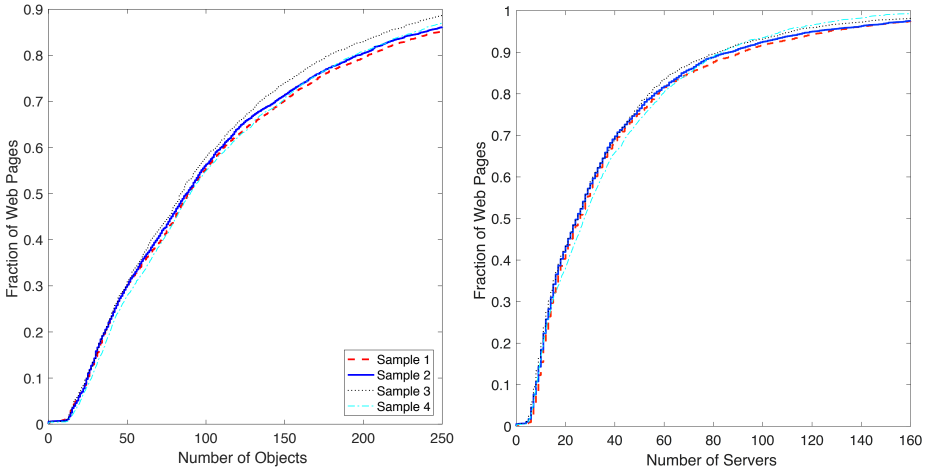
- Web page performance analysis: State of the art web page performance analysis tools, such as Pagespeed [Google], are not robust to the traffic differences we observe across browsers. Thus, different browsers will identify different performance issues for the *same* web page — this is important because there are currently no mechanisms in such tools to identify browser-specific performance issues, which makes it more difficult for web page designers to optimize web page performance.
- Privacy analysis: The fact that different browsers may contact different servers, particularly ads and tracking services, raises privacy concerns for users who may want to limit the number of hosts that their device contacts and may be unaware that choice of browser may influence such behavior [Chaabane et al., 2014, Li et al., 2015a].

- We find that choice of operating system has a minor, and mostly negligible, influence over web page traffic as compared to browsers. The only significant difference is the frequency with which the RE-



(a) **(b)**

Figure 3.54: Cumulative distributions showing that time does not significantly impact many traffic features across 4 samples which are ≈ 1 month apart each (Number of TCP connections and bytes).



(a) **(b)**

Figure 3.55: Cumulative distributions showing that time does not significantly impact many traffic features across 4 samples which are ≈ 1 month apart each (Number of web objects and servers).

SET and FIN flags are set in segments. We show, however, that these differences can be used to classify operating systems. We developed a classification model that can identify the operating system of a user with an accuracy that is as high as 97%. Operating system detection can be useful for security applications that are operating system specific (e.g., detecting operating system specific malware) [Canali et al., 2011], traffic classification [Yen et al., 2009], and general purpose characterization of network use with coarse traffic features such as anonymized TCP/IP headers [Sanders and Kaur, 2015b, Newton et al., 2013, Smith et al., 2001].

- As expected, device type has a significant impact on web page download traffic. We also find that tablets may serve pages that are designed for smartphones (i.e., small screen), tablets (i.e., moderate screen size), or even laptops (i.e., large screen). This observation suggests that there is a current lack of consensus among web designers on the properties of a web page that is designed for a tablet. Though, contrary to our browser-based analysis, the differences in web page traffic we observe across devices is primarily due to differences in the source HTML served when different types of devices request the same web page (servers will redirect web requests by mobile devices to mobile-optimized versions of the web page). In fact, when we ensure that the *same* source HTML is downloaded for different devices, we do not observe any differences in the traffic across devices.

- Vantage point has a significant impact on temporal traffic features such as average RTT. We find that the RTT observed across different vantage points are likely influenced by delays incurred by middleboxes within a network in addition to the distance between end-hosts. For example, the average RTT observed in China is over 4 times the average RTT observed in Japan despite being in similar regions of the world and requesting similar content. We observe similar, yet less extreme, differences in RTT even within the United States. This observation is important because it shows that moving servers closer to end-users may not improve web performance as much as expected because other factors within the network may serve as a bottleneck in network latency. We also find that vantage point has a minor, and mostly negligible, impact on non-temporal traffic features. Most of these minor differences are due to when hosts in certain countries download objects that are unique to that region. Again, such observations are rare and do not impact web page traffic features in a statistically significant manner.

CHAPTER 4: WEB PAGE CLASSIFICATION[1]

Traffic classification using only anonymized TCP/IP headers has received much attention in the literature, because emerging privacy legislation [Sicker et al., 2007] and increase in obfuscated traffic [White et al., 2013] make packet payloads inaccessible. The intended application domains of such classification are fairly diverse — including traffic engineering [Kim et al., 2008, Erman et al., 2007b], network characterization [Hernández-Campos et al., 2003b], and privacy analysis [Dyer et al., 2012, Panchenko et al., 2011, Liberatore and Levine, 2006, Sun et al., 2002]. The granularity at which these different domains classify traffic, is fairly diverse as well. At one extreme, for instance, traffic engineering applications have traditionally classified traffic according to coarse-grained labels such as application type (e.g., peer-to-peer or Web or Email), in order to provide differentiated services to different types of applications [Kim et al., 2008, Erman et al., 2007b, Sen and Wang, 2004, Lim et al., 2010, Schatzmann et al., 2010, Xie et al., 2012, Erman et al., 2007a]. At the other extreme of granularity, are applications such as security and privacy analysis, which may classify web traffic at the fine granularity of even identifying exactly which web page is being downloaded (also referred to as web page identification) [Dyer et al., 2012, Panchenko et al., 2011, Liberatore and Levine, 2006, Sun et al., 2002].

Both of these granularity extremes, however, face practical issues. First, classifying traffic according to the application type is too coarse of a label for modern traffic. Indeed, many applications such as file sharing, social networking, email, and video streaming, continue to migrate to the Web — in fact, recent studies show that web traffic accounts for over 80% of Internet traffic and is still gaining traffic share [Popa et al., 2010, Ihm and Pai, 2011]. As such, there is little benefit to simply classifying a given traffic sample as Web or non-Web — a more fine-grained label would be more useful. On the other hand, classifying web traffic according to the granularity of the exact web page being downloaded (i.e., web page identification), has limited applicability "in the wild" because it does not scale well to a large number of web pages — there are simply too many web pages to measure, fingerprint, store, and reliably identify. In fact, most prior studies show that the effectiveness of this fine-grained classification decreases as the number of web pages

[1] This chapter includes material from an article published in the proceedings of IEEE Infocom 2015 [Sanders and Kaur, 2015b] — the differences in the methodology and results between these two versions are summarized in Section 4.5.

considered increases [Dyer et al., 2012, Liberatore and Levine, 2006].

In this chapter, we adopt a middle-ground and consider classification of web traffic according to labels that correspond to the type of web page downloaded using only anonymized TCP/IP headers — we refer to this problem as web page classification. Web page classification is different from web page identification because multiple web pages may have the same classification label — thus, web page classification will likely scale better "in the wild" than web page identification because there will be a substantially smaller number of classes of web pages to characterize and subsequently label. The applicability of web page classification, however, depends directly on the type of labels that are used to categorize web pages. For example, categorizing a web page as being malicious or not is useful for security applications while categorizing a web page as bandwidth sensitive is useful for traffic engineering applications. Thus, it is important to study multiple different labeling schemes when evaluating the feasibility and applicability of web page classification. Some of the labeling schemes that we consider in this chapter are:

- Video-streaming based: A labeling scheme that is based on whether video streaming content is the primary focus on the page. For example, a Netflix or a Youtube video is considered a video page. A page with a banner ad that plays a video or does not play a video is not considered a video page. The ability to distinguish between bandwidth-hungry video and non-video streams, at critical traffic aggregation points, can be used for traffic engineering applications [Rao et al., 2011]. For instance, a network manager may be able to prevent network abuse and/or rate-limit video streams that use a large amount of network resources.

- Targeted-device based: A labeling scheme that is based on the type of device that the page is intended for. In the Chapter 3, we found that the device that is used to request a web page influences the HTML source that is downloaded and ultimately the traffic that is generated. Here, we consider mobile optimized web pages and traditional web pages (everything else). The ability to identify web page downloads targeted for mobile devices can help in: (i) building profiles of mobile web usage within an enterprise (for capacity planning, modeling, and forecasting purposes), and (ii) delivering personalized content and advertisements that are customized for constrained displays, power, and connectivity [Xu et al., 2011, Butkiewicz et al., 2011].

- Navigation based: A labeling scheme that is based on the type of page a content provider uses to navigate users to their desired content. We consider 3 types of pages here: 1) Homepages; 2) Search result pages; and 3) Clickable content pages (everything else). Such navigation-based labels can be useful for identifying network misuse. For instance, web crawlers can be misused for purpose of scraping search result pages [Jacob et al., 2012].

- Content-genre based: A labeling scheme that is based on the genre of a web page. The genre is obtained from the third party analytics service Alexa [Inc]. Example genres include News, Business, Computer, and Shopping. Knowledge of the genre of web pages downloaded by a given user can be used for gauging user interest, which is invaluable for delivering personalized content and targeted advertisements [Yan et al., 2009, Maciá-Fernández et al., 2010].

In the rest of this chapter, we present our data collection methodology in Section 4.1; feature selection methodology, classification results, and an applicability study in Sections 4.2 -4.4; comments regarding the differences between the results presented in this Chapter and the results published in an earlier version of this work in Section 4.5 [Sanders and Kaur, 2015b]; and our conclusions in Section 4.6.

4.1 Data Collection Methodology

This study requires the collection of web page traffic that has been (i) filtered on a per web page basis, (ii) labeled according to the client platform that was used to load the page, and (iii) labeled according to the web page that was downloaded. These requirements are the same as those for the study provided in Chapter 3. Thus, we adopt a client side measurement methodology for similar reasons. Details of the client-side measurement methodology that is specific to this work is provided in this section.

Web Page Selection The methodology used to select the web pages that are downloaded is the same as the methodology used in Chapter 3 — we focus on the top 250 web sites in the world [Inc.] and browse each of these web sites to collect a list of URLs for their landing pages, as well as non-landing pages, including search results and media content.[2] A list of these 250 web sites is provided in Appendix 1. Overall, we include a list of 3345 web page URLs — these web pages are listed in Appendix 2.[3]

[2] Our methodology does not capture the fact that some web sites present different landing pages to users who are logged in (e.g., facebook.com) — study of such "personalized" web pages is left for future work.

[3] Please note that these web pages correspond to the first 3345 URLs provided in Appendix 2. The number of web pages studied in this chapter is different from Chapter 3 because there were more web page load failures for the outdated Safari browser for the Windows 7 platform — we only include web pages that download at least one object with a status code of 200 for each browser in our analysis.

Ground Truth Labels Web pages must be labeled in order for us to train and evaluate classification methods. We assign labels to each web page (Table 4.1), according to the four labeling schemes as follows:

- VSL: The *video-streaming* (vs. non-video streaming) label is *manually* assigned to web pages where a video has played — these web pages include samples from top video streaming providers such as Netflix, Youtube, and Hulu. The non-video category corresponds to all other web pages. These include web applications that are fairly bandwidth-intensive, including radio sites (soundcloud and pandora) and file transfer sites with large files (dropbox and thepiratebay) — these web pages are included to make the classification of video pages more challenging and realistic.

- TDL: This set of labels correspond to *mobile-optimized* or traditional pages and are also *manually* assigned. We only include mobile web pages that also have a traditional web page that serves the same content — e.g., a superbowl article on bleacherreport.com that also appears on its mobile web site. Only including web pages with two distinct versions eliminates the ambiguity of determining whether a page has been optimized for mobile or not, since some web pages may have only a single version that is efficient for all devices [Sanders and Kaur, 2015a].

- WNL: A *navigation-based* label is *manually* assigned based on whether the web page was a landing page, a search-result page (obtained by entering random keywords in a search box), or a clickable content page (including news articles, video content, and social networking pages).[4]

- AGL: A *content-genre* based label is assigned to each web page, using the top-level Alexa genre for the corresponding web site. Alexa assigns multiple top-level genres to some web pages — we categorize each web page according to each genre that Alexa specifies. Thus, the total number of web pages for the the content-genre based labels is greater than 3345. There are also web pages that are not assigned a label by Alexa — we label these as “unknown”.

Trace Collection Methodology In Chapter 3, we performed a study to understand some of the properties of modern web page traffic and to determine if it is impacted by client platforms. We found that browser platform was the only client platform that had a consistent impact on web page traffic when the source

[4] There may be several landing pages per web site — e.g., www.yahoo.com and www.finance.yahoo.com. We classify each of these as landing pages.

TABLE 4.1: Distribution of Class Labels

Labeling Scheme	Class Names	# Web Pages
Video Streaming	Video page	169 (5.05%)
(2 Classes)	Non-Video page	3176 (94.95%)
Targeted Device	Traditional page	2481 (74.17%)
(2 Classes)	Mobile optimized page	864 (25.83%)
Web page	Clickable content page	1505 (44.99%)
Navigation	Search result page	1226 (36.65%)
(3 Classes)	Landing page	614 (18.36%)
Alexa Genres	Computers	821 (24.54%)
(18 Classes)	Business	568 (16.98%)
	Shopping	470 (14.05%)
	News	409 (12.23%)
	Arts	391(11.69%)
	Games	32 (0.96%)
	Adult	103 (3.08%)
	Health	32 (3.96%)
	Home	140 (4.19%)
	Kids and Teens	42 (1.26%)
	Recreation	31 (0.93%)
	Reference	96 (2.87%)
	Regional	401 (11.99%)
	Science	45 (1.35%)
	Society	84 (2.51%)
	Sports	80 (2.39%)
	World	68 (2.03%)
	Unknown	321 (9.60%)

TABLE 4.2: Browser Usage Statistics [Sta, b]

Chrome	Internet Explorer	Firefox	Safari	Opera
42.09%	25.32%	20.58%	9.49%	2.52%

HTML is the *same*. Thus, in this chapter, we explicitly consider the impact of multiple browsers when developing and evaluating web page classification methods. The 5 browsers — Internet Explorer (IE) v 9.0.8112.16502, Firefox v 23.0.1, Google Chrome v 29.01547.66m, Safari v 5.1.7, and Opera v 12.16 — are run on a Windows 7 desktop.[5] We focus on these 5 browsers and the Windows 7 operating system because they are the most popular for desktop platforms [Sta, b]. TCP/IP traces are *automatically* collected for each download as:

1. Start packet capture tool (windump)
2. Start a browser with a web page URL as an argument
3. Close the browser and packet capture tool after 120 seconds
4. Clear the local DNS resolver and browser cache
5. Go to Step 1 using a new URL

Web pages are also updated over time [Ihm and Pai, 2011]. To study which TCP/IP features remain stable over time for a given web page, we also repeat the above 3345×5 downloads 6 times each, over a period of 20 weeks (Mar 10 - July 24, 2014). Overall, this results in 100,350 web page downloads.

4.2 Feature Selection

4.2.1 Quantitative Feature Extraction

We process tcpdump traffic logs to extract TCP/IP header-based features — these features are provided in Appendix 8. Some of these include many bidirectional traffic features — such as the number of PUSH flags or the size of web objects transmitted in a TCP connection — that are not available from other sources such as NetFlow logs.[6] These features also include *multi-flow* features which span the multiple TCP connections characterizing a given page download — these include, the number of TCP connections, the number of distinct servers contacted (i.e., IP pairs), the TCP connection inter-arrival times, and the total number of

[5] Apple does not support Safari on Windows. Thus, the version of Safari used for this study is outdated compared to the version used on OSX.

[6] We use the method by Weigle et al. [2006] to identify application data units in TCP/IP traces — these generally correspond to objects.

segments transmitted. We also include statistical derivatives — such as the minimum, maximum, and several percentiles — of the occurrence of a given feature. In total, we extract 162 quantitative features derived from anonymized TCP/IP headers for processing.

The success of classification models relies critically on the selection of *informative*, *uncorrelated*, and *robust* features [Hall, 1999]. Prior traffic classification studies have focused on the first two properties by using automated correlation-based feature selection algorithms (e.g., [Lim et al., 2010]) — robustness of features has not been considered though. Given the diversity and dynamism present in the Internet (and especially in the World Wide Web), this lack of attention to feature selection is a rather serious issue [Fetterly et al., 2003, Cho and Garcia-Molina, 1999, Douglis et al., 1997]. Specific to our goal, it is important to consider the impact of at least two factors:

- Time: Modern web pages may change several times a day [Fetterly et al., 2003]. It is important to study how this impacts the stability over time of the TCP/IP features generated when the page is downloaded — indeed, classifiers that are trained on features that are stable over time are more likely to perform well on unseen data and do not need to be retrained often.

- Browsers: Browser platforms differ in their configurations and may generate different TCP/IP features when downloading the same web page (as shown by our analysis in Chapter 3). It is important to study which features are consistent (similar) across different browsers — else, classifiers trained on one browser will not perform well on unseen data that may have been generated by a different browser.

We do not consider the impact of client location, device type, or operating system in this chapter. Our results in Chapter 3 show that these factors do not impact the majority of non-temporal web page traffic features when the source HTML is the same.[7] Our analysis in Chapter 3 also found that temporal traffic features can differ in a statistically significant manner across client platforms (most notably vantage points). Thus, we do not consider temporal traffic features for web page classification.

In order to incorporate the above aspects, browser and time, we use a 3-step process for feature selection: (i) identify a set of the most *informative* features for web page classification; (ii) group the most informative features into subsets of highly *correlated* features; and (iii) select the most *stable* (over time) and *consistent* (across browsers) features from each of the above subsets.

[7] As a summary, vantage point and device did not impact non-temporal TCP/IP features when referencing the same source HTML. Operating system only impacted the number of FIN and RESET flags set features — these features, along with there statistical derivatives, are not considered in this analysis.

We elaborate on these steps below. In what follows, for each feature i, let $M^i_{n,b,t}$ represent an $N \times B \times T$ matrix populated with the measurements of feature i across the N (= 3345) web pages, B (= 5) browsers, and T (= 6) repeated web page downloads over time.

Identifying and Grouping Informative Features For selecting informative features, we first minimize noise due to browser selection or time of measurement by computing the *average* of the $B \times T$ measurements of a feature for a given web page. We then use the RELIEF method [Hall, 1999] to rank the 162 averaged features according to their ability to classify the 3345 web pages. We select the top 40 ($\sim$ top 25%) most informative features for *each* of the four labeling schemes — of the total 160 features, we find that only 57 are unique (many features were informative for multiple labeling schemes).

We then group the 57 features into correlated subsets. For this, we use the pearson correlation, ρ, to identify 7 groups of highly correlated features (listed in Appendix 9). The features within each group have $\rho \geq 0.75$, whereas the correlation between features from different groups is less than 0.35.

4.2.2 Selection of Stable and Consistent Features

Next, we estimate how stable these 57 features are over time. Recent statistics show that some browsers are more widely used than others "in the wild" [Sta, b]. In order to incorporate this into our analysis, and hence make our study more realistic, we explicitly consider the scenario where the **usage fraction** for browsers is not evenly distributed. Given this scenario, for each feature i we define an $N \times T$ matrix: $S^i_{n,t} = \sum_b w(b) M^i_{n,b,t}$, where $w(b) \in [0,1]$ represents the usage fraction for browser b (Table 4.2).

We define the *stability* over time for each feature i and web page n as:

$$DS^i_n = 100 \cdot \frac{\sum_{t=1}^{T} \left| S^i_{n,t} - \mu^i_n \right|}{T \mu^i_n} \tag{4.1}$$

where, $\mu^i_n = \frac{\sum_t S^i_{n,t}}{T}$. For each feature i, we then extract the median, 10- and 90-percentile values of the stability DS^i_n observed across the 3345 web pages — these values are plotted in Figure 4.1(a). The features are first grouped according to the 7 correlated subsets, and then sorted according to the median value of DS^i_n. Appendix 9 lists these features in the same order.

We use a similar formulation to estimate the *consistency* across browsers for each feature i and web page

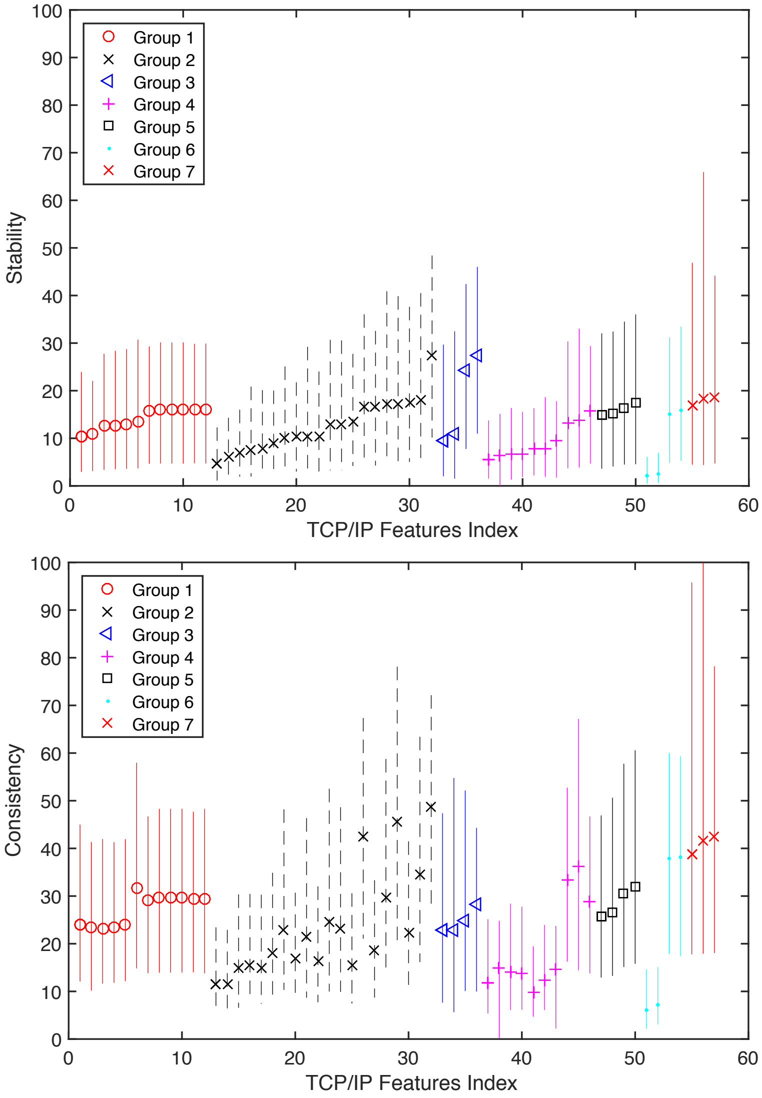


Figure 4.1: Stability(a) and Consistency(b) of features in Groups 1-7.

n. This is computed as:

$$DC_n^i = 100 \cdot \frac{\sum_{b=1}^{B} w(b) \left| C_{n,b}^i - v_n^i \right|}{v_n^i} \tag{4.2}$$

where, $C_{n,b}^i = \frac{\sum_t M_{n,b,t}^i}{T}$ is an $N \times B$ matrix, each element of which represents the *average* measurement of feature i when browser b downloads web page n repeatedly; and $v_n^i = \sum_b w(b) C_{n,b}^i$.

Figure 4.1(b) plots the median, 10- and 90-percentile values of DC_n^i observed across the 3345 web pages — the x-axis uses the same feature index as Figure 4.1.

We use Figure 4.1(a) and Figure 4.1(b) to select the most time agnostic and browser agnostic features from each of the 7 groups of correlated features. The groups of correlated features are provided in Appendix

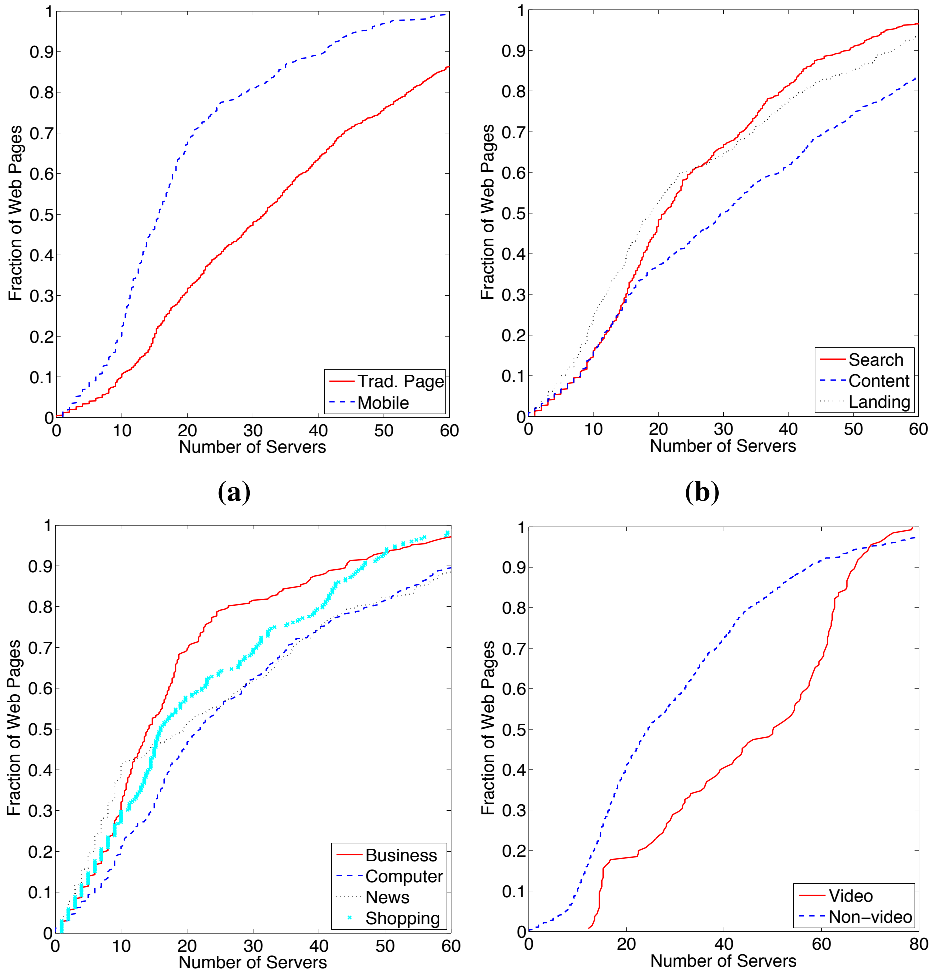
9. By comparing these two plots we find that the median, 10- and 90-percentiles for nearly all features in the feature consistency plot are larger than the corresponding values in the feature stability plot. This clearly implies that the *TCP/IP features generated by the download of a given web page vary more across client browser platforms than over time* — the analysis in Chapter 3 provides details for these observed differences. Based on this, we select the most consistent feature from each group (7 features total) to be the features that we include in our classification model — these are bolded in Appendix 9 and are discussed in the next section.

Although most of our selected features vary significantly across browsers, some features vary much more across browsers than others. For example, feature 6, the number of non-secure TCP connections is relatively stable over time like other features within its group (Group 1). However, this feature changes much more dramatically across browsers.

4.2.3 Which Features are Informative as Well as Robust?

Number of servers contacted The *total number of distinct servers contacted* for downloading a web page is discriminatory for several classes across the 4 labeling schemes. Cumulative distribution plots for the number of servers feature are shown in Figure 4.2(a), Figure 4.2(b), Figure 4.2(c), and Figure 4.2(d) for the TDL, WNL, AGL, and VSL labels, respectively. We find that mobile optimized web pages contact significantly fewer servers than traditional web pages — this result is presumably because they are designed for devices with constrained resources. We also find that video pages contact more servers than non-video pages — these extra servers correspond to increased number of ads, images, and comment boxes. This is especially true for Youtube pages, which establish around 400 TCP connections (whereas Netflix uses 60 connections). We also find that search results generally display less content from multiple servers than do clickable content pages. In the genre-based category, we find that News pages contact significantly more servers than other categories — this was also previously observed in a study by Butkiewicz et al. [2011], and is presumably because News sites tend to summarize multiple types of topics (sports, weather, finance, etc) on the same page.

Total number of bytes transferred The number of *bytes* is a feature that approximates the amount of data transferred to render a web page. Cumulative distribution plots for the maximum number of bytes sent by the client per TCP connection feature are shown in Figure 4.3(a), Figure 4.3(b), Figure 4.3(c), and Figure 4.3(d)

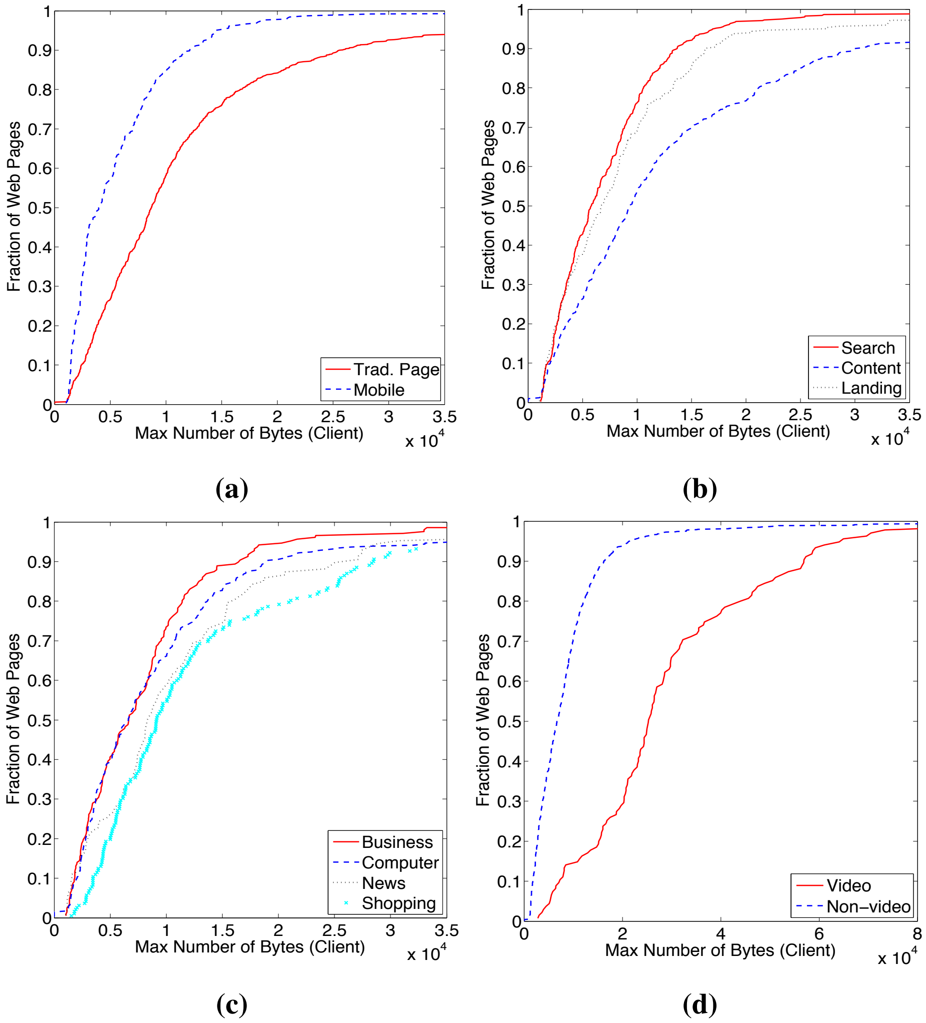


(a) **(b)**

(c) **(d)**

Figure 4.2: Discriminatory Power of Non Temporal Features (Number of servers contacted)

for the TDL, WNL, AGL, and VSL labels, respectively. When considering the TDL labeling scheme, the cumulative distribution for the mobile optimized web pages is consistently to the left of the cumulative distribution for the traditional web pages — this result is expected since, as shown in Chapter 3, mobile web pages usually reference less data than traditional web pages [Sanders and Kaur, 2015a, Johnson and Seeling, 2014, Huang et al., 2010]. We observe similar results when comparing the cumulative distributions for the non-video and video streaming web pages for the VSL labeling scheme. For the WNL labeling scheme, we find that the cumulative distributions for the landing and search result web pages overlap, while the cumulative distribution for clickable content pages is consistently to the right of these — thus, it is likely that clickable content pages are the most distinguishable/unique among the navigation-based labels. We observe similar results for the cumulative distributions that correspond to the classes for the AGL labeling scheme — specifically, the cumulative distributions for the Business and Computer pages overlap while the cumulative distributions for the Shopping and News pages are consistently to the right in this case.



(a)

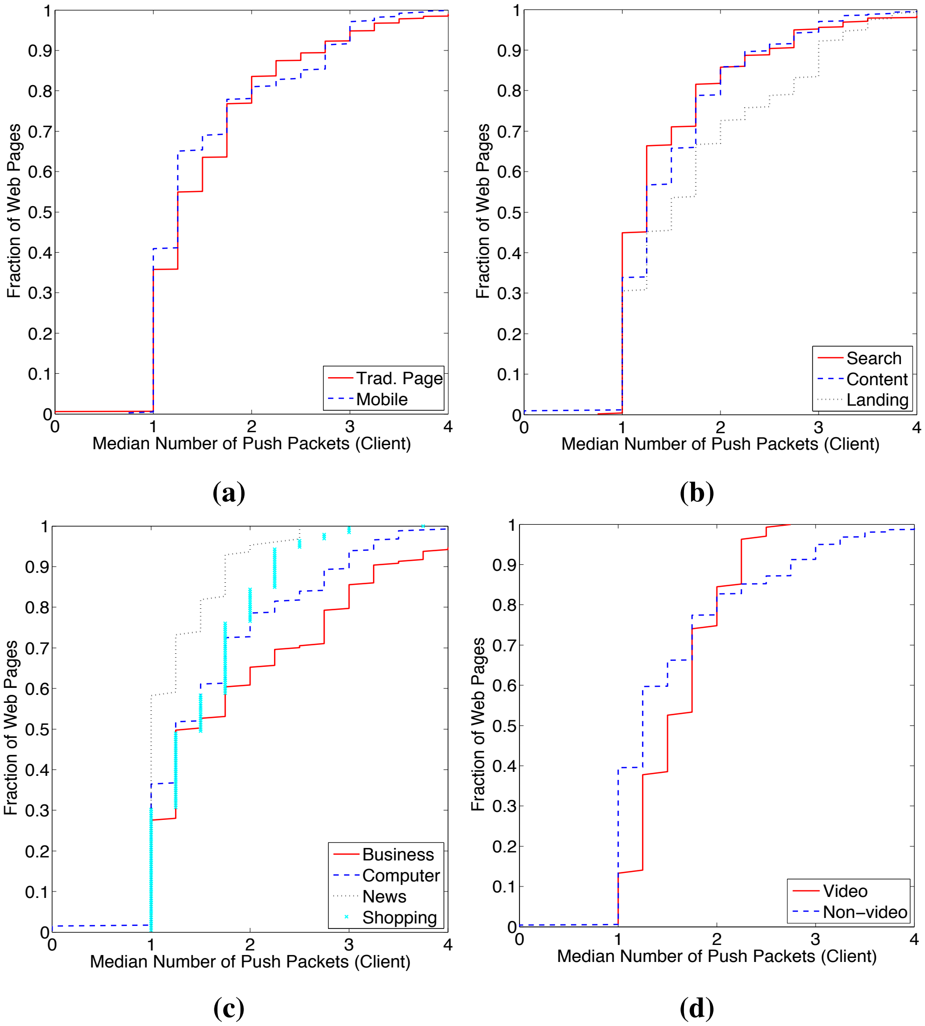
(b)

(c)

(d)

Figure 4.3: Discriminatory Power of Non Temporal Features (Maximum number of bytes transferred in a connection)

Number of PUSH flags per TCP connection Figure 4.4(a), Figure 4.4(b), Figure 4.4(c), and Figure 4.4(d) shows the *median number of PUSH segments sent per TCP connection by the client* for the TDL, WNL, AGL, and VSL labels, respectively. We find that the different types of web pages for the TDL labeling scheme have cumulative distributions that significantly overlap — given this observation, it is likely that this feature will not be critical for the classification of TDL labeled web pages. We also find that landing pages use PUSH flags slightly more often than non-landing or search pages. This is presumably because landing pages are likely to collect many more objects summarizing the web site; these objects are co-located on a small number of servers — this may be done to help reduce the load time for the "entry page" of the web site by using persistent connections and by contacting fewer servers. We observe similar trends for the AGL labels as we did for the WNL labels — that is, PUSH segments are consistently sent more often for some types of web pages than others. For instance, Figure 4.4(c) shows that the cumulative distribution for the number of PUSH segments do not completely overlap, where the News pages are the leftmost, the Shopping and Computer pages are in the middle, and the Business pages are the rightmost. For the VSL labeling scheme, we find that video streaming pages rarely have a median number of PUSH segments that is above 3 — from this, we can infer that any web page that has a median number of PUSH segments ¿ 3 is likely not a video page.



(a)

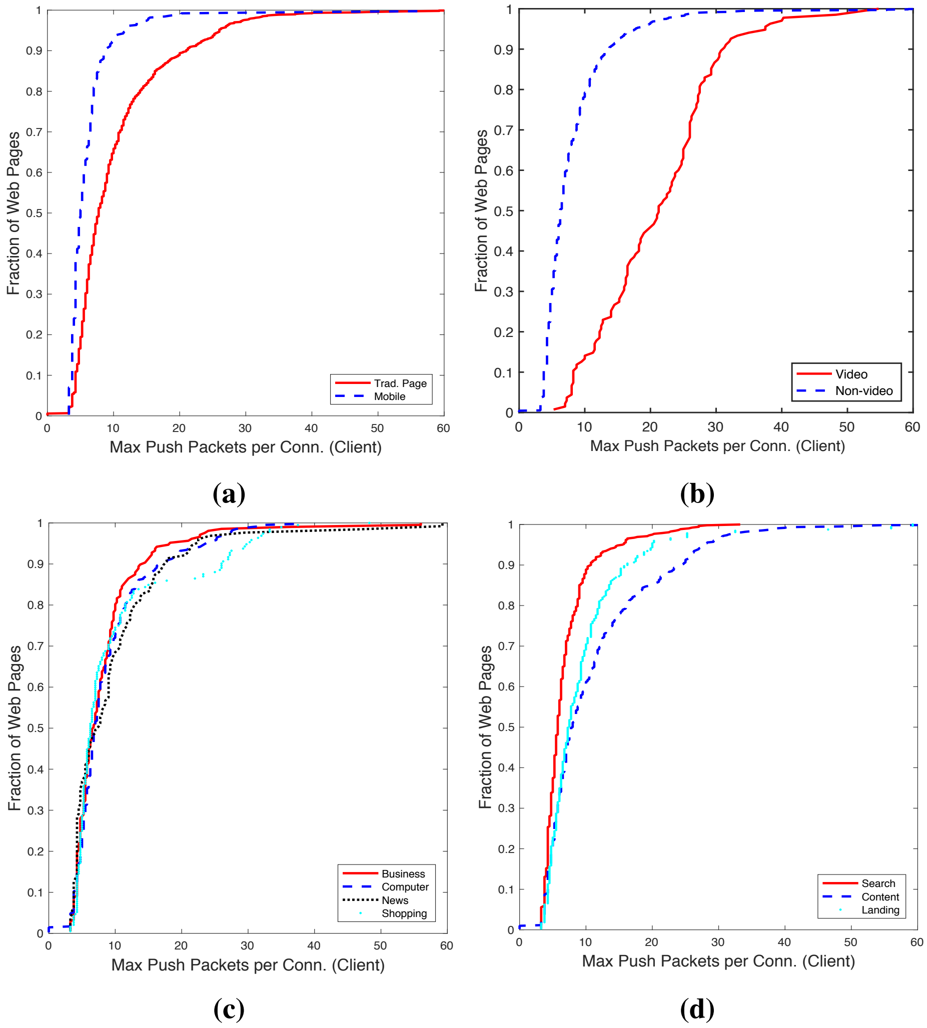
(b)

(c)

(d)

Figure 4.4: Discriminatory Power of Non Temporal Features (Median number of PUSH segments sent)

The *maximum number of push segments sent in a TCP connection* for a given web page is also an informative feature for all labeling schemes — cumulative distributions for this feature are shown in Figure 4.5(a), Figure 4.5(b), Figure 4.5(c), and Figure 4.5(d) for the TDL, VSL, AGL, and WNL labels, respectively. Previous studies show that the PUSH flag corresponds to an HTTP object [Maciá-Fernández et al., 2010] — our data also yields a high correlation ($\rho = .98$) between the two. Thus, it is likely that the multiple PUSH segments sent in a TCP connection are associated with TCP connection reuse. We find that this is more popular for traditional (non-mobile) web pages and video pages as compared to their counterparts (Figure 4.5(a) and Figure 4.5(b)). We also find that shopping web pages transmit slightly more PUSH segments per connection than the other genre-based labels shown (Figure 4.5(c)), while clickable content pages transmit more PUSH segments per connection than landing and search result pages (Figure 4.5(d)).



(a)

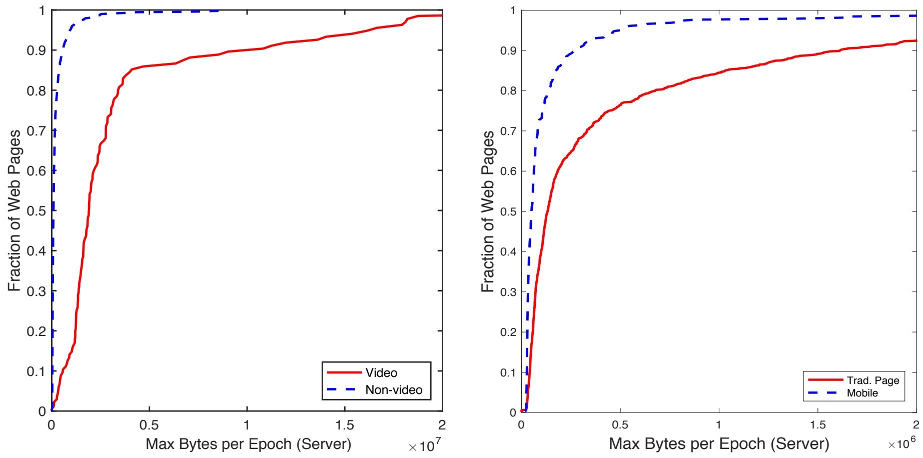
(b)

(c)

(d)

Figure 4.5: Discriminatory Power of Non Temporal Features (Maximum number of PUSH segments sent)

Epoch size The number of bytes transferred by the server for each epoch (or web object) is a valuable feature for discerning video and mobile traffic and is not useful for any other labeling scheme — that is, such features were not selected when considering the genre-based and navigation-based labeling schemes. Figure 4.6 shows cumulative distribution plots of the size of the largest epoch downloaded per web page (i.e., largest object) for the video streaming and targeted device labeling schemes. In the case for video traffic, Figure 4.6(a) shows that it is common for video web pages to transfer epochs that are larger than 200 KB, which is rare for other types of traffic — please note that the x-axis is scaled by a factor of 10^7. Figure 4.6(b) shows a similar difference between mobile and non-mobile traffic — please note that the x-axis is scaled by a factor of 10^6. These results are expected because (i) video traffic accounts for large amount of traffic and that (ii) mobile web page are designed to be more efficient that traditional web pages — that is, they transmit fewer bytes.



(a) **(b)**

Figure 4.6: Discriminatory Power of Non Temporal Features (Number of objects).

Are port numbers and first few segments helpful? Prior work on traffic classification (identifying the application layer protocol) has found port numbers and the sizes of the first few segments to be the most informative features [Lim et al., 2010]. Our analysis of web page classification, which also incorporates features that span multiple flows finds a completely different set of informative features — while port numbers are not even an applicable feature for web-only traffic, we find that even the first few packet sizes does *not* help distinguish between different types of web downloads. We believe that this is because the first few segments may capture *handshaking* mechanisms that are application protocol/application specific, but do not capture the differences between different *types* of web pages which are transmitted over the same application.

4.3 Web Page Classification Performance Evaluation

4.3.1 Background on Machine Learning Approaches Tested

We compare diverse types of machine learning methods to classify web pages including parametric and non-parametric methods [Bishop, 2007]. We use the classification trees (CT) and K-Nearest Neighbors (KNN) as our non-parametric methods, and the Naive Bayes (NB) and linear discriminant analysis (LDA) as our parametric methods.

KNN Procedure We compute the K nearest neighbors of an unseen web page with features, Y, with labeled web pages that comprise the training dataset. The unseen web page is classified with the same label as the most frequent label that is present among its K nearest neighbors — here, K is a parameter for the KNN classification method. We break ties by selecting the label that includes the web pages that have the shortest cumulative distance between the unseen web page in question. KNN also has a parameter for the type of distance metric used for the computation. We evaluate performance using the euclidean, city block, cosine, and correlation distance functions — for each of these distance functions, we also tune the parameter K for all integers from 1 to 10. We use the MATLAB *knnclassify* procedure for our KNN classification [MAT].

Classification Trees Procedure Classification trees are non-parametric models that construct decision trees by recursively splitting a node, according to a split criterion function. We use the classification tree model because it is known to be robust, requires very little data preparation, and generally performs well on large datasets. We use the MATLAB *classificationtree* procedure to train the classification tree [MAT]. We do not specify a prior distribution (i.e., each class is equally probable) because we do not know how frequent each class of web page occurs “in the wild”. We do, however, evaluate the performance of the classification trees procedure using different split criteria — these serve as the single parameter for the classification trees procedure that we vary. The three split criteria that we consider are the Gini’s diversity index, twoing rule, and deviance (or cross entropy).

Naive Bayes Procedures We use the NaiveBayes.Fit procedure in MATLAB to fit our Naive Bayes classifier [MAT]. This procedure approximates probability distributions for each class in the input dataset in order to develop a robust probabilistic model that can be used for classification. One of the parameters that we vary for the Naive Bayes classifier is the type of distribution used to model the features in the dataset. We evaluate the performance of our Naive Bayes classifier by fitting a number of different distributions to the feature set. These include the normal distribution, the multinomial distribution, and the kernel distribution with a normal, triangle, box, and epanechnikov distributed kernel smoother. Similar to previous methods we assume a constant prior distribution.

Linear Discriminant Analysis We use the MATLAB *classify* procedure to train our LDA classifier [MAT]. Contrary to the Naive Bayes assumption, LDA does not assume that the features are conditionally independent. We also reduce our feature space using principal component analysis (PCA) and use these features as input to the *classify* procedure. If we do not reduce the feature space using PCA, the procedure will fail because LDA has an assumption, that the data matrix is positive definite, that is violated by our dataset. We consider the influence of one parameter, the discriminant function, which characterizes the LDA classifier during training and evaluation. We train and evaluate the LDA classifier using with the linear, quadratic, and mahalanobis discriminant functions.

4.3.2 Evaluation Methodology

We evaluate the classification performance using the different parameters for each classification method that we discussed before. To ensure that the dataset used in this section is consistent with prior knowledge of browser usage in real-world traces, we randomly sample our 100,350 captured web page downloads by browser (using weights from Table 4.2). This data is then used to evaluate web page classification using the 4 independent labeling schemes: video streaming-based (VSL), web page navigation-based (WNL), target device-based (TDL), and Alexa genre-based (AGL). We conduct 10 independent 10-fold cross validation trials (90% of the dataset is used for training and 10% is used for testing). Please note that during this process, we ensure that a sample of the same web page is not included in both the training and test data set. Also note that some web pages may have multiple genre-based labels (e.g., Cnn.com web page is classified as both Arts and News). In such scenarios, we randomly select a single label for each cross validation trial.

We next consider which metrics are best suited for evaluating the different classification methods. The most basic metric that has been used for classification problems is perhaps the accuracy metric [Erman et al., 2006, Dyer et al., 2012]. This metric is defined as:

$$Accuracy = \frac{\sum_{i=1}^{L} tp_i}{N} \tag{4.3}$$

where L is the number of classes considered, N is the number of samples, and tp_i is the number of true positives for class i [Erman et al., 2006]. While the accuracy metric is applicable to classification problems, and is also an easy metric to interpret, the accuracy metric itself is not ideal for performance evaluation because it does not incorporate the impact that other factors, such as the number of false negatives and false positives, have on the performance of different methods. Other prominent metrics that consider these include the precision, recall, and F-score [Chase Lipton et al., 2014, Erman et al., 2006, Schatzmann et al., 2010]. These metrics, which are functions of the number of true positives (tp), false positives (fp), and false negatives (fn), are defined below:

$$Precision = \frac{tp}{tp + fp} \tag{4.4}$$

$$Recall = \frac{tp}{tp + fn} \tag{4.5}$$

$$Fscore = \frac{Precision \times Recall}{Precision + Recall} \tag{4.6}$$

Most prior work that evaluates the performance of classification methods uses the F-score as the single metric for determining the overall best performing method. This is because the F-score is the harmonic mean of the precision and recall metrics which, when considered together, encapsulate information about the number of true positives, false positives, and false positives into a single metric. One weakness of using the F-score is that it is difficult to interpret since it is a function of multiple metrics. It is because of this difficulty that most prior work that report the F-score also report the precision and recall metrics to provide intuition on how the F-score is influenced by each [Schatzmann et al., 2010, Ihm and Pai, 2011]. Another weakness of the F-score is that it is only defined for binary classification problems — that is, the F-score (and also the precision and recall metrics) only exists for problems where the number of classes considered is two. This is a problem for our evaluation because we consider two labeling schemes that include more than two classes (the WNL and AGL labels). Instead, we opt to use modified versions of the F-score, called the micro and macro F-scores, that are also applicable to classification problems with two or more classes [Chase Lipton et al., 2014]. These metrics are defined below:

- *Micro F-score:* The Micro F-score, MiF, is a metric that is the function of the micro precision, MiP, and micro recall, MiR, of each of the L classes in a classification problem. For each class i, we define tp_i as the number of true positives, fp_i as the number of false positives, and fn_i as the number of false negatives. MiP, MiR, and MiF are defined below:

$$MiP = \frac{\sum_{i=1}^{L} tp_i}{\sum_{i=1}^{L} tp_i + fp_i} \tag{4.7}$$

$$MiR = \frac{\sum_{i=1}^{L} tp_i}{\sum_{i=1}^{L} tp_i + fn_i} \tag{4.8}$$

$$MiF = 2 \cdot \frac{MiR \times MiP}{MiP + MiR} \tag{4.9}$$

Due to the way MiR and MiP are computed, MiF is biased towards classes that make up a large fraction of the dataset.

- *Macro F-score:* The Macro F-score, MaF, is a metric that is the function of the macro precision, MaP, and macro recall, MaR. The macro precision and recall is a function of the individual precisions and recalls for each of the L classes in the classification problem.

$$MaP = \frac{\sum_{i=1}^{L} \frac{tp_i}{tp_i + fp_i}}{L} \tag{4.10}$$

$$MaR = \frac{\sum_{i=1}^{L} \frac{tp_i}{tp_i + fn_i}}{L} \tag{4.11}$$

$$MaF = 2 \cdot \frac{MaR \times MaP}{MaP + MaR} \tag{4.12}$$

MaF gives an average performance across each class in the classification problem and is not biased for datasets that may have labels that make up a large fraction of the data set.

Please note that while we determine the best classification method using the macro and micro F-score metrics, we also report the precision, recall, and accuracy metrics across each labeling scheme for the best performing classification method since they are easier to interpret than F-scores. The mean performance for each metric above, $\bar{M}$, for the 10 cross validation trials is computed as: $\bar{M} = \sum_{i=1}^{10} M_i/10$.

For comparison, we also include results for a baseline heuristic called *random guessing* (RG), which randomly assigns a label to a web page.[8] We also add another, more challenging baseline heuristic called *apriori guessing* (AG), which relies on prior knowledge of the frequency of the most common class. Apriori guessing always assigns a label to the class that appears most often in the dataset — that is, the expected accuracy of apriori guessing can be expressed as $max_{i\in[1,L]}\{F_i\}$, where F_i is the fraction of times class i appears in a dataset (obtained from Table 4.1). Outperforming the apriori guessing baseline heuristic, using any metric, means that the classification performance is not biased towards dominant classes in the dataset. Outperforming these baseline heuristics also shows that web page classification using anonymized TCP/IP headers is feasible.

4.3.3 Classification Results

In this section, we present the results of our performance evaluation in three steps. First, we determine which classification method performs the best by using the micro F-score and macro F-score metrics. Second, we compare the performance of the best performing classification method across different labeling schemes to determine how labeling scheme choice impacts classification performance. And lastly, we use the precision and recall metrics to further analyze the performance of the best performing classification method.

For the stable tcpdump derived features we find that:

- Table 4.3 summarizes the mean classification performance for the micro F-score for each labeling scheme and classification method tested. The micro F-scores of the non-parametric KNN and Classification Trees models are comparable (usually differ by less than .08), where KNN with the City Block

[8] Thus, the expected accuracy of random guessing is $1/L$, where L is the number of classes for a given labeling scheme.

TABLE 4.3: Web Page Classification Performance: Micro F-score

Classification Model	VSL	TDL	AGL	WNL
Stable Tcpdump features				
K Nearest Neighbors (KNN)				
KNN - Euclidean (K=1)	.9969	.8931	.7195	.8137
KNN - City Block/L1 distance (K=1)	.9976	.9020	.7371	.8298
KNN - Cosine (K=1)	.9938	.8921	.7188	.8131
KNN - Correlation (K=1)	.9962	.8922	.7176	.8100
Classification Trees (CT)				
CT - Gini Diversity Index	.9957	.8526	.6009	.7441
CT - Twoing Rule	.9962	.8529	.6071	.7479
CT - Deviance	.9978	.8698	.5951	.7671
Naive Bayes (NB)				
NB - Normal	.9182	.3626	.2212	.3216
NB - Kernel: Normal	.9804	.4418	.2018	.3808
NB - Kernel: Triangle	.9727	.4678	.2106	.3835
NB - Kernel: Box	.9688	.4983	.2273	.3930
NB - Kernel: Epanechnikov	.9728	.4735	.2077	.3844
NB - Multinomial/Histogram	.9884	.7340	.3198	.4264
Linear Discriminant Analysis (LDA)				
LDA - Linear	.9878	.5457	.4362	.4297
LDA - Quadratic	.9760	.3907	.3845	.3679
LDA - Mahalanobis	.9436	.6645	.1047	.4340
Random Guessing (RG)	.5000	.5061	.0613	.3414
Apriori Guessing (AG)	.9495	.7417	.2454	.4499

TABLE 4.4: Web Page Classification Performance: Macro F-score

Classification Model	VSL	TDL	AGL	WNL
Stable Tcpdump features				
K Nearest Neighbors (KNN)				
KNN - Euclidean (K=1)	.9908	.8637	.6624	.8210
KNN - City Block/L1 distance (K=1)	.9928	.8679	.6785	.8367
KNN - Cosine (K=1)	.9893	.8625	.6614	.8131
KNN - Correlation (K=1)	.9887	.8629	.6624	.8173
Classification Trees (CT)				
CT - Gini Diversity Index	.9877	.8100	.5040	.7468
CT - Twoing Rule	.9887	.8105	.4855	.7515
CT - Deviance	.9934	.8325	.4989	.7722
Naive Bayes (NB)				
NB - Normal	.8465	.5823	.2788	.4719
NB - Kernel: Normal	.9477	.6278	.2789	.5325
NB - Kernel: Triangle	.9280	.6365	.2859	.5311
NB - Kernel: Box	.9182	.6466	.2793	.5335
NB - Kernel: Epanechnikov	.9267	.6396	.2817	.4607
NB - Multinomial/Histogram	.9514	.8128	.3040	.5827
Linear Discriminant Analysis (LDA)				
LDA - Linear	.9631	.6241	.0815	.4498
LDA - Quadratic	.9366	.5999	.1021	.4715
LDA - Mahalanobis	.8799	.6192	.0588	.3685

Table 4.5: Web Page Classification Performance: Accuracy

Classification Model	VSL	TDL	AGL	WNL
Stable Tcpdump features				
K Nearest Neighbors (KNN)				
KNN - Euclidean (K=1)	.9969	.8927	.7198	.8140
KNN - City Block/L1 distance (K=1)	.9976	.8959	.7374	.8300
KNN - Cosine (K=1)	.9964	.8625	.7190	.8132
KNN - Correlation (K=1)	.9962	.8919	.7177	.8102
Classification Trees (CT)				
CT - Gini Diversity Index	.9959	.8557	.6010	.7398
CT - Twoing Rule	.9960	.8557	.6050	.7477
CT - Deviance	.9975	.8662	.5993	.7703
Naive Bayes (NB)				
NB - Normal	.9183	.3628	.2005	.3213
NB - Kernel: Normal	.9779	.4403	.2245	.3805
NB - Kernel: Triangle	.9624	.4614	.2072	.3744
NB - Kernel: Box	.9557	.4872	.2245	.3855
NB - Kernel: Epanechnikov	.9615	.4666	.2050	.3784
NB - Multinomial/Histogram	.9788	.4581	.2340	.4014
Linear Discriminant Analysis (LDA)				
LDA - Linear	.9877	.5460	.0432	.4296
LDA - Quadratic	.9760	.3902	.0384	.3684
LDA - Mahalanobis	.9436	.6640	.1044	.4344

distance metric and with $K = 1$ performs the best for all labeling schemes. We only show the best performing K value for the KNN methods provided in Table 4.3 — please refer to Appendix 10 for results with other values of K. The different distance functions for the non-parametric methods do not impact the micro F-scores by more than .02. These non-parametric methods perform much better than the parametric Naive Bayes and LDA models in all cases — in fact, the Naive Bayes and LDA classifiers sometimes perform worse than apriori guessing (though all methods outperform random guessing in most cases). This is likely because the non-parametric methods do not rely on assumptions about the distribution of the features and are able to handle arbitrary feature distributions — that is, parametric methods assume specific theoretical distributions of features (e.g., Normal distribution) which is not typically the case for network traffic [Lim et al., 2010, Barford and Crovella, 1998]. This result is also consistent with the observations made in [Lim et al., 2010]. While data transformation and other preprocessing techniques may help improve the performance of parametric machine learning methods, say by transforming features such that they follow a theoretical distribution, we do not perform such an analysis in this work. Instead, we focus on examining the performance of the different machine learning methods by treating them as a black-box and do not perform any extra preprocessing to make the features more suitable for a particular method.[9] As noted before, using this approach, KNN is the best performing method while Classification Trees performs second. Similar results were obtained when using the macro F-score and accuracy metrics— these are shown in Table 4.4 and Table 4.5, respectively. Thus, for the rest of the analysis in this section we use KNN as the classification method for evaluation.

- The performance for the best performing method, KNN with the City Block distance, differs greatly across the different labeling schemes. Table 4.3 shows that the micro F-score for the VSL labeling scheme is the best of all of the labeling schemes at .9976 — the TDL labeling scheme is the second-best at .9020, the WNL labeling scheme is third at .8298, and the genre-based labeling scheme is last with a micro F-score of .7371. Table 4.4 and Table 4.5 show that we observe *very* similar results when using the macro F-score and accuracy metrics — we do not discuss these metrics further since they do not contribute additional insight on the performance of KNN.

[9] We do perform a simple standardization procedure to normalize our features (that is, each feature in a feature set is subtracted by its mean and divided by its standard deviation).

- With respect to the distance functions used by KNN, we find that the City block distance function performs the best, while the Euclidean distance function performs the second-best — though, the distance functions for KNN do not significantly impact the performance of KNN. We also find that, for the VSL labeling scheme, all classification methods can achieve a micro F-score above .980 and are fairly similar (within .018 of KNN). However, we observe much more considerable differences in performance between the classification methods for the other labeling schemes. For instance, the micro F-score for the best classification method for the AGL labeling scheme is .7374, while the micro F-score for the NB-Normal method is .2005. As described before, we believe that much of the performance differences between these methods is due to the distributional assumptions of parametric methods. It is also important to note that KNN performs better than classification trees despite the fact that both methods are non-parametric and significantly outperform the parametric approaches — more specifically, the micro F-scores for KNN are over 5% higher than the micro F-scores for classification trees for the WNL and AGL labeling schemes. It is likely that this is due to known issues with the optimization of classification tree models which may produce locally-optimal models instead of the, much more preferred, globally-optimal models. This issue usually worsens as classification/decision trees become larger and/or the number of classes considered increases [Rokach and Maimon, 2014]. KNN does not have this issue [Mikolajczyk and Schmid, 2005].

- We next discuss the metrics of precision and recall, shown in Table 4.6, for the KNN method with the City Block distance function — we do this to provide more details on web page classification performance than is present when analyzing micro and macro F-scores alone. For the video streaming labeling scheme, the precision and recall are higher than .99 for all labels except for the video label whose recall is .9834. This result shows that while the video streaming labeling scheme can be used to classify pages at a high rate, there is a slightly higher false negative rate for the video label itself — that is, the recall is smaller for the video label because the number of false negatives for this label is higher.[10] We believe this occurs because the classification method may, at times, confuse a video streaming page with other types of bandwidth-intensive web pages that are included in our dataset (such as audio streaming web pages). For the targeted device labeling scheme, we find (i) that the precision and recall are above .79 and (ii) that precision is always higher than the recall. We also find

[10] Please refer to the definition of precision and recall.

that the precision for the traditional web page label is over .10 higher than the precision of the mobile optimized web page labels. This result shows that the traditional web page label has a much lower false positive rate than the mobile optimized web page label — that is, mobile optimized web pages are more likely to be misclassified as traditional web pages than the reverse. This is likely because there are many traditional web pages that are efficient across all devices despite not being labeled as a mobile optimized web page.[11] This result also implies that there are more efficient traditional web pages than inefficient mobile web pages. Observations for the navigation-based labeling scheme are fairly similar to the video streaming and targeted device-based labeling schemes where the precision and recall tend to be higher than .80. In this case, the landing page label has a precision and recall that is higher than the search result and clickable content labels. This result shows that the landing page label can be classified slightly more effectively than the others. Overall, the precision and recall are fairly high (i.e., consistently above .79) and relatively consistent across the video streaming, targeted device, and navigation-based labeling schemes. This result shows that the KNN classification method is able to classify web pages according to these labeling schemes without being excessively biased towards these classes.

Within the Alexa genre labeling scheme, there are labels, particularly the games and adult labels, that have precision and recall values that are above .80 while there are other labels, particularly the sports and health labels, that have precision and recall values that are approximately .55. Most of the other labels within the Alexa genre labeling scheme (e.g., business, computers, science, etc) have precision and recall values that are between .68-.80. These results imply that while each label can be classified at a rate higher than random guessing (i.e., precision and recall above .50), some web page categories can be classified much more effectively than others (i.e., precision and recall above .75).

4.3.4 Sensitivity Analysis

Importance of Feature Stability We next study whether the features selected in Section 4.2.1 actually outperform features that are less robust. Specifically, we compare classification accuracy when the most *unstable* (over time) features are selected from each of the 10 feature groups in Section 4.2.1, instead of the most stable ones — the most unstable features correspond to the last feature listed in each group in Appendix

[11] Please refer to Section 4.1 for our definition of a mobile optimized page.

TABLE 4.6: Precision and Recall (KNN - City Block Distance, Stable Tcpdump Features)

Labeling Scheme	Class Names	Precision	Recall
Video Streaming	Video	.9913	.9834
	Non-video	.9984	.9992
Targeted Device	Traditional	.9342	.7933
	Mobile optimized	.8200	.7993
Web page Navigation	Clickable content	.8440	.8084
	Search result	.7992	.8211
	Landing	.8579	.9054
Alexa Genres	Computers	.7710	.7750
	Business	.7690	.7939
	Shopping	.5592	.6788
	News	.7214	.7283
	Games	.9089	.8335
	Adult	.9051	.8373
	Arts	.6914	.7088
	Health	.5355	.5541
	Home	.6891	.7235
	Kids and Teens	.5467	.5054
	Recreation	.6855	.5459
	Reference	.5785	.6395
	Regional	.6714	.6295
	Science	.7526	.7409
	Society	.5509	.6799
	Sports	.5712	.5121
	World	.6864	.6370

TABLE 4.7: KNN - City Block Distance Classification Performance for Different Features and Training Data Sets: Micro F-score

Features Used	VSL	TDL	AGL	WNL
Stable Tcpdump features	.9977	.9100	.7380	.8355
Unstable Tcpdump features	.9969	.8410	.6140	.7680
Stable Netflow features	.9945	.8816	.6782	.7805
Stable Tcpdump: different browsers	.9837	.8468	.5690	.7280
Stable Tcpdump: different time	.9940	.8920	.7056	.8051

9. Recall that all features in each group are fairly informative (for classification) and are highly correlated with each other. These results are summarized in Table 4.7. Table 4.7 shows that the micro F-score obtained when using unstable features can be up to .10 lower than when using stable features. Thus, we conclude that it is important to include not just informative features for classification (as most prior work on traffic classification does), but to also consider the stability of features.

Classification with NetFlow-based Features Our results above are obtained with classification performed based on fine-grained features derived from per-packet TCP/IP headers. Sometimes, access to such packet traces may be infeasible or costly. We next ask: what accuracy can be achieved if only coarse-grained features that are obtainable from Netflow logs are used for classification? For this, we consider those (stable) features from each group that can be derived from Netflow logs. For instance, instead of the maximum number of PUSH segments sent by the client (Group 2), we include the maximum number of bytes sent by the client per TCP connection. None of the features in Group 6 and 7 qualify, though.

Table 4.7 shows that while video-streams can still be identified with high accuracy, netflow-derived features yield lower classification accuracy by up to .06 for the other classes. It is important to note that the performance with even coarse-grained netflow features is *better* than with unstable tcpdump features — this further underscores the importance of considering stability in selecting fine-grained features.

Sensitivity to Time and Browser Our dataset includes 6 repeated downloads of each web page, using 5 different browsers for each. While we have explicitly identified features that are the most robust across time and browsers, it is important to understand the impact of training on one portion of a dataset and testing on another. We first consider the impact of time on classification performance, controlling for browser — this is done by training our classifier using data obtained at an instance in time, say the first measurement taken for each web page which includes measurements across all browsers, and testing on data obtained at a later time sample, say the first repeated measurement.[12] We repeat this process by training on the data obtained using each repeated measurement and testing on all others (occurs 6 times total, once for each repeated measurement) — the results for each train and test procedure were averaged. Table 4.7 shows that this hardly impacts classification performance at all on average. This result is promising, because it implies that classifiers do not have to be trained on data every day. In fact, our dataset includes measurements spaced out over a period of nearly 20 weeks.

We next consider the impact of browser on classification performance, controlling for time — this is done by training our classifier using data obtained using a single browser, say Firefox, and testing on data obtained using all other browsers. We repeat this process by training on the data obtained using each browser and testing on all other browsers (occurs 5 times total, once for each browser) — the results for each train and test procedure were averaged. Table 4.7 shows that while video streams can still be identified with the same rate, the micro F-scores for the mobile-targeted and navigation-labels reduces by about .06-.10. The most significant impact, however, is on the genre-based labels, which have a micro F-score of .58 as compared to .73. These results imply that traffic classification performance is much more browser-dependent than time-dependent — our analysis of repeatability and consistency of traffic features in Section 4.2.1 supports this observation. We conclude that it is important to train models on data that is representative of browser mixes found in real-world traces.

Miscellaneous Comments on Web Page Classification Results Traffic classification studies from recent literature boast of classification accuracies higher than even 94% [Kim et al., 2008, Schatzmann et al., 2010]. In comparing our results from this section to those, it is important to keep in perspective several fundamental differences:

- Our classification framework is subject to the web page design decisions of developers. Standards do not exist that ensure that web pages of a particular category yield similar traffic. We find that many web sites that host similar content tend to follow similar web page design trends. For example, mobile web sites tend to design their pages to be more resource conscious than traditional web sites.[13] Modern web sites also use web page templates and content management systems. Thus, many of their

[12] Recall, our data includes 6 repeated measurements.

[13] Also, modern search engines include similar search options such as web search, image search, and news search.

corresponding web pages follow a predefined *structure* that can be observed in traffic. We stress the need of strategically sampling web sites that are likely to be included in a real data set — we focus on popular web sites in this work. It is also necessary to keep the training data set up to date, since web pages evolve over time.

- The Alexa genre classification may be particularly noisy. Web page designers and third party analytics services have the freedom to arbitrarily assign labels to web pages. Even when web pages serve a particular purpose, its label may be unclear. For example, what type of web site is the gaming review web site www.ign.com – a news web site, an entertainment web site, or a game web site? Should social networking sites be considered news sites? These factors dramatically increase the variance and noise in each of the class labels.

- There is also legitimate room for improvement in performance by incorporating prior information (about how often a particular class appears "in the wild") into the classification models. We elect to not use prior information because our dataset is synthetic — any prior information would give an inappropriate and non-representative increase in performance. A real-world dataset that was collected "in the wild", would be able to benefit from such information.

- Our feature selection methodology identified a few temporal traffic features that were particularly informative for the video labeling scheme. However, our analysis in Chapter 3 shows that temporal traffic features is impacted by time and is significantly impacted by vantage point. Thus, we recommend training web page classification methods using data collected from the vantage points in which they are being used.

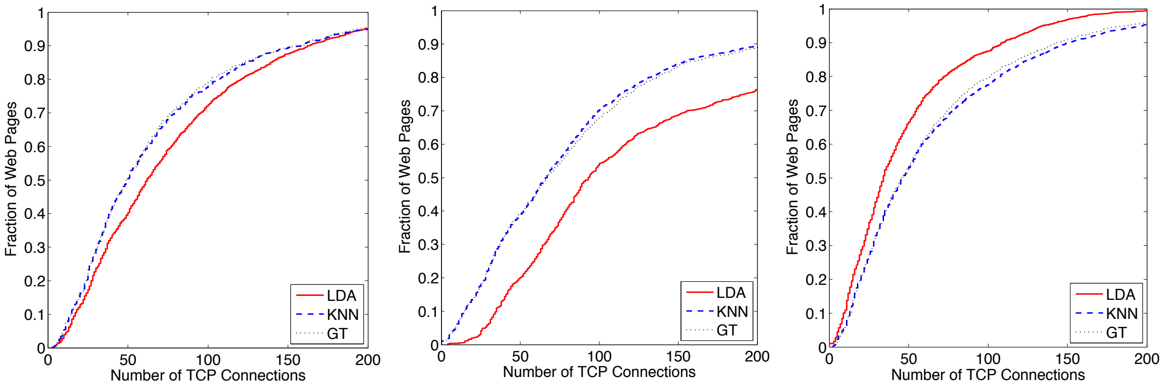
4.4 Applicability of Web Page Classification

In the previous section, we found that the performance of the classification methods differed across the different labeling schemes. Specifically, we observe micro F-scores of over .90 for the video streaming and targeted device-based labeling schemes and somewhat lower performance for the others. Given these results, it is natural to ask — is the classification performance that we observe for the different labeling schemes good enough? The answer to this question depends on the intended application of web page classification. Many applications rely on measuring statistical properties of metrics that are observable in traffic — common metrics include the fraction of traffic contributed by different types of applications [Butkiewicz et al., 2011, Ihm and Pai, 2011], distributions of file sizes [Smith et al., 2001, Callahan et al., 2010, Schneider et al., 2008], and usage profiles [Liu et al., 2010, Maciá-Fernández et al., 2010]. In this section, we conduct three case studies to determine whether web page classification can be realistically applied to different application domains. An overview of the application domains in which these three case studies will be useful is provided below:

- Network Usage Characterization: In many scenarios it is imperative to understand the traffic characteristics of the applications that are observed in a network [Hernández-Campos et al., 2003b, Smith et al., 2001]. For instance, network administrators may need to understand the extent to which emerging applications, such as mobile and video streaming applications, are used in their network [Mob, c]. Understanding the properties and statistical distributions of the traffic features for different types of web applications is necessary for domains such as web page performance analysis [Liu et al., 2010, Butkiewicz et al., 2011] and click bot detection [Wang et al., 2013, Gu et al., 2008]. In one case study, we evaluate whether web page classification can be used to estimate the statistical distributions of different types of web page traffic.

- Simulation Modeling and Network Forecasting: The properties of the network traffic that is observed today are often used to predict the network traffic that will be observed in the future [Moore and Zuev, 2005, Karagiannis et al., 2004, McGregor et al., 2004]. These predictions are commonly used to determine when capacity issues in a network will occur — the idea being that these issues will be addressed preemptively. Network forecasting models require (i) simulation models and (ii) data on the statistical properties of previously observed traffic. We perform a case study that investigates whether network forecasts produced when using ground truth data is *statistically equivalent* to the forecasts produced when using data derived from our web page classification procedure.

- Behavioral Ad Targeting: Accounting for approximately 170 billion US dollars in spending in 2015, digital advertising is an important driver of the Internet economy [Statista]. Behavioral ad targeting is an advertising approach that uses *user profiles* to recommend ads that are more likely to be relevant to users than when using traditional advertising approaches [Chen and Stallaert, 2014]. These user profiles, however, are typically created a user's browsing history which is generally only available to analytics companies and content providers [Maciá-Fernández et al., 2010]. We conduct a case study to determine whether web page classification can be used to approximate these user profiles using only anonymized TCP/IP headers. This study is useful because it will show whether ISPs, or similar entities, can leverage web page classification to assist in expanding into the digital advertising industry.



(a) Homepage **(b) Clickable Content** **(c) Search Result**

Figure 4.7: Distribution of the number of TCP connections across 3 navigation-based classes

4.4.1 Approximating Labeled Traffic Distributions

We want to study if our classification results are useful in extracting the *true* distributions of features within a given class — that is, for a given class, do the feature distributions observed across classified web pages match those observed across ground-truth labels? If they do, then our classification methodology can be used to derive ground-truth feature distributions from real traces — which can then be used for traffic modeling and simulation studies. They can also be used to monitor and understand the general usage profile of a user population.

For studying this issue, we first divide traffic into two groups according to (i) the ground-truth labels and (ii) the classified labels. We then compare the distributions of features obtained from these two groups of traffic. We first present our analysis using two hypothesis testing approaches (and later visually). The first test, the Wilcoxon sum ranked test, tests the hypothesis that the *medians* of the two distributions in question are the same. We use the Wilcoxon sum ranked test because it does not rely on strong distribution assumptions such as normality. The second hypothesis test method, the Kolmogorov-Smirnov test, tests whether the two *empirical distributions* are the same. Recall that a p-value that is larger than .05 validates the concerned hypothesis.

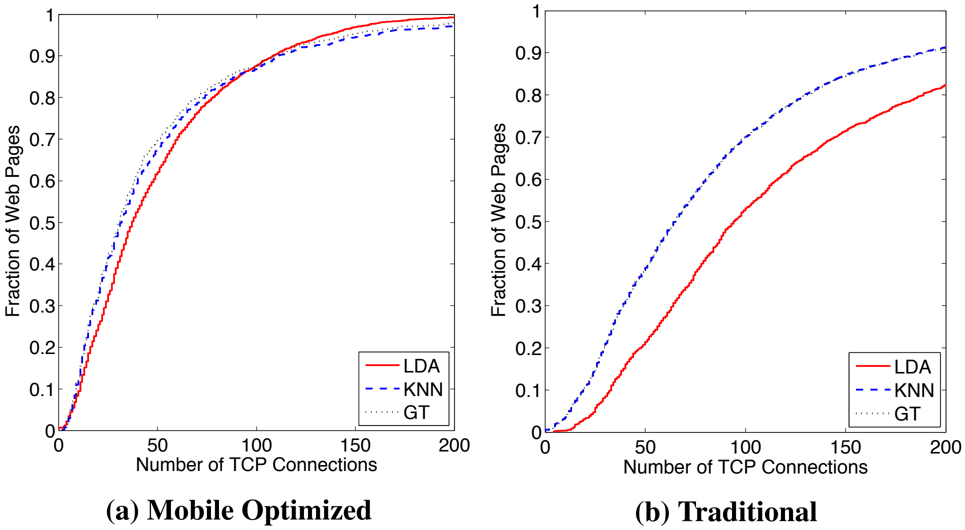
Table 4.8 shows the p-values for two traffic features, *the number of TCP Connections* and *the number of bytes transmitted*, when the traffic is classified using either KNN or LDA. We show results for the number of TCP connections and number of bytes transmitted features in this table because they are important when modeling and simulating TCP/IP traffic [Weigle et al., 2006, Barford and Crovella, 1998] — results for other features such as the number of segments and objects (epochs) transferred are provided in Appendix 11. We also only show results for the KNN and LDA methods in Table 4.8 because they were the best and worst performing classification methods, respectively — Appendix 11 provides results for all classification methods considered in this chapter. We find that:

- With the KNN classifier, the Wilcoxon sum ranked and Kolmogorov-Smirnov tests yield p-values that indicate that both the median as well as the empirical distributions of these two features are the same across classes identified using either classified labels or ground truth labels. More importantly, the tests for the results for KNN are favorable with p-values that are larger than .05 for *all* classes for each feature shown — this is true even for the AGL labeling scheme (results shown only for the 4 most common genres).

- This result is not true for all classification methods. In fact, LDA usually outputs p-values that favor the alternative hypothesis in which the distributions of the classified traffic *differ* from the ground truth — these p-values are generally much lower than 10^{-10}. There are exceptions, where the p-values for LDA yield results in favor of the null hypothesis for some classes, but this is not true for an overwhelming majority of classes in each respective labeling scheme.

We have also analyzed other features that are relevant for traffic generation and simulation modeling, including the number of servers contacted, the number of objects transferred, and the number of segments transferred, and arrived at the same statistical conclusions — a list of the p-values for these features for each labeling scheme and classification method considered in this chapter is provided in Appendix 11.

Figures 4.7 and 4.8 plot the cumulative distributions for the number of TCP connections feature, as yielded using the KNN and LDA classification methods — we show these plots as a visual representation of the results in Table 4.8. Figure 4.7 shows the results for the navigation-based labeling scheme. We observe that in most cases, KNN is able to classify web pages into classes that closely match the distribution of the ground truth dataset. Similar to the hypothesis testing results, LDA is not able to consistently achieve this. In particular, LDA exaggerates the number of TCP connections required for the search result and clickable



(a) Mobile Optimized **(b) Traditional**

Figure 4.8: Distribution of the number of TCP connections across 2 target device-based classes

content pages. These results are more clear in Figure 4.8, which shows distributions corresponding to the targeted device-based labels. LDA, essentially classifies the data such that it maximizes the separation between the two classes — this behavior is similar to that of a clustering method [Erman et al., 2006]. Hence, the large shift of the number of TCP connections for web pages for the traditional web page class. This shift is unrealistic and does not align with the ground truth distribution. The nonparametric KNN method does not have this problem.

We conclude that classification methods that perform well, such as KNN, can be used to extract true distributions of traffic features within classes (matching the trained distribution), while other methods, such as LDA, have difficulty doing so.

4.4.2 Simulation Modeling and Network Forecasting

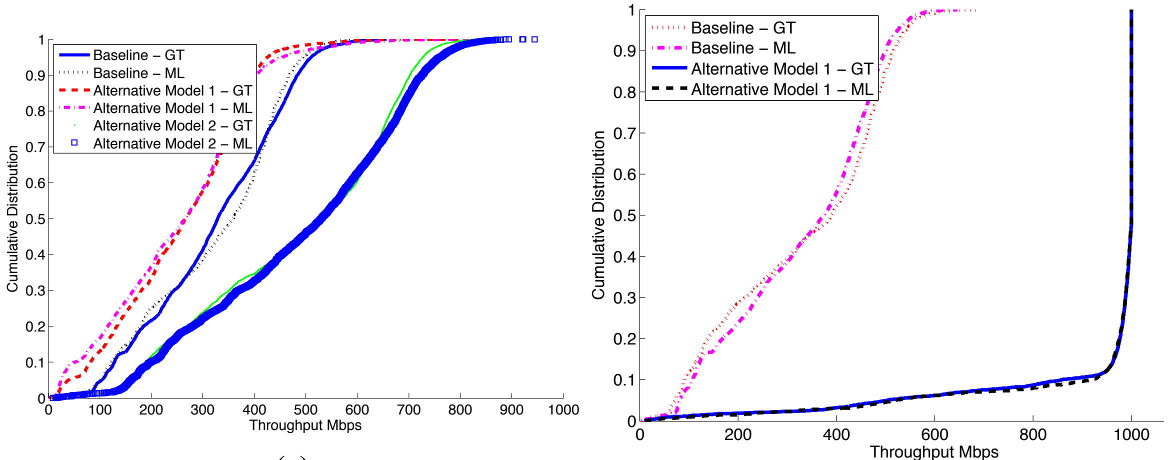
We next compare traffic generated using ground-truth and classified labels. We use the ns-2 network simulator to simulate the web browsing behavior of 400 web users where all traffic gets aggregated on a shared 1Gbps link — the traffic for each individual web page download is taken from our data set. Each user behaves independently and randomly visits a web page. The inter-arrival time for web page downloads by a given user is gaussian distributed with a mean of 30s and standard deviation of 15s — this distribution is chosen for simplicity (and is adequate for our purpose of comparing distributions).

The download of each web page itself is simulated using TMIX, which provides a source-level traffic generation interface in ns-2 [Weigle et al., 2006]. Specifically, we provide this tool with the TCP/IP trace

TABLE 4.8: Comparing Feature Distributions from Classified Web Pages and Ground Truth Labels

Feature -	Statisical Test -	Label -	p-value (KNN/LDA)
Number of TCP Connections	Ranked Sum	Mobile web page	.6066/.0230
	Kolmogorov-Smirnov	Mobile web page	$.7969/1.728 \times 10^{-9}$
	Ranked Sum	Traditional web page	$.9998/1.66 \times 10^{-136}$
	Kolmogorov-Smirnov	Traditional web page	$1.000/1.846 \times 10^{-49}$
Number of TCP Connections	Ranked Sum	Video web page	$.8764/1.13 \times 10^{-8}$
	Kolmogorov-Smirnov	Video web page	$1.000/8.1467 \times 10^{-12}$
	Ranked Sum	Non-video web page	.9583/.4879
	Kolmogorov-Smirnov	Non-video web page	1.000/.2954
Number of TCP Connections	Ranked Sum	Computers	$.6405/3.26 \times 10^{-6}$
	Kolmogorov-Smirnov	Computers	$.8773/3.211 \times 10^{-12}$
	Ranked Sum	Business	$.8193/7.9043 \times 10^{-11}$
	Kolmogorov-Smirnov	Business	$.9828/4.5604 \times 10^{-1}$
	Ranked Sum	Shopping	.6660/.0019
	Kolmogorov-Smirnov	Shopping	$.9997/1.8681 \times 10^{-10}$
	Ranked Sum	News	.7675/0.0115
	Kolmogorov-Smirnov	News	$.6255/3.4481 \times 10^{-6}$
Number of TCP Connections	Ranked Sum	Homepage	$.3018/3.2032 \times 10^{-74}$
	Kolmogorov-Smirnov	Homepage	$.9828/8.3124 \times 10^{-9}$
	Ranked Sum	Search	$.9262/2.88 \times 10^{-106}$
	Kolmogorov-Smirnov	Search	$.6814/4.5578 \times 10^{-18}$
	Ranked Sum	Clickable Content	$.5667/6.9477 \times 10^{-31}$
	Kolmogorov-Smirnov	Clickable Content	$.4800/1.0665 \times 10^{-28}$
Number of Bytes	Ranked Sum	Mobile web page	$.3133/1.449 \times 10^{-4}$
	Kolmogorov-Smirnov	Mobile web page	$.4775/1.5063 \times 10^{-19}$
	Ranked Sum	Traditional web page	$.8597/6.413 \times 10^{-84}$
	Kolmogorov-Smirnov	Traditional web page	$.9999/4.6465 \times 10^{-89}$
Number of Bytes	Ranked Sum	Video web page	.9173/.4364
	Kolmogorov-Smirnov	Video web page	$1.00/4.2076 \times 10^{-4}$
	Ranked Sum	Non-video web page	.9924/.3151
	Kolmogorov-Smirnov	Non-video web page	1.00/.7021
Number of Bytes	Ranked Sum	Computers	$.7127/6.0806 \times 10^{-13}$
	Kolmogorov-Smirnov	Computers	$.9999/2.122 \times 10^{-09}$
	Ranked Sum	Business	$.9440/6.2451 \times 10^{-10}$
	Kolmogorov-Smirnov	Business	$.9941/2.573 \times 10^{-1}$
	Ranked Sum	Shopping	.2248/.0045
	Kolmogorov-Smirnov	Shopping	$.9821/4.933 \times 10^{-13}$
	Ranked Sum	News	$.9108/3.1335 \times 10^{-5}$
	Kolmogorov-Smirnov	News	$.5558/9.5609 \times 10^{-10}$
Number of Bytes	Ranked Sum	Homepage	$.1847/4.9872 \times 10^{-33}$
	Kolmogorov-Smirnov	Homepage	$.8981/3.3592 \times 10^{-17}$
	Ranked Sum	Search	$.6982/6.0573 \times 10^{-99}$
	Kolmogorov-Smirnov	Search	$4978/1.4042 \times 10^{-5}$
	Ranked Sum	Clickable Content	.2476/0.0765
	Kolmogorov-Smirnov	Clickable Content	$.9998/1.4113 \times 10^{-99}$

of a web page download (selected randomly from the 100,350 downloads we collect in Section 4.1). TMIX then derives from the trace, application-level descriptors of the corresponding traffic sources — including request sizes, response sizes, user think times, and server processing times. It then generates corresponding traffic in ns-2 by reproducing these source-level events. Thus, this tool allows us to faithfully produce realistic source-level behavior for each web page download. We use this traffic generation methodology in the context of the forecasting application below.



(a)

(b)

Figure 4.9: Distribution of aggregate throughput for mobile model (a) and video model (b)

Modeling Growth in Mobile Web Usage We first construct a *baseline model*, in which each user visits a mobile-optimized web page 20% of the time and a traditional page 80% of the time — nearly 20% of current web traffic is considered mobile [Mob, a]. The TMIX input for each user is obtained by randomly selecting a mobile (or traditional) page download from our set of 100,350 downloads — we conduct two experiments, in which the mobile or traditional pages are selected based on either ground-truth (GT) labels or KNN labels (ML). The throughput on the 1 Gbps aggregated link is observed every 1ms — its cumulative distribution is plotted in Figure 4.9(a).

We next conduct two sets of experiments that incorporate growth in mobile traffic. In the first set, referred to as *alternate model 1*, we envision the scenario in which all users increase their reliance on mobile devices — specifically, in this model, each user visits a mobile-optimized web page 50% of the time (labelled using either GT or ML). In the second set of experiments, we envision growth in the number of users that rely *solely* on mobile devices. In this model, referred to as *alternate model 2*, we retain the behavior of the

400 baseline users, but simulate an additional 200 users that browse only (GT or ML-identified) mobile-optimized web pages (100% of the time).

The distribution of the aggregate throughput for each of the forecasting experiments is also plotted in Figure 4.9(a). We find that the distributions yielded by the ground-truth (GT) and the classified (ML) labels are quite similar to each other. In fact, we run the hypothesis testing approaches mentioned earlier to confirm that the distributions are, in fact, statistically equivalent. This is true for the baseline traffic, as well as each of the forecasted alternative models. This confirms that *web page classification, based only on anonymized TCP/IP headers, can be used to effectively conduct traffic modeling studies involving mobile web traffic*.

Modeling Growth in Video Streaming We use the same approach as above to construct a *baseline model*, in which a user downloads pages with video traffic 20% of the time, and an *alternate model 1*, in which users access video-based web pages 50% of the time — please note that these percentages were arbitrarily chosen to compare the distributions of the different models. The aggregate throughput is plotted in Figure 4.9(b). We find that, as before, the distributions obtained by relying on classified (ML) labels are nearly identical to the ones derived using ground-truth (GT) labels. This is true even when the forecasted traffic drives the network to nearly full-utilization (alternate model 1).

We emphasize that our intention is *not* to make forecasting claims, but simply to illustrate that our classification work can very well facilitate such traffic modeling applications.

4.4.3 Behavioral Ad Targeting

Applications such as behavioral ad targeting use browsing profiles to infer what the user is likely to be *most* interested in. For instance, if the user has been recently visiting sports sites, it may be a good idea to pop up jersey ads for him/her. Our goal here is to analyze the accuracy with which web page classification (for the genre-based labeling scheme) can help build a useful browsing profile for the user browsing sessions.

We define a browsing session as a sequence of *N* web page downloads by a single user. We synthetically generate browsing sessions by randomly selecting a sequence of *N* web pages using a Markov model of web browsing — we obtain a list of the transition probabilities of the top 5 web sites that a user is most likely to visit while on a given web site from Alexa [Chierichetti et al., 2012, Inc.]. We use these transition probabilities and a simple Markov model to generate browsing sessions that include *N* web page downloads by doing the following steps:

TABLE 4.9: Accuracy in Identifying Most-visited Genres.

Labeling scheme	$N = 20$	$N = 50$	$N = 200$
Genre-based labels			
$K = 1$	86.1%	85.1%	89.2%
$K = 2$	89.0%	89.2%	89.2%
$K = 3$	83.2%	87.1%	87.1%
$K = 4$	84.9%	83.0%	81.8%
$K = 5$	80.5%	82.2%	91.8%

1. Select a single browser to download all web pages for a single browsing session.

2. Randomly select and browse a "seed" web page from our list of 3345 web pages using the selected browser.

3. Given the previous web page, select a web site to browse according to the transition probabilities obtained from Alexa — if the web site selected is not included in our dataset of 250 web sites, randomly select a web site.

4. Randomly browse a web page in our dataset that is hosted by the selected web site.

5. Repeat steps 3 and 4 until N web pages have been browsed.

We perform the above steps to generate 1000 browsing sessions for N = 20, N = 50, and N = 200 — that is, 3000 browsing sessions total. Please note that the steps above do not explicitly download web pages. We are simply grouping previously downloaded web pages into browsing sessions with N web pages — thus, we do not consider the scenario of overlapping traffic which requires using a web page segmentation approach.

We collect statistics on the top-K (in terms of frequency) AGL genres observed in each browsing session — we collect two sets of statistics, one based on ground-truth genre labels, and the other based on classified genre labels.

In Table 4.9, we list the percent of users for which the unordered set of top-K genres based on classified-labels matches with the ground-truth labels. We find that the top-5 genres that a user is interested in can be estimated for more than 80% of the users for each of the parameters tested (K and N). The impact of these parameters on the performance of the user-profiles is discussed below:

- We find that the impact of K on the performance of the user profiles depends on two cases: 1) $K \leq 2$; and 2) $K > 2$.
  - When $K \leq 2$, the performance for the profiles usually increases as K increases. This result occurs because there are browsing streams where the two most frequently observed classes were classified in the wrong order — that is, the most frequently observed label was classified second and the second most frequently observed label was classified first. In such scenarios, the profile for $K = 1$ would be incorrect whereas it would be correct for when $K = 2$ — recall that our metric for top K labels preferences is based on unordered sets.
  - When $K > 2$, the performance of the user profiles usually decreases as K increases — we believe that this result is reasonable because adding more labels to a profile will increase the chances for errors.[14]
- We find that increasing N from 20 to 200 usually improves the performance of the user profiles anywhere from 0-10%. We believe that this modest performance improvement for increasing N is due to using more samples of web pages to build more reliable user profiles — indeed, additional samples provide more chances for the web page classification method to correctly classify web pages in a browsing stream. While we observe a modest performance decline of 3.1% when $K = 4$, we believe that this decline is an anomaly (essentially noise).

Overall, these results are highly encouraging and suggest that targeted ad delivery can significantly benefit from web page classification based on just anonymized TCP/IP headers.[15]

Another piece of useful information that may be needed from a user browsing profile is the *fraction* of web pages visited by a user that correspond to a particular genre. For instance, if a user visits the top genre for 95% of the web pages that they browse, he/she is unlikely to be interested in ads related to any of the other top genres. We collect statistics on the fraction of web pages that a user visits for each of their respective top-5 genres (both based on ground-truth labels as well as classified labels). Figure 4.10 plots the median and 95% confidence intervals of these per-user fractions, for their top-5 genre and navigation preferences. We find that the top genre-based and navigation-based browsing frequencies yielded by classified labels align

[14] Though, the performance will increase as K approaches the number of possible classes.

[15] It is important to note that even though the classification micro F-score for the AGL class were only around .75, the user browsing profiles being constructed here are simply relying on a comparison of the *sets* of top-K genres (and not correct classification of each of the N web-page visited).

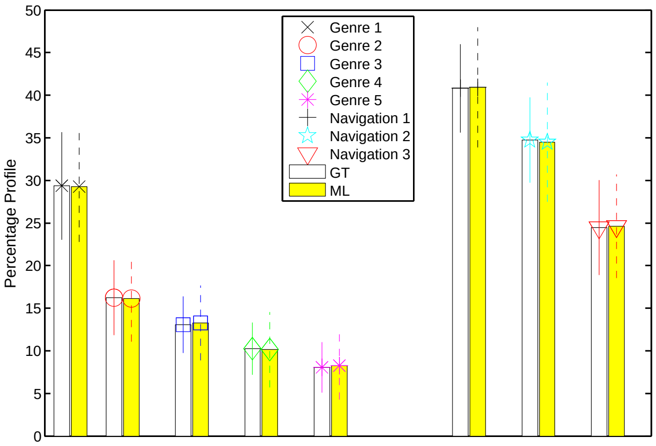


Figure 4.10: Frequency of Labels (Ground Truth and KNN classified).

extremely well with those based on ground-truth labels for both genre-based and navigation-based labels. We conclude that web page classification is fairly well suited for building frequency-based user browsing profiles, even for content-genre based labels.[16]

4.5 Comparison with the Previously Published Version of this Work

An earlier version of the work presented in this Chapter was published in the proceedings of IEEE Infocom 2015 [Sanders and Kaur, 2015b]. While most of the methodology used and the results presented in this Chapter are the same as those in the previously published version, there are a few key differences. These include the following:

- *Features considered for web page classification:* The previous version of this work did not consider the impact that operating systems, device type, and vantage point may have on the features used for classification. Thus, features that we found in our analysis in Chapter 3 that differ significantly across these client platforms — such as the number of FIN flags, number of RESET flags, and temporal traffic features — were included in the feature set of the previously published version [Sanders and Kaur, 2015b]. We do not include these features in this chapter since they are not likely to be consistent across the different client platforms — hence, the number of features considered was reduced from 216 to 162 in this chapter. We find that reducing this feature set had a minimal impact on the results in this

[16] It is important to note that identifying individual users via traffic analysis may be difficult due to network proxies and other technology. However, previous work shows that other methods for identifying malicious users behind proxies are still effective despite this limitation [Jacob et al., 2012].

study. We believe this occurs for two reasons. First, the non-temporal traffic features that we removed were not selected for classification in the previous version. And second, the temporal traffic features that we removed were included because they were among the most informative features for *only* the video streaming labeling scheme. When comparing the results between these two versions, we find that removing the temporal traffic features did not significantly impact the performance of the video streaming labeling scheme.[17] We believe that this occurs because the other features included in the classification methods (e.g., number of servers and number of bytes transferred per TCP connection) are informative enough to perform well without the temporal features.

- *Parameters used for classification methods:* The previous version of this work did not consider the many different parameters for the classification methods used in this Chapter — these include the split criteria functions for the Classification Trees classifier, the distance functions for the KNN classifier, the discriminant functions for the LDA classifier, and the distribution functions for the Naive Bayes classifier. The best performing classification method in this Chapter was the KNN method with the *City block distance function*, while the best performing method in the previous version was the KNN method with the *Euclidean distance function* — both methods performed best when $K = 1$. Please note that the City Block distance function was not considered in the previously published version of this work. Thus, if it were considered, it is likely that KNN would have performed best with the City block distance function in the previous version as well.

- *Metrics used for performance evaluation:* This Chapter considers the metrics of macro F-score and micro F-score for performance evaluation in addition to the accuracy, precision, and recall metrics used in the previous version. The addition of these extra metrics did not change the conclusions of this study.

Overall, while the methodology used in this Chapter is more comprehensive than the previous version, the high-level conclusions drawn from both versions of this work are the same — that is, KNN is the best performing classification method and web page classification can be applied to different web-related application domains including traffic characterization and simulation modeling.

[17] Or rather, any labeling scheme.

4.6 Contributions and Concluding Remarks

This chapter advances the state of the art in traffic classification both methodologically as well as by offering new insights. We make the following key contributions:

1. *Data collection:* We use five different modern browser platforms to conduct and analyze downloads of 3345 web pages, all belonging to the top-250 most popular web sites. Overall, we analyze more than 100,000 web page downloads. For each download, we process TCP/IP data as well as collect the ground-truth about the *type* of the corresponding web page, based on four classification schemes.

2. *Feature extraction and selection:* We process the TCP/IP headers to derive 216 features, including temporal and multi-flow features as well as their statistical derivatives. We then conduct a systematic analysis of these features to identify robust and discriminatory features — to the best of our knowledge, this is the first work that argues for, and explicitly considers, *consistency* (across different browser platforms) and *stability* (over time) while selecting robust features.

3. *Web page classification:* Using the selected robust features, we then evaluate how effectively can these help classify web page downloads according to each of the four diverse labeling schemes. We find that while mobile-targeted and video downloads can be KNN-classified with more than 90% accuracy, the genre- and navigation-based categories can be classified with a somewhat lower accuracy.

4. *Applicability of classification:* We then evaluate the impact of our work on traffic modeling applications. Specifically, we study the distributions of (i) traffic modeling parameters, as well as (ii) properties of the generated traffic — we find that these are *statistically indistinguishable* from distributions derived using ground-truth labels.

In summary, methodologically, we (i) establish the need for (and present metrics for) finding consistent (across browsers) and stable (over time) informative features, (ii) use features that span *multiple* TCP/IP flows, and (iii) use a statistical framework to study the applicability of classification results in the context of real-world applications. Our analysis leads to new insights on which multi-flow TCP/IP features are robust and informative for web page classification, as well as what type of web page classes that can be successfully identified using these.

CHAPTER 5: WEB PAGE SEGMENTATION

In the previous chapter, we explored web page classification as a method for analyzing modern web page traffic using anonymized TCP/IP headers [Maciá-Fernández et al., 2010, Sanders and Kaur, 2015b]. While the results were promising, we only considered traffic traces that were (i) associated with a known user, (ii) Web-only, and (iii) separated according to individual web page downloads — also, the web page downloads were allowed to continue for 60s. This is in stark contrast to real traffic traces which are comprised of a mixture of traffic from different applications (e.g., HTTP, P2P, and FTP), from multiple users that may use different client platforms, and from multiple web page downloads.[1] These additional characteristics make it non-trivial to separate/demultiplex real traffic on a per-user, Web-only, and individual web page download basis.[2] Some of these limitations have been explored in the literature. For instance, filtering traffic according to different application types when using only anonymized TCP/IP headers has been explored in the traffic classification literature [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b]. Other limitations may be addressed by strategically measuring traffic at vantage points where traffic is less obfuscated. For example, demultiplexing traffic from multiple users is more feasible in environments where IP addresses are not shared (e.g., NATs) — in such an environment, it is reasonable to assume that each IP address corresponds to a single user [Smith et al., 2001].[3]

One limitation, or problem area, that has not been recently explored and is not likely addressable by strategic traffic measurement is that of demultiplexing web page traffic according to individual web page downloads. Thus, an approach that is able to separate, or segment, web traffic into individual web page downloads is needed in order for web page classification to be applicable "in the wild." In fact, many other web page traffic analysis applications, such as web page performance analysis [Huang et al., 2010, Butkiewicz et al., 2011, Dhawan et al., 2012] and privacy analysis [Yen et al., 2009, Dyer et al., 2012, Chierichetti et al., 2012], also assume the availability of an approach that can do this.

[1] It is likely that real traffic traces include client platforms that we did not consider in previous chapters.

[2] These web page downloads may not have been allowed to load for 60s.

[3] Although, it is important to note that strategic measurement of traffic may not be possible in all environments.

These approaches, referred to as web page segmentation approaches, are needed in order for these traffic analysis applications to be usable "in the wild." For example, in the area of web page performance analysis, web page traffic measurements lose meaning in the absence of a web page segmentation approach because the traffic for the individual web page downloads are not grouped [Ihm and Pai, 2011] — this limits the performance analysis to the web object-level, which is, at times, insufficient [Newton et al., 2013]. Similar issues occur in the area of privacy analysis which focuses on making inferences on obfuscated traffic [Yen et al., 2009, Dyer et al., 2012, Chierichetti et al., 2012]. Indeed, the absence of a web page segmentation approach limits what can be inferred from web traffic including the (i) number of web pages a user downloaded and (ii) the inter-arrival time between web page downloads — these are useful for both web performance and privacy applications [Neasbitt et al., 2014, Chierichetti et al., 2012, Wang et al., 2013].

Web page segmentation using anonymized TCP/IP headers is usually done by identifying the beginning of a web page download in traffic — the assumption being that all traffic detected between the beginning of the current download and the beginning of the next download correspond to the same web page [Maciá-Fernández et al., 2010, Newton et al., 2013, Mah, 1997, Barford and Crovella, 1998]. Detecting the beginning of a web page download using only anonymized TCP/IP headers is challenging because identifiers/references do not exist in TCP/IP headers that map each segment to a web page download. Instead, the characteristics of the TCP/IP traffic must be studied in order to determine robust heuristics that are able to detect new web page downloads — a reliable method should be able to do this while limiting the number of false positives (i.e., incorrectly identified web page downloads) and false negatives (i.e., missed web page downloads). Please note, however, that HTTP headers include fields, such as the host and referrer fields, that have been shown to be useful for grouping web objects that belong to the same web page download [Neasbitt et al., 2014, Xie et al., 2013]. We do not consider these methods in this work because we assume that only anonymized TCP/IP headers are available.

Web page segmentation using anonymized TCP/IP headers has been addressed in the past using idle time-based approaches [Mah, 1997, Barford and Crovella, 1998]. These approaches estimate the beginning of a web page download by detecting if network activity levels exceed a threshold after a predefined amount of idle time [Maciá-Fernández et al., 2010, Newton et al., 2013, Ihm and Pai, 2011, Mah, 1997, Barford and Crovella, 1998]. The assumption of these methods is that new web page downloads usually transmit a certain amount of traffic (a minimum threshold) and occur after the previous web page download has finished (a minimum idle time) [Barford and Crovella, 1998]. These assumptions are reasonable for static web pages that do not transmit additional traffic after they have been rendered by a browser — thus, an increase in traffic after an idle-time delay usually corresponds to a user initiating a new web page download. However, idle time-based methods were proposed over a decade ago and have not been comprehensively evaluated on modern web page traffic which consists of dynamic web pages that may transmit traffic after the web page has rendered. This increased transmission of traffic by modern web pages, reduces the amount of idle time on a link — this reduced idle time will likely make it more difficult for idle time-based methods to detect new web page downloads [Mah, 1997, Barford and Crovella, 1998, Newton et al., 2013].[4] It is not known whether idle time-based methods will work for modern web traffic where idle time periods are reduced.

In this chapter, we bridge this knowledge gap by conducting the first comprehensive evaluation of web page segmentation approaches on web page traffic in over a decade. We do this by evaluating both idle time-based methods (which have been previously proposed) and change point detection methods (which have not been considered before) on synthetically generated and real-user browsing data. We also conduct case studies that investigate whether web page segmentation methods can be applied to domains that are restricted to using only anonymized TCP/IP headers — the two application domains that we consider are web page classification and user behavior modeling. For the web page classification case study, we investigate whether the performance of web page segmentation approaches impact the applicability of web page classification to real traffic traces. We do this (i) to "put together" the concepts researched in this dissertation[5] and (ii) to determine the strengths and limitations of using web page segmentation and web page classification in tandem. For the user behavior modeling case study, we investigate whether web page segmentation methods can approximate web browsing metrics such as the number of web pages a user downloaded and the average inter-arrival time between web page downloads.

In the rest of this chapter, we describe the web page segmentation problem, existing idle time-based approaches, and results from a preliminary analysis of the time series characteristics of modern web page traffic in Section 5.1; details on the change point detection methods and evaluation methodology in Section 5.2 and Section 5.3, respectively; results of our performance evaluation using synthetic and real user browsing data in Section 5.3.2 - 5.4; results of our case studies and high-level conclusions in Section 5.5 -

[4] Please note that other modern web features, such as multi-tab browsing and automatically generated traffic, may also reduce the amount of idle time observed in traffic.

[5] These include incorporating diversity of client platform when designing traffic analysis techniques (Chapter 3), web page classification (Chapter 4), and web page segmentation (Chapter 5).

5.6.

5.1 High-Level Analysis of Time Series Characteristics of Modern Web Page Traffic

5.1.1 Defining Web Page Segmentation using Anonymized TCP/IP Headers

In order to evaluate web page segmentation approaches we must first define a “web page download.” This is non-trivial because modern web pages are complex and may consist of many components including: objects from different servers, flash content, advertisements, text, css technology, AJAX technology, javascript, and more. In the context of this work, a web page download is initiated when a user either (i) enters a URL into the browser window, (ii) clicks on a hyperlink, and/or (iii) enters text in a browser that results in a new web page such as a search query and its corresponding result web page — please note that the web page download itself is the traffic that results from these actions.[6] This is the same definition for a web page download that was used in [Neasbitt et al., 2014]. The web page segmentation problem that we address in this work involves identifying the time in which a web page download starts when only anonymized TCP/IP headers are available for analysis. Some of the implications of these definitions are described below:

- *What about automatically generated traffic?* Web page interactions *must* initiate a new web page download or reload a web page[7] in order for us to consider it as a web page download. Dynamic and automated web page interactions such as instant search or video auto-play that *update* an already downloaded web page or simply download additional web objects are not categorized as a web page download — the traffic generated by these interactions correspond to the web page in which it originated. Please note that some web page interactions, such as instant search, may either download a new a web page or update an already downloaded page — in this work, we use changes in the URL in the browser as a trigger for new web page downloads.
- *When does a web page download end?* A web page download ends when the last packet is transmitted between the client and web servers involved *that* web page download event. This means that web page

[6] Please refer to Chapter 2 for specifics regarding our definition of a web page, web object, and other terms frequently used in this chapter.

[7] Please note that a web page download correspond to traffic being transmitted over the network. So, while some web objects can be satisfied by a local cache on the client, at least one object must be obtained from a web server or proxy in order to be considered a web page download.

downloads that transmit traffic after the web page has rendered on a browser may continue for an indefinite amount of time. It is difficult to predict when a web page download ends unless the browser window in which the web page is rendered is closed. This is because many web pages leverage technologies that automatically generate traffic (e.g., AJAX technologies).

- *What about web page downloads that have overlapping traffic?* Lets assume that web page download A starts at time t_1 and web page download B starts at a later time t_2. Indeed, it is possible for traffic from web page download A to be transmitted after web page download B starts — this results in overlapping traffic. Precisely determining which traffic corresponds to which web page download has not been addressed in the literature when only anonymized TCP/IP headers are available [Maciá-Fernández et al., 2010, Newton et al., 2013, Mah, 1997, Barford and Crovella, 1998].[8] This is likely because there are many scenarios where there is simply not enough information available to do so properly. For instance, it is possible that a user will browse multiple web pages from the same web site that may contact servers that share the same IP address. In this case, TCP connections that are established by the first web page, but after the second web page starts downloading, may easily get confused with the second web page. Such realistic and challenging scenarios are not likely addressable with the limited information present in anonymized TCP/IP headers — although, methods have been proposed which address such challenges when HTTP headers are available [Ihm and Pai, 2011, Xie et al., 2013]. Instead of precisely determining which traffic corresponds to which web page download, prior work usually estimate the web page download start time and make assumptions on which traffic corresponds to different web page downloads [Maciá-Fernández et al., 2010, Newton et al., 2013, Mah, 1997, Barford and Crovella, 1998]. This essentially relaxes a very difficult problem into something that is feasible.

  In this work we simply assume that all traffic that is transmitted via TCP connections established after time t_1 and before time t_2 correspond to web page download A, while all traffic transmitted via TCP connections established after time t_2 and before the next web page download correspond to web page download B and so on.[9] This assumption, while not true, is reasonable since web applications are usually designed to load as quickly as possible [Gavaletz et al., 2012, Ihm and Pai, 2011]. This

[8] This past work only considers identifying the start of a web page download.

[9] Most prior work make similar assumptions [Maciá-Fernández et al., 2010, Newton et al., 2013, Mah, 1997, Barford and Crovella, 1998].

assumption is also reasonable since human users are likely only actively browsing a single web page at a time [Jansen et al., 2008] — in fact, some browsers limit the traffic transmitted by web pages that are inactive (e.g., pages in background tabs) [CCr]. We also assume that each user has a unique IP address — this assumption eliminates the possibility of observing overlapping traffic which may occur when multiple users share an IP address (e.g., NATs). Please note that the work in this chapter is likely not applicable to scenarios in which these assumptions are unreasonable.

- *Why explicitly consider when a web page download starts?* Knowing the time in which a web page download starts is critical for a number of applications. For example, in the area of user modeling the difference in the time between two consecutive web page downloads is needed for gauging user interest in a page [Neasbitt et al., 2014] — this can be easily derived from the times in which web page downloads are initiated. Also, understanding when a web page download starts is an important step in determining which portions of traffic (TCP segments) corresponds to which web page download [Mah, 1997, Ihm and Pai, 2011, Wang et al., 2013] — this may be useful for web page classification and is explored in this chapter.

A *web page segmentation approach* is a method which processes web traffic in a manner that detects the beginning of a web page download. Indeed, the performance of a web page segmentation approach is influenced by the characteristics of web page traffic. In Chapters 3 and 4, we discussed the characteristics of modern web page traffic using cumulative distributions of different web page traffic features. However, cumulative distributions do not illustrate the time series characteristics of web page traffic features. For instance, they do not reveal how many bytes are typically transferred during the first 5, 10, or 15 seconds of a web page download. In this section, we address this lack of understanding by performing a high-level analysis of the time series characteristics of modern web page traffic. We do this by first providing a brief overview of existing web page segmentation approaches and describing some of the reasoning behind them. We then analyze the time series characteristics of modern web page traffic. And lastly, we discuss the impact that such characteristics have on web page segmentation.

5.1.2 Idle time-based Approaches

Existing web page segmentation methods use idle time heuristics to detect web page downloads. Details on the approaches that use such heuristics are described below:

- *Idle time method:* Most prior methods identify web page downloads when a traffic feature of interest (i.e., a byte or a TCP segment with the SYN flag set) is observed after *I* seconds of idle network activity. These methods assume that any increase in network activity after *I* seconds is the result of a new web page download. Typical values for *I* are 1 second [Mah, 1997, Maciá-Fernández et al., 2010, Hernández-Campos et al., 2003a] and 10 seconds [Yen et al., 2009].

- *Idle time method with a threshold for activity:* [Newton et al., 2013] proposed a method that detects web page downloads only when the number of observed SYN segments exceeds *T* (a predefined threshold for network activity), while enforcing a minimum idle time of *I* between consecutive page detections. The number of SYN segments feature is used instead of the number of bytes feature to represent network activity, because of the high amount of variance in the latter. The recommended values for *T* and *I* are 2 and 2.5s, respectively. This approach can be interpreted as a generalization of the previously discussed idle time method with an additional threshold for network activity. It is also possible to apply this method using the number of bytes feature instead of the number of SYN segments feature.

The above approaches have not been thoroughly evaluated to determine their strengths and weaknesses. In the next section, we take a closer look at modern web page traffic to determine the challenges associated with (i) web page segmentation using idle time-based approaches and (ii) web page segmentation in general.

5.1.3 Analysis of Time Series Behavior of Modern Web Page Traffic

We next perform a preliminary analysis of the time series characteristics of modern web page traffic. We do this to (i) understand the time series characteristics of the web page traffic and (ii) determine some factors that may impact web page segmentation performance including traffic feature selection (e.g., number of bytes or number of SYN segments) and browser choice. The data collection methodology, details of the time series representation of web traffic, and the implications of this analysis are provided in this section.

Data Collection Methodology We adopt a client-side measurement methodology for the same reasons provided in Chapters 3 and 4. More specifically, we use scripts to *automatically* download a diverse set of web pages using the Google Chrome v 37.0.2062.124, Firefox v 32.0.2, and Safari v 7.1 (9537.85.10.17.1) browsers on a Mac desktop running OSX 10.9.4.[10] We focus on Mac OS X 10.9.4 over the Windows 7 and Linux operating systems in this work because there is a tool exclusive to Mac OS X platforms, Applescript, which enables the automated control of *all* browsers and other applications — other tools that were available on different operating systems, including Selenium Web driver, are limited by browser type and version [SeleniumHQ]. We do, however, collect and analyze time series data generated using multiple browsers running Windows 7 to determine whether the high-level time series characteristics of web page traffic is impacted by operating system.

Details of the traffic generation and capture process is provided below:

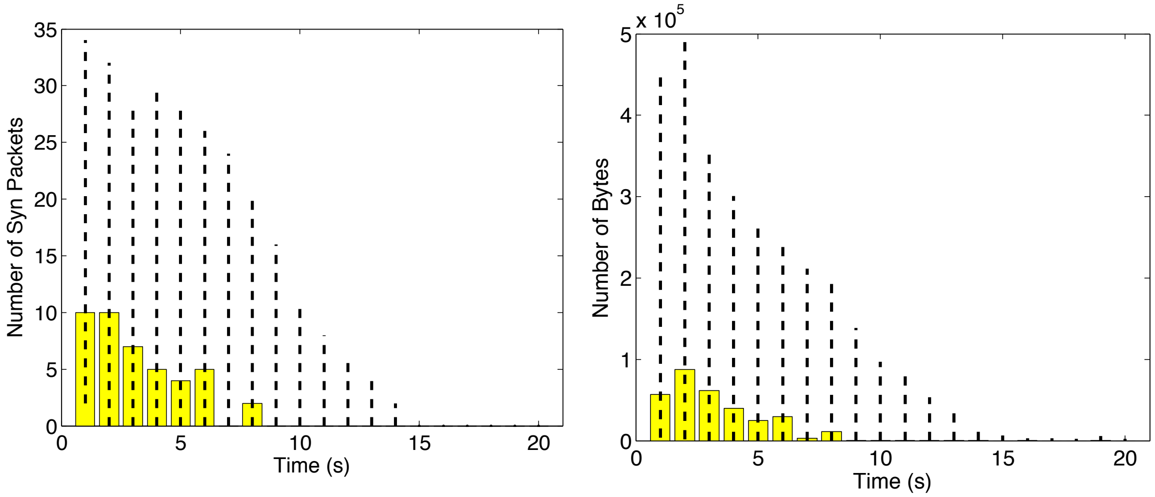
1. Start packet capture tool (tcpdump)
2. Start a browser with a web page URL as an argument
3. Close the browser after 60 seconds
4. Clear the local DNS resolver cache
5. Clear the browser cache
6. Go to Step 1 using a new URL

Time Series Representation of Traffic We then process the resulting TCP/IP header traces to generate two types of time series, Xb_t and Xs_t. The resolution (the time interval over which bytes/SYNs are counted), r, is a tunable parameter of Xb_t and Xs_t. For simplicity, we use $r = 1$ second in this section.[11] Xb_t is a time series that represents the *number of application layer bytes* that were observed within time $[t, t+r)$ for each t, while Xs_t is a time series that represents the *number of SYN segments* (including SYN+ACKs) that were observed within time $[t, t+r)$ — please note that we only consider HTTP and HTTPS traffic. These features are the two primary features that have been used for web page segmentation when using TCP/IP headers in past work [Maciá-Fernández et al., 2010, Newton et al., 2013, Ihm and Pai, 2011, Mah, 1997].

[10] A complete list of the web pages downloaded is provided in Appendix 2.

[11] We discuss the influence that r has on web page segmentation performance in Section 5.3.2.

Why Use SYNs Over Bytes? Figure 5.1(a) shows the time series for the number of SYNs observed during each of the first 20 seconds of a page download.[12] The bars correspond to the median number of SYN segments observed for all web pages (across all browsers), while the dashed lines correspond to the 10 and 90 percentiles. We find that there is a large amount of variation in the number SYN segments transmitted when loading a web page. In particular, the median number of SYN segments observed during the 1st second of the download is less than a third of its corresponding 90th percentile. This general observation holds true for each subsequent second of the download. Figure 5.1(b) shows similar results for the number of bytes feature except the difference between the median and the 90th percentile observation is much higher (∼ up to 10x). Thus, the SYN feature is likely the more stable metric to use for web page segmentation.



(a) Number of Initial SYNs (first 20s)

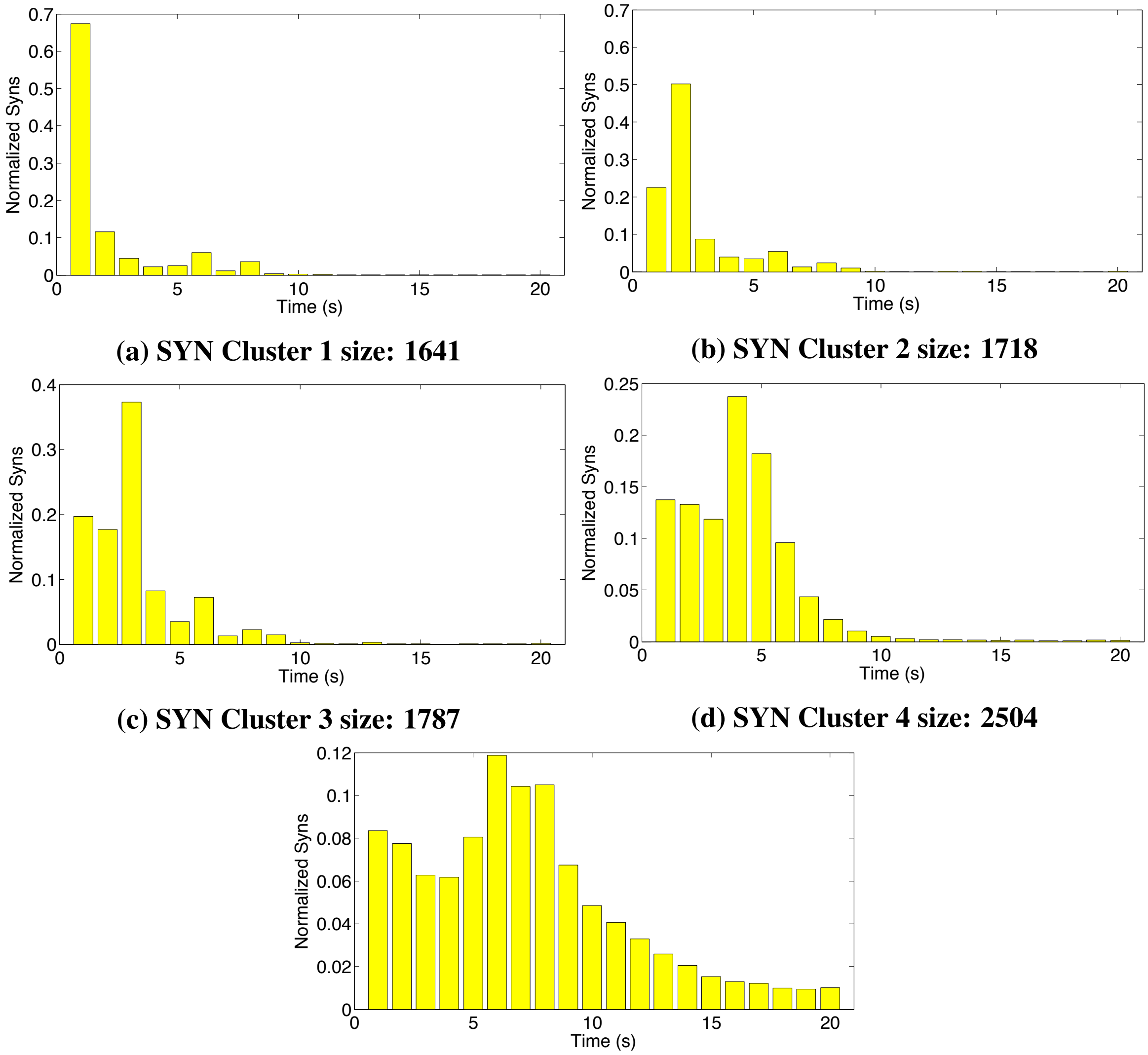
(b) Number of Initial Bytes (first 20s)

Figure 5.1: Overview of the variability in the magnitude of the traffic features across web pages.

Why Idle-time Based Methods May Not Work? Figures 5.1(a) and 5.1(b) show that, over time, there is a general monotonically decreasing behavior in the traffic metrics for a page download that peaks fairly quickly. If this behavior is universal, a simple peak detection heuristic may suffice as a web page segmentation approach.

However, the high 90 percentile values in several of the first 8-9 seconds suggest that peaks may occur anytime in this duration. To study this, we analyze the 20-second SYNs/bytes time series for *each* of the 10872 web page downloads (3 browsers × 3614 web pages). We use cluster analysis to analyze the

[12] We restrict our analysis to the first 20 seconds of the web page download process because most of the traffic occurs within this time-frame — similar results were observed in [Ihm and Pai, 2011].



(a) SYN Cluster 1 size: 1641

(b) SYN Cluster 2 size: 1718

(c) SYN Cluster 3 size: 1787

(d) SYN Cluster 4 size: 2504

(e) SYN Cluster 5 size: 3212

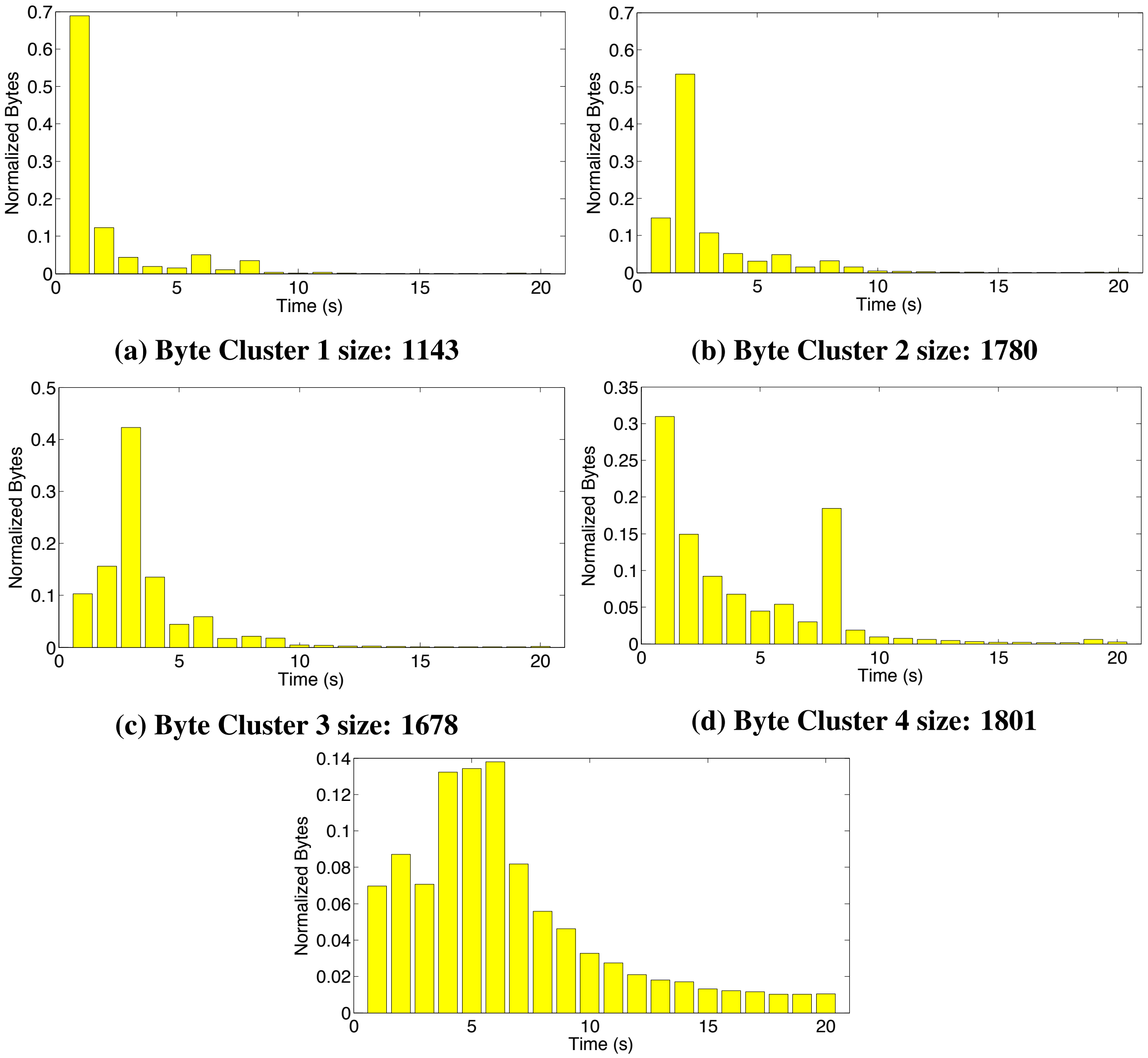
Figure 5.2: K-mean cluster centroids for web page download time series (SYNs).

shapes of these time series — we normalize each time series by the magnitude of the number of SYNs/bytes transferred for its download so that the clusters will be dominated by shape instead of absolute differences in traffic metrics. We use K-means, a popular clustering approach, for this analysis [Hartigan, 1975]. The input to the K-means algorithm is essentially N web pages with 20 features (i.e., one feature for each second). K-means also takes a parameter, k, that specifies the number of clusters to generate. We empirically test multiple values of k and find that $k = 5$ works well for our purposes.[13]

The centroids for the 5 clusters that we find for the number of SYNs and number of bytes time series are

[13] The values of k that we tested were 2, 3, 4, 5, 6, 7, 8, 9, and 10.

shown in Figure 5.2 and Figure 5.3, respectively.



(a) Byte Cluster 1 size: 1143

(b) Byte Cluster 2 size: 1780

(c) Byte Cluster 3 size: 1678

(d) Byte Cluster 4 size: 1801

(e) Byte Cluster 5 size: 4474

Figure 5.3: K-mean cluster centroids for web page download time series (Bytes).

We find that:

- Figure 5.2(e) shows that some web pages may transmit SYN segments for over 20 seconds — similar results are shown for the number of bytes time series provided in Figure 5.3(e). It is possible that this long load time is likely due to javascript and other traffic that occurs in the background after the page renders. This is an important observation because it shows that there are cases where two web pages can be requested within 20 seconds of each other without having an idle time period. In such a scenario, an idle time-based approach would simply guess that these two web page downloads correspond to a single page.

- The above suggests that peak detection, instead of idle time detection, may be a more robust approach. These clusters shown in Figure 5.2 and Figure 5.3 show that the time series for web page downloads tend to have either 1 or 2 peaks. For the number of SYNs time series centroids, approximately half of the web page downloads for the exhibit two peaks (Figure 5.2(e) and to a lesser extent Figure 5.2(d)), while the other time series only exhibit one primary peak (Figures 5.2(a), 5.2(b), and to a lesser extent 5.2(c)). We observe similar results for the number of bytes time series — that is, two of the centroids (Figure 5.3(e) and Figure 5.3(d)) exhibit two peaks and the other centroids (Figures 5.3(a), 5.3(b), and extent 5.3(c)) exhibit only one peak. These results imply that peak detection approaches themselves must be designed to be robust to this diversity — else, two consecutive web page requests with a single peak each may get confused by a single web page request that has two peaks. Consequently, a segmentation method may consider the two peaks as a single web page request (depending on how close the peaks occur) or identify each of the two peaks as an individual web page download (essentially two pages).

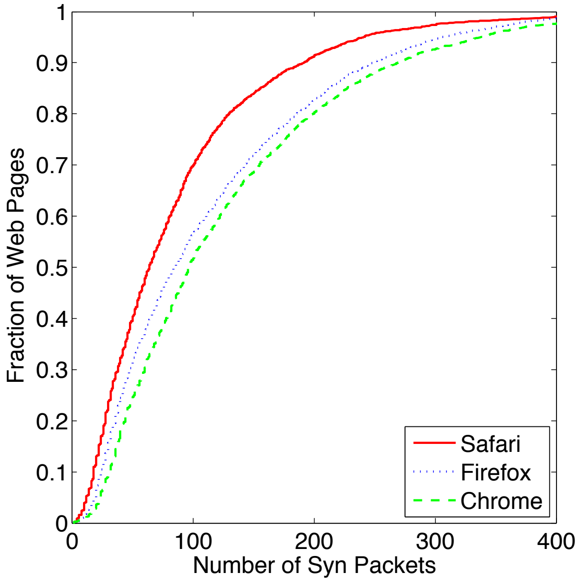


Figure 5.4: Overview of the variability in the magnitude of the traffic features across different browsers and web pages.

Do Browsers Impact Web Page Segmentation? Figure 5.4 shows a plot of the cumulative distribution of the number of SYN segments transferred when downloading a page. We find that browser choice has an impact on the traffic features, where the number of SYNs transferred by the Chrome browser is larger than the other browsers — these observations are similar to those made in Chapter 3. This observation means that it is likely that the differences in the traffic across browser may impact the performance of different web page segmentation approaches.

Does Operating System Choice Impact Time Series Characteristics? As noted before, we investigated whether operating systems impact the time series characteristics that we observe. A summary of the similarities and minor differences between the time series characteristics across operating system is provided below:

- The variance in the values for the number of bytes time series is higher than the variance in the values for the number of SYNs time series. This result is consistent for the Mac OSX 10.9.4 and Windows 7 operating systems and is shown in Figure 5.1 and Figure 6.7 (in Appendix 12), respectively.
- The time series that we observe usually increases and decreases at least once — the number of times this occurs is equivalent to the number of peaks present in the time series. In general, the time series that we observe exhibit one or two peaks. This result is consistent for the Mac OSX 10.9.4 and Windows 7 operating systems and is shown, for the number of SYNs time series, in Figure 5.2 and Figure 6.8, respectively — please refer to Figure 5.3 and Figure 6.9 for the number of bytes time series for these operating systems.
- Most of the web page time series ($\approx$ 50%) are non-zero for *only* the first 10 seconds. That is, most web page downloads finish in less than 10 seconds. However, a significant fraction of web page time series (over 30%) is non-zero for over 20 seconds. This result shows that web page download time can vary significantly across different web pages. We observe these results for both operating systems and for both the number of SYNs time series (Figure 5.2 and Figure 6.8) and the number of bytes time series (Figure 5.3 and Figure 6.9).
- In Chapter 3, we found that the total number of SYNs and bytes observed for each web page download is similar across operating systems. Despite this, we observe minor disparities in the number of bytes and SYNs transmitted on a per second basis across operating system. For instance, on average there are slightly more bytes and SYNs transmitted (10-15%) during the first second of the web page

download on the Windows 7 client than the Mac OSX 10.9.4 client — the Mac OSX 10.9.4 client compensates for this increase later during the web page download. Please refer to Figure 5.1 and Figure 6.7 (in Appendix 12) for a visual representation of such minor differences.

In the next section, we consider some change point detection techniques to achieve the above peak detection based segmentation. In subsequent sections, we comprehensively evaluate different web page segmentation approaches with modern web traffic to understand the extent to which the above challenges impact performance in practice.

5.2 Change Point Detection Methods

In the previous section we found that there may be very little idle time during modern web page downloads — thus, idle time-based methods are not likely to be effective for the modern traffic mix. Given this, it is clear that modern web page segmentation methods should be robust to traffic that is rarely idle. We believe that change point detection methods, a class of time series approaches that are able to detect significant increases or decreases in time series data (essentially peak detection), will be effective for modern web traffic since they do not require idle activity to work. There are three primary types of change point detection methods that are used in the literature: 1) Heuristic/Sliding window methods; 2) Regression methods; and 3) Probabilitistic/Stochasitic methods [Newton et al., 2013, Rabiner, 1989, Tibshirani and Wang, 2008, Bleakley and Vert, 2011, Rabiner, 1989]. We provide background on these approaches and describe how they are applicable to the web page segmentation problem in this section.

Time series Definition and Methods Overview: Let X_t be the time series of the number of SYN segments or bytes observed at a timescale of r, for a given source IP — that is, at each time t, X_t represents the number of SYNs or bytes observed in $[t, t+r)$. This raw time series signal, X_t, has a high degree of noise, which makes it difficult for change point detection methods to be applied directly. A moving average filter (i.e., a simple lowpass filter) is used to reduce noise in the traffic signal. Given X_t, a moving average is defined as: $M_t = \sum_{i=0}^{k} X_{t-i}/k$ for all $t \in (k, N)$, where N is the number of elements in the time series and k is the lag of the moving average. We use M_t as an input for the change point detection approaches investigated in this work because it has less noise than X_t — this is a common preprocessing step for analyzing time series data [Nose-Filho et al., 2011, Chen and Chen, 2003].[14]

5.2.1 Heuristic Method for Change Point Detection

We first consider a simple heuristic method for change point detection that can be applied to web page segmentation. At a high-level, this method detects web pages when the thresholds for several parameters including idle-time and the number of bytes/SYNs observed are exceeded — this type of method is perhaps the simplest method for change point detection. More specifically, we propose a heuristic that takes three parameters, $T_{threshold}$, L_{delay}, and I_{idle}. Web page detections are identified when X_t is greater than or equal to $T_{threshold}$ [Newton et al., 2013], there has been no traffic observed in the last I_{idle} time units, and the time since the previous detection is more than L_{delay} — all of these thresholds must be exceeded in order to detect new web pages. Please note that this method is different from idle time-based approaches because it can work realistically even with zero idle-time (i.e., I_{idle} can be 0).[15] Intuitively, an idle time value of 0 for the idle time method will identify every part of traffic activity as a new page. In other words, the number of web pages will be equal to the number of non-zero instances in traffic — clearly, this will result in a high false positive rate. We add the L_{delay} term so that methods can detect new web pages in scenarios in which traffic activity is rarely idle. Because of this, we consider this heuristic approach more of a threshold-based change point detection method with an option for idle-time.

5.2.2 Fused Lasso Regression

Regression and stochastic models have been used to solve change point detection problems in other fields. In particular, fused lasso, a regression model, and the hidden Markov model (HMM), a stochastic model, have been used in the fields of computational biology and speech recognition [Rabiner, 1989, Tibshirani and Wang, 2008, Bleakley and Vert, 2011].

Fused lasso regression performs a very specific task: outputs an “optimal” denoised version of an input signal that approximates a piecewise constant function (i.e., a step function) [Tibshirani and Wang, 2008] — such piecewise constant (or in some areas piecewise linear) approximation is perhaps the most popular regression approach for segmenting time series data [Keogh et al., 2004]. Although this method has been

[14] Please note that we initially explored using X_t instead of M_t for web page segmentation using change point detection methods. However, the performance that we observed when using X_t was inferior to the performance when using M_t.

[15] Similarly, L_{delay} can also be 0.

widely used for computational biology and clustering applications, to the best of our knowledge it has not been applied to network traffic analysis.

The convex optimization problem for fused lasso regression is given by:

$$argmin_Y 1/2||M-Y||_F^2 + \lambda||Y||_1 + \mu||DY||_1 \tag{5.1}$$

where $||A||_1 = \sum_{i,j} |A_{i,j}|$ is the L1 norm and $||A||_F = \sqrt{\sum_{i,j} A_{i,j}^2}$ is the frobenious norm for arbitrary A, D is an $N \times N$ matrix such that DY is a differenced Y vector (i.e., $DY_i = Y_i - Y_{i+1}$), M is the input moving average vector, and Y is the denoised version of M which solves the above optimization problem. This convex optimization problem computes an optimal denoised signal, Y, that minimizes the L1 norms of Y and DY. Intuitively, minimizing the L1 norm of Y reduces measurements that are small, while minimizing the L1 norm of DY groups values of similar levels together. In other words, this problem (i) eliminates small traffic signatures (background noise or AJAX traffic) and (ii) groups traffic that are likely to be linked to a single event together (individual web page requests). The λ and μ terms in the above equation are parameters that impact the solution of the optimization problem (Y) by influencing the L1 norms of Y and DY — in short, increasing λ and μ increases the degree in which traffic measurements are reduced and grouped together, respectively.

The convex optimization problem in Equation 5.1 can be solved via Alternating Directions Method of Multipliers (ADMM) [Boyd et al., 2011]. Algorithm 1, denoted FLADMM, provides the pseudocode for this, where $st(B,\alpha) = max(B-\alpha,0)$ for any non-negative B.[16]

Web page detections are identified as the point where Y_t changes behavior from monotonically non-increasing to increasing — this function is denoted as DetectChangeFL(A). The complete regression method, SegmentationFL, is summarized in Algorithm 2. It is important to note, however, that SegmentationFL is identifying increases in web traffic — it is conceivable that events other than new web page downloads (such as scrolling down on some pages) may result in similar increases in traffic.

[16] Line 2-3 of FLADMM includes a pseudo-inverse operation which is implemented in MATLAB using gaussian elimination (i.e., $A^{-1}B$ = A\B).

Algorithm 1 FLADMM

```
input: $M$, $\lambda$, $\mu$
initialize: $\gamma$, $\delta$, $\zeta$, $Y_1$, $Y_2$, $Y_3$ to a $N \times 1$ vector of zeros
**for** $k = 0, 1, ...., MAXITER$ **do**
    $Y = \begin{bmatrix} I \\ I \\ D \end{bmatrix}^{-1} \begin{bmatrix} Y_1 + \zeta \\ Y_2 + \delta \\ Y_3 + \gamma \end{bmatrix}$
    $Y_1 = \begin{bmatrix} I \\ I \end{bmatrix}^{-1} \begin{bmatrix} M \\ Y - \zeta \end{bmatrix}$
    $Y_2 = st(Y - \delta, \lambda)$
    $Y_3 = st(DY - \gamma, \mu)$
    $\zeta = \zeta + (Y_1 - Y)$
    $\delta = \delta + (Y_2 - Y)$
    $\gamma = \gamma + (Y_3 - DY)$
**end for**
output: $Y$
```

Algorithm 2 SegmentationFL

```
input: $X$, $k$, $\lambda$, $\mu$
$M = MovingAverage(X, k)$
$Y = FLADMM(M, \lambda, \mu)$
$B = DetectChangeFL(Y)$
output: $B$
```

5.2.3 Hidden Markov Model

Stochastic models incorporate random variables and probability distributions in the modeling process. Hidden Markov Models (HMMs), perhaps the most common stochastic method for change point detection, are commonly used for change point detection in speech processing [Rabiner, 1989, Fridlyand et al., 2004]. An HMM is essentially a Markov model where the current state is unknown. An overview of the elements of an HMM and how it can be applied to web page segmentation is provided below:

- *Finite number of States:* An HMM has a finite number of states, N, that are denoted as S_0, S_1, S_2, ..., S_{N-1}. A state at a given time t is denoted as X_t. The state transition probabilities of an HMM follow the Markov property (i.e., $Pr(X_{t+1} = S_j | X_t = S_i)$).[17] The thing that makes an HMM different

[17] The Markov property is a simplifying assumption that is rarely true. However, it has tremendous benefits and utility for real applications.

from a regular Markov model is that the states are hidden. Although the states are hidden, there is some significance attached to the states that stem from observable variables. In terms of web page segmentation, it makes sense to identify two hidden states:1) S_0 = User idle; and 2) S_1 = User downloading page.

- *Observation symbols per state:* As noted before, the transition between states, or rather the state itself, is hidden in an HMM. Instead, the current state must be observed from some measurable phenomena, or more formally, observation symbols. An HMM has Q observation symbols that are denoted as V_0, V_1, V_2, ..., V_{Q-1}. In terms of web page segmentation, we identify two observation symbols: 1) V_0 = A non-increase in a traffic feature (i.e., $M_t \leq M_{t-1}$); and 2) V_1 = An increase in a traffic feature (i.e., $M_t > M_{t-1}$).[18]

- *Observation symbol probability in state j:* Observation symbols, V, are the only information available to infer the hidden state, S, of an HMM. A key assumption of an HMM is that hidden states exhibit probability distributions of observation symbols. The observation symbol probability distribution in S_j, is B_{jk}, where $B_{jk} = P$ [V_k at t — $X_t = S_j$], $1 \leq j \leq N$, $1 \leq k \leq Q$.

- *State transition probability:* HMMs also include a state transition probability matrix, A_{ij}. A_{ij} encodes the probability of going from S_i to S_j.

Given appropriate values of N, M, B, and A, the HMM can be used to determine the state sequence X_1, X_2, ... , X_T that was most probable provided an observation sequence O_1,O_2, ..., O_T, where T is the length of the sequence. Here, each O_t corresponds to one of the observation symbols, V. In the context of the web page segmentation problem, O_1,O_2, ..., O_T are measurements in time that correspond to traffic features, and X_1, X_2, ... , X_T are the hidden states that correspond to web page segments. The Viterbi algorithm can be used to solve for the most probable state sequence, X_1, X_2, ... , X_T, given N, M, B, A, and O_1,O_2, ..., O_T [Forney Jr, 1973] — this procedure is denoted as HMMViterbi(N,M,B,A,O). *Web page detections* are identified as the point when the states in X_{t-1} and X_t changes from S_0 to S_1 — this function is denoted as DetectChangeHMM(X). The complete SegmentationHMM(X) is provided in Algorithm 3.[19]

[18] M_t is the moving average of the input time series described previously.

[19] MovingAverage(X,k) outputs a moving average of raw time series X with lag k that has been processed to yield a sequence of observational symbols consisting of V_0 and V_1.

Algorithm 3 SegmentationHMM

```
input: $X$, $k$, $A$, $B$, $N$, $M$
$J = MovingAverage(X,k)$
$Z = HMMViterbi(N,M,B,A,J)$
$C = DetectChangeHMM(Z)$
output: $C$
```

5.3 Evaluation With Synthetic Data

5.3.1 Synthetic Data Traffic Generation

We first conduct a controlled evaluation of segmentation approaches using web traffic generated by web requests made by automated scripts (non-humans). Ideally, if good models existed for how a typical user browses the web, the scripts could use these for generating per-user browsing streams to evaluate the ability of different approaches to segment the resulting traffic. However, such models do not exist.[20] Thus, we use a simple approach to study, and control for, the key factors that are likely to impact web page segmentation performance — these include web page inter-arrival time, the number of tabs that are simultaneously open in a browser, and browser choice [Sanders and Kaur, 2014a, Miller et al., 2014]. This approach is described in this section.

Web page inter-arrivals: To the best of our knowledge, previous evaluations of web page segmentation methods do not explicitly investigate the impact of web page inter-arrival times (IATs) on performance [Ihm and Pai, 2011, Xie et al., 2013, Newton et al., 2013, Hernández-Campos et al., 2003a]. This is a problem since, from our analysis in Section 5.1, it is clear that inter-arrival time may impact web page segmentation performance. Since we do not have any clear evidence that web page IATs follow a particular theoretical distribution, we generate traffic using multiple theoretical distributions. These include: uniform(0,60); uniform(0,30); normal(20,5); normal(10,3); weibull(30,.5); and weibull(20,.5).[21] We selected these distributions to incorporate a good mix of inter-arrival times that: (i) are evenly spread among a wide range of

[20] Markov models have been used in past work to model web site navigation patterns [Chen et al., 2005]. However, recent work has challenged the notion that web users are Markovian [Chierichetti et al., 2012]. Indeed, features like the "back" button and multi-window browsing, introduce "memory" into a browsing session.

[21] Uniform(a,b) is a uniform distribution defined for lower bound a and upper bound b in seconds; normal(μ,σ) is a normal distribution defined for mean μ and standard deviation σ in seconds; weibull(λ,k) is an weibull distribution with mean rate parameter λ and shape parameter k.

values (i.e., the uniform distributions), (ii) concentrate around a mean inter-arrival time (i.e., the normal distributions), and (iii) includes a balance of both large and small inter-arrival times (i.e., the weibull distributions). We select the parameters of the distributions such that their mean is between 10-30 seconds, but also include fairly small inter-arrival times. Note that we do not claim that these distributions approximate IAT distributions observed "in the wild." Instead, we use this diverse set of distributions to increase the coverage of IATs in our empirical evaluation.

Number of Simultaneous Browser Tabs: Modern browsers support multi-tab browsing — this essentially allows for multiple browser windows to be open at a given time. This may present a problem because increasing the number of tabs will increase the number of web page requests whose traffic overlaps in the network (and will reduce idle time observed in the traffic), thereby negatively impacting web page segmentation performance [Miller et al., 2014]. To account for this, we evaluate the impact that the number of simultaneous tabs — 1, 2, 4, or 8 — may have on web page segmentation performance.

Diverse Browsers and Web Pages: Our analysis in Chapter 3, Chapter 4, and Section 5.1 shows that web page traffic is diverse and is influenced by browser choice. Thus, it is important to include this diversity in our evaluation. We sample web pages from the same list of 3614 URLs provided in Appendix 2, which is composed of a diverse set of web pages from the top-250 US web sites [Inc.] — the process to obtaining this list of URLs is also discussed in Chapter 3. These URLs correspond to a diverse mix of web pages including landing pages, search results, media content, and mobile web pages. As discussed before, web pages are downloaded using multiple modern browsers including Chrome v 37.0.2062.124, Firefox v 32.0.2, and Safari v 7.1 (9537.85.10.17.1) on Mac OS X 10.9.4.

Browsing Stream Generation: Web page downloads are initiated at times defined by the inter-arrival time distribution — for each, a random web page is selected from the list of 3614 URLs on the current tab. The tabs cycle from tab number 1 to tab number N for each web page download for simplicity.[22] For example, if $N = 2$, we load the first random page on the first tab, load the second random page on the second tab, and load the third random page on the first tab.

For each combination of inter-arrival time distribution, number of tabs, and browser, we generate web

[22] Here, N corresponds to the number of tabs tested for the particular browsing stream. In this study, N is either 1, 2, 4, or 8.

traffic as described for a duration of 15 minutes — this is referred to as a *browsing stream*. These browsing streams are automatically generated using Applescript. Applescript is a scripting language that is built into Mac OS X that can be used to automatically control any application, including browsers. The traffic generated by these browsing streams are captured using tcpdump. We log the time and URL for each web page downloaded in each browsing stream — this information is kept in our *operating system log* and is later used to match web page downloads initiated by the operating system with the web pages observed in traffic. The browser cache is also cleared after each browsing stream.

We repeat this process 40 times for each combination of inter-arrival time distribution and browser — so, there are 6 IAT distributions × 3 browsers × 4 tab configurations × 40 repetitions for a total of 2880 *browsing streams* lasting 12 minutes each. Thus, there is a total of 43,200 minutes of browsing streams that are used for evaluation — this includes over 100,000 web page downloads.

Ground Truth Determination: Ground truth times for web page segmentation are determined by correlating the web page request time observed in the operating system log with the web page traffic times that are observed in our *traffic log* — our traffic log includes the time in which each HTTP message, DNS message, and TCP segment was captured on the link and is generated using tcpdump and pcap2har [Jacobson et al., 1989, pcap2har].[23] A web page j is determined to have been *detected* if we observe (i) an HTTP request in our traffic log that has a URL field that matches the URL in our operating system log or (ii) a DNS request that has a hostname field that matches the hostname in our operating system log. These matching HTTP and/or DNS messages in the traffic log must occur between the times web page j and web page $j+1$ were observed in the operating system log in order to be identified as the ground truth time in which a web page was requested in our traffic log — the resulting *ground truth web page request time* is the time observed in the *traffic log*. We allow this small window between detecting web page requests in the traffic log because there are delays between the web page request time at the operating system level and at the network level — that is, the times in the operating system and traffic logs do not completely match. Statistics regarding how often web pages are (i) detected using an HTTP message that matches the operating system log, (ii) detected using a DNS message that matches the operating system log, or (iii) not detected when using neither HTTP nor DNS messages but present in our operating system log are shown in Table 5.1.

[23] Tcpdump is used to generate logs with TCP-level and DNS-level information while pcap2har is used to generate logs with HTTP-level information.

TABLE 5.1: Number of Web Pages Detected and Missed by Different Approaches

Detected Using HTTP	Detected Using DNS	Not Detected	Total Number of Requests
92814 (84.19%)	102989 (93.42%)	2918 (2.64%)	110240

We also explicitly allow a small deviation of s seconds between the time a web page segmentation method detects a web page and the ground truth time because these values do not always match — this occurs because the ground truth DNS and HTTP request times do not necessarily occur at the same time the thresholds for the number of SYN and/or bytes transmitted are satisfied for the different web page segmentation methods. We arbitrarily test several values of s including .250, .500, 1, and 5 seconds to determine if it impacts the conclusions drawn from our study. We discuss this impact in more detail in Section 5.3.2.

Performance Metrics: We must also consider the metrics to use when comparing the performance of the different web page segmentation approaches. Metrics that are appropriate for evaluating web page segmentation methods are the same metrics that are used for evaluating binary classification problems — these include precision, recall, F-scores, and accuracy [Schatzmann et al., 2010, Ihm and Pai, 2011, Erman et al., 2006]. These metrics, which are defined in chapter 4, can be difficult to interpret since they are functions of the number of true positives, false positives, and false negatives. For problems where there are only two classes (in this case, the presence of a detection or not) it is informative to report the rates in which true positives, false positives, and false negatives occur in addition to the other metrics, such as an F-score, which are more difficult to interpret. The metrics that we present in this section are defined below:

- *True positive rate (TPR)*: We identify a true positive if the web page segmentation method is able to detect a web page that corresponds to a ground truth time. The true positive rate, TPR, is then calculated as the number of True Positives divided by the number of Web Page Downloads.

- *False positive rate (FPR)*: We identify a false positive for detections that do not align with the ground truth. The false positive rate, FPR, is calculated as the number of False Positives divided by the number of Web Page Downloads.

- *False negative rate (FNR)*: A false negative occurs when the ground truth indicates that a detection should occur and the web page segmentation method fails to detect it. For web page segmentation,

TABLE 5.2: Parameter Search Space for Optimization of Web Page Segmentation Approaches

Segmentation Approach	Parameter	Search Space
Idle-time method (SYNs)	I	250ms to 20s in steps of 250ms
Idle-time method (Bytes)	I	250ms to 20s in steps of 250ms
Idle-time method with threshold (SYNs)	I	250ms to 20s in steps of 250ms
Idle-time method with threshold (SYNs)	T	1 to 50 in steps of 1
Idle-time method with threshold (Bytes)	I	250ms to 20s in steps of 250ms
Idle-time method with threshold (Bytes)	T	1 to 50000 in steps of 1000
Heuristic Change Detection Method (SYNs)	I	250ms to 20s in steps of 250ms
Heuristic Change Detection Method (SYNs)	T	1 to 50 in steps of 1
Heuristic Change Detection Method (SYNs)	L	250ms to 20s in steps of 250ms
Heuristic Change Detection Method (Bytes)	I	250ms to 20s in steps of 250ms
Heuristic Change Detection Method (Bytes)	T	1 to 50000 in steps of 1000
Heuristic Change Detection Method (Bytes)	L	250ms to 20s in steps of 250ms
Fused Lasso (SYNs)	λ	0 in steps of .05 to 3
Fused Lasso (SYNs)	μ	0 in steps of .05 to 3
Fused Lasso (Bytes)	λ	0 in steps of 1000 to 30000
Fused Lasso (Bytes)	μ	0 in steps of 1000 to 30000
HMM (SYNs and Bytes)	$A_{i,j}$	0 in steps of .01 to 1,where $\sum_{j=1}^{2} A_{i,j} = 1$
HMM (SYNs and Bytes)	$B_{i,j}$	0 in steps of .01 to 1,where $\sum_{j=1}^{2} B_{i,j} = 1$

the false negative rate, FNR, is the number of False Negatives divided by the number of Web Page Downloads. This metric is equivalent to $1 - TPR$ — hence, we do not report it since it is redundant.

- *Precision:* The precision metric is defined as $\frac{TP}{TP+FP}$, where TP is the number of true positives and FP is the number of false positives.

- *Recall:* The recall metric is defined as $\frac{TP}{TP+FN}$, where TP is the number of true positives and FN is the number of false negatives. Please note that the recall metric is equivalent to the true positive rate and accuracy metrics in this analysis.

- *F-score:* The F-score metric is defined as the harmonic mean between the precision and recall — this is expressed as $2 \times \frac{precision \times recall}{precision + recall}$.

Web Page Segmentation Method Parameterization: All of the segmentation methods that we use for our analysis have parameters that must be configured. We perform an explicit exhaustive search for the parameters for each method and select the settings that result in the "best" performance — one that yields a high TPR and a low FPR. For this study, we do this by selecting the parameters that result in the highest F-

TABLE 5.3: Optimal Parameter Settings Used for Web Page Segmentation Approaches for r = 250ms

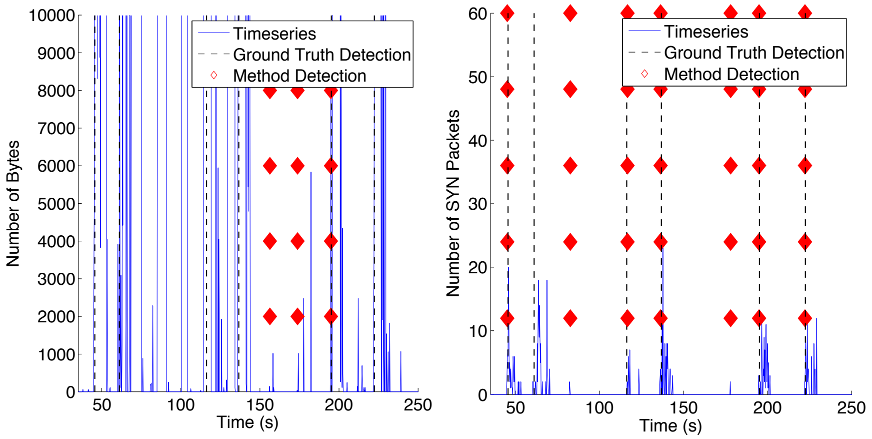
Segmentation Approach	Optimal Parameters
Idle-time method (Number of SYNs) Idle-time method (Number of Bytes)	I = 9s I = 10s
Idle-time method with threshold (Number of SYNs) Idle-time method with threshold (Number of Bytes)	I = 9s; T=6 I = 3s; T=3000
Heuristic Change Detection Method (Number of SYNs) Heuristic Change Detection Method (Number of Bytes)	I=0s; T=6; L=8s I=0s; T=7000; L=9s
Fused Lasso (Number of SYNs) Fused Lasso (Number of Bytes)	λ = .2; μ=2.25; k = 5s λ = 2000; μ=15000; k = 5s
HMM (Number of SYNs)	k = 5s; $A_{1,1}$=.9; $A_{1,2}$=.1; $A_{2,1}$ =.05; $A_{2,2}$ =.95 $B_{1,1}$=.99; $B_{1,2}$=.01; $B_{2,1}$ =.8; $B_{2,2}$ =.2
HMM (Number of Bytes)	k = 5s; $A_{1,1}$=.87; $A_{1,2}$=.13; $A_{2,1}$ =.03; $A_{2,2}$ =.97 $B_{1,1}$=.98; $B_{1,2}$=.02; $B_{2,1}$ =.77; $B_{2,2}$ =.23

score (which is a measure that incorporates in true positives, false positives, and false negatives into a single metric) for each web page segmentation method by processing a random sample of half of our browsing streams — please note that we use the F-score metric to determine the best methods in this work since it is a function of each metric that we consider. The parameter space that we searched during this optimization is in Table 5.2.

We searched this parameter space for when the resolution parameter, r, is 100ms, 250ms, 500ms, and 1000ms. The optimal parameters selected for each web page segmentation method when r = 250ms is shown in Table 5.3 — please refer to Appendix 13 for the optimal parameters for when r is 100ms, 500ms, and 1000ms. We discuss the impact that r has on the performance of web page segmentation methods in the next section.

5.3.2 Evaluation Results

Before we go into details about our overall performance results, we highlight some of the strengths and weaknesses of the different segmentation approaches. Figure 5.5 plots the time series of a synthetically generated web page browsing stream with annotations for when web pages are detected by different segmentation approaches. The vertical dashed bars in these plots correspond to the ground truth web page download start time, while the vertical "diamond" markers correspond to the time that the web page segmentation

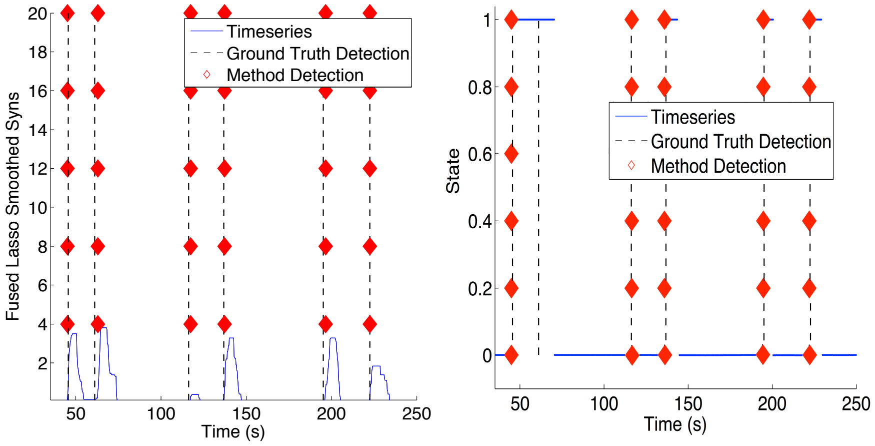


(a) **Number of Bytes Time Series: Idle time**

(b) **Number of SYN segments Time Series: Idle time**

Figure 5.5: Time series plots showing how idle time-based methods work.

method specified in Figure 5.5 detected new pages — we use this notation for illustrating the performance of different web page segmentation methods in this section.[24] Some key observations are outlined below.



(a) **Number of SYNs: Fused Lasso Time Series**

(b) **Number of SYNs: HMM Time Series**

Figure 5.6: Time series plots showing how change point detection methods work

Bytes are more volatile than SYNs: Figure 5.5(a) shows a time series for the number of bytes, while Figure 5.5(b) shows a time series for the number of SYN segments. There are many instances in Figure 5.5(a), for example around 60-100s, where there are many bytes being transferred when a page is not being explicitly downloaded. These instances makes it difficult for idle time-based approaches to distinguish web page

[24] The settings for each web page segmentation method plotted is provided in Table 5.3.

downloads from background traffic. This is particularly an issue because the example in Figure 5.5 is a fairly "easy" web page browsing stream to process because there are large web page inter-arrival times between subsequent downloads. This is not as common in the number of SYN segments time series in Figure 5.5(b) — this highlights the need to use the number of SYN segments as the primary traffic feature over the number of bytes for web page segmentation.

Performance of methods on a simple example: It is clear from the raw time series plots in Figure 5.5(a) and Figure 5.5(b) that idle time-based approaches will be unable to correctly identify all web page downloads without identifying false positives — this is due to the background traffic. Figure 5.6(a) shows an example of how fused lasso, a change point detection method, will smooth the number of SYN segments time series to help detect the beginning of a web page download — note that a web page is "detected" using fused lasso at the point when the smoothed time series changes from being monotonically non-increasing to increasing. For this simple case, fused lasso is able to identify all the web page downloads without any false positives. Similar observations can be made for the HMM smoothed time series in Figure 5.6(b) — note that a web page is detected here when the state transitions from State 0 to State 1. The primary difference is that the HMM approach groups two web page downloads together into a single download (around the 50-60s portion of the plot), when it should be two distinct pages. From this simple example, it is clear that HMMs will have more difficulty than fused lasso in distinguishing between two web pages with small inter-arrival time. This occurs despite an "optimal" parameterization of each method.

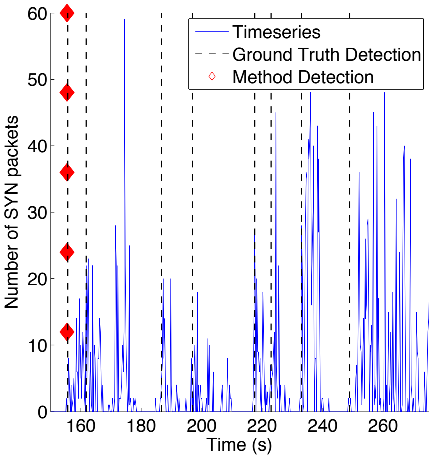


Figure 5.7: Time series plot showing some of the difficulties of applying idle time-based methods on synthetic data.

A more noisy example: Figure 5.7 shows a time series for the number of SYN segments for a web browsing stream where some features (i.e., number of SYNs) are rarely idle. In this scenario, most web pages will not be identified using idle time-based methods because the idle time threshold is rarely met. Figure 5.8 shows the fused lasso smoothed time series. Here, we can see that fused lasso has more difficulty identifying web pages than in the past example. For instance, we notice that fused lasso does not immediately detect the web page request at around 162s due to the small inter-arrival time between the 157s download and the 162s download. Instead, this page is detected late at around the 170s portion of the plot — this *delayed detection* is technically a false positive, but it is clear that it could correspond to the earlier missed request. The 215-240s portion of the plot shows that fused lasso is able to detect some pages when there is little idle time period — it essentially identifies areas where the time series peaks as a new page.

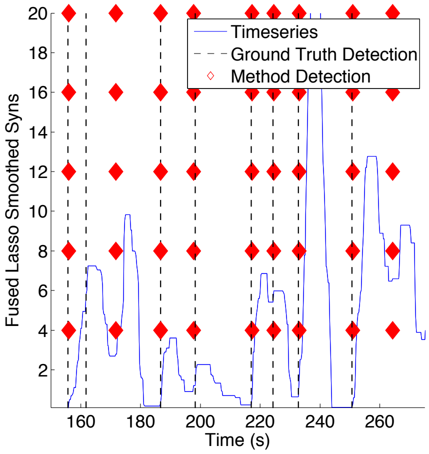


Figure 5.8: Time series plot showing some of the difficulties of applying Fused Lasso on synthetic data.

There are times when detecting pages in this manner can backfire. For example, at around the 250-270s portion of the time series, we observe a web page is requested that takes dozens of seconds to load and exhibits multiple peaks. Here, fused lasso correctly identifies the first peak (around 252s) as a true positive and incorrectly identifies the second peak (around 262s) as a new page (i.e., a false positive). We also find that the HMM time series, shown in Figure 5.9, tends to group these active periods of traffic into a single web page download — this result is not surprising given what we observed on the easier example in Figure 5.6(b). This results in fewer true positives and fewer false positives.

Overall Performance Results: We next discuss the overall performance results for all web page segmentation methods. Table 5.4 summarizes the results for the different web page segmentation methods when

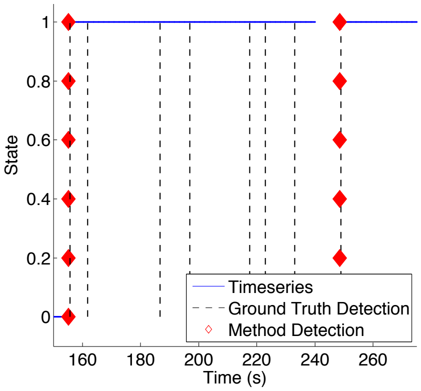


Figure 5.9: Time series plot showing some of the difficulties of applying HMM on synthetic data.

TABLE 5.4: Overall Performance of Segmentation Methods for Synthetic Data

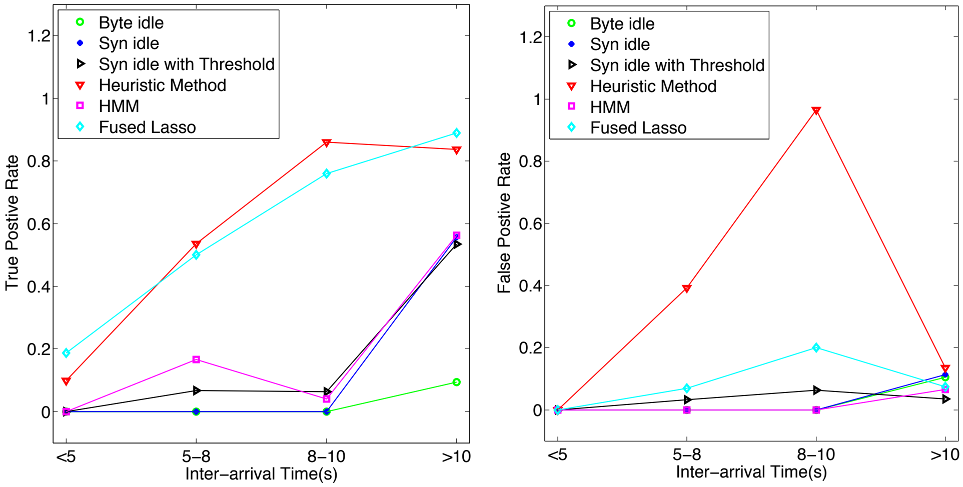
Segmentation Approach	SYNS		Bytes	
	TPR/Recall	FPR	TPR/Recall	FPR
Idle-time method	.4148	.0869	.0889	.0770
Idle-time method with threshold	.4016	.0328	.8951	.7623
Heuristic Change Detection Method	.7320	.2038	.8393	.5476
Fused Lasso	.8217	.0629	.8115	.1863
HMM	.5079	.0584	.0257	.0107
	Precision	F-score	Precision	F-score
Idle-time method	.8268	.5524	.5359	.1525
Idle-time method with threshold	.9245	.5600	.5401	.6737
Heuristic Change Detection Method	.7822	.7563	.6052	.7033
Fused Lasso	.9289	.8720	.8133	.8124
HMM	.8969	.6485	.7060	.0496

using either the number of bytes or the number of SYN segments as the traffic feature. We find that:

- Most web page segmentation methods perform better overall in terms of a higher F-score (which is a single metric that is a function of the number of true positives, false positives, and false negatives) when the number of SYNs is used as the traffic feature than when the number of bytes is used. This result is not surprising given that our analysis in Section 5.1 showed that the number of bytes feature has a much higher variance and exhibits more noise than the number of SYNs feature. The idle-time method with threshold is the only approach that had a F-score that was higher when the number of bytes feature was used. Upon further inspection, we find that the best parameterization for this approach follows two different extremes with respect to the TPR and FPR metrics. In particular, Table 5.4 shows that the TPR is modest (.4016) and the FPR is low (.0328) when the number of

SYNs feature is used, while the TPR and FPR are both high (.8951 and .7623) when the number of bytes feature is used. In fact, no other approach shown in Table 5.4 yielded a TPR and FPR that was this high. Recall that the F-score is a metric that encapsulates multiple metrics in an attempt to find the best balance between TPR, FPR, and FNR. However, the best balance between these is debatable when one approach is very conservative (i.e., has a modest to high TPR with a low to modest FPR) and a competing approach is very aggressive (i.e., has a TPR and FPR that are both high). Indeed, a web page segmentation approach that has a high TPR is usually not preferred if the FPR is also high.

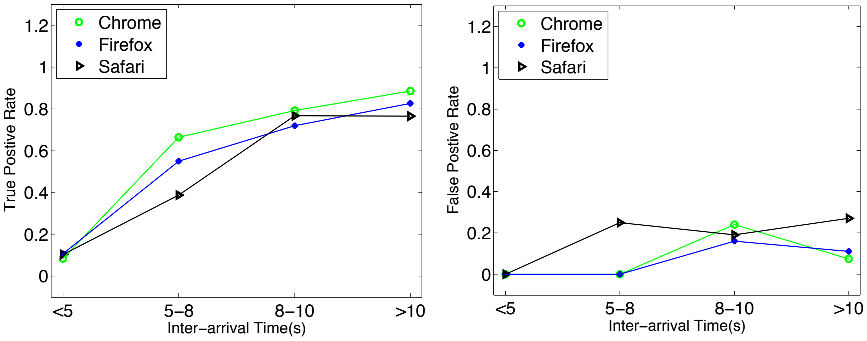
- The change point detection methods (i.e., fused lasso, HMM, and heuristic change detection method) are able to achieve a true positive rate that is at least .09 higher than the idle time-based methods when using the number of SYN segments traffic feature. We believe that this is largely because the idle time-based methods have difficulty finding an idle period that is long enough to detect a new page, while being short enough to not detect many false positives. We find that change point detection methods have F-scores that are at least .08 higher than idle-time based methods. We also find that the fused lasso method achieves the best overall performance with a F-score of .8720 when the number of SYNs feature is used.



(a) Overall Results: True positive rate **(b) Overall Results: False positive rate**

Figure 5.10: Overall of performance of web page segmentation approaches by inter-arrival time.

We further investigate the performance of the different web page segmentation methods when considering the impact of other factors such as web page inter-arrival time, browser choice, and the number of simultaneous tabs. We present results here with respect to different ranges of web page inter-arrival times to show its impact on performance in all cases. We also restrict our analysis to the number of SYN segments traffic feature since it outperforms the number of bytes feature for most web page segmentation methods tested. We find that:



(a) Impact of Browser: True positive rate

(b) Impact of Browser: False positive rate

Figure 5.11: Impact that the browser choice have on web page segmentation performance.

- *Impact of Inter-arrival Time:* We first discuss the impact that web page inter-arrival time has on some performance metrics for different web page segmentation methods. Plots for this impact are provided in Figures 5.10(a) and 5.10(b). The web page inter-arrival times, which are divided into four groups ($< 5s$, $5-8s$, $8-10s$, and $> 10s$), are provided on the x-axis of these plots while the true positive rate and false positive rate are provided on the y-axis of Figures 5.10(a) and 5.10(b), respectively — please note that these plots include the cumulative measurements taken from each of the 40 repetitions for all browsers, interarrival time distributions, and tab configurations considered. Figure 5.10(a) shows that the true positive rate usually increases as the inter-arrival time increases. At one extreme, when the inter-arrival time is less than 5s, all methods have difficulty detecting web pages — that is, the TPR for each web page segmentation method is the lowest in this range. This result is expected because pages with small inter-arrival time are more likely to be considered as a part of the previous web page. At the other extreme, the TPR for most web page segmentation methods is the highest when the inter-arrival time is $> 10s$ — the only exception being the heuristic method which has a slightly higher TPR of .836 in the 8-10s range as compared to a TPR of .811. A higher TPR when the inter-arrival between web pages is high (>10) is expected since there is additional time for web page traffic from previous web page downloads to end — this makes it easier for web page segmentation methods to detect new web page downloads. The TPR between 5-10s for each method is between the performance achieved when the inter-arrival time is < 5s and > 10s. More specifically, the performance for each method, excluding the HMM approach, either increases or remains unchanged in the 5-10s range. This result suggests that the performance of web page segmentation methods will usually increase or remain constant as web page inter-arrival times increases. Although, the magnitude of the increase in performance depends on the web page segmentation approach. For example, Figure 5.10(a) shows that the TPR of the fused lasso and the heuristic change detection methods increased by over 0.20 in the 5-10s range, while the performance of the idle time-based methods are largely unchanged. It is not surprising that the TPR for the idle time-based methods do not change significantly in this range since many web page downloads continue for a significant amount of time (over 10s) — that is, it is unlikely that the idle time periods for these methods are satisfied in less than 10s after a web page download starts. This is also a possible explanation for the poor TPR (< 0.06) for idle time-based methods when the inter-arrival time is less than 10s.[25] The HMM method is unique in that the TPR moderately decreases ($\approx .12$) in the 5-10s range. This moderate decrease in performance between the 5-8s and 8-10s range is difficult to explain. Nevertheless, the high TPR when the inter-arrival time is > 10s (0.58) as compared to the TPR when the inter-arrival time is $<$ than 10s (< 0.18) suggests that the HMM approach struggles when the inter-arrival time is small.

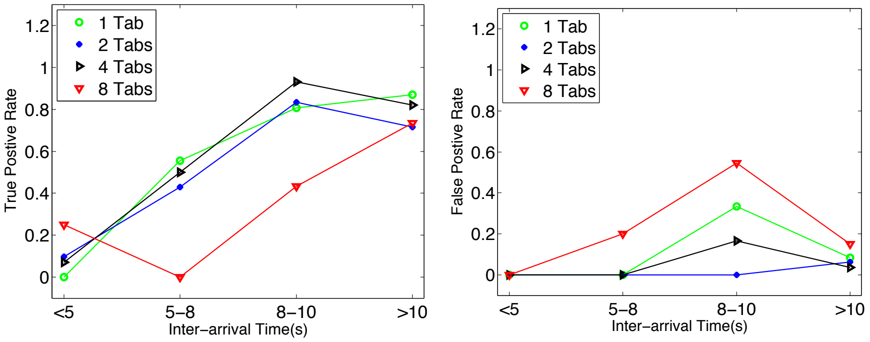
Figure 5.10(b) shows that the false positive rate is also low for inter-arrival times less than 5s. Similar to the case with the true positive rate, web page segmentation approaches have difficulty detecting anything, even false positives, when the inter-arrival time is small. For the 5-8s and the 8-10s inter-arrival times, we observe an increase in false positives. This is likely due to the "two peak" behavior we observed for some pages in Figure 5.2 (i.e., detecting the same page download twice). This observation is most pronounced with the heuristic method and, to a lesser extent, the fused lasso approach.[26] The idle time-based and HMM approaches are not impacted by this as much because they tend to group multiple web page downloads into a single download. Most methods have lower false positive rates when the inter-arrival time is greater than 10s. Overall the fused lasso method has

[25] Please refer to Section 5.1 for a discussion on the time series characteristics of time series.

[26] Recall that Figure 5.2 shows that traffic features for web pages may peak between the 7-11 second range after the web page download process started — this observation means that it is possible for change point detection methods to detect web pages more than once since the web page traffic feature peaks multiple times. For the heuristic change detection method, it is likely that web pages are being detected twice since $L_{delay} = 8$, which falls within the range where web pages may peak — this is a possible explanation for the surge in false positive rate during this period. Despite this issue, we still recommend this heuristic method over the idle-time methods because it can achieve a much higher true positive rate.

the best performance with respect to being able to achieve both a high true positive rate and a low false positive rate.

- *Impact of Browsers:* We next discuss the impact that browser choice has for on web page segmentation performance. Plots that show the true positive rate and false positive rate across different browsers for the fused lasso regression method (i.e., the overall best approach) are provided in Figures 5.11(a) and 5.11(b), respectively — please note that these figures include the cumulative results for the different interarrival time distributions, tab configurations, and browsers for the fused lasso method. Figure 5.11(a) and Figure 5.11(b) shows that the true positive rate and false positive rate for the fused lasso regression method (i.e., the overall best approach) is not heavily influenced by browser choice. This result implies that browser choice should not significantly impact web page segmentation performance "in the wild".



(a) Impact of Tabs: True positive rate

(b) Overall Results: False positive rate

Figure 5.12: Impact that the number of tabs have on web page segmentation performance.

- *Impact of Multiple Tabs:* We next consider the impact that the number of simultaneously open tabs in browser window has on web page segmentation performance. Plots that show the impact that the number of tabs have on the true positive rate and false positive rate are provided in Figures 5.12(a) and 5.12(b), respectively. Please note that while these figures include the cumulative measurements from the different tab configurations, interarrival time distributions, and browsers considered in this study, it only includes the performance of the fused lasso regression method (i.e., the overall best approach). The results in Figure 5.12(a) and Figure 5.12(b) show that the number of tabs that are open may have an impact on performance. In particular, the true positive and false positive rates for the data collected when there are 8 simultaneous tabs open are consistently worse than for the case when there are 1, 2, and 4 tabs. This result implies that it is expected that increasing the number of tabs beyond a modest number (in this case 4) will decrease the performance of web page segmentation methods. This is likely due to a higher amount of background traffic from multiple web pages that essentially decreases the effective inter-arrival time of web requests and increases the interleaving of traffic from different pages.

Impact of Resolution Parameter r: The parameters in Table 5.3 were optimized for when the resolution, r, is 250ms. We also evaluated against several values for r, including 100ms, 500ms, and 1000ms. We found that r has little impact on the performance of idle time methods and a modest impact on change point detection methods. The little impact on idle time approaches is to be expected — r will only impact the idle time-based approaches if the optimal value for idle time is less than r, which is not the case (greater than 9 seconds). However, r does have an impact on change point detection methods. This is mainly because changing the resolution of a time series will impact its perceived burstiness (i.e., lower values of r will appear more bursty). We find that we can achieve similar performance for all values of r tested by simply modifying the other parameters of the change point detection methods. This includes either increasing or decreasing the threshold for the heuristic change detection method, changing the amount of smoothing for fused lasso regression (i.e., λ and μ), or changing the transition probabilities of the HMM.[27] The optimal parameterization and performance of each web page segmentation method for different r is provided in Appendix 13.

Impact of Ground Truth Parameter s: Please recall that the ground truth times used for our performance analysis allows for a delay of s seconds which accounts for the minor differences between ground truth times and the web page segmentation method times. All of the results shown in this section were computed where $s = 1$ second. To determine whether s impacts the conclusions from the analysis conducted in this chapter, we recompute results using different values of s. More specifically, we compute results where s is .250, .500, 1, and 5 seconds — these results are shown in Table 5.5. Table 5.5 shows results for the TPR and FPR metrics — results for the precision and F-score metrics are provided in Appendix 13.[28] We find that,

[27] The value of r may also be influenced by environmental factors such as link speed and end-host performance. We do not consider these factors in this chapter.

[28] Please note that the conclusions drawn from these metrics are similar since they are functions of the same primary metrics (i.e., number of true positives, false positives, and false negatives).

TABLE 5.5: Impact of *s* Parameter on Web Page Segmentation Performance (*r* = 250ms)

Segmentation Approach	*s* (seconds)	TPR (SYNs)	FPR (SYNs)	TPR (Bytes)	FPR (Bytes)
Idle-time method	.250	.4148	.0869	.0889	.0770
Idle-time method (threshold)	.250	.4016	.0328	.8951	.7623
Heuristic Change Detection	.250	.7320	.2038	.8393	.5476
Fused Lasso	.250	.8217	.0629	.8115	.1863
HMM	.250	.5079	.0584	.0257	.0107
Idle-time method	.500	.4213	.0804	.0896	.0763
Idle-time method (threshold)	.500	.4112	.0231	.8956	.7618
Heuristic Change Detection	.500	.7411	.1947	.8410	.5459
Fused Lasso	.500	.8250	.0596	.8147	.1831
HMM	.500	.5214	.0449	.0260	.0104
Idle-time method	1	.4338	.0679	.0998	.0661
Idle-time method (threshold)	1	.4150	.0194	.8998	.7576
Heuristic Change Detection	1	.7613	.1745	.8574	.5295
Fused Lasso	1	.8415	.0431	.8230	.1748
HMM	1	.5354	.0309	.0280	.0084
Idle-time method	5	.4576	.0441	.1104	.0555
Idle-time method (threshold)	5	.4173	.0171	.9543	.7031
Heuristic Change Detection	5	.8141	.1217	.8812	.5057
Fused Lasso	5	.8522	.0324	.8754	.1224
HMM	5	.5531	.0132	.0294	.0070

as expected, increasing *s* increases the true positive rate and decreases the false positive rate for all methods. This occurs because the range for ground truth time becomes more "loose" which increases the number of detections that are classified as a true positive — that is, false positives that were previously outside the ground truth time range with small *s* become true positives with larger *s*. While increasing *s* modifies the true positive and false positives rates of all methods, it does not change the conclusions drawn from the performance analysis conducted in this section. In other words, fused lasso outperforms all other methods and change point detection methods outperform idle time-based methods.

5.4 Evaluation With Real User Data

5.4.1 Real User Data Collection

We next evaluate the segmentation approaches using data collected from browsing sessions by real users. Our motivation in doing so is to: (i) include in our dataset, personalized pages as well as those with user-interactive content; (ii) include pages that are not restricted to only the top-250 U.S. web sites; and (iii) use real web browsing data to study the impact of segmentation performance on application domains that rely on web traffic analysis [Ihm and Pai, 2011].

TABLE 5.6: Overview of Real User Browsing Dataset (40 users)

Property/Metric	Min	Median	Mean	Std	Max	Total
# Web page downloads (per user)	5	41	39	17	90	1567
# Web page downloads in Top 250 (per user)	0	22	24	16	71	980
Web page IAT statistics (among all users)	1s	10s	24s	42s	507s	N/A

We collect data by recruiting *40 real users* in an IRB approved study [Sanders and Kaur, 2015c]. We provide the users with a desktop machine running Mac OSX 10.9.4. We instruct the users to browse the Internet using Chrome v 37.0.2062.124 for 15-20 minutes. The users are encouraged to do whatever type of web activity they desire, including visiting sites that host dynamic content such as Facebook and Youtube — in fact, many users visited such sites. We record the browser history and the tcpdump trace for the web browsing sessions. The browser history is used to determine the ground truth web page download times by matching the browsing logs with the traffic observed in the traffic logs, similar to how it was performed for the synthetic data collection — please note, however, that we are not able to track tab usage for the real user dataset. A summary of some of the properties of this dataset, such as the number of web page downloads observed, the number of web pages that were among the top 250 web sites, and web page interarrival time-related statistics, are provided in Table 5.6.[29]

It is also important to note that we do *not* re-parameterize the web page segmentation approaches on the real user data — in practice, we will not be able to tune our methods for real traffic observed "in the wild." Thus, we choose to use the optimal parameters determined from the synthetic data analysis for the real data analysis — this reduces the odds of overfitting.

5.4.2 Real User Browsing Data Results

Overall Performance Results: Table 5.7 shows the overall true positive and false positive rates for the real data, while Figure 5.13(a) and Figure 5.13(b) shows the impact inter-arrival time have on true positive and false positive rates, respectively. We find that:

- Similar to the analysis using synthetic data, change point detection methods usually have a higher

[29] We use abbreviations for a number of terms in Table 5.6 for conciseness — specifically, std is short for standard deviation, min is short for minimum, and max is short for maximum. Also, please refer to Section 5.5.1 for a plot of the cumulative distribution of the interarrival times for the real user browsing data.

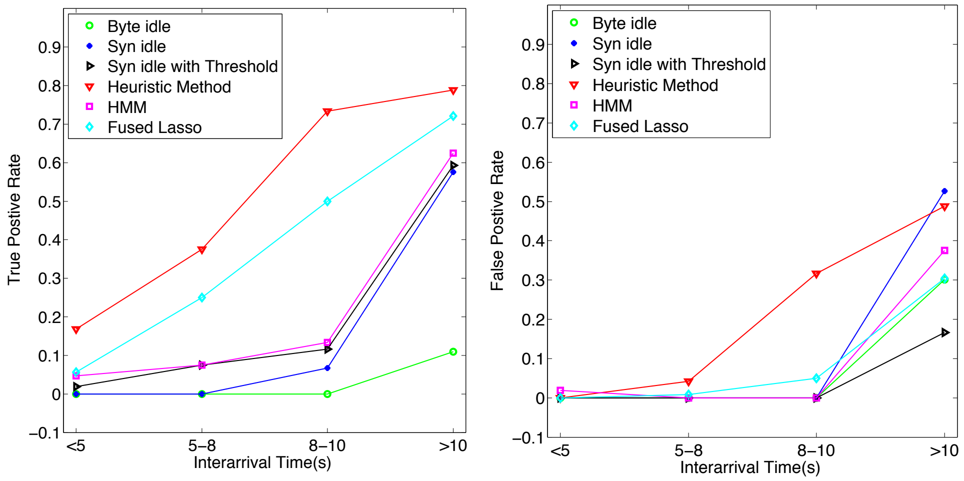
TABLE 5.7: Overall Performance of Segmentation Methods for Real Data

Segmentation Approach	TPR/Recall	FPR	Precision	F-score
Idle-time method (Number of SYNs)	.4107	.3690	.5267	.4615
Idle-time method (Number of Bytes)	.0759	.2128	.2629	.1178
Idle-time method with threshold	.4299	.1355	.7603	.5492
Heuristic Change Detection Method	.6610	.3636	.6451	.6530
Fused Lasso	.5701	.2150	.7261	.6387
HMM	.4567	.2620	.6354	.5314

F-score metric than idle time-based approaches. More specifically, the best performing methods are fused lasso and the heuristic change method which have F-scores of .6387 and .6530, respectively — though, the overall performance observed when considering real data is worse than the performance for synthetic data. Generally speaking, many of the issues we identified in the synthetic data evaluation occur in real data. Figure 5.14(a) shows an example of where a change point detection method (fused lasso) performs in a manner that is fairly expected. This includes grouping web pages with small inter-arrival time together (i.e., around the 310s part of the plot).

- Overall, web page segmentation approaches perform worse on real user data than on the synthetically generated web browsing data. We believe that this is due to a number of factors, many of which boil down to an increase in automatically downloaded traffic that does not correspond to a "new web page download", and the presence of web pages with traffic characteristics that makes web page segmentation difficult. Despite this, the best segmentation methods can still identify web page segments with inter-arrival times higher than 10s with true positive rate of approximately 70%. However, the false positive rate is also higher. We discuss some of the possible causes for this behavior below.

Issues with Automatically Generated Traffic: Figure 5.14(b) shows an example of where a user visited a web page that played a video. Based on the this user's browsing history, and visiting the page that the user browsed, we believe that the user played a video within a web page at around the 80s portion of the plot. Once we played the only video that was present on this page, we found that a new video automatically plays after the previous video finishes. Thus, we believe that the multiple increases in traffic, at approximately the 150s, 250s, 390s, and 430s parts of the plot, are likely due to a new video loading immediately after each video ends — please note that we manually verified that (i) these new videos generate new traffic and



(a) Real Data:True positive rate

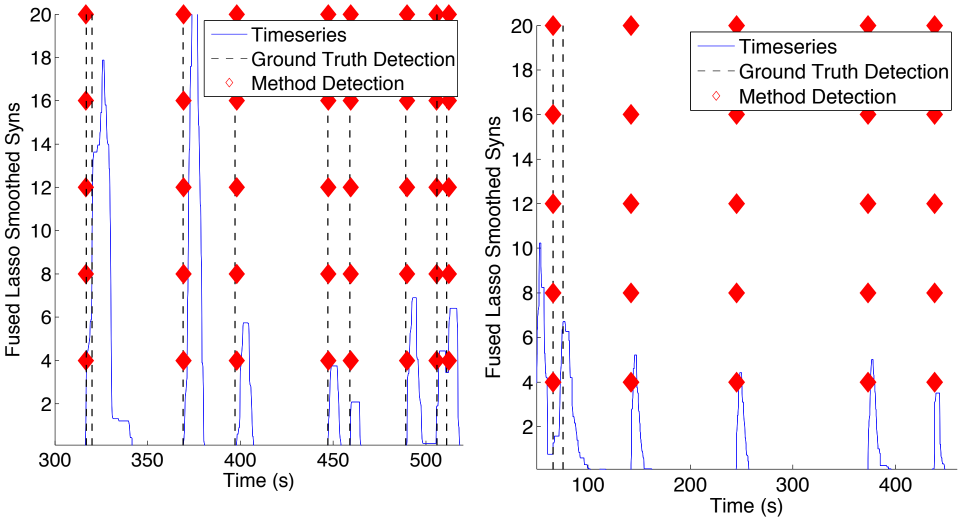
(b) Real Data: False positive rate

Figure 5.13: Plot showing the performance of using web page segmentation approaches using real data.

(ii) the user did not click on a new web page during these time periods. The segmentation approaches are technically doing what they are designed to do, which is detect increases in traffic. However, in this case, an increase in traffic does not correspond to a web page download as requested by the user (which is our definition of a web page download). Thus, this behavior of automatically starting a new video within an already loaded page gets classified as a false positive which negatively impacts segmentation performance.

We also find a more troublesome case where a web page can generate such random and sporadic network activity that web page segmentation approaches essentially become unusable. Figure 5.15(a) shows an example of a browsing stream that is rarely idle. This is surprising given that this is a plot of the number of SYN segments traffic feature — similar behavior is observed for the number of bytes traffic feature. Figure 5.15(b) shows that this high degree of noise makes web page segmentation via fused lasso particularly difficult (i.e., a high false positive rate). Upon further inspection, we find that this user was browsing on a forums web page which continuously communicates with the client long after the page has loaded. This is the only example web page in the real user data that we found that had characteristics that were this sporadic for both the number of SYNs and number of bytes time series.

Discussion of Web Page Download Definition In Section 5.1, we define a web page download as an event that is initiated when a user either (i) enters a URL into the browser window, (ii) clicks on a hyperlink, and/or (iii) enters text in a browser that results in a new web page such as a search query and its corresponding result



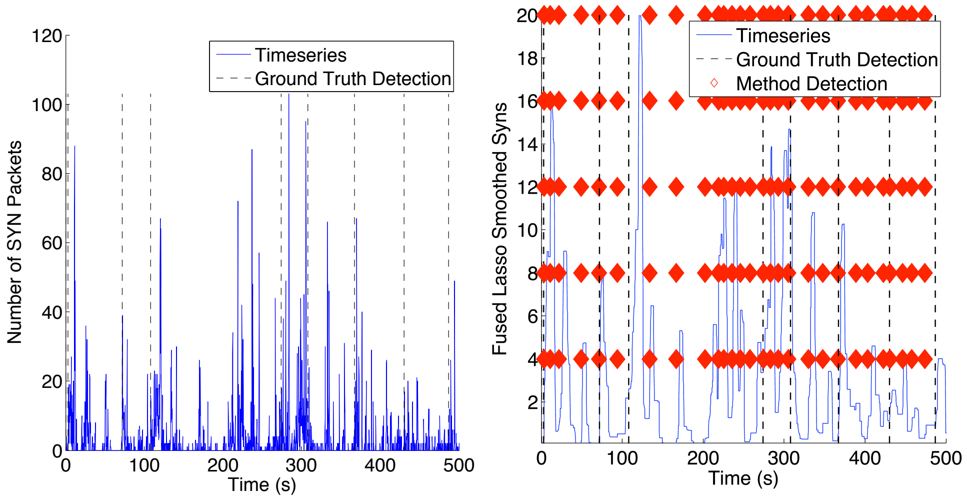
(a) Example when segmentation "works"

(b) Auto generated traffic: Video

Figure 5.14: Examples of where web page segmentation works as expected for real data.

web page. The web page segmentation methods that we consider are evaluated based on their ability to detect the web page download event as it is defined. However, we found that there are scenarios where a web page segmentation method may confuse an increase in traffic as a new web page download when it actually corresponds to some other type of event — these include cases where traffic is automatically generated. It is important to note that, in such scenarios, an increase in traffic that do not correspond to a new web page download may correspond to an event that may be of interest. For instance, being able to detect when a new video is playing within a web page may be useful for monitoring user-interactions with a web page [Neasbitt et al., 2014]. One approach for addressing scenarios where web page segmentation methods identify false detections is to assume that methods exist, essentially a classifier, that can correctly label detections — such methods must also be able to differentiate between web page downloads, user-interactions, noise, and other events of interest.[30] The definition of a web page, or rather an event, when using a web page segmentation method will depend on the classifier's ability to label detections according to the correct event. In this work, we only consider the case where we assume all detections correspond to web pages.

[30] This classification is beyond the scope of this work.



(a) Automated Traffic: Forums (SYNs) **(b) Automated Traffic: Forums (fused lasso)**

Figure 5.15: Examples of where web page segmentation fails for real data.

5.5 Impact of Web Page Segmentation Performance on Web-related Applications

In the previous section, we found that fused lasso outperforms all other methods and change point detection methods outperform idle time-based methods. While the metrics we base our performance analysis on are effective for determining which web page segmentation method performs the best overall, they do not provide insight on whether web page segmentation method choice has a measurable impact on different web traffic analysis domains. In this section, we explicitly consider this by conducting two case studies to ascertain if web page segmentation choice impact the application domains of (i) user behavior modeling and (ii) web page classification. An overview of these application domains, along with some details about how web page segmentation is applied to each, is provided below:

- *User behavior modeling:* In many scenarios, organizations are interested in the click behavior of users to understand how they use different web-related services [Neasbitt et al., 2014, Chierichetti et al., 2012, Wang et al., 2013]. For example, in security applications, the number of user-invoked clicks (that is, the number of web page requests) within a browsing stream can be used to help determine whether a user has malicious-intent or not [Wang et al., 2013], while in web performance applications this information can be used to determine whether users are engaged in a web page. Web page segmentation methods, which estimate when a web page is requested, can be applied directly on traffic to obtain two important metrics that are useful for such problems: 1) the number of clicks observed in a browsing stream; and 2) the average inter-arrival time between clicks. In our first case study, we use statistical tests to determine whether the web page segmentation approaches that we study are able to estimate these metrics in a manner that is statistically equivalent to the ground truth. If successful, then these methods can be used to model the click behavior of users.

- *Web page classification:* In Chapter 4, we discuss and evaluate the problem of web page classification. We showed that web pages can be classified according to several orthogonal labeling schemes [Sanders and Kaur, 2015b]. These include labeling schemes that can be used for traffic engineering applications such as the video streaming-based labels, and labeling schemes that can be used for user profiling such as the genre-based labels. However, a critical step in applying web page classification, or even web page identification [Liberatore and Levine, 2006], to traffic "in the wild" is to first use a web page segmentation approach to separate browsing streams into individual web page downloads. If this step is ignored, classification methods may incorrectly classify traffic by treating multiple web page downloads as if it were single web page — this is detrimental to the interpretation of the classification results. In this section, we explicitly study whether the choice of web page segmentation approach impacts the performance of web page classification by first applying web page segmentation on the browsing streams and then performing web page classification on the segmented traffic.

We use both synthetic and real user browsing data to evaluate the impact that web page segmentation methods have on these different applications. We consider the domains of estimating user browsing behavior and web page classification in Section 5.5.1 and Section 5.5.2, respectively. We also discuss the implications of applying web page segmentation to these application in Section 5.5.4.

5.5.1 Estimating User Browsing Behavior

Two key metrics of user browsing behavior that should be obtainable from an effective web page segmentation approach are the number of clicks made by a user (i.e., number of web page downloads) and the average inter-arrival times of those web pages. In our first case study, we compare the ground truth values for these metrics with those calculated after applying web page segmentation to determine whether there is a statistically significant difference between the two. We use paired t-tests for this analysis. A paired t-test yields a p-value that indicates whether the two sets of measurements (i.e., ground truth vs estimated web page segmentation) are similar. Lower p-values (generally p-values $< .05$), correspond to statistically significant differences, while higher p-values indicate that the two sets are statistically equivalent.

TABLE 5.8: Table of P-values for Testing Whether Web Page Segmentation Can Approximate User Browsing Metrics (Synthetically Generated Web Browsing Data)

Segmentation Approach	Number of Clicks	Average IAT	IAT Distribution
Idle-time method (Number of SYNs)	1.05×10^{-51}	2.94×10^{-19}	6.13×10^{-109}
Idle-time method (Number of Bytes)	9.36×10^{-106}	4.78×10^{-15}	4.65×10^{-74}
Idle-time method with threshold	1.23×10^{-37}	1.06×10^{-27}	3.28×10^{-100}
Heuristic Change Detection Method	1.50×10^{-31}	5.68×10^{-35}	7.31×10^{-55}
Fused Lasso	2.42×10^{-4}	7.02×10^{-43}	4.84×10^{-31}
HMM	4.31×10^{-5}	3.98×10^{-57}	3.85×10^{-91}

Evaluation Using Synthetic Data Table 5.8 shows the calculated p-values for this analysis using synthetically generated browsing streams — this is the same synthetic dataset that was used in the previous section. We find that none of the web page segmentation approaches are able to approximate the number of clicks or the average inter-arrival time metrics — that is, not a single approach yielded a p-value that was greater than .05 for either metric.

We next investigate whether the web page segmentation methods can approximate the empirical distribution of the interarrival time. We used the Kolmogorov-Smirnov test, a well-known statistical test that yields a p-value that indicates whether the two empirical distributions are similar, to determine this. The p-values that correspond to this test are also shown in Table 5.8. These p-values show that these empirical distributions are not statistically equivalent (i.e., p-values $< .001$) — that is, it is likely that web page segmentation methods *cannot* approximate the empirical distribution of interarrival times.

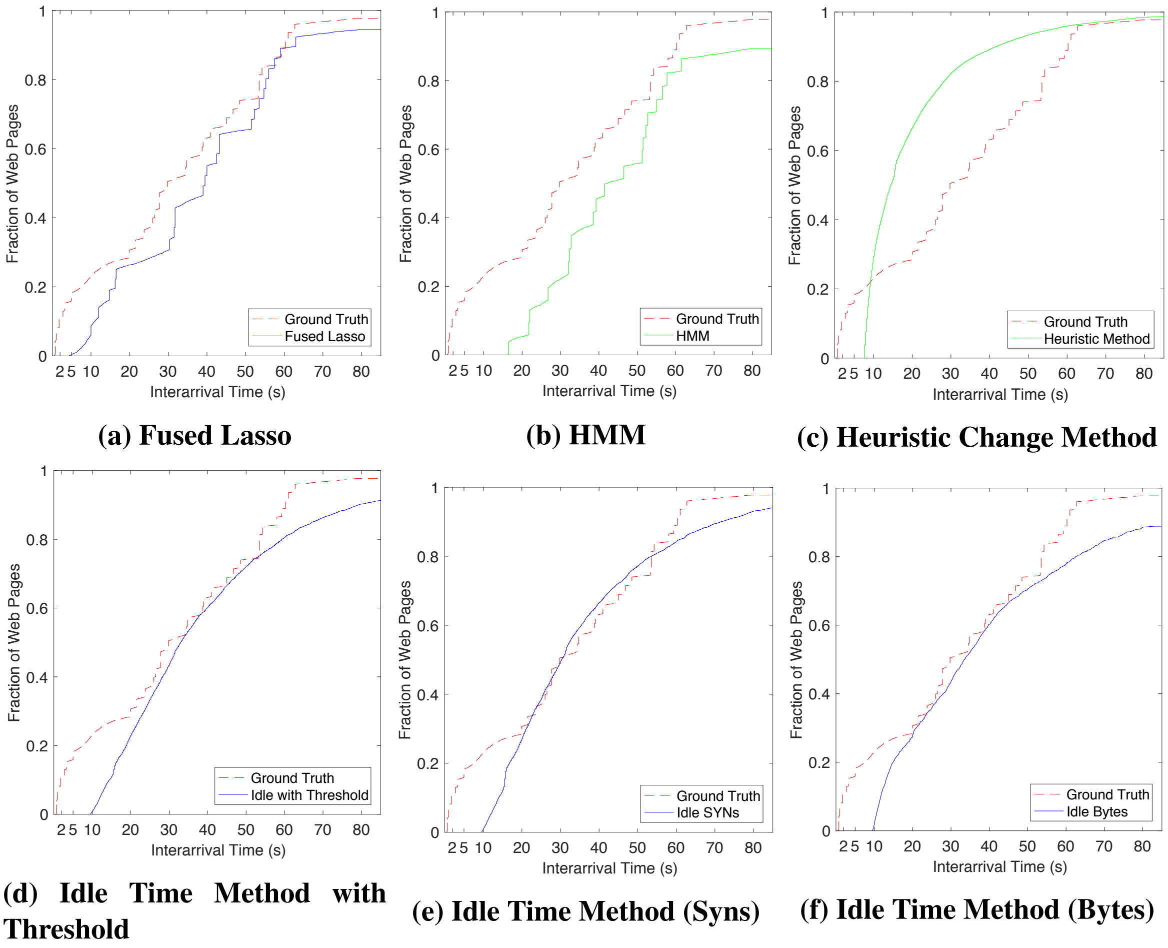
Figure 5.16 shows cumulative distribution plots which compare the interarrival time distributions of the web page segmentation methods to the ground truth — the different web page segmentation approaches used are labeled in the captions of the Figure. We find that the cumulative distributions for each of the web page segmentation methods do not significantly overlap with the cumulative distribution of the ground truth. This result is expected since the results of our distribution comparison tests, shown in Table 5.8, all yielded p-values that were very small ($< 10^{-10}$).

We also find that the cumulative distributions for most of the change point detection methods considered are consistently to the right of the ground truth distribution. This is likely because the interarrival times for web pages that are “missed” by web page segmentation approaches are added to the interarrival time of the next page that is detected. That is, every web page that is missed increases the interarrival time for the next web page that is subsequently detected. This cumulative increase in inter-arrival time is likely a contributing factor for why none of the web page segmentation methods yield interarrival time distributions that are statistically equivalent to the ground truth. The web page segmentation approaches that are not consistently to the right of the ground truth distribution are the heuristic change detection method and the idle time method with SYNs. We believe these exceptions are related to the fact that these methods have the highest false positive rates among the methods considered — more specifically, Table 5.4 shows that heuristic change detection method has a FPR of .2038 while the idle time method with SYNs has a FPR of .0869. Indeed, each additional false positive resets the inter-arrival time to zero which negates the cumulative increase in inter-arrival time that we previously discussed — that is, higher false positive rates are likely correlated to a reduction in interarrival time. This is also a possible explanation why the cumulative distribution for the heuristic change detection method, which has a FPR that is $> 2\times$ the FPR of all other methods, the farthest to the left of the ground truth distribution as compared to the cumulative distributions for the other web page segmentation methods.

TABLE 5.9: Table of P-values for Testing Whether Web Page Segmentation Can Approximate User Browsing Metrics (Real User Web Browsing Data)

Segmentation Approach	Number of Clicks	Average IAT	IAT Distribution
Idle-time method (Number of SYNs)	2.27×10^{-4}	7.93×10^{-5}	6.7252×10^{-30}
Idle-time method (Number of Bytes)	5.22×10^{-8}	1.10×10^{-4}	3.7947×10^{-29}
Idle-time method with threshold	4.63×10^{-7}	1.04×10^{-6}	4.7784×10^{-50}
Heuristic Change Detection Method	.7174	.5421	2.4585×10^{-10}
Fused Lasso	.1182	.3726	4.6371×10^{-13}
HMM	5.40×10^{-3}	4.74×10^{-2}	1.1652×10^{-54}

Evaluation Using Real User Data Table 5.9 shows the calculated p-values for this analysis using real user browsing data — this is the same real user browsing dataset that we used in the previous section. In this scenario, we find that the p-values for the Heuristic Change Detection Method and the Fused Lasso method are above .05 for both the number of clicks and average inter-arrival time metrics. This is different from our analysis using synthetic data where neither method yielded p-values higher than .05 for these metrics. Based on our analysis in the previous section, it is likely that small interarrival times are the root cause



(a) Fused Lasso

(b) HMM

(c) Heuristic Change Method

(d) Idle Time Method with Threshold

(e) Idle Time Method (Syns)

(f) Idle Time Method (Bytes)

Figure 5.16: Distribution comparison between ground truth web page inter-arrival time and the web page segmentation approximated inter-arrival times (Synthetic Dataset).

of the disparity in the performance for web page segmentation approaches — that is, it is likely that the performance difference between the real user browsing data and the synthetically generated data is due to differences in their interarrival time distributions. To investigate this, we plot the cumulative distributions of the interarrival times for these two datasets in Figure 5.17. We find that:

- The interarrival time distributions for the real user data and the synthetic data have different characteristics. In fact, the p-value resulting from performing a Kolmogorov-Smirnov test between these is statistically significant (p = 1.2084×10^{-78}).
- Approximately 10% of the interarrival times for the synthetic data is $\leq$ 2s, while it is rare for the interarrival time for the real data to be $\leq$ 2s. This is a significant fraction of very small interarrival times that is exclusive to the synthetic data. It is likely that these small interarrival times negatively impact

the results for the web page segmentation methods since their corresponding web page downloads will be difficult to detect using the approaches that we consider.

- In summary, the web page segmentation methods have more difficulty approximating the number of clicks and average interarrival time metrics for the synthetic dataset than the real dataset.

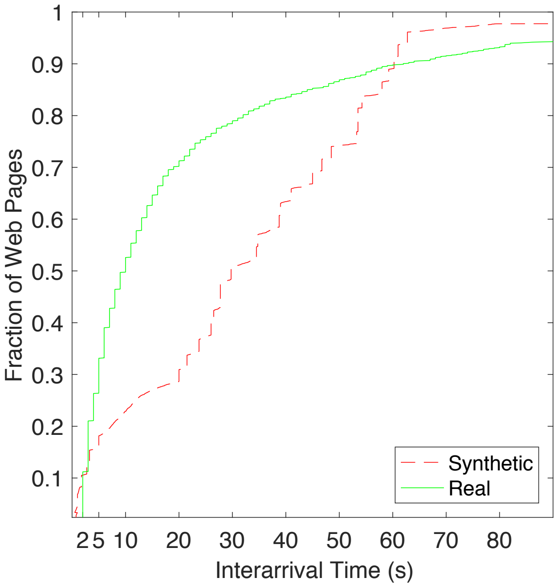
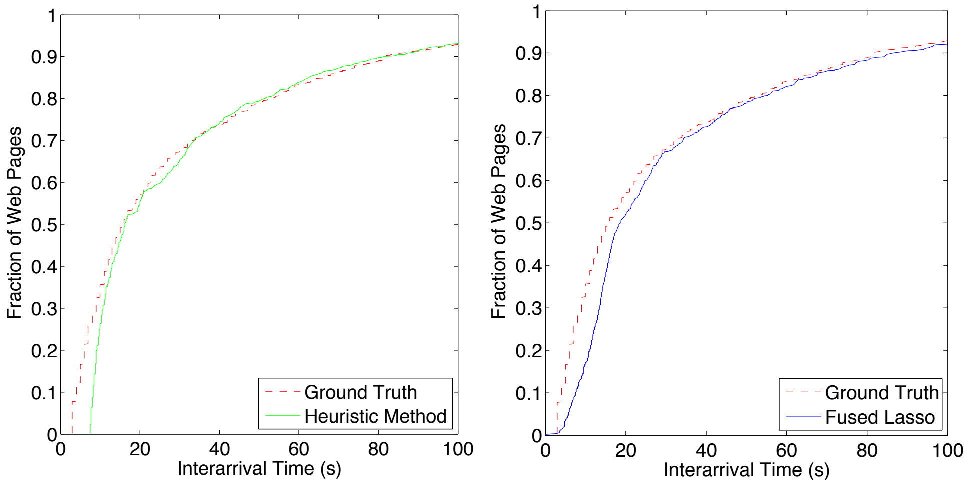


Figure 5.17: Distribution comparison between synthetic and real user browsing interarrival time distributions.

Next, we used the Kolmogorov-Smirnov to determine whether web page segmentation methods can approximate the complete empirical distribution of web page inter-arrival times using real user data — the p-values that correspond to this are shown in Table 5.9. These p-values show that these empirical distributions are not statistically equivalent. The cumulative distributions for the web page interarrival times as computed using the heuristic change method and the fused lasso regression method are provided in Figure 5.18(a) and Figure 5.18(b), respectively — these plots also include the ground truth cumulative distribution as a reference. These plots show that even the best performing web page segmentation approaches have difficulty identifying web pages with low inter-arrival time. Despite this limitation, being able to approximate the *average* web page inter-arrival time is a significant enabler for driving simulations and modeling problems in many application domains.

These results imply two things: 1) some user browsing metrics can be derived from *real user browsing data* using web page segmentation approaches; and 2) the web page segmentation approach used can make a significant difference in estimating such metrics.



(a) Comparison with Heuristic **(b) Comparison with Fused Lasso**

Figure 5.18: Distribution comparison between ground truth web page inter-arrival time and the web page segmentation approximated inter-arrival times (Real Dataset).

5.5.2 Impact Web Page Segmentation Methods have on Web Page Classification Performance

We next investigate whether web page classification methods are impacted by the performance of web page segmentation approaches. To do this, we apply web page classification on browsing streams that have been segmented using different web page segmentation approaches. More specifically, we use the KNN algorithm,[31] the best performing classification method in Chapter 4, to classify web pages according to different labeling schemes when applied to browsing streams that are segmented using different web page segmentation methods. For clarification we refer to the browsing streams that are segmented using a web page segmentation approach as *web page segments* — each web page segment includes all traffic that was transferred within each TCP connection established between web page inter-arrivals. This definition of a web page segment means that traffic that is transmitted in persistent TCP connections after a new web page download starts is included in the previous web page download.

Metrics used for evaluation In this case study, it is possible for the performance of the web page segmentation method to indirectly impact the performance of the web page classification method that is applied to the segmented traffic. This is because the web page segments that are input to the web page classification

[31] Where K = 1 and the distance function is city block.

method are determined by the web page segmentation method — these may not align with the ground truth web page segments. Thus, there are scenarios where the web page segments that are being classified differ from the ground truth web page segments. This is an issue because it impacts the interpretability and value of traditional metrics such as the number of false positives and the number of true positives. For example, if we consider a web page traffic trace that includes a single web page download of class A but the web page segmentation method detects two web page downloads — how should we evaluate the performance? Clearly, we are classifying two web pages where there is only one web page considered in the ground truth. It is possible that we correctly classify the web page as class A in one segment, and that we incorrectly classify the web page as class B in the other segment — one may consider such a classification as being one true positive and one false positive. It is also possible that we correctly classify both web page segments as class A. For this case, should we consider the results as two true positives since they both categorize the same page correctly, or should we consider it as one true positive and one false positive since, according to the ground truth, there is only one web page to classify? Indeed, it is important to clearly define how we address such scenarios.

We address this issue by ensuring that *at most* one web page segment q_j, as identified using a web page segmentation approach, can be considered as either a true positive or a false positive per ground truth web page segment t_i — other web page segments q_j that overlap with a ground truth segment t_i are considered duplicates.[32] In practice, this is done by first checking whether any of the segments q_j, which overlap with ground truth segment t_i, share the same label as ground truth segment t_i. If at least one segment q_j shares the same label as a segment t_i, a single segment q_j is selected as a true positive – if not, a single segment q_j is chosen to be a false positive.[33] We prioritize the results primarily so that the labeling procedure of true positives and false positives is consistent across browsing streams. Prioritizing the selection of true positives over false positives biases results towards a higher number of true positives. We do this because we assume that true positives are a better match with the ground truth. We do not believe that this is a significant problem since (i) our performance evaluation is largely relative to the performance of different web page segmentation methods and (ii) we use other metrics such as the number of duplicate false positives and the number of duplicate true positives to quantify the degree of overlapping segments in our analysis. Details

[32] Either duplicate true positives, duplicate false positives, or duplicate positives. We described these in more detail later.

[33] In both cases, the segment that is selected prioritized according to ascending temporal order — that is, segments that occur first take priority in labeling.

Algorithm 4 Pseudocode for Labeling Segmented Web Pages (Phase 1 and 2)

```
input: $q, t$                                          ▷ $q$ and $t$ are ordered temporally
initialize: $Q$ as length of q, and $T$ as length of t
initialize: $qlabels$ to array of length Q
initialize: $tlabels$ to array of length T                                   ▷ Phase 1:
for $i = 0, 1, ...., T$ do
    for $j = 0, 1, ...., Q$ do
        if $qlabels_j$ equals $TruePositive$ OR $tlabels_i$ equals $TruePositive$ then
            $continue$
        end if
        if $t_i$ overlaps with $q_j$ AND Label($t_i$) equals Label($q_j$) then
            $qlabels_j = TruePositive$
            $tlabels_i = TruePositive$
            $break$
        end if
    end for
end for                                                                      ▷ Phase 2:
for $i = 0, 1, ...., T$ do
    if $tlabels_i$ equals $TruePositive$ then
        $continue$
    end if
    for $j = 0, 1, ...., Q$ do
        if $qlabels_j$ equals $TruePositive$ OR $qlabels_j$ equals $FalsePositive$ then
            $continue$
        end if
        if $t_i$ overlaps with $q_j$ AND Label($t_i$) not equals Label($q_j$) then
            $qlabels_j = FalsePositive$
            $tlabels_i = FalsePositive$
            $break$
        end if
    end for
end for
```

on how we define such metrics (e.g., number of duplicate true positives and duplicate false positives) for a number of different scenarios are provided below:

- *True positives:* A true positive occurs when a web page segment q_j, as identified using a web page segmentation method, is classified using the same label associated with *at least one* ground truth web page segment t_i in which it overlaps with in time. Please note, however, that multiple ground truth web page segments, t_i, may overlap with a single web page segment q_j. This is because, at times, web page segmentation methods may not detect web page downloads. Also note, that multiple web page segments q_j may overlap with a single ground truth web page segment t_i. In either case, there is at most *one* true positive per ground truth segment. The true positive rate (TPR) is calculated as the number of true positives divided by the number of web page downloads.

- *Duplicate true positives:* A duplicate true positive occurs when multiple web page segments q_j are (i) correctly classified and (ii) overlap in time with a web page segment t_i — this occurs when a web page segmentation method is overly aggressive when segmenting web page traffic. For example, a scenario where there are three q_j segments that are correctly classified with the same label as t_i, will yield 1 true positive and 2 duplicate true positives. The duplicate true positive rate (DTPR) is calculated as the number of duplicate true positives divided by the number of web page downloads.

- *False positives:* A false positive may occur when a web page segment q_j is incorrectly classified for *at least one* of the segments t in which it overlaps with in time. There can be at most *one* segment q_j that can be categorized as a false positive per ground truth segment. Also, a ground truth segment t_i cannot have multiple segments q_j that are a labeled as a true positive or false positive — that is, the sum of the number of true positives and false positives for all segments q_j that overlap with a single web page segment t must be less than or equal to 1. The false positive rate (FPR) is calculated as the number of false positives divided by the number of web page downloads.

- *Duplicate false positives:* A duplicate false positive occurs when (i) a web page segment q_j does not share the same label with all of the segments t in which it overlaps with in time and (ii) another web page segment q_j exists that is either a true positive or false positive for segment t — please note that a duplicate false positive cannot exist for web page segment t unless another segment q is categorized as a true positive or a false positive. Duplicate false positives and duplicate true positives are assigned only after all true positives and false positives have been assigned. The duplicate false positive rate (DFPR) is calculated as the number of duplicate false positives divided by the number of web page downloads.

- *Duplicate positives:* It is important to note that it is possible for a single web page segment q_j to serve as both a duplicate false positive and a duplicate true positive. This may occur when a single web page segment q_j overlaps with multiple web page segments t_i — specifically, when segment q_j is a duplicate true positive for one segment t_i in which it overlaps with in time and is a duplicate false positive for the other segment. In such scenarios, we denote the web page segment q_j as a duplicate positive. The duplicate positive rate (DPR) is calculated as the number of duplicate positives divided by the number of web page downloads.

- *False negatives:* A false negative occurs when a web page segment q does not exist to be classified as any of the above metrics for a ground truth web page segment t_i.[34] This occurs when a web page segmentation method is too conservative and fails to detect web page downloads. The false negative rate (FNR) is calculated as the number of false negatives divided by the number of web page downloads.

Pseudocode for the procedures we use to label web page segments according to the above metrics is provided in Algorithms 4 and 5. These procedures includes 4 main phases that are denoted as Phase 1, Phase 2, Phase 3, and Phase 4, respectively. Phase 1 and Phase 2 are provided in Algorithm 4, while Phase 3 and Phase 4 are provided in Algorithm 5. The purpose of Phase 1 is to determine which segments q_j and t_i should be labeled as true positives — please note that the $Label(A)$ method outputs the classification label of A (either a ground truth label or web page classification-based label). Phase 2 is similar to Phase 1, except the purpose is to determine which segments q_j and t_i should be labeled as false positives. Phase 3 determines which segments q_j should be labeled as duplicates (either duplicate true positives, duplicate false positives, or duplicate positives), while Phase 4 determines which segments t_i should be labeled as false negatives.

A summary of some of the primary ways in which these metrics should be interpreted is provided below:

- Web page segments q and web page segments t have the same number of true positives and false positives. That is, there is a one-to-one match between each true positive and false positive between web page segments q and t.

- Duplicate true positives, duplicate false positives, and duplicate positives only occur for web page segments q — these are best viewed as extra or alternative segments that were not matched as a true positive or false positive with a web page segment t_i. The sum of these is a measure of the number of times a web page segmentation method detected extra web pages. Thus, high values for $DTPR$, $DFPR$, and DPR are primarily due to overly aggressive web page segmentation. Please note that this means that the $DTPR$, $DFPR$, and DPR for the ground truth segmentation are 0.

[34] Each web page segment q_j must be categorized as either a true positive, false positive, duplicate false positive, duplicate true positive, or duplicate positive.

Algorithm 5 Pseudocode for Labeling Segmented Web Pages (Phase 3 and 4)

```
input: $q$, $t$, $qlabels$, $tlabels$, $Q$, $T$          ▷ Phase 3: Continued from Phase 2
for $j = 0, 1, ...., Q$ do
    if $qlabels_j$ equals $TruePositive$ OR $qlabels_j$ equals $FalsePositive$ then
        $continue$
    end if
    $ftrue = 0$
    $ffalse = 0$
    for $i = 0, 1, ...., T$ do
        if $t_i$ overlaps with $q_j$ AND Label($t_i$) not equals Label($q_j$) then
            $ffalse = 1$
        else if $t_i$ overlaps with $q_j$ AND Label($t_i$) equals Label($q_j$) then
            $ftrue = 1$
        end if
    end for
    if $ffalse$ equals 1 $AND$ $ftrue$ equals 1 then
        $qlabels_j = DuplicatePositive$
    else if $ffalse$ equals 1 then
        $qlabels_j = DuplicateFalsePositive$
    else if $ftrue$ equals 1 then
        $qlabels_j = DuplicateTruePositive$
    end if
end for
                                                          ▷ Phase 4:
for $i = 0, 1, ...., T$ do
    if $tlabels_i$ equals $TruePositive$ OR $tlabels_i$ equals $FalsePositive$ then
        $continue$
    else
        $tlabels_i = FalseNegative$
    end if
end for
$output$ : $qlabels, tlabels$
```

- False negatives only occur for ground truth web page segments t. The number of false negatives is a measure for the number of times a web page segmentation method failed to detect a web page download — thus, the FNR for the ground truth segmentation is 0.

- Web page segments t can only be labeled as true positives, false positives, and false negatives. Thus, the sum of the rates for these is equal to 1.

We are able to derive traditional classification performance metrics, such as precision, recall, and F-score, using the number of true positives, false positives, and false negatives observed in our results. The number of duplicate true positives and duplicate false positives are used to give a measure of the impact that excessive segmentation of web page traffic may have on the interpretation of the web page classification results. Conversely, the number of false negatives is used to give a measure of the impact of overly conservative segmentation.

Evaluation Using Synthetic Browsing Data For the evaluation using synthetic data, we consider the same synthetic dataset used in the previous Section 5.3. For each web page browsing stream in this dataset, we classify each web page segment (as determined using a web page segmentation approach) according to each of the 4 orthogonal labeling schemes — the performance for the genre-based, navigation-based, video streaming-based, and target device-based labeling schemes are provided in Table 5.10, Table 5.11, Table 5.12, and Table 5.13, respectively. We find that:

TABLE 5.10: Classification Performance Using Web Page Segmentation for Synthetic Data (Genre-based Labels)

Segmentation Approach	TPR/Recall	FPR	Precision	F-score
Idle-time method (Number of SYNs)	.2765/.3148	.1216	.6945	.4332
Idle-time method (Number of Bytes)	.0513/.0576	.1093	.3194	.0976
Idle-time method with threshold	.2416/.2671	.0954	.7169	.3892
Heuristic Change Detection Method	.4401/.6611	.3338	.5687	.6114
Fused Lasso	.4812/.6673	.2789	.6331	.6497
HMM	.3140/.3697	.1507	.6757	.4779
Ground Truth Segmentation	.6234/1	.3766	.6234	.7680
	DTPR	DFPR	DPR	FNR
Idle-time method (Number of SYNs)	.0319	.0591	.0126	.6019
Idle-time method (Number of Bytes)	.0012	.0030	.0011	.8394
Idle-time method with threshold	.0052	.0288	.0023	.6630
Heuristic Change Detection Method	.0571	.0845	.0203	.2261
Fused Lasso	.0619	.0608	.0018	.2399
HMM	.0322	.0521	.0173	.5353
Ground Truth Segmentation	0	0	0	0

- *Change point detection methods outperform idle time-based methods:* Table 5.10 shows that, according to the F-score metric, the change point detection methods outperform the idle time-based methods (.09-.43 vs .47-.64), where the fused lasso method performs the best (.6497) — we observe similar results when considering the navigation-based, video streaming-based, and target device-based labeling schemes. This result is expected since change point detection methods more accurately segments

TABLE 5.11: Classification Performance Using Web Page Segmentation for Synthetic Data (Navigation-based Labels)

Segmentation Approach	TPR/Recall	FPR	Precision	F-score
Idle-time method (Number of SYNs)	.3172/.3451	.0810	.7966	.4816
Idle-time method (Number of Bytes)	.1208/.1258	.0398	.7522	.2156
Idle-time method with threshold	.2763/.2942	.0607	.8199	.4330
Heuristic Change Detection Method	.5728/.7170	.2011	.7401	.7284
Fused Lasso	.6007/.7146	.1534	.7966	.7534
HMM	.3780/.4139	.0867	.8134	.5486
Ground Truth Segmentation	.6811/1	.3189	.6811	.8103
	DTPR	DFPR	DPR	FNR
Idle-time method (Number of SYNs)	.0443	.0341	.0252	.6019
Idle-time method (Number of Bytes)	.0037	.0009	.0006	.8394
Idle-time method with threshold	.0244	.0086	.0033	.6630
Heuristic Change Detection Method	.0980	.0414	.0225	.2261
Fused Lasso	.0783	.0345	.0117	.2399
HMM	.0406	.0450	.0160	.5353
Ground Truth Segmentation	0	0	0	0

web page browsing sessions than idle time-based methods.[35] It is also important to note that other metrics also usually perform better for the change point detection methods than for idle time-based methods. For instance, the true positive rates for change point detection methods are higher than the true positive rates for the idle time-based methods, while the false negative rates are lower.[36] One important exception to this is the observation that the false positive rate is higher for change point detection methods than for the idle time-based methods (.09-.20 vs .06-.08). We believe that this is due to the fact that change point detection methods are able to detect more web page downloads than idle time-based methods — thus, there are more opportunities for web page classification methods to both correctly classify and incorrectly classify web page segments. Overall, the higher false positive rates (and true positive rates) for the change point detection methods is a benefit to the performance of these techniques — we know this because, as noted before, the F-score calculated for these methods is higher than the F-score for idle time-based methods.

- *Ground truth segmentation outperforms all methods:* Table 5.10, Table 5.11, Table 5.12, and Table 5.13 shows that the ground truth web page segmentation outperforms all the other web page

[35] Please refer to Table 5.4.

[36] We discuss the false negative rate in more detail later.

segmentation methods when using the F-score metric. Indeed, this result is expected since web page classification should perform better when more accurate web page segments are provided. Nevertheless, this result is important because it supports the premise that accurate web page segmentation improves web page classification performance — please note that this is also supported by the observation that change point detection methods outperform idle time-based methods.

The difference in the classification performance between the ground truth web page segmentation and the other web page segmentation methods is non-negligible. In particular, the decline in the F-score metric between the ground truth segmentation and the next best performing web page segmentation method ranges from 7-17% depending on which labeling scheme is used — please note that the genre-based labeling scheme exhibits the largest performance decline while the navigation-based labeling scheme exhibits the smallest. It is likely that this 7-17% drop in F-score is significant enough to impact the utility of web page classification in the "real world" where the ground truth segmentation is not known.

- *High false negative rates:* We find that every web page segmentation method for the synthetic dataset has a high false negative rate that is at least .22 — in fact, some web page segmentation methods have false negative rates that are as high .84. This high false negative rate is due to the difficulty that web page segmentation approaches have in detecting web pages in the synthetic dataset. We specifically highlight the false negative rate as a key metric for reduced classification performance because it defines an upper bound for the true positive and false positive rates when classification methods are used in tandem with web page segmentation — indeed, a web page must first be detected in order to be classified by as classification method.[37] Thus, the fairly low F-scores and true positive rates (less than .75) across all segmentation methods shown in the tables is largely due to the difficulty of web page segmentation methods to detect web page downloads (i.e, an *FNR* that is at least .22).

- *The number of duplicates is lower than the number of true positives and false positives:* We find that the sum of the TPR and FPR metrics is less than the sum of the DTPR, DFPR, and DPR metrics for each web page segmentation approach — this result occurs for every labeling scheme considered. This result means that, for all web page segmentation methods, it is more likely that web page segments correspond to a single ground truth segment than multiple ground truth segments.

[37] For example, a false negative rate of .25 means that the highest possible true positive rate is .75.

- *Labeling scheme choice impacts classification performance:* Labeling scheme choice has an impact on web page classification performance when used in tandem with web page segmentation methods. For example, when we consider the ground truth segmentation, the best performing labeling scheme according to the F-score metric is the video streaming-based labeling scheme (.9333) — this is followed by the target device-based labels (.8992), navigation-based labels (.8103), and the genre-based labels (.7680). This result is similar to those discussed in Chapter 4 in that the video streaming labeling scheme performs the best, while the target device-based, navigation-based, and genre-based labeling schemes perform second, third, and fourth, respectively.

  We also find that the F-score for each of these labeling schemes when web page segmentation and web page classification are used in tandem are within .06 of the F-scores presented in Chapter 4. Though, it is important to note that the F-scores presented in this section are *inflated* as compared to the F-scores presented in Chapter 4 — this is due to the fact that the recall metric for the ground truth segmentation is always one.[38] When we compare the TPR (accuracy)[39] between these two sets of results, we observe that the web page segmentation-based results are 9-15% lower than the results in Chapter 4. The navigation-based labeling scheme suffers the largest performance decline of 15% — this is followed by performance declines of 12%, 11%, and 8% for the video streaming-based, genre-based, and target device-based labeling schemes, respectively. It is difficult to determine a concrete explanation for the different degree of performance declines across the different labeling schemes since many key factors, such as the impact of overlapping traffic, are difficult to account for.

Evaluation Using Real User Browsing Data For the evaluation using real user browsing data, we focus on the genre-based (17 classes) and targeted device-based (2 classes) labeling schemes considered in Chapter 4. We consider the genre-based labeling scheme since it is the only labeling scheme that we considered in Chapter 4 that is straightforward to categorize using web browser history logs. More specifically, all that is needed is to extract the hostname from the URL and categorize the page according to the genre provided by Alexa for that hostname — if Alexa does not classify the web page, the label assigned is "Unknown." The navigation-based, video-based, and targeted device-based labeling schemes that we use in Chapter 4

[38] This is because the false negative rate, according to our definition in this section, is zero for the ground truth segmentation — thus, the recall, r, calculation simplifies from $r = \frac{TP}{TP+FN}$ to 1.

[39] which does not suffer from such inflation issues

Table 5.12: Classification Performance Using Web Page Segmentation for Synthetic Data (Video streaming-based Labels)

Segmentation Approach	TPR/Recall	FPR	Precision	F-score
Idle-time method (Number of SYNs)	.3435/.3633	.0547	.8626	.5113
Idle-time method (Number of Bytes)	.0841/.0911	.0765	.5237	.1551
Idle-time method with threshold	.2719/.2908	.0651	.8068	.4275
Heuristic Change Detection Method	.6874/.7525	.0865	.8882	.8147
Fused Lasso	.6942/.7431	.0599	.9206	.8224
HMM	.3586/.4012	.1061	.7717	.5279
Ground Truth Segmentation	.8750/1	.1250	.8750	.9333
	DTPR	DFPR	DPR	FNR
Idle-time method (Number of SYNs)	.0218	.0499	.0319	.6019
Idle-time method (Number of Bytes)	.0029	.0020	.0009	.8394
Idle-time method with threshold	.0047	.0279	.0037	.6630
Heuristic Change Detection Method	.1246	.0175	.0198	.2261
Fused Lasso	.0794	.0241	.0210	.2399
HMM	.0642	.0288	.0086	.5353
Ground Truth Segmentation	0	0	0	0

Table 5.13: Classification Performance Using Web Page Segmentation for Synthetic Data (Targeted device-based Labels)

Segmentation Approach	TPR/Recall	FPR	Precision	F-score
Idle-time method (Number of SYNs)	.3275/.3524	.0707	.8224	.4934
Idle-time method (Number of Bytes)	.0719/.0789	.0887	.4477	.1341
Idle-time method with threshold	.1641/.1984	.1729	.4869	.2819
Heuristic Change Detection Method	.6703/.7485	.1036	.8666	.8032
Fused Lasso	.6444/.7287	.1097	.8545	.7866
HMM	.3491/.3947	.1156	.7512	.5175
Ground Truth Segmentation	.8169/1	.1831	.8169	.8992
	DTPR	DFPR	DPR	FNR
Idle-time method (Number of SYNs)	.0781	.0120	.0135	.6019
Idle-time method (Number of Bytes)	.0041	.0004	.0008	.8394
Idle-time method with threshold	.0308	.0032	.0023	.6630
Heuristic Change Detection Method	.1302	.0257	.0060	.2261
Fused Lasso	.0818	.0301	.0126	.2399
HMM	.0747	.0148	.0121	.5353
Ground Truth Segmentation	0	0	0	0

TABLE 5.14: Classification Performance Using Web Page Segmentation for Real Data (Genre-based Labels)

Segmentation Approach	TPR/Recall	FPR	Precision	F-score
Idle-time method (Number of SYNs)	.2317/.3362	.3108	.4271	.3762
Idle-time method (Number of Bytes)	.0824/.0976	.1561	.3455	.1522
Idle-time method with threshold	.1998/.2259	.1155	.6337	.3331
Heuristic Change Detection Method	.4674/.5314	.2646	.6604	.5889
Fused Lasso	.4582/.5001	.2338	.6895	.5797
HMM	.3810/.4607	.1729	.6878	.5518
Ground Truth Segmentation	.6162/1	.3838	.6162	.7625
	DTPR	DFPR	DPR	FNR
Idle-time method (Number of SYNs)	.1204	.0944	.0224	.4575
Idle-time method (Number of Bytes)	.0218	.0227	.0057	.7615
Idle-time method with threshold	.0772	.1243	.0486	.6847
Heuristic Change Detection Method	.1098	.1554	.0321	.2680
Fused Lasso	.0503	.0320	.0108	.3080
HMM	.0626	.0947	.0075	.4461
Ground Truth Segmentation	0	0	0	0

cannot be consistently categorized using a URL alone. Please recall that we addressed this problem in Chapter 4 by manually inspecting each web page download. This is not possible using real user browsing logs because many users (60%) visited at least one web site that requires login information (so we do not know specific details such as whether the user was searching for content, watching a video, or even visiting a mobile-optimized web page).[40] However, for the case of the targeted device-based labeling scheme, we know that the users in our study were using a desktop machine during their browsing session. Thus, we use the targeted-device labeling scheme since it is reasonable to assume that all of the web pages downloaded by the real users can be labeled as traditional web pages.

The web page classification results when using the genre-based and target device-based labeling schemes are shown in Table 5.14 and Table 5.15, respectively. Table 5.14 shows that the F-scores for the change point detection methods (i.e., heuristic change detection method, fused lasso, and HMM) are higher than the F-scores for the idle time-based methods for the genre-based labeling scheme — similar observations can be made for the classification results for the target device-based labeling scheme (Table 5.15). More specifically, for the genre-based labels, the F-scores for the change point detection methods are between .55-

[40] 28.16% of the total number of web pages visited by the users in our study requires login information. Also, we found many cases during our analysis where search queries could not be classified using simple heuristics such as having a "q=" in the URL. Similar observations were made for mobile optimized web pages which may not include the "m dot" prefix.

.59, while the F-score for the idle-time based methods are between .15-.37 — the gap between these ranges is .18. Other key observations that were made from analyzing the other metrics provided in Table 5.14 and Table 5.15 are similar to the observations made from the synthetic dataset (e.g., high false negative rates and ground truth segmentation outperforms all other methods).

It is also important to note that the overall performance of web page classification is modest even when we use the "best" web page segmentation methods for the real browsing streams. In fact, the highest true positive rate, or accuracy, that we observe for the genre-based labels is less than .42, while the highest true positive rate that we observe for the target device-based labels is less than .60. Recall, that the classification performance that we observed for the genre-based and target device-based labeling schemes in Chapter 4 was .71 and .89, respectively. Overall, this corresponds to approximately a .30 decline in TPR when we consider the realistic scenario where web pages must be segmented before being classified — other metrics such as F-score, precision, and recall are impacted in a similar manner. Overall, these performance declines show that using web page classification in a more realistic scenario than considered in Chapter 4 faces significant practical challenges. Improvements are needed in order for the web page classification to be more robust to more realistic traffic. This includes improving the web page segmentation approaches such that the number of true positives increase, the number of false negatives decrease, and the number of duplicate positives (or overlapping segments) decrease. This also includes designing web page classification methods that are more robust to error-prone web page segments.

5.5.3 Evaluation Using User Profiles

Pure classification performance is not the only way to interpret the performance of web page classification. For instance, the user browsing profiles considered in Chapter 4, which are constructed using the frequency of which web pages are classified in a browsing stream, may be useful for gauging user interest. Thus, we investigate the efficacy of using web page classification to approximate a user browsing profile. We provide details on our methodology for building user profiles and present classification results that are based on these profiles in this section.

Building User Profiles In order to build profiles of user browsing behavior we must (i) categorize the web pages that were visited by a user during their web browsing session and (ii) weight the preference of each category for each user. We use the Alexa genre to categorize web pages for this study since it

TABLE 5.15: Classification Performance Using Web Page Segmentation For Real Data (Target device-based Labels)

Segmentation Approach	TPR/Recall	FPR	Precision	F-score
Idle-time method (Number of SYNs)	.5202/.4787	.0223	.7746	.5917
Idle-time method (Number of Bytes)	.2068/.2136	.0317	.8671	.3428
Idle-time method with threshold	.2768/.2879	.0385	.8779	.4336
Heuristic Change Detection Method	.6434/.6170	.0886	.9389	.7447
Fused Lasso	.5966/.5726	.0954	.9233	.7068
HMM	.4787/.5176	.0752	.8642	.6474
Ground Truth Segmentation	.7418/1	.2582	.7418	.8518
	DTPR	DFPR	DPR	FNR
Idle-time method (Number of SYNs)	.1804	.0568	0	.4575
Idle-time method (Number of Bytes)	.0258	.0244	0	.7615
Idle-time method with threshold	.1639	.0862	0	.6847
Heuristic Change Detection Method	.2076	.0850	0	.2680
Fused Lasso	.0507	.0424	0	.3080
HMM	.1105	.0543	0	.4461
Ground Truth Segmentation	0	0	0	0

straightforward to categorize using web browser history logs.[41] Profiles for each user are then generated by sorting the frequency in which each Alexa genre occurred in the web browsing history for each user. For example, a user that downloaded 8 web pages in which four web sites were categorized as "News", 3 web sites categorized as "Business", and 1 categorized as "Weather" would have a profile of {News, Business, Weather}, where "News" is the most frequent genre and "Weather" is the least frequent.[42] We build profiles according to the each other the four labeling schemes when considering the synthetic data and for the genre-based and target device-based labeling schemes when considering the real data.[43]

Performance of Profile-based Classification We first consider the performance of the profile-based classification when using the real user browsing data. The user profile-based classification performance for the genre-based and target device-based labels are shown in Table 5.16 and Table 5.17, respectively.[44] The

[41] All that is needed is to extract the hostname from the URL and categorize the page according to the genre provided by Alexa for that hostname — the navigation-based, video-based, and device-based labels that we use in Chapter 4 cannot be consistently/reliably categorized using a URL alone for the same reasons discussed earlier in this section.

[42] Web pages without an Alexa genre are ignored.

[43] Please note that we only considered the genre-based and target device-based labeling schemes for the real user browsing data for the same reasons provided in the previous section.

[44] The percentages only include 36 of the 40 users. This is because a few users browsed pages that Alexa did not have a label for. More specifically, four users generated browsing streams where over 90% of the web pages observed in the browser history did not a genre-based label. So we are unable to reliably profile their browsing behavior.

TABLE 5.16: Table Showing the Impact that Web Page Segmentation Approach has on Web Page Classification User Profiles (Genre-based labels).

Segmentation Approach	1st	2nd	3rd
Ground Truth Segmentation	76%	70%	38%
Fused Lasso	76%	62%	28%
Heuristic Change Detection Method	68%	62%	28%
HMM	52%	48%	25%
Idle-time method with threshold	58%	38%	22%
Idle-time method (Number of SYNs)	50%	30%	22%
Idle-time method (Number of Bytes)	41%	9%	0%

TABLE 5.17: Table Showing the Impact that Web Page Segmentation Approach has on Web Page Classification User Profiles (Targeted device-based labels).

Segmentation Approach	1st
Ground Truth Segmentation	100%
Fused Lasso	97%
Heuristic Change Detection Method	97%
HMM	97%
Idle-time method with threshold	97%
Idle-time method (Number of SYNs)	100%
Idle-time method (Number of Bytes)	100%

columns in Table 5.16 correspond to the rate in which the classification and segmentation method can classify the 1st, 2nd, and 3rd most popular content genres visited by a user in a given web browsing stream — please note that Table 5.17 only includes a single column since there are only two categories considered for the target device-based labeling scheme. Table 5.16 shows that the selection of web page segmentation technique has a large impact on the performance of web page classification methods. In particular, the heuristic change detection method and fused lasso can identify the top 2 most popular categories visited in over two-thirds of the users. This is comparable to the performance that the ground truth segmentation yields. We also observe that the performance achieved when using the heuristic change detection method and fused lasso are better than the other methods. Overall, the results for the user browsing profiles show that web page segmentation method choice can have a significant impact on the performance while profiling user behavior.

The conclusions made when analyzing the results for the genre-based labeling scheme do not hold when considering the target device-based labeling scheme. Instead, Table 5.17 shows that the performance for the different web page segmentation methods are comparable. More specifically, each web page segmentation method identifies traditional web pages as the dominant web page category for either 97% or 100% of the browsing streams. While we expected promising results from the change point detection methods, since they are more effective at segmenting browsing streams than idle time-based methods, the high performance of the idle time-based approaches are somewhat unexpected. Upon further inspection, we find that the poor web page segmentation performance of the idle time-based methods is *surprisingly* beneficial for web page classification in this scenario. This is because fewer web page segments detected yields larger web page segments on average. Thus, in this scenario, larger web page segments will likely be classified as "traditional web pages" which is the only category of web pages present in our browsing stream — please note that the low FPR for the idle time-based methods in Table 5.15 supports this. Because of this, we believe that the high performance of the idle-time based methods for the target device-based labels is due to coincidence and, therefore, is not a positive reflection of the performance of web page classification when used in tandem with these methods.

We next consider, the performance of the profile-based classification when using synthetically generated using browsing data — this is the same synthetic dataset considered in the previous section. The profile-based classification results for synthetically generated browsing schemes for the genre-based, navigation-based, target device-based, and video streaming based labels are shown in Table 5.18, Table 5.19, Table 5.21,

TABLE 5.18: Table Showing the Impact that Web Page Segmentation Approach has on Web Page Classification User Profiles for Synthetic Data (Genre-based labels).

Segmentation Approach	1st	2nd	3rd
Ground Truth Segmentation	75%	67%	53%
Fused Lasso	71%	64%	51%
Heuristic Change Detection Method	63%	55%	43%
HMM	59%	54%	32%
Idle-time method with threshold	52%	48%	33%
Idle-time method (Number of SYNs)	45%	29%	20%
Idle-time method (Number of Bytes)	32%	23%	16%

TABLE 5.19: Table Showing the Impact that Web Page Segmentation Approach has on Web Page Classification User Profiles for Synthetic Data (Navigation-based labels).

Segmentation Approach	1st	2nd
Ground Truth Segmentation	78%	71%
Fused Lasso	71%	66%
Heuristic Change Detection Method	70%	61%
HMM	70%	59%
Idle-time method with threshold	62%	56%
Idle-time method (Number of SYNs)	54%	44%
Idle-time method (Number of Bytes)	53%	42%

TABLE 5.20: Table Showing the Impact that Web Page Segmentation Approach has on Web Page Classification User Profiles for Synthetic Data (Targeted device-based labels).

Segmentation Approach	1st
Ground Truth Segmentation	94%
Fused Lasso	91%
Heuristic Change Detection Method	92%
HMM	90%
Idle-time method with threshold	87%
Idle-time method (Number of SYNs)	91%
Idle-time method (Number of Bytes)	85%

TABLE 5.21: Table Showing the Impact that Web Page Segmentation Approach has on Web Page Classification User Profiles for Synthetic Data (Video streaming-based labels).

Segmentation Approach	1st
Ground Truth Segmentation	92%
Fused Lasso	92%
Heuristic Change Detection Method	91%
HMM	89%
Idle-time method with threshold	92%
Idle-time method (Number of SYNs)	65%
Idle-time method (Number of Bytes)	56%

and Table 5.21, respectively. The results shown in these tables are similar to the results shown in the real user evaluation when considering the genre-based labeling scheme. That is, the user profiles generated using the change point detection methods approximate the ground truth profiles at least as effectively as the idle time-based methods for all labeling schemes investigated and the ground truth segmentation performs the best. Though, for the synthetically generated data the video streaming-based and target device-based labeling schemes approximates profiles above a rate of 90%, while the navigation-based and genre-based labeling schemes do not perform better than 78% — that is, for the dataset and methods tested, the target device-based and video streaming-based labeling schemes outperform the others. This result is rather expected since we observe similar results previously in this chapter and in chapter 4.

5.5.4 Discussion: Importance and Implications of Accurate Web Page Segmentation

In this section, we evaluate whether web-related application domains are impacted by the web page segmentation method used. We find that higher performing web page segmentation approaches are beneficial for multiple web-related applications — particularly, web page classification and user behavior modeling.

This result is important because it is common for web measurement and traffic classification efforts to either adopt methods which have not been recently evaluated or do not consider the web page segmentation problem altogether. Thus, current and future efforts would likely benefit by utilizing web page segmentation methods which are more effective than prior idle time-based methods. For example, web measurement studies, which characterize user browsing behavior and/or web page traffic, would yield more accurate characterizations with more effective web page segmentation [Newton et al., 2013, Ihm and Pai, 2011]. Similarly, web page classification and web page identification will likely perform more effectively "in the wild" if the web page segmentation methods improve — in fact, if the performance improves enough, more advanced techniques for web traffic measurement and classification may not be needed [Maciá-Fernández et al., 2010, Yen et al., 2009, Dyer et al., 2012]. Nevertheless, we believe that the results in this section suggests that the web page segmentation problem warrants significant attention in order to successfully study the web in the future.

5.6 Contributions and Concluding Remarks

In this chapter, we conduct an empirical evaluation of approaches for web page segmentation using only anonymized TCP/IP headers. We make the following key contributions:

- We perform an extensive empirical evaluation of time series approaches for web page segmentation. Our analysis includes both synthetically generated and real user browsing data. This work is the first comprehensive empirical evaluation of time series-based web page segmentation approaches in over a decade.

- We find that web page segmentation is a challenging problem given the properties of modern web page traffic. The primary factors that influence this are the inter-arrival times of web pages, the number of simultaneous tabs open in a client browser, and the presence of automatically generated traffic (due to AJAX-based technology). The resulting increase in network activity is so high that idle time-based approaches are unable to reliably differentiate between new web page requests and this background noise traffic because the links are rarely idle.

- We also evaluate a new class of segmentation approaches that rely on *change point detection*, which instead identify if/when a time series exhibits a substantial increase in network activity. We find that change point detection methods can more robustly segment modern web traffic than existing idle time-based approaches.

- We extend our analysis by explicitly addressing the issue of whether web page segmentation approach choice has an impact on application domains that conduct web traffic analysis. We first consider the domain of user behavior modeling, and find that the performance of a web page segmentation approach has a significant impact on estimating several parameters, including (i) the number of clicks a user made and (ii) the average inter-arrival time of web page downloads.
- We then consider the domain of web page classification, in the context of estimating the content genre of web pages visited most often by a given user. We find that the classification performance is also impacted by the choice of the web page segmentation method. These are important insights because segmentation methods are widely used for web page fingerprinting and web traffic characterization. Thus, we conclude that web traffic segmentation is a problem that requires significant attention and will impact the performance of multiple web-related applications.

We also highlight a few limitations of our methodology and provide some possible areas for further study. First, our methodology assumes that a single user is represented by a single IP address. Based on our analysis, we believe that an environment where multiple users share a single IP address (e.g., NATs) will negatively impact the performance of the web page segmentation approaches evaluated in this chapter. This decrease in performance is due to the expected increase in background/noisy traffic. Also, time series approaches are likely to be sensitive to the performance of the network and the clients. Thus, web page segmentation methods should be tuned in an environment that has comparable performance characteristics to the environment in which they will be used.

CHAPTER 6: CONTRIBUTIONS AND CONCLUDING REMARKS

Traffic trace analysis is a network analysis methodology that has been widely adopted for studying the Web for several decades. However, recent privacy legislation and the increasing amount of encrypted traffic has made traffic trace analysis more challenging than it was in the past because the amount of content data that is available for analysis is decreasing [Law, Sicker et al., 2007, Register, Belshe et al., 2015, Jerome]. These challenges imply that, in some scenarios, only anonymized TCP/IP headers are available for traffic trace analysis. This dissertation explores web page classification when only anonymized TCP/IP headers (a type of non-content data) are available. Web page classification is important because it can be used as a building block for numerous web measurement-related applications including privacy analysis, user behavior modeling, traffic forecasting, and potentially behavioral ad-targeting.

In addition to studying web page classification when only anonymized TCP/IP headers are available, this dissertation also explicitly investigates two realistic issues which likely impact classification performance "in the wild" — these include classifying web page download traffic which (i) consists of web traffic that is generated using different client platforms and (ii) consists of web traffic that is not separated according to individual web page downloads (i.e., traffic traces which include a mixture of multiple web page downloads). These issues are not usually considered in the traffic classification literature [Lim et al., 2010, Kim et al., 2008, Erman et al., 2007b, Dyer et al., 2012, Herrmann et al., 2009, Miller et al., 2014, Schatzmann et al., 2010]. We conclude our discussion of this research by summarizing the contributions, limitations, open issues, and possible future directions of this work. This summary is provided below.

- *Summary of contributions:* This dissertation makes three primary contributions. These are summarized below:
  - This dissertation first comprehensive measurement study of modern web page traffic that investigates the impact that client platforms have on the measured traffic. We found that client platform choice (i.e., browsers, operating system, device, and vantage point) has an impact on the traffic that web pages generate. While some of these observations are anecdotally known, this research is the first that quantitively measures the prevalence of these influences. The results

of this study suggests that web page classification methods should be designed to be robust to client platform-specific differences.

  - This dissertation is the first to perform a comprehensive empirical evaluation of web page segmentation methods that uses anonymized TCP/IP headers. We evaluate both prior idle time-based methods and a new class of techniques known as change point detection methods. Our results explicitly show that the change point detection methods outperformed idle time-based methods.
  - This dissertation evaluates web page classification and the impact that client platform and web page segmentation method choice may have on its performance. We found that selecting features that were more consistent across client platforms is more effective for web page classification than selecting features that were not. We also found that web page segmentation method choice impacts the performance of web page classification — specifically, change point detection methods tend to outperform idle time-based methods when used for web page classification.

- *Summary of limitations:* A list of the primary limitations of this work is provided below:
  - This dissertation only considers web page traffic. Thus, the methods that we propose and the measurement results that we presented are not necessarily applicable to other types of HTTP applications including mobile applications and smart watches.
  - The results presented in this dissertation are biased for the top 250 web sites. Traffic characteristics that are exclusive to emerging or less popular web sites are not captured.
  - The web page segmentation approaches that we consider, which work for anonymized TCP/IP headers, are designed to determine when a web page download starts and do not determine when a web page download ends. This is different from web page segmentation approaches that inspect HTTP headers and packet payloads which determine the web objects that correspond to each individual web page download — this helps in determining when a web page downloads starts and ends.
  - Each web page segmentation method that we consider struggle with detecting web page downloads with small inter-arrival time ($<$ 10s) and may confuse automatically generated traffic as new web page downloads.

- *Open Issues not addressed in this dissertation:* There are a number of issues that are not considered in this dissertation. Some of these are described below:
  - We assume that all web traffic in the traffic traces that we consider is web page traffic. Real traces may include traffic from emerging technologies (such as speech recognition systems such as Google Home, and mobile applications) that also use HTTP. Traffic from such applications will likely exhibit characteristics that were not considered in this dissertation — thus, the performance of the web page classification and segmentation approaches will likely decrease when the traffic from such emerging technologies are included in the traffic mix.
  - We do not explicitly characterize the traffic generated by AJAX technology or personalized web pages — the methods proposed in this dissertation will likely be impacted by the traffic characteristics of these. We did not consider characterizing AJAX technology and personalized web pages because it requires the development of more advanced tools that would enable us to perform more complex web browsing behavior such as login, logout, scroll, and more.
  - Scenarios where individual users share the same IP address have not been considered. Such scenarios will increase the chances that the traffic from multiple web page downloads will overlap. This extra overlapping of traffic will make it more difficult for web page segmentation approaches to identify and separate each respective web page download because the perceived inter-arrival times between web pages will be reduced — please recall that a key limitation of the web page segmentation methods studied in this dissertation is that they do not perform well for browsing streams with small web page inter-arrival times (less than 10s).

- *Possible directions for further work:* Some possible directions for further work include:
  - Investigating the impact that traffic generated by AJAX technology, personalized web pages, and/or mobile applications have on web page classification and web page segmentation performance would help further determine the applicability of these approaches.
  - In this dissertation, we explore 4 orthogonal labeling schemes for web page classification. However, there are other labeling schemes that exist that may also be useful. Additional studies that investigate whether additional labeling schemes (such as a labeling scheme which identifies a web page as malicious or not) are useful will help further demonstrate the overall applicability

of web page classification.

- Investigating whether other types of web page segmentation methods, besides change point detection and idle time-based methods, can more accurately segment modern web page traffic would likely be useful for further improving web page classification “in the wild.”

APPENDIX 1: LIST OF WEB SITES STUDIED

This appendix includes a list of the web sites studies in this dissertation.

1 -http://www.godaddy.com, 2 -http://www.bankofamerica.com, 3 -http://www.wikipedia.org, 4 -http://www.java.com, 5 -http://www.twitch.tv, 6 -http://www.zedo.com, 7 -http://www.wunderground.com, 8 -http://www.okcupid.com, 9 -http://www.craigslist.org, 10 -http://www.pandora.com, 11 -http://www.sogou.com, 12 -http://www.jcpenney.com, 13 -http://www.linkedin.com, 14 -http://www.bestbuy.com, 15 -http://www.swagbucks.com, 16 -http://www.stumbleupon.com, 17 -http://www.fidelity.com, 18 -http://www.amazon.de, 19 -http://www.reddit.com, 20 -http://www.jackhenrybanking.com, 21 -http://www.dailymotion.com, 22 -http://www.ettoday.net, 23 -http://www.oracle.com, 24 -http://www.yelp.com, 25 -http://www.ebay.com, 26 -http://www.kohls.com, 27 -http://www.homedepot.com, 28 -http://www.drudgereport.com, 29 -http://www.microsoftstore.com, 30 -http://www.youtube.com, 31 -http://www.tmall.com, 32 -http://www.reference.com, 33 -http://www.163.com, 34 -http://www.ndtv.com, 35 -http://www.conduit.com, 36 -http://www.forbes.com, 37 -http://www.qq.com, 38 -http://www.priceline.com, 39 -http://www.life.com.tw, 40 -http://www.pornhub.com, 41 -http://www.amazon.in, 42 -http://www.blogger.com, 43 -http://www.douban.com, 44 -http://www.accuweather.com, 45 -http://www.southwest.com, 46 -http://www.zappos.com, 47 -http://www.leboncoin.fr, 48 -http://www.wikimedia.org, 49 -http://www.mapquest.com, 50 -http://www.buzzfeed.com, 51 -http://www.dailyfinance.com, 52 -http://www.amazon.cn, 53 -http://www.dailymail.co.uk, 54 -http://www.uol.com.br, 55 -http://www.constantcontact.com, 56 -http://www.yellowpages.com, 57 -http://www.deviantart.com, 58 -http://www.wikia.com, 59 -http://www.bycontext.com, 60 -http://www.rakuten.co.jp, 61 -http://www.bbc.co.uk, 62 -http://www.nba.com, 63 -http://www.groupon.com, 64 -http://www.adobe.com, 65 -http://www.idrudgereport.com, 66 -http://www.livejournal.com, 67 -http://www.aweber.com, 68 -http://www.pogo.com, 69 -http://www.netflix.com, 70 -http://www.amazon.es, 71 -http://www.staples.com, 72 -http://www.woot.com, 73 -http://www.usatoday.com, 74 -http://www.nih.gov, 75 -http://www.costco.com, 76 -http://www.overstock.com, 77 -http://www.jd.com, 78 -http://www.surveymonkey.com, 79 -http://www.webmd.com, 80 -http://www.babylon.com, 81 -http://www.online.wsj.com, 82 -http://www.people.com, 83 -http://www.taobao.com, 84 -http://www.soso.com, 85 -http://www.answers.com, 86 -http://www.pornhublive.com, 87 -http://www.ask.com, 88 -http://www.flipkart.com, 89 -http://www.cbslocal.com, 90 -http://www.stylelist.com, 91 -http://www.zillow.com, 92 -http://www.nytimes.com, 93 -http://www.cbs.com, 94 -http://www.aili.com, 95 -http://www.reimageplus.com, 96 -http://www.coupons.com, 97 -http://www.youporn.com, 98 -http://www.hulu.com, 99 -http://www.csnphilly.com, 100 -http://www.gmx.net, 101 -http://www.indeed.com, 102 -http://www.chinadaily.com.cn, 103 -http://www.blogspot.com, 104 -http://www.gap.com, 105 -http://www.bloomberg.com, 106 -http://www.dmm.co.jp, 107 -http://www.time.com, 108 -http://www.mama.cn, 109 -http://www.npr.org, 110 -http://www.tumblr.com,

111 -http://www.naver.com, 112 -http://www.bankrate.com, 113 -http://www.shopathome.com, 114 -http://www.xcar.com.cn, 115 -http://www.addthis.com, 116 -http://www.empowernetwork.com, 117 -http://www.search-results.com, 118 -http://www.nfl.com, 119 -http://www.ancestry.com, 120 -http://www.daum.net, 121 -http://www.facebook.com, 122 -http://www.china.com.cn, 123 -http://www.earthlink.net, 124 -http://www.tubecup.com, 125 -http://www.foodnetwork.com, 126 -http://www.redtube.com, 127 -http://www.adnxs.com, 128 -http://www.tripadvisor.com, 129 -http://www.espn.go.com, 130 -http://www.americanexpress.com, 131 -http://www.avg.com, 132 -http://www.google.com, 133 -http://www.chase.com, 134 -http://www.photobucket.com, 135 -http://www.yahoo.com, 136 -http://www.about.com, 137 -http://www.goodreads.com, 138 -http://www.aliexpress.com, 139 -http://www.outbrain.com, 140 -http://www.boston.com, 141 -http://www.amazon.co.jp, 142 -http://www.etsy.com, 143 -http://www.siteadvisor.com, 144 -http://www.slideshare.net, 145 -http://www.bing.com, 146 -http://www.pch.com, 147 -http://www.target.com, 148 -http://www.lowes.com, 149 -http://www.kayak.com, 150 -http://www.weather.gov, 151 -http://www.nbc.com, 152 -http://www.irs.gov, 153 -http://www.huffingtonpost.com, 154 -http://www.yaolan.com, 155 -http://www.jabong.com, 156 -www.abcnews.go.com, 157 -http://www.orange.fr, 158 -http://www.pof.com, 159 -http://www.xvideos.com, 160 -http://www.ca.gov, 161 -http://www.usmagazine.com, 162 -http://www.today.com, 163 -http://www.att.com, 164 -http://www.hp.com, 165 -http://www.youku.com, 166 -http://www.verizonwireless.com, 167 -http://www.newegg.com, 168 -http://www.barnesandnoble.com, 169 -http://www.intuit.com, 170 -http://www.concast.com, 171 -http://www.zol.com.cn, 172 -http://www.wikihow.com, 173 -http://www.flickr.com, 174 -http://www.examiner.com, 175 -http://www.alibaba.com, 176 -http://www.skype.com, 177 -http://www.patch.com, 179 -http://www.theblaze.com, 180 -http://www.imdb.com, 181 -http://www.legacy.com, 182 -http://www.realtor.com, 183 -http://www.ups.com, 184 -http://www.ticketmaster.com, 185 -http://www.amazon.co.uk, 186 -http://www.t-mobile.com, 187 -http://www.snapdeal.com, 188 -http://www.baidu.com, 189 -http://www.weather.com, 190 -http://www.drudgereportArchives.com, 191 -http://www.washingtonpost.com, 192 -http://www.tube8.com, 193 -http://www.capitalone.com, 194 -http://www.pconline.com.cn, 195 -http://www.retailmenot.com, 196 -http://www.hao123.com, 197 -http://www.msn.com, 198 -http://www.doublepimp.com, 199 -http://www.reuters.com, 200 -http://www.tmall.hk, 201 -http://www.mozilla.org, 202 -http://www.amazon.fr, 203 -http://www.trulia.com, 204 -http://www.worldstarhiphop.com, 205 -http://www.guardian.co.uk, 206 -http://www.nicovideo.jp, 207 -http://www.ehow.com, 208 -http://www.backpage.com, 209 -http://www.amazon.it, 210 -http://www.tudou.com, 211 -http://www.xhamstercams.com, 212 -http://www.warriorforum.com, 213 -http://www.ign.com, 214 -http://www.apple.com, 215 -http://www.tagged.com, 216 -http://www.rr.com, 217 -http://www.sina.com.cn, 218 -http://www.indiatimes.com, 219 -http://www.ebay.de, 220 -http://www.globo.com, 221 -http://www.macys.com, 222 -http://www.aol.com, 223 -http://www.whitepages.com, 224 -http://www.goo.ne.jp, 225 -http://www.latimes.com, 226 -

http://www.walmart.com, 227 -http://www.salesforce.com, 228 -http://www.fool.com, 229 -http://www.microsoft.com, 230 -http://www.match.com, 231 -http://www.xnxx.com, 232 -http://www.bbc.com, 233 -http://www.thefreedictionary.com, 234 -http://www.sohu.com, 235 -http://www.ebay.co.uk, 236 -http://www.incredibar.com, 237 -http://www.expedia.com, 238 -http://www.onet.pl, 239 -http://www.tmz.com, 240 -http://www.sears.com, 241 -http://www.yandex.ru, 242 -http://www.monster.com, 243 -http://www.amazon.com, 244 -http://www.youjizz.com, 245 -http://www.xhamster.com, 246 -http://www.roblox.com, 247 -http://www.fedex.com, 248 -http://www.directrev.com, 249 -http://www.meetup.com, 250 -http://www.usps.com,

APPENDIX 2: LIST OF WEB PAGES STUDIED

This appendix includes a list of the web pages studied in this dissertation.

1-http://www.google.com/#q=golden&biw=1280&bih=596&fp=f2197e5262a0e58e, 2-http://www.google.com/#q=ray\%20lewis&biw=1280&bih=596&fp=f2197e5262a0e58e, 3-http://www.google.com/#q=mcdonalds&biw=1280&bih=596&fp=f2197e5262a0e58e, 4-http://www.google.com/#q=stone\%20mountain&biw=1280&bih=596&fp=f2197e5262a0e58e, 5-http://www.google.com/search?q=golden&biw=1280&bih=596&um=1&ie=UTF-8&hl=en&tbm=isch&source=og&sa=N&tab=wi&ei=_sW8UYvUI4LQ9ASY_IH4Aw, 6-http://www.google.com/search?q=ray+lewis&biw=1280&bih=596&um=1&ie=UTF-8&hl=en&tbm=isch&source=og&sa=N&tab=wi&ei=_sW8UYvUI4LQ9ASY_IH4Aw, 7-http://www.google.com/search?q=mcdonalds&biw=1280&bih=596&um=1&ie=UTF-8&hl=en&tbm=isch&source=og&sa=N&tab=wi&ei=_sW8UYvUI4LQ9ASY_IH4Aw, 8-http://www.google.com/search?q=stone+mountain&biw=1280&bih=596&um=1&ie=UTF-8&hl=en&tbm=isch&source=og&sa=N&tab=wi&ei=_sW8UYvUI4LQ9ASY_IH4Aw, 9-https://www.facebook.com/, 10-http://m.facebook.com/, 11-http://www.facebook.com/search.php?q=golden, 12-http://www.facebook.com/search.php?q=ray+lewis, 13-http://www.facebook.com/search.php?q=stone+mountain, 14-http://www.facebook.com/search.php?q=mcdonalds, 15-https://www.facebook.com/officialraylewis, 16-https://m.facebook.com/officialraylewis, 17-https://www.facebook.com/McDonalds, 18-https://m.facebook.com/McDonalds, 19-https://www.facebook.com/stonemountainpark, 20-https://m.facebook.com/stonemountainpark, 21-http://www.youtube.com/, 22-http://m.youtube.com/, 23-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=golden, 24-http://www.youtube.com/results?search_query=golden, 25-http://www.youtube.com/results?search_query=ray+lewis, 26-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=ray+lewis, 27-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=mcdonalds, 28-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=stone+mountain, 29-http://www.youtube.com/results?search_query=mcdonalds, 30-http://www.youtube.com/results?search_query=stone+mountain, 31-http://www.yahoo.com/, 32-http://m.yahoo.com/, 33-http://search.yahoo.com/search;_ylt=Am.1j255Am1uvsUH9qmVz12bvZx4?p=golden, 34-http://search.yahoo.com/search;_ylt=Am.1j255Am1uvsUH9qmVz12bvZx4?p=ray+lewis, 35-http://search.yahoo.com/search;_ylt=Am.1j255Am1uvsUH9qmVz12bvZx4?p=stone+mountain, 36-http://search.yahoo.com/search;_ylt=Am.1j255Am1uvsUH9qmVz12bvZx4?p=mcdonalds, 37-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=golden&x=0&y=0, 38-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=mcdonalds&x=0&y=0, 39-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=ray+lewis&x=0&y=0, 40-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?

submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=stone+mountain&x=0&y=0, 41-http://images.search.yahoo.com/search/images;_ylt=Ak7Fnn5y_wsfbJPGilB4lI2bvZx4?p=golden&toggle=1&cop=mss&ei=UTF-8&fr=yfp-t-748, 42-http://images.search.yahoo.com/search/images;_ylt=Ak7Fnn5y_wsfbJPGilB4lI2bvZx4?p=ray+lewis&toggle=1&cop=mss&ei=UTF-8&fr=yfp-t-748, 43-http://images.search.yahoo.com/search/images;_ylt=Ak7Fnn5y_wsfbJPGilB4lI2bvZx4?p=stone+mountain&toggle=1&cop=mss&ei=UTF-8&fr=yfp-t-748, 44-http://images.search.yahoo.com/search/images;_ylt=Ak7Fnn5y_wsfbJPGilB4lI2bvZx4?p=mcdonalds&toggle=1&cop=mss&ei=UTF-8&fr=yfp-t-748, 45-http://omg.yahoo.com/blogs/celeb-news/kim-kardashian-gives-birth-baby-girl-195355964.html, 46-http://m.yahoo.com/w/ygo-frontpage/lp/story/us/3348478/coke.bp\%3B_ylt=A2KL8xzg1bxR2UwAAggp89w4\%3B_ylu=X3oDMTFybGdnODRoBGNwb3MDMQRjc2VjA21vYmlsZS10ZARpbnRsA3VzBHBrZ=wNpZC0zMzQ4NDc4BHBvcwMzBHNsawNtb3Jl?ref_w=frontdoors&view=today&.intl=US&.lang=en, 47-http://m.yahoo.com/w/ygo-frontpage/lp/story/us/3346216/coke.bp\%3B_ylt=A2KL8xv.1bxRYTMAMg4p89w4\%3B_ylu=X3oDMTFzc21sYmhiBGNwb3MDMwRjc2VjA21vYmlsZS10ZARpbnRsA3VzBHBrZwNpZ=C0zMzQ2MjE2BHBvcwMxBHNsawN0aHVtYg--?ref_w=frontdoors&view=today&.intl=US&.lang=en, 48-http://sports.yahoo.com/blogs/nascar-from-the-marbles/carl-edwards-locks-keys-car-still-wins-pole-222723347.html, 49-http://m.yahoo.com/w/ygo-frontpage/lp/story/us/3345618/coke.bp\%3B_ylt=A2KL8yBo17xRO3UA9QIp89w4\%3B_ylu=X3oDMTFzMjZpYWIwBGNwb3MDNARjc2VjA21vYmlsZS10ZARpbnRsA3Vz=BHBrZwNpZC0zMzQ1NjE4BHBvcwMxBHNsawN0aHVtYg--?ref_w=frontdoors&view=today&.intl=US&.lang=en, 50-http://movies.yahoo.com/blogs/movie-talk/lois-lane-man-steel-makeover-superman-squeeze-different-204334078.html, 51-http://www.amazon.com/, 52-http://www.amazon.com/gp/aw, 53-http://www.amazon.com/s/ref=nb_sb_noss?url=search-alias\%3Daps&field-keywords=golden, 54-http://www.amazon.com/gp/aw/s/ref=is_box_?k=golden, 55-http://www.amazon.com/Golden-Jessi-Kirby/dp/1442452161/ref=sr_1_1?ie=UTF8&qid=1371330689&sr=8-1&keywords=golden, 56-http://www.amazon.com/gp/aw/d/1442452161/ref=mp_s_a_1_1?qid=1371330635&sr=8-1, 57-http://www.amazon.com/gp/aw/s/ref=is_s_?ie=UTF8&k=ray+lewis&url=i\%3Daps, 58-http://www.amazon.com/s/ref=nb_sb_noss?url=search-alias\%3Dstripbooks&field-keywords=ray+lewis, 59-http://www.amazon.com/Rays-Ride-Amazing-Journey-ebook/dp/B00BEZQJ5A/ref=sr_1_1?s=books&ie=UTF8&qid=1371330767&sr=1-1&keywords=ray+lewis, 60-http://www.amazon.com/gp/aw/d/B00BEZQJ5A/ref=mp_s_a_1_10?qid=1371330819&sr=8-10, 61-http://www.amazon.com/gp/aw/s/ref=is_s_?ie=UTF8&k=stone+mountain&url=i\%3Daps, 62-http://www.amazon.com/s/ref=nb_sb_noss?url=search-alias\%3Dstripbooks&field-keywords=stone+mountain, 63-http://www.amazon.com/gp/aw/d/1596296828/ref=mp_s_a_1_1?qid=1371331323&sr=8-1, 64-http://www.amazon.com/Atlantas-

Stone-Mountain-Multicultural-History/dp/1596296828/ref=sr_1_1?s=books&ie=UTF8&qid=1371331418&sr=1-1&keywords=stone+mountain, 65-http://www.bing.com/, 66-http://m.bing.com/, 67-http://www.bing.com/search?q=golden, 68-http://m.bing.com/search?q=golden, 69-http://m.bing.com/search?q=ray+lewis, 70-http://www.bing.com/search?q=ray+lewis, 71-http://m.bing.com/search?q=stone+mountain, 72-http://m.bing.com/search?q=mcdonalds, 73-http://www.bing.com/search?q=stone+mountain, 74-http://www.bing.com/search?q=mcdonalds, 75-http://m.bing.com/images/search?q=golden, 76-http://m.bing.com/images/search?q=ray+lewis, 77-http://m.bing.com/images/search?q=stone+mountain, 78-http://m.bing.com/images/search?q=mcdonalds, 79-http://www.bing.com/images/search?q=golden, 80-http://www.bing.com/images/search?q=ray+lewis, 81-http://www.bing.com/images/search?q=stone+mountain, 82-http://www.bing.com/images/search?q=mcdonalds, 83-http://www.ebay.com/hp?_feedexp=0, 84-http://www.ebay.com/sch/i.html?_trksid=p5197.m570.l1313.TR10.TRC1.A0&_nkw=golden&_sacat=0&_from=R40, 85-http://www.ebay.com/itm/Lords-Prayer-Pendant-/111092951748?pt=Fashion_Jewelry&hash=item19dda7eac4, 86-http://www.ebay.com/sch/i.html?_trksid=p2047675.m570.l1313.TR0.TRC0&_nkw=ray+lewis&_sacat=0&_from=R40, 87-http://www.ebay.com/itm/Ray-Lewis1996-Fleer-Impact-Rookie-Card-Gem-Mt-10-/321143465615?pt=US_Football&hash=item4ac5a4668f, 88-http://www.ebay.com/sch/i.html?_trksid=p2047675.m570.l1313.TR0.TRC0&_nkw=mcdonalds&_sacat=0&_from=R40, 89-http://www.ebay.com/itm/McDonalds-Batman-Toys-/121125962607?pt=LH_DefaultDomain_0&hash=item1c33ab836f, 90-http://www.ebay.com/sch/i.html?_trksid=p2047675.m570.l1313.TR11.TRC1.A0&_nkw=stone+mountain&_sacat=0&_from=R40, 91-http://www.ebay.com/itm/STONE-MOUNTAIN-BLACK-LEATHER-FLAP-TOTE-HANDBAG-GORGEOUS-/400510271322?pt=US_CSA_WH_Handbags&hash=item5d4045bf5a, 92-http://www.wikipedia.org/, 93-http://en.m.wikipedia.org/wiki/Main_Page, 94-http://en.m.wikipedia.org/wiki/Golden, 95-http://en.m.wikipedia.org/wiki/Ray_Lewis, 96-http://en.m.wikipedia.org/wiki/Stone_Mountain, 97-http://en.m.wikipedia.org/wiki/Mcdonalds, 98-http://en.wikipedia.org/wiki/Golden, 99-http://en.wikipedia.org/wiki/Ray_Lewis, 100-http://en.wikipedia.org/wiki/Stone_Mountain, 101-http://en.wikipedia.org/wiki/Mcdonalds, 102-http://mobile.craigslist.org/, 103-http://mobile.craigslist.org/search/?areaID=200&subAreaID=&query=golden&catAbb=sss, 104-http://mobile.craigslist.org/search/sss?zoomToPosting=&query=ray+lewis&srchType=A&minAsk=&maxAsk=, 105-http://mobile.craigslist.org/search/sss?zoomToPosting=&query=stone+mountain&srchType=A&minAsk=&maxAsk=, 106-http://mobile.craigslist.org/search/sss?zoomToPosting=&query=mcdonalds&srchType=A&minAsk=&maxAsk=, 107-http://mobile.craigslist.org/ctd/3870400185.html, 108-http://mobile.craigslist.org/spo/3868723679.html, 109-http://pensacola.craigslist.org/cbd/3811138503.html, 110-http://pensacola.craigslist.org/clt/3854676346.html, 111-https://www.linkedin.com/nhome/, 112-https://touch.www.linkedin.com/login.html, 113-

http://www.linkedin.com/company/mcdonald’s-corporation, 114-http://www.linkedin.com/pub/raymond-lewis/8/552/944, 115-http://www.linkedin.com/company/northrop-grumman-corporation, 116-https://login.live.com/login.srf?wa=wsignin1.0&rpsnv=11&ct=1371400782&rver=6.1.6206.0&wp=MBI_SSL_SHARED&wreply=https:\%2F\%2Fmail.live.com\%2Fdefault.aspx\%3Frru\%3Dhome\%26livecom\%3D1&lc=1033&id=64855&mkt=en-us&cbcxt=mai, 117-https://login.live.com/?wa=wsignin1.0&rpsnv=11&ct=1368921021&rver=6.1.6206.0&wp=MBI_SSL_SHARED&wreply=https\%3a\%2f\%2fmail.live.com\%2fm\%2f\%3ffl\%3d635045178212091183&lc=1033&id=64855&mspco=1&pcexp=false, 118-https://twitter.com/, 119-https://mobile.twitter.com/signup, 120-https://twitter.com/search?q=mcdonalds&src=typd, 121-https://mobile.twitter.com/search?q=mcdonalds&src=typd, 122-https://mobile.twitter.com/search?q=ray+lewis&src=typd, 123-https://mobile.twitter.com/search?q=golden&src=typd, 124-https://mobile.twitter.com/search?q=stone+mountain&src=typd, 125-https://twitter.com/search?q=ray+lewis&src=typd, 126-https://twitter.com/search?q=golden&src=typd, 127-https://twitter.com/search?q=stone+mountain&src=typd, 128-https://twitter.com/raylewis, 129-https://mobile.twitter.com/raylewis, 130-https://twitter.com/McDonalds, 131-https://mobile.twitter.com/McDonalds, 132-https://twitter.com/StoneMtnPark, 133-https://mobile.twitter.com/StoneMtnPark, 134-https://twitter.com/Ashton5SOS, 135-https://mobile.twitter.com/Ashton5SOS, 136-http://www.blogspot.com, 137-http://www.blogger.com/mobile-start.g, 138-http://www.aol.com/, 139-http://m.aol.com/portal/, 140-http://search.aol.com/aol/search?enabled_terms=&s_it=comsearch51&q=golden, 141-http://search.aol.com/aol/search?enabled_terms=&s_it=comsearch51&q=ray+lewis, 142-http://search.aol.com/aol/search?enabled_terms=&s_it=comsearch51&q=stone+mountain, 143-http://search.aol.com/aol/search?enabled_terms=&s_it=comsearch51&q=mcdonalds, 144-http://m.aol.com/search/aol/search?q=mcdonalds&s_it=srch_entr, 145-http://m.aol.com/search/aol/search?q=ray+lewis&s_it=srch_entr, 146-http://m.aol.com/search/aol/search?q=stone+mountain&s_it=srch_entr, 147-http://m.aol.com/search/aol/search?q=mcdonalds&s_it=srch_entr, 148-http://m.aol.com/search/aol/images?q=ray+lewis&v_t=srch_entr, 149-http://m.aol.com/search/aol/images?q=mcdonalds&v_t=srch_entr, 150-http://m.aol.com/search/aol/images?q=stone+mountain&v_t=srch_entr, 151-http://m.aol.com/search/aol/images?q=golden&v_t=srch_entr, 152-http://search.aol.com/aol/image?q=ray+lewis&v_t=comsearch51&s_it=searchtabs, 153-http://search.aol.com/aol/image?q=mcdonalds&v_t=comsearch51&s_it=searchtabs, 154-http://search.aol.com/aol/image?q=stone+mountain&v_t=comsearch51&s_it=searchtabs, 155-http://search.aol.com/aol/image?q=golden&v_t=comsearch51&s_it=searchtabs, 156-http://www.huffingtonpost.com/2013/06/16/pope-francis-blessing, 157-http://www.huffingtonpost.com/2013/06/16/pope-francis-blessing-of-the-bikes_n_3449601.html, 158-http://www.huffingtonpost.com/2013/06/14/the-best-summer-ice-cream_n_3417468.

html?icid=maing-grid7\%7Cmain5\%7Cdl2\%7Csec1_lnk2\%26pLid\%3D329265, 159-http://www.huffingtonpost.com/2013/06/14/the-best-summer-ice-cream_n_3417468.html?icid=maing-grid7\%7Cmain5\%7Cdl2\%7Csec1_lnk2\%26pLid\%3D329265, 160-http://www.sportingnews.com/mlb/story/2013-06-15/alex-cobb-tampa-bay-rays-hit-by-line-drive-eric-hosmer-royal-skull-stretcher?icid=maing-grid7\%7Cmain5\%7Cdl3\%7Csec1_lnk2\%26pLid\%3D330099, 161-http://www.sportingnews.com/mlb/story/2013-06-15/alex-cobb-tampa-bay-rays-hit-by-line-drive-eric-hosmer-royal-skull-stretcher?icid=maing-grid7\%7Cmain5\%7Cdl3\%7Csec1_lnk2\%26pLid\%3D330099, 162-http://go.com/, 163-http://pinterest.com/, 164-http://m.pinterest.com/, 165-http://pinterest.com/search/pins/?q=golden, 166-http://pinterest.com/search/pins/?q=ray+lewis, 167-http://pinterest.com/search/pins/?q=stone+mountain, 168-http://pinterest.com/search/pins/?q=mcdonalds, 169-http://m.pinterest.com/search/pins/?q=ray+lewis, 170-http://m.pinterest.com/search/pins/?q=mcdonalds, 171-http://m.pinterest.com/search/pins/?q=stone+mountain, 172-http://m.pinterest.com/search/pins/?q=golden, 173-http://pinterest.com/pin/9922061651802862/, 174-http://m.pinterest.com/pin/9922061651802862/, 175-http://pinterest.com/pin/510103095264410678/, 176-http://m.pinterest.com/pin/510103095264410678/, 177-http://pinterest.com/pin/34199278393284055/, 178-http://m.pinterest.com/pin/34199278393284055/, 179-http://pinterest.com/pin/29766047510066706/, 180-http://m.pinterest.com/pin/29766047510066706/, 181-http://www.msn.com/, 182-http://m.now.msn.com/, 183-http://now.msn.com/SiteSearch?q=golden&x=0&y=0&form=MSNTRE, 184-http://m.now.msn.com/SiteSearch?q=ray+lewis&x=0&y=0&form=MSNTRE, 185-http://m.now.msn.com/SiteSearch?q=stone+mountain&x=0&y=0&form=MSNTRE, 186-http://m.now.msn.com/SiteSearch?q=mcdonalds&x=0&y=0&form=MSNTRE, 187-http://firstread.nbcnews.com/_news/2013/06/16/18987472-cheney-says-nsa-monitoring-could-have-prevented-911?lite, 188-http://worldnews.nbcnews.com/_news/2013/06/16/18987285-holy-rollers-pope-blesses-hundreds-of-harley-davidsons, 189-http://autos.msn.com/research/compare/default.aspx?c=0&i=0&tb=0&ph1=t0&ph2=t0&dt=0&v=t117617&v=t116741&v=t117318&icid=autos_4411, 190-http://www.cnn.com/, 191-http://www.cnn.com/search/?query=golden&x=0&y=0&primaryType=mixed&sortBy=relevance&intl=false, 192-http://www.cnn.com/search/?query=ray+lewis&x=0&y=0&primaryType=mixed&sortBy=relevance&intl=false, 193-http://www.cnn.com/search/?query=stone+mountain&x=0&y=0&primaryType=mixed&sortBy=relevance&intl=false, 194-http://www.cnn.com/search/?query=mcdonalds&x=0&y=0&primaryType=mixed&sortBy=relevance&intl=false, 195-http://www.cnn.com/2013/06/16/us/gay-rights-immigration/index.html?hpt=hp_t1, 196-http://us.cnn.com/2013/06/12/us/california-naked-bart-man/?iref=obinsite, 197-http://www.cnn.com/2013/06/12/world/meast/dubai-twisted-tower/?iref=obnetwork, 198-http://www.huffingtonpost.com/, 199-http://search.huffingtonpost.com/search?q=golden&s_it=header_form_v1, 200-http://search.huffingtonpost.com/search?q=ray+lewis&s_

it=header_form_v1, 201-http://search.huffingtonpost.com/search?q=stone+mountain&s_it=header_form_v1, 202-http://search.huffingtonpost.com/search?q=mcdonalds&s_it=header_form_v1, 203-http://www.huffingtonpost.com/2013/06/16/obama-nsa_n_3450211.html, 204-http://www.huffingtonpost.com/2013/06/15/kim-kardashians-baby-girl-kanye-west_n_3416230.html?utm_hp_ref=mostpopular, 205-http://www.huffingtonpost.com/2013/06/15/steve-irwin-daughter-death-hoax_n_3446760.html?utm_hp_ref=mostpopular, 206-http://www.huffingtonpost.com/jim-wallis/losing-control-and-learni_b_3436506.html, 207-http://www.ask.com/?o=0&l=dir, 208-http://www.ask.com/answers/362997801/how-do-i-make-my-hair-look-wavy?qsrc=4034, 209-http://www.ask.com/answers/362919141/what-is-the-best-and-cheapest-laptop-i-can-buy?qsrc=4034, 210-http://www.ask.com/answers/362918661/what-is-a-good-photo-editing-app?qsrc=4034, 211-https://signup.netflix.com/, 212-https://www.paypal.com/, 213-https://mobile.paypal.com/us/cgi-bin/wapapp?cmd=_wapapp-homepage, 214-http://espn.go.com/, 215-http://m.espn.go.com/wireless/, 216-http://search.espn.go.com/golden/, 217-http://search.espn.go.com/ray_lewis/, 218-http://search.espn.go.com/stone_mountain/, 219-http://search.espn.go.com/mcdonalds/, 220-http://m.espn.go.com/wireless/search/results?q=mcdonalds&fromForm=true, 221-http://m.espn.go.com/wireless/search/results?q=ray+lewis&fromForm=true, 222-http://m.espn.go.com/wireless/search/results?q=stone+mountain&fromForm=true, 223-http://m.espn.go.com/wireless/search/results?q=golden&fromForm=true, 224-http://espn.go.com/mlb/story/_/id/9392118/discharge-expected-sunday-alex-cobb-tampa-bay-rays, 225-http://espn.go.com/boston/nfl/story/_/id/9392090/vladimir-putin-denies-stealing-new-england-patriots-owner-robert-kraft-super-bowl-ring, 226-http://espn.go.com/tennis/story/_/id/9392035/roger-federer-takes-wimbledon-tuneup-end-title-drought, 227-http://espn.go.com/racing/nascar/cup/story/_/id/9392286/roger-penske-says-fully-supports-brad-keselowski, 228-http://wordpress.com/, 229-http://www.weather.com/, 230-http://www.weather.com/search/enhancedlocalsearch?where=golden&loctypes=1003\%2C1001\%2C1000\%2C1\%2C9\%2C5\%2C11\%2C13\%2C19\%2C20&from=hdr_localsearch, 231-http://www.weather.com/search/enhancedlocalsearch?where=ray+lewis&loctypes=1003\%2C1001\%2C1000\%2C1\%2C9\%2C5\%2C11\%2C13\%2C19\%2C20&from=hdr_localsearch, 232-http://www.weather.com/weather/today/Stone+Mountain+GA+USGA0540:1:US, 233-http://www.weather.com/weather/today/Chapel+Hill+TN+USTN0080:1:US, 234-http://www.weather.com/weather/today/Washington+DC+USDC0001:1:US, 235-http://conduit.com/, 236-http://www.conduit.com/search?q=golden&cx=010301873083402539744\%3Anxaq5wgrtuo&cof=forid\%3A11&ie=utf-8&sa=search, 237-http://www.conduit.com/search?q=ray+lewis&cx=010301873083402539744\%3Anxaq5wgrtuo&cof=forid\%3A11&ie=utf-8&sa=search, 238-http://www.conduit.com/search?q=stone+mountain&cx=010301873083402539744\%3Anxaq5wgrtuo&cof=forid\%3A11&ie=utf-

8&sa=search, 239-http://www.conduit.com/search?q=mcdonalds&cx=010301873083402539744\%3Anxaq5wgrtuo&cof=forid\%3A11&ie=utf-8&sa=search, 240-http://www.conduit.com/products, 241-http://www.conduit.com/aboutus/events, 242-http://www.conduit.com/aboutus, 243-https://www.bankofamerica.com/, 244-https://www.bankofamerica.com/mobile/banking.go, 245-https://www4.bankofamerica.com/search/Search.do?questionbox=golden&searchSourceSite=dotcom&searchSourceDir=\%2Fpbi-homepage\%2Foverview&locale=en_US, 246-https://www4.bankofamerica.com/search/Search.do?questionbox=ray+lewis&searchSourceSite=dotcom&searchSourceDir=\%2Fpbi-homepage\%2Foverview&locale=en_US, 247-https://www4.bankofamerica.com/search/Search.do?questionbox=stone+mountain&searchSourceSite=dotcom&searchSourceDir=\%2Fpbi-homepage\%2Foverview&locale=en_US, 248-http://instagram.com/, 249-http://www.microsoft.com/en-us/default.aspx, 250-http://m.microsoft.com/en-us/default.mspx, 251-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=golden, 252-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=ray+lewis, 253-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=stone+mountain, 254-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=mcdonalds, 255-http://www.microsoft.com/surface/en-us, 256-http://www.skype.com/en/, 257-http://m.microsoft.com/en-us/Products/default.mspx?prodtype=office, 258-http://office.microsoft.com/en-us/buy/?WT.mc_id=mscom_en-us_bnr_0365CArefresh_HP-FEATURE-, 259-http://www.microsoftstore.com/store?Action=DisplayPage&Env=BASE&Locale=en_US&SiteID=msusa&id=ThreePgCheckoutShoppingCartPage, 260-http://www.chase.com/, 261-https://mobilebanking.chase.com/, 262-https://www.chase.com/, 263-https://www.chase.com/ccp/index.jsp?pg_name=ccpmapp/generic/shared/page/chase_search&q=golden&emptyQueryText=false, 264-https://www.chase.com/ccp/index.jsp?pg_name=ccpmapp/generic/shared/page/chase_search&q=ray+lewis&emptyQueryText=false, 265-https://www.chase.com/ccp/index.jsp?pg_name=ccpmapp/generic/shared/page/chase_search&q=stone+mountain&emptyQueryText=false, 266-https://www.chase.com/ccp/index.jsp?pg_name=ccpmapp/generic/shared/page/chase_search&q=mcdonalds&emptyQueryText=false, 267-http://www.foxnews.com/, 268-http://www.foxnews.mobi/, 269-http://www.foxnews.com/politics/2013/06/16/lawmakers-press-obama-to-implement-no-fly-zone-over-syria/, 270-http://www.foxnews.com/politics/2013/06/16/officials-nsa-programs-broke-plots-in-20-nations-1706735692/, 271-http://www.foxnews.mobi/quickPage.html?page=38321&content=94097423, 272-http://www.foxnews.mobi/quickPage.html?page=38321&content=94098112, 273-http://www.foxnews.com/us/2013/06/16/north-carolina-medical-examiner-resigns-after-3-hotel-guests-die-months-apart/, 274-http://politics.foxnews.mobi/quickPage.html?page=23888&external=2194212.proteus.fma, 275-http://www.foxnews.com/politics/2013/06/16/plane-carrying-george-w-bush-makes-emergency-landing/?intcmp=HPBucket, 276-http://www.foxnews.com/search-results/search?q=golden&submit=Search, 277-http://www.foxnews.com/search-results/search?q=ray+lewis&submit=Search, 278-http://www.foxnews.com/search-results/search?q=stone+mountain&submit=Search, 279-http://www.foxnews.com/search-results/search?q=mcdonalds&submit=Search, 280-http://www.imdb.com/, 281-http://m.imdb.com/, 282-http://m.imdb.com/news/ni55804425, 283-http://www.imdb.com/news/ni55804425/?ref_=hm_nw_tp_t1, 284-http://m.imdb.com/title/tt0944947/, 285-http://www.imdb.com/title/tt0944947/, 286-http://www.imdb.com/name/nm0227759/?ref_=tt_cl_t1, 287-http://m.imdb.com/name/nm0227759/?ref_=tt_cl_t1, 288-http://m.imdb.com/title/tt0340377/, 289-http://www.imdb.com/title/tt0340377/, 290-http://m.imdb.com/find?q=golden&button.x=0&button.y=0&button=Search, 291-http://www.imdb.com/find?q=ray+lewis&button.x=0&button.y=0&button=Search, 292-http://m.imdb.com/find?q=ray+lewis&button.x=0&button.y=0&button=Search, 293-http://www.imdb.com/find?q=golden&button.x=0&button.y=0&button=Search, 294-http://m.imdb.com/find?q=stone+mountain&button.x=0&button.y=0&button=Search, 295-http://m.imdb.com/find?q=mcdonalds&button.x=0&button.y=0&button=Search, 296-http://www.imdb.com/find?q=stone+mountain&button.x=0&button.y=0&button=Search, 297-http://www.imdb.com/find?q=mcdonalds&button.x=0&button.y=0&button=Search, 298-http://www.about.com/, 299-http://search.about.com/?q=golden, 300-http://search.about.com/?q=ray+lewis, 301-http://search.about.com/?q=stone+mountain, 302-http://search.about.com/?q=mcdonalds, 303-http://baltimore.about.com/od/prorecsports/ss/Ravensphotos_8.htm, 304-http://thaifood.about.com/od/thaicookingessentials/a/goldenmountainsauce.htm, 305-http://japanese.about.com/od/namikosbloglessons/a/lesson60.htm, 306-http://americanfood.about.com/, 307-http://www.apple.com/, 308-http://www.apple.com/search/?q=golden&section=global&geo=us, 309-http://www.apple.com/search/?q=ray\%20lewis&section=global&geo=us, 310-http://www.apple.com/search/?q=stone\%20mountain&section=global&geo=us, 311-http://www.apple.com/search/?q=mcdonalds&section=global&geo=us, 312-http://store.apple.com/us/browse/home/shop_mac/family/macbook_pro, 313-http://store.apple.com/us/browse/home/shop_mac/family/mac_mini, 314-http://store.apple.com/us/browse/home/shop_mac/mac_accessories, 315-http://store.apple.com/us/questions/mac, 316-http://www.pornhub.com/, 317-http://m.pornhub.com/, 318-http://www.pornhub.com/video/search?search=golden, 319-http://www.pornhub.com/video/search?search=ray+lewis, 320-http://www.pornhub.com/video/search?search=stone+mountain, 321-http://www.pornhub.com/video/search?search=mcdonalds, 322-http://m.pornhub.com/video/search?query=mcdonalds, 323-http://m.pornhub.com/video/search?query=stone+mountain, 324-http://m.pornhub.com/video/search?query=ray+lewis, 325-http://m.pornhub.com/video/search?query=golden, 326-http://xfinity.comcast.net/, 327-http://m.comcast.net/m/, 328-http://search.comcast.net/?cat=web&con=betac&q=golden, 329-http://search.comcast.net/

?cat=web&con=betac&q=ray+lewis, 330-http://search.comcast.net/?cat=web&con=betac&q=stone+mountain, 331-http://search.comcast.net/?cat=web&con=betac&q=mcdonalds, 332-http://m.comcast.net/m/articles/news-general/20130616/US-The-Secret-Government/, 333-http://xfinity.comcast.net/articles/news-general/20130616/US-The-Secret-Government/, 334-http://www.csnphilly.com/article/stefani-records-first-us-open-ace-merion, 335-http://m.comcast.net/m/articles/news-general/20130616/GLF--US.Open-Stefani.Ace/, 336-http://xfinity.comcast.net/articles/news-politics/20130616/US-Cheney/, 337-http://m.comcast.net/m/articles/news-politics/20130616/US-Cheney/, 338-http://home.mywebsearch.com/, 339-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=golden&ptb=&n=&tpr=hpsb&ts=1371413446586&st=hp, 340-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=ray+lewis&ptb=&n=&tpr=hpsb&ts=1371413446586&st=hp, 341-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=mcdonalds&ptb=&n=&tpr=hpsb&ts=1371413446586&st=hp, 342-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=stone+mountain&ptb=&n=&tpr=hpsb&ts=1371413446586&st=hp, 343-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=ray+lewis&ts=1371413446586&n=&ss=sub&st=hp&ptb=&tpr=sbt, 344-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=stone+mountain&ts=1371413446586&n=&ss=sub&st=hp&ptb=&tpr=sbt, 345-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=mcdonalds&ts=1371413446586&n=&ss=sub&st=hp&ptb=&tpr=sbt, 346-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=golden&ts=1371413446586&n=&ss=sub&st=hp&ptb=&tpr=sbt, 347-http://www.nytimes.com/, 348-http://mobile.nytimes.com/, 349-http://www.nytimes.com/2013/06/17/world/europe/turkey.html?hp&_r=0, 350-http://mobile.nytimes.com/2013/06/17/world/europe/turkey.html?from=homepage, 351-http://www.nytimes.com/2013/06/17/business/economy/for-g-8-meeting-talk-of-economy-but-syria-looms-large.html?hp, 352-http://mobile.nytimes.com/2013/06/17/business/economy/for-g-8-meeting-talk-of-economy-but-syria-looms-large.html?from=homepage, 353-http://www.nytimes.com/2013/06/16/world/asia/chinas-great-uprooting-moving-250-million-into-cities.html?hp, 354-http://mobile.nytimes.com/2013/06/16/world/asia/chinas-great-uprooting-moving-250-million-into-cities.html?from=homepage, 355-http://mobile.nytimes.com/travel/2013/06/16/travel/looking-for-clementine-hunters-louisiana.html?from=homepage, 356-http://travel.nytimes.com/2013/06/16/travel/looking-for-clementine-hunters-louisiana.html?hp&_r=0, 357-http://mobile.nytimes.com/search?query=golden&sort=rel, 358-http://mobile.nytimes.com/search?query=ray+lewis&sort=rel, 359-http://mobile.nytimes.com/search?query=stone+mountain&sort=rel, 360-http://mobile.nytimes.com/search?query=mcdonalds&sort=rel, 361-http://query.nytimes.com/search/sitesearch/#/golden, 362-http://query.nytimes.com/search/sitesearch/#/ray+lewis/, 363-http://query.nytimes.com/search/sitesearch/#/stone+mountain/, 364-http://query.nytimes.com/search/sitesearch/#/mcdonalds/, 365-http://www.nbcnews.com/, 366-http://news.mobile.msn.com/en-us/default.aspx, 367-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=golden&go=Go, 368-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=ray+lewis&go=Go, 369-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=stone+mountain&go=Go, 370-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=mcdonalds&go=Go, 371-http://news.mobile.msn.com/en-us/articles.aspx?aid=18987063&afid=11, 372-http://firstread.nbcnews.com/_news/2013/06/16/18987063-gop-hawks-question-obamas-syria-strategy?lite, 373-http://news.mobile.msn.com/en-us/articles.aspx?aid=6C10313219&afid=11, 374-http://www.today.com/money/modern-dads-day-has-become-longer-more-hectic-6C10313219, 375-http://news.mobile.msn.com/en-us/articles.aspx?aid=18987359&afid=11, 376-http://usnews.nbcnews.com/_news/2013/06/16/18987359-colorado-wildfire-evacuees-return-to-charred-neighborhoods-devastation?lite, 377-http://firstread.nbcnews.com/_news/2013/06/16/18987395-rubio-95-percent-of-immigration-bill-in-perfect-shape-still-needs-border-fixes?lite, 378-http://news.mobile.msn.com/en-us/articles.aspx?aid=18987395&afid=11, 379-http://www.xhamster.com/, 380-http://m.xhamster.com/, 381-http://xhamster.com/search.php?q=golden&qcat=video, 382-http://m.xhamster.com/search.html?search=golden, 383-http://m.xhamster.com/search.html?search=ray+lewis, 384-http://m.xhamster.com/search.html?search=stone+mountain, 385-http://m.xhamster.com/search.html?search=mcdonalds, 386-http://xhamster.com/search.php?q=ray+lewis&qcat=video, 387-http://xhamster.com/search.php?q=stone+mountain&qcat=video, 388-http://xhamster.com/search.php?q=mcdonalds&qcat=video, 389-http://www.adobe.com/, 390-http://www.adobe.com/cfusion/search/index.cfm?term=golden&loc=en_us&siteSection=home, 391-http://www.adobe.com/cfusion/search/index.cfm?term=ray+lewis&loc=en_us&siteSection=home, 392-http://www.adobe.com/cfusion/search/index.cfm?term=stone+mountain&loc=en_us&siteSection=home, 393-http://www.adobe.com/cfusion/search/index.cfm?term=mcdonalds&loc=en_us&siteSection=home, 394-http://www.adobe.com/products/creativecloud.html?promoid=KCHGR, 395-http://www.adobe.com/products/acrobatpro.html?promoid=KCHGV, 396-http://www.adobe.com/products/photoshop.html?promoid=KCHGZ, 397-http://www.adobe.com/products/creativecloud/teams.html, 398-http://www.avg.com/us-en/homepage, 399-http://www.avg.com/us-en/avg-premium-security, 400-http://www.avg.com/us-en/buy-antivirus, 401-http://www.avg.com/us-en/avg-email-server-edition, 402-http://www.avg.com/us-en/antivirus-for-android, 403-http://www.walmart.com/, 404-http://mobile.walmart.com/, 405-http://www.walmart.com/search/search-ng.do?ic=16_0&Find=Find&search_query=golden&Find=Find&search_constraint=0, 406-http://mobile.walmart.com/m/searchr;jsessionid=75DCE9533EB1E2432ADBE096151DAA6B?search_query=golden, 407-

http://www.walmart.com/ip/23914822, 408-http://mobile.walmart.com/ip/23914822, 409-http://www.walmart.com/search/search-ng.do?search_query=ray+lewis&ic=16_0&Find=Find&search_constraint=4104, 410-http://mobile.walmart.com/m/searchr?search_query=ray+lewis, 411-http://www.walmart.com/ip/Inspector-Lewis-Series-4-Blu-ray-Widescreen/17126113, 412-http://mobile.walmart.com/ip/Inspector-Lewis-Series-4-Blu-ray-Widescreen/17126113, 413-http://www.walmart.com/search/search-ng.do?search_query=stone+mountain&ic=16_0&Find=Find&search_constraint=0, 414-http://mobile.walmart.com/m/searchr;jsessionid=0D4DCE2A266C7F64307D26911AE25830?search_query=stone+mountain, 415-http://www.walmart.com/ip/Mountain-Jewels-Gravel-Aquariums-25-lb/10449968, 416-http://mobile.walmart.com/ip/Mountain-Jewels-Gravel-Aquariums-25-lb/10449968, 417-http://mobile.walmart.com/m/searchr;jsessionid=C9B61EC47010A7B9F8988B0BA7F10003?search_query=mcdonalds, 418-http://www.walmart.com/search/search-ng.do?search_query=mcdonalds&ic=16_0&Find=Find&search_constraint=0, 419-http://mobile.walmart.com/ip/Mary-McDonald-Interiors-The-Allure-of-Style/13430747, 420-http://www.walmart.com/ip/Mary-McDonald-Interiors-The-Allure-of-Style/13430747, 421-http://imgur.com/, 422-http://imgur.com/?q=golden, 423-http://imgur.com/?q=ray+lewis, 424-http://imgur.com/?q=stone+mountain, 425-http://imgur.com/?q=mcdonalds, 426-http://imgur.com/gallery/0tf2xFt, 427-http://imgur.com/gallery/7nPiJ3t, 428-http://imgur.com/gallery/hkLTOr4, 429-http://imgur.com/gallery/cjkAx, 430-https://www.wellsfargo.com/, 431-https://m.wellsfargo.com/, 432-http://www.yelp.com/, 433-http://www.yelp.com/search?find_desc=golden&find_loc=Cary\%2C+NC&ns=1, 434-http://www.yelp.com/search?find_desc=stone+mountain&find_loc=Cary\%2C+NC&ns=1, 435-http://www.yelp.com/search?find_desc=ray+lewis&find_loc=Cary\%2C+NC&ns=1, 436-http://www.yelp.com/search?find_desc=mcdonalds&find_loc=Cary\%2C+NC&ns=1, 437-http://www.yelp.com/biz/mountain-stone-masonry-new-hill#query:stone\%20mountain, 438-http://www.yelp.com/biz/mountain-stone-masonry-new-hill#query:stone\%20mountain, 439-http://www.yelp.com/biz/lewis-chiropractic-creedmoor#query:ray\%20lewis, 440-http://www.yelp.com/biz/mcdonalds-cary-5#query:mcdonalds, 441-http://www.cnet.com/, 442-http://www.cnet.com/1770-5_1-0.html?query=golden&tag=srch, 443-http://www.cnet.com/1770-5_1-0.html?query=ray+lewis&tag=srch, 444-http://www.cnet.com/1770-5_1-0.html?query=stone_mountain&tag=srch, 445-http://www.cnet.com/1770-5_1-0.html?query=mcdonalds&tag=srch, 446-http://reviews.cnet.com/headphones/jvc-gumy-plus-black/4505-7877_7-35054542.html, 447-http://www.cnet.com/apple-macbook-air-13-inch/, 448-http://news.cnet.com/8301-17938_105-57588673-1/galaxy-s4-upsets-iphone-5-in-our-brutal-destruction-test-video/, 449-http://reviews.cnet.com/8301-33199_7-57589240-221/best-low-lag-hdtvs-for-serious-gamers/, 450-http://www.xvideos.com/, 451-http://www.xvideos.com/?k=golden, 452-http://www.xvideos.com/?k=stone+mountain, 453-http://www.xvideos.com/?k=ray+lewis, 454-http://www.xvideos.com/?k=mcdonalds, 455-http://www.xvideos.com/video962788/sharon_stone_hot_sexy_hollywood_celebrities_nude_porn_movie_clip, 456-http://www.xvideos.com/video943841/juliette_lewis_hot_sexy_hollywood_celebrity_nude_porn_movie_clip, 457-http://www.xvideos.com/video944072/kater_bosworth_hot_sexy_hollywood_celebrity_nude_porn_movie_clip, 458-http://www.xvideos.com/video961315/penelope_cruz_hot_sexy_hollywood_celeb_nude_porn_movie_clip, 459-http://www.ehow.com/, 460-http://www.ehow.com/search.html?s=golden&skin=corporate&t=all, 461-http://www.ehow.com/search.html?s=ray+lewis&skin=corporate&t=all, 462-http://www.ehow.com/search.html?s=stone+mountain&skin=corporate&t=all, 463-http://www.ehow.com/search.html?s=mcdonalds&skin=corporate&t=all, 464-http://www.ehow.com/info_8373449_skills-needed-inside-linebacker.html, 465-http://www.ehow.com/how_6631187_play-linebacker-football.html, 466-http://www.ehow.com/slideshow_12255839_home-makeover-meaningful-home-improvements-sellers.html, 467-http://www.ehow.com/list_7517458_landscaping-ideas-small-budgets.html#page=0, 468-http://www.reddit.com/, 469-http://m.reddit.com/, 470-http://www.reddit.com/search?q=golden, 471-http://www.reddit.com/search?q=ray+lewis, 472-http://www.reddit.com/search?q=stone+mountain, 473-http://www.reddit.com/search?q=mcdonalds, 474-http://i.imgur.com/7deC2lV.png, 475-http://i.imgur.com/7deC2lV.png, 476-http://www.reddit.com/r/HistoricalWhatIf/comments/1getdg/what_if_for_whatever_reason_britain_had_sided/cajnmax?context=3, 477-http://i.imgur.com/lOoepNr.jpg?1, 478-http://www.zedo.com/, 479-http://www.zedo.com/publisher-products/ad-server-technology/, 480-http://www.zedo.com/publisher-products/ad-network-revenue-maximization/, 481-http://www.zedo.com/publisher-products/outsourced-advertising-operations-ad-ops/, 482-http://www.zedo.com/publisher-products/rich-media-formats/, 483-http://www.etsy.com/, 484-http://www.etsy.com/search?q=golden&view_type=gallery&ship_to=US, 485-http://www.etsy.com/search?q=ray+lewis&view_type=gallery&ship_to=US, 486-http://www.etsy.com/search?q=stone+mountain&view_type=gallery&ship_to=US, 487-http://www.etsy.com/search?q=mcdonalds&view_type=gallery&ship_to=US, 488-http://www.etsy.com/listing/105195408/hand-cut-card-neon-geometry-custom-text?ref=fp_treasury_1, 489-http://www.etsy.com/listing/115547098/landscape-photograph-modern-neon-field?ref=fp_treasury_2, 490-http://www.etsy.com/listing/77611017/vintage-bowl-lime-green-milkglass-bright?ref=fp_treasury_3, 491-http://www.etsy.com/listing/113928673/geometric-chain-earrings-diamond-pattern?ref=fp_treasury_4, 492-http://m.flickr.com/#/home, 493-http://www.flickr.com/, 494-http://m.flickr.com/#/search/advanced/_QM_q_IS_golden, 495-http://www.flickr.com/search/?q=golden, 496-http://m.flickr.com/#/search/advanced/_QM_q_IS_ray+lewis, 497-http://www.flickr.com/search/?q=ray+lewis, 498-http://www.flickr.com/search/?q=stone+mountain, 499-http://www.flickr.com/search/?q=mcdonalds, 500-http://m.flickr.com/#/search/advanced/

_QM_q_IS_stone+mountain, 501-http://m.flickr.com/#/search/advanced/_QM_q_IS_mcdonalds, 502-http://m.flickr.com/#/photos/luis_villablanca/9059399766/in/explore-1371382723/, 503-http://www.flickr.com/photos/luis_villablanca/9059399766/in/explore-2013-06-16, 504-http://m.flickr.com/#/photos/snowyturner/9055322469/in/explore-1371382723/, 505-http://www.flickr.com/photos/snowyturner/9055322469/in/explore-2013-06-16, 506-http://m.flickr.com/#/photos/vineetradhakrishnan/9058440177/in/explore-1371382723/, 507-http://www.flickr.com/photos/vineetradhakrishnan/9058440177/in/explore-2013-06-16, 508-http://m.flickr.com/#/photos/petezelewski/9058318873/in/explore-1371382723/, 509-http://www.flickr.com/photos/petezelewski/9058318873/in/explore-2013-06-16, 510-http://www.hulu.com/, 511-http://www.hulu.com/search?q=golden, 512-http://www.hulu.com/search?q=ray+lewis, 513-http://www.hulu.com/search?q=stone+mountain, 514-http://www.hulu.com/search?q=mcdonalds, 515-http://www.hulu.com/watch/450058, 516-http://www.hulu.com/watch/492494#i1,p0,d1, 517-http://www.hulu.com/watch/486595#i2,p0,d2, 518-http://www.hulu.com/watch/306771#i4,p0,d2, 519-http://www.pch.com/unrecognized, 520-http://www.pandora.com/, 521-http://www.outbrain.com/, 522-http://optmd.com/, 523-http://www.indeed.com/, 524-http://www.indeed.com/jobs?q=golden&l=, 525-http://www.indeed.com/q-ray-lewis-jobs.html, 526-http://www.indeed.com/q-stone-mountain-jobs.html, 527-http://new.livejasmin.com/en/, 528-http://m.livejasmin.com/en/, 529-http://new.livejasmin.com/en/girls/golden, 530-http://new.livejasmin.com/en/girls/ray+lewis, 531-http://new.livejasmin.com/en/girls/stone+mountain, 532-http://new.livejasmin.com/en/girls/mcdonalds, 533-http://www.zillow.com/, 534-http://www.zillow.com/homes/golden_rb/, 535-http://www.zillow.com/homes/ray+lewis_rb/, 536-http://www.zillow.com/homes/stone+mountain_rb/, 537-http://www.zillow.com/homes/mcdonalds_rb/, 538-http://www.zillow.com/homedetails/101-Easy-St-Troy-AL-36081/2112340720_zpid/, 539-http://www.zillow.com/homedetails/102-N-Hillcrest-Blvd-Troy-AL-36081/104528401_zpid/, 540-http://www.zillow.com/homedetails/101-Mallard-Dr-Troy-AL-36081/104525504_zpid/, 541-http://www.zillow.com/b/7115-County-Road-7707/31.8737,-85.9537_ll/, 542-http://www.target.com/, 543-http://m.target.com/, 544-http://www.target.com/s?searchTerm=golden&category=0\%7CAll\%7Cmatchallpartial\%7Call+categories&lnk=snav_sbox_golden, 545-http://www.target.com/s?searchTerm=ray+lewis&category=0\%7CAll\%7Cmatchallpartial\%7Call+categories&lnk=snav_sbox_golden, 546-http://www.target.com/s?searchTerm=stone+mountain&category=0\%7CAll\%7Cmatchallpartial\%7Call+categories&lnk=snav_sbox_golden, 547-http://www.target.com/s?searchTerm=mcdonalds&category=0\%7CAll\%7Cmatchallpartial\%7Call+categories&lnk=snav_sbox_golden, 548-http://m.target.com/s/golden#keywords=golden, 549-http://m.target.com/s/ray+lewis, 550-http://m.target.com/s/stone+mountain, 551-http://m.target.com/s/mcdonalds, 552-http://www.target.com/p/nfl-player-throw-ray-lewis/-/A-

10148918#prodSlot=medium_1_1&term=ray+lewis, 553-http://m.target.com/p/nfl-player-throw-ray-lewis/-/A-10148918, 554-http://www.target.com/p/sun-maid-golden-raisins-bag-10oz/-/A-13207041#prodSlot=medium_1_1&term=golden, 555-http://m.target.com/p/sun-maid-golden-raisins-bag-10oz/-/A-13207041, 556-http://m.target.com/p/funky-monkey-bananamon-1oz/-/A-12935987, 557-http://www.target.com/p/funky-monkey-bananamon-1oz/-/A-12935987?reco=Rec\%7Cpdp\%7C12935987\%7CClickCP\%7Citem_page.new_vertical_1&lnk=Rec\%7Cpdp\%7CClickCP\%7Citem_page.new_vertical_1, 558-http://www.target.com/p/ortega-yellow-corn-taco-shells-12-ct/-/A-13388903#prodSlot=medium_1_1&term=yellow, 559-http://m.target.com/p/ortega-yellow-corn-taco-shells-12-ct/-/A-13388903, 560-http://www.shopathome.com/, 561-http://m.shopathome.com/, 562-http://m.shopathome.com/Search?sf=golden, 563-http://m.shopathome.com/Search?sf=ray+lewis, 564-http://search.shopathome.com/search?sf=golden&Submit=Search, 565-http://search.shopathome.com/search?sf=ray+lewis&Submit=Search, 566-http://search.shopathome.com/search?sf=stone+mountain&Submit=Search, 567-http://search.shopathome.com/search?sf=mcdonalds&Submit=Search, 568-http://m.shopathome.com/Search?sf=stone+mountain, 569-http://m.shopathome.com/Search?sf=mcdonalds, 570-http://www.answers.com/, 571-http://www.answers.com/topic/golden, 572-http://www.answers.com/topic/ray+lewis, 573-http://www.answers.com/topic/stone_mountain, 574-http://www.answers.com/topic/mcdonalds, 575-http://www.redtube.com/, 576-http://m.redtube.com.brazzersmobile.com/, 577-http://www.redtube.com/?search=golden, 578-http://www.redtube.com/?search=ray+lewis, 579-http://www.redtube.com/?search=stone+mountain, 580-http://www.redtube.com/?search=mcdonalds, 581-http://www.xnxx.com/, 582-http://mobile.xnxx.com/, 583-http://www.homedepot.com/, 584-http://www.homedepot.com/webapp/catalog/servlet/Search?storeId=10051&langId=-1&catalogId=10053&keyword=golden&Ns=None&Ntpr=1&Ntpc=1&selectedCatgry=Search+All, 585-http://www.homedepot.com/webapp/catalog/servlet/Search?storeId=10051&langId=-1&catalogId=10053&keyword=ray+lewis&Ns=None&Ntpr=1&Ntpc=1&selectedCatgry=Search+All, 586-http://www.homedepot.com/webapp/catalog/servlet/Search?storeId=10051&langId=-1&catalogId=10053&keyword=stone+mountain&Ns=None&Ntpr=1&Ntpc=1&selectedCatgry=Search+All, 587-http://www.homedepot.com/webapp/catalog/servlet/Search?storeId=10051&langId=-1&catalogId=10053&keyword=mcdonalds&Ns=None&Ntpr=1&Ntpc=1&selectedCatgry=Search+All, 588-http://www.homedepot.com/p/RoomMates-Modern-Baby-Peel-Stick-Wall-Decal-RMK1777SCS/203302003#.UdIjMxy4wcM, 589-http://m.homedepot.com/p/RoomMates-Modern-Baby-Peel-Stick-Wall-Decal-RMK1777SCS/203302003/, 590-http://www.homedepot.com/p/Fathead-50-in-x-78-in-Ray-Lewis-Baltimore-Ravens-Wall-Decal-FH12-20014/202076175#.UdIjXxy4wcM, 591-http://m.homedepot.com/p/Fathead-50-in-x-78-in-Ray-Lewis-Baltimore-Ravens-Wall-Decal-FH12-20014/202076175/, 592-http://m.homedepot.com/p/TrafficMaster-6-in-x-36-in-Golden-Maple-Resilient-Vinyl-Plank-Flooring-24-sq-ft-case-161215/100595231/, 593-http:

//www.homedepot.com/p/TrafficMaster-6-in-x-36-in-Golden-Maple-Resilient-Vinyl-Plank-Flooring-24-sq-ft-case-161215/100595231#.UdIjjhy4wcM, 594-http://www.homedepot.com/p/Delray-Plants-8-in-Golden-Pothos-HB-in-Plastic-Pot-8POTHB/202204580#.UdIjwRy4wcM, 595-http://m.homedepot.com/p/Delray-Plants-8-in-Golden-Pothos-HB-in-Plastic-Pot-8POTHB/202204580/, 596-http://www.att.com/global-search/search.jsp?App_ID=HOME&autoSuggest=FALSE&tabPressed=FALSE&q=golden#!/All/, 597-http://www.att.com/global-search/search.jsp?App_ID=HOME&autoSuggest=FALSE&tabPressed=FALSE&q=ray+lewis#!/All/, 598-http://www.att.com/global-search/search.jsp?App_ID=HOME&autoSuggest=FALSE&tabPressed=FALSE&q=stone+mountain#!/All/, 599-http://www.att.com/global-search/search.jsp?App_ID=HOME&autoSuggest=FALSE&tabPressed=FALSE&q=mcdonalds#!/All/, 600-http://m.att.com/, 601-http://www.att.com/#fbid=RpFOsFt1Wx0, 602-http://m.usps.com/, 603-https://www.usps.com/, 604-http://m.usps.com/MobileTrackPackage.aspx, 605-http://m.usps.com/MobilePOLocator.aspx, 606-http://m.usps.com/about.aspx, 607-https://tools.usps.com/go/TrackConfirmAction!input.action, 608-https://tools.usps.com/go/POLocatorAction!input.action, 609-http://about.usps.com/, 610-https://www.usps.com/search.htm?q=golden, 611-https://www.usps.com/search.htm?q=ray\%20lewis, 612-https://www.usps.com/search.htm?q=stone\%20mountain, 613-https://www.usps.com/search.htm?q=mcdonalds, 614-https://m.ups.com/mobile/home, 615-http://www.ups.com/, 616-https://m.ups.com/mobile/trackhome?loc=en_US, 617-http://www.ups.com/WebTracking/track?loc=en_US&WT.svl=PriNav, 618-https://m.ups.com/mobile/locator?loc=en_US, 619-https://www.ups.com/dropoff?loc=en_US&WT.svl=PriNav, 620-http://www.ups.com/content/us/en/shipping/index.html?WT.svl=PriNav, 621-https://m.ups.com/one-to-one/mdotlogin?returnto=https\%3a//m.ups.com/ums/m.ship\%3floc\%3den_US&reasonCode=-1&appid=UIS, 622-http://www.ups.com/search/quick?loc=en_US&results=25&view=both&query=golden&searchButton=, 623-http://www.ups.com/search/quick?loc=en_US&results=25&view=both&query=ray+lewis&searchButton=, 624-http://www.ups.com/search/quick?loc=en_US&results=25&view=both&query=stone+mountain&searchButton=, 625-http://www.ups.com/search/quick?loc=en_US&results=25&view=both&query=mcdonalds&searchButton=, 626-http://www.usatoday.com/, 627-http://m.usatoday.com/, 628-http://www.usatoday.com/search/golden/, 629-http://www.usatoday.com/search/ray\%20lewis/, 630-http://www.usatoday.com/search/stone\%20mountain/, 631-http://www.usatoday.com/search/mcdonalds/, 632-http://www.usatoday.com/story/money/personalfinance/2013/07/01/retirement-savings-shortfall-crisis-catch-u0p/2464119/, 633-http://m.usatoday.com/article/news/2464119, 634-http://www.usatoday.com/story/news/2013/07/01/arizona-firefighters-mourned/2481207/, 635-http://m.usatoday.com/article/news/2481207, 636-http://www.usatoday.com/story/news/nation/2013/07/01/trayvon-martin-zimmerman-trial-voice-analysis/2479185/, 637-http://m.usatoday.com/article/news/2479185, 638-http://www.reference.com/, 639-http://m.

dictionary.com/r/, 640-http://www.reference.com/browse/golden?s=t, 641-http://m.dictionary.com/r/?q=golden&submit-result-SEARCHR=Search, 642-http://m.dictionary.com/r/?q=ray+lewis&submit-result-SEARCHR=Search, 643-http://www.reference.com/browse/ray+lewis?s=t, 644-http://www.reference.com/browse/stone+mountain?s=t, 645-http://www.reference.com/browse/mcdonalds?s=t, 646-http://m.dictionary.com/r/?q=stone+mountain&submit-result-SEARCHR=Search, 647-http://m.dictionary.com/r/?q=mcdonalds&submit-result-SEARCHR=Search, 648-http://m.dictionary.com/r/?q=computerAD, 649-http://www.reference.com/browse/computer-aided+design?s=t, 650-http://m.dictionary.com/r/?q=Guise, 651-http://www.reference.com/browse/guise?s=t, 652-http://m.dictionary.com/r/?q=family&submit-result-SEARCHR=Search, 653-http://www.reference.com/browse/family?s=t, 654-http://www.dailymail.co.uk/ushome/index.html, 655-http://www.dailymail.co.uk/home/search.html?sel=site&searchPhrase=golden, 656-http://www.dailymail.co.uk/home/search.html?sel=site&searchPhrase=ray+lewis, 657-http://www.dailymail.co.uk/home/search.html?sel=site&searchPhrase=stone+mountain, 658-http://www.dailymail.co.uk/home/search.html?sel=site&searchPhrase=mcdonalds, 659-http://www.dailymail.co.uk/news/article-2353272/New-Mexico-abortion-doctor-caught-audio-telling-woman-just-sit-toilet-27-week-fetus-comes-out.html, 660-http://www.dailymail.co.uk/news/article-2353311/Wikileaks-releases-letter-blasting-Obama-claims-NSA-whistleblower-Edward-Snowden.html, 661-http://www.dailymail.co.uk/femail/article-2352727/What-makes-The-Perfect-Man-Hes-educated-successful-40-doesnt-drunk-probably-doctor.html, 662-http://m.godaddy.com/, 663-http://www.godaddy.com/, 664-http://support.godaddy.com/search/all/golden/, 665-http://support.godaddy.com/search/all/ray+lewis/, 666-http://support.godaddy.com/search/all/stone+mountain/, 667-http://support.godaddy.com/search/all/mcdonalds/, 668 -http://www.washingtonpost.com, 669-http://www.washingtonpost.com/newssearch/search.html?st=golden&submit=Submit, 670-http://www.washingtonpost.com/newssearch/search.html?st=ray+lewis&submit=Submit, 671-http://www.washingtonpost.com/newssearch/search.html?st=stone+mountain&submit=Submit, 672-http://www.washingtonpost.com/newssearch/search.html?st=mcdonalds&submit=Submit, 673-http://www.washingtonpost.com/world/europe/edward-snowden-applies-for-asylum-in-russia-news-reports-say/2013/07/01/cc3daf20-e26e-11e2-aef3-339619eab080_story.html?hpid=z1, 674-http://www.washingtonpost.com/politics/immigration-deal-would-boost-defense-manufacturers/2013/07/01/d1c115e4-df63-11e2-b2d4-ea6d8f477a01_story.html?hpid=z1, 675-http://www.washingtonpost.com/world/the_americas/with-mexican-auto-manufacturing-boom-new-worries/2013/07/01/10dd57e8-d7d9-11e2-b418-9dfa095e125d_story.html?hpid=z1, 676-http://www.washingtonpost.com/blogs/fact-checker/post/sarah-palins-misreading-of-polling-data/2013/06/29/4f164476-e0ed-11e2-8ae9-5db15d3c0fca_blog.html?tid=pm_pop, 677-http://www.youporn.com/, 678-http://mobile.youporn.com/, 679-http://www.youporn.

com/search/?query=golden, 680-http://www.youporn.com/search/?query=ray+lewis, 681-http://www.youporn.com/search/?query=stone+mountain, 682-http://www.youporn.com/search/?query=mcdonalds, 683-http://abcnews.go.com/, 684-http://abcnews.go.com/m/, 685-http://abcnews.go.com/2020/george-zimmerman-recalled-trayvon-martin-gosh/story?id=19543886#.UdI6gRy4wcM, 686-http://abcnews.go.com/m/story?id=19543886&sid=359&ts=true, 687-http://abcnews.go.com/m/blogEntry?id=19545572&sid=7623874&cid=7623874&ts=true, 688-http://abcnews.go.com/blogs/headlines/2013/07/putin-edward-snowden-can-stay-in-russia-on-one-strange-condition/, 689-http://abcnews.go.com/m/blogEntry?id=19549487&sid=7623874&cid=7623874&ts=true, 690-http://abcnews.go.com/blogs/headlines/2013/07/edward-snowden-blasts-obama-deception-wikileaks/, 691-http://abcnews.go.com/search?searchtext=golden, 692-http://abcnews.go.com/search?searchtext=ray+lewis, 693-http://abcnews.go.com/search?searchtext=stone+mountain, 694-http://abcnews.go.com/search?searchtext=mcdonalds, 695-http:www.babylon.com, 696-http://www.babylon.com/products, 697-http://store.babylon.com/category/38/8/2/0/1/learn, 698-http://store.babylon.com/product/dictionary/9501/7/1/4/1/2/Babylon+for+Mac, 699-https://store.babylon.com/?trid=HPBUY, 700-http://m.bestbuy.com/m/e/digital/, 701-http://www.bestbuy.com/, 702-http://www.bestbuy.com/site/Harry+Potter+and+the+Half-Blood+Prince+-+Widescreen+Special+-+Blu-ray+Disc/9615177.p?id=2056395&skuId=9615177&st=ray\%20lewis&lp=1&cp=1, 703-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=9615177&pid=&ev=prodView, 704-http://www.bestbuy.com/site/Help!+-+Blu-ray+Disc/9406171.p?id=22204&skuId=9406171&st=help&lp=1&cp=1, 705-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=9406171&pid=&ev=prodView, 706-http://www.bestbuy.com/site/Amazing+Spider-Man+(3+Disc)+(W/Dvd)+-+Widescreen+AC3+-+Blu-ray+Disc/6836663.p?id=2592986&skuId=6836663&st=spiderman\%20blu\%20ray&lp=2&cp=1, 707-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=6836663&pid=&ev=prodView, 708-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=9136379&pid=&ev=prodView, 709-http://www.bestbuy.com/site/Batman+Begins+-+Blu-ray+Disc/9136379.p?id=1484301&skuId=9136379&st=batman\%20begins&lp=3&cp=1, 710-http://m.groupon.com/raleigh-durham?z=skip, 711-http://www.groupon.com/browse/raleigh-durham?z=skip, 712-http://m.groupon.com/deals/shutterfly-334-raleigh-durham, 713-http://www.groupon.com/deals/cincinnati-lubes-22, 714-http://www.groupon.com/deals/shutterfly-334-raleigh-durham?c=dnb&p=2, 715-http://m.groupon.com/deals/the-golf-warriors, 716-http://www.groupon.com/deals/lonnie-poole-golf, 717-http://m.groupon.com/deals/cinellis-3, 718-http://www.groupon.com/deals/cinellis-3?c=dnb&p=3, 719-http://www.groupon.com/deals/amf-bowling-centers-nat-5-raleigh-durham?c=dnb&p=2, 720-http://m.groupon.com/deals/amf-bowling-centers-nat-5-raleigh-durham, 721-http://www.bbc.co.uk/, 722-http://www.bbc.co.uk/search/?q=golden, 723-http://www.bbc.co.uk/search/?q=ray+lewis, 724-http://www.bbc.co.uk/search/?q=stone+mountain, 725-http://www.bbc.co.uk/search/?q=mcdonalds, 726-http://www.bbc.co.uk/news/world-africa-23148749, 727-http://www.bbc.co.uk/sport/0/cricket/23154209, 728-http://www.bbc.co.uk/sport/0/tennis/23145785, 729-http://www.bbc.co.uk/news/business-23150656, 730-http://www.wikia.com/Wikia, 731-http://www.wikia.com/index.php?search=golden&fulltext=Search, 732-http://www.wikia.com/index.php?search=ray+lewis&fulltext=Search, 733-http://www.wikia.com/index.php?search=stone+mountain&fulltext=Search, 734-http://www.wikia.com/index.php?search=mcdonalds&fulltext=Search, 735-http://www.wikia.com/Video_Games, 736-http://www.wikia.com/Lifestyle, 737-http://pokemon.wikia.com/?redirect=no, 738-http://www.deviantart.com/, 739-http://browse.deviantart.com/?q=golden, 740-http://browse.deviantart.com/?q=ray+lewis, 741-http://browse.deviantart.com/?q=stone+mountain, 742-http://browse.deviantart.com/?q=mcdonalds, 743-http://browse.deviantart.com/art/Ray-Lewis-148098394, 744-http://browse.deviantart.com/art/Golden-372963051, 745-http://www.deviantart.com/art/Neverland-s-Grand-Finale-382307154, 746-http://www.deviantart.com/art/original-costumes-magical-girls-FM-comics-382238492, 747-http://www.buzzfeed.com/, 748-http://www.buzzfeed.com/search?q=golden, 749-http://www.buzzfeed.com/search?q=ray+lewis, 750-http://www.buzzfeed.com/search?q=stone+mountain, 751-http://www.buzzfeed.com/search?q=mcdonalds, 752-http://www.buzzfeed.com/stouffers/9-reasons-why-winter-is-way-better-than-summer, 753-http://www.buzzfeed.com/amyodell/20-sweatshirts-you-need-in-your-life-immediately, 754-http://www.buzzfeed.com/lyapalater/temp-title-1372802401082, 755-http://www.buzzfeed.com/gidthekid/crazy-footage-from-egyptian-protests-a06o, 756-https://www.capitalone.com/, 757-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www, 758-http://www.capitalone.com/search/?qt=golden&cg2=&search-btn=Search&cg2=\%5Bobject+HTMLInputElement\%5D&refer=https\%3A\%2F\%2Fwww.capitalone.com\%2F, 759-http://www.capitalone.com/search/?qt=ray+lewis&cg2=&search-btn=Search&cg2=\%5Bobject+HTMLInputElement\%5D&refer=https\%3A\%2F\%2Fwww.capitalone.com\%2F, 760-http://www.capitalone.com/search/?qt=stone+mountain&cg2=&search-btn=Search&cg2=\%5Bobject+HTMLInputElement\%5D&refer=https\%3A\%2F\%2Fwww.capitalone.com\%2F, 761-http://www.capitalone.com/search/?qt=mcdonalds&cg2=&search-btn=Search&cg2=\%5Bobject+HTMLInputElement\%5D&refer=https\%3A\%2F\%2Fwww.capitalone.com\%2F, 762-http://www.capitalone.com/credit-cards/?Log=1&EventType=Link&ComponentType=T&LOB=MTS::L0RT6ME8Z&SubLob=&PageName=Home\%20Page\%20C&PortletLocation=2&ComponentName=primary_nav&ComponentStrategy=&ContentElement=5\%3BCredit+Cards&TargetLob=MTS\%3A\%3ALCTMMQC4S&TargetPageName=Credit+Cards+Home&linkid=&email_delivery_id=&referer=http://www.capitalone.com/homepage&external_id=, 763-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/

EnterpriseMobileBanking.html#www/cards/login?redirect=www/cards/accounts, 764-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www/atm, 765-http://maps.capitalone.com/locator/, 766-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www/contact, 767-http://www.capitalone.com/contact/, 768-http://www.drudgereport.com/, 769-http://www.idrudgereport.com/, 770-http://mlb.mlb.com/home, 771-http://m.mlb.com/, 772-http://mlb.mlb.com/search/?query=golden&c_id=mlb, 773-http://mlb.mlb.com/search/?query=ray+lewis&c_id=mlb, 774-http://mlb.mlb.com/search/?query=stone+mountain&c_id=mlb, 775-http://mlb.mlb.com/search/?query=mcdonalds&c_id=mlb, 776-http://mlb.mlb.com/mlb/scoreboard/index.jsp?tcid=nav_mlb_scoreboard, 777-http://m.mlb.com/scores/, 778-http://m.mlb.com/news/article/2013070252487424/, 779-http://mlb.mlb.com/news/article.jsp?ymd=20130702&content_id=52487424&vkey=news_mlb&c_id=mlb, 780-http://m.mlb.com/news/article/2013070252452294/, 781-http://mlb.mlb.com/news/article.jsp?ymd=20130702&content_id=52451388&notebook_id=52452294&vkey=notebook_mil&c_id=mil, 782-http://m.mlb.com/news/article/2013070252456948/, 783-http://mlb.mlb.com/news/article.jsp?ymd=20130702&content_id=52456400&notebook_id=52456948&vkey=notebook_nyy&c_id=nyy, 784-http://bleacherreport.com/, 785-http://bleacherreport.com/search?q=golden, 786-http://bleacherreport.com/search?q=ray+lewis, 787-http://bleacherreport.com/search?q=stone+mountain, 788-http://bleacherreport.com/search?q=mcdonalds, 789-http://bleacherreport.com/articles/1692214-winners-and-losers-of-clippers-suns-bucks-three-way-deal-involving-eric-bledsoe, 790-http://bleacherreport.com/articles/1691823-grading-baltimore-orioles-chicago-cubs-trade-sending-scott-feldman-to-baltimore, 791-http://bleacherreport.com/articles/1692054-awesome-power-of-video-game-bo-jackson-reportedly-returns-for-ncaa-football-14, 792-http://bleacherreport.com/articles/1691830-dwight-howard-rumors-latest-on-d12s-free-agency-decision, 793-http://www.match.com/home/mymatch.aspx?lid=2, 794-http://www.match.com/, 795-http://www.aweber.com/, 796-http://www.aweber.com/search.htm?q=golden&submit=Go, 797-http://www.aweber.com/search.htm?q=ray+lewis&submit=Go, 798-http://www.aweber.com/search.htm?q=stone+mountain&submit=Go, 799-http://www.aweber.com/search.htm?q=mcdonalds&submit=Go, 800-http://www.aweber.com/subscriber-management.htm, 801-http://www.aweber.com/html-email-templates.htm, 802-http://www.aweber.com/customer-solutions.htm, 803-http://www.aweber.com/blog-newsletters.htm, 804-http://m.fedex.com/mt/www.fedex.com/us/?un_zip_uat=&un_jtt_redirect, 805-http://www.fedex.com/us/, 806-http://www.fedex.com/Search/search?q=golden&output=xml_no_dtd&sort=date\%3AD\%3AL\%3Ad1&client=fedex_us&ud=1&oe=UTF-8&ie=UTF-8&proxystylesheet=fedex_us&hl=en&site=us&headerFooterDir=us, 807-http://www.fedex.com/

Search/search?q=ray+lewis&output=xml_no_dtd&sort=date\%3AD\%3AL\%3Ad1&client=fedex_us&ud=1&oe=UTF-8&ie=UTF-8&proxystylesheet=fedex_us&hl=en&site=us&headerFooterDir=us, 808-http://www.fedex.com/Search/search?q=stone+mountain&output=xml_no_dtd&sort=date\%3AD\%3AL\%3Ad1&client=fedex_us&ud=1&oe=UTF-8&ie=UTF-8&proxystylesheet=fedex_us&hl=en&site=us&headerFooterDir=us, 809-http://www.fedex.com/Search/search?q=mcdonalds&output=xml_no_dtd&sort=date\%3AD\%3AL\%3Ad1&client=fedex_us&ud=1&oe=UTF-8&ie=UTF-8&proxystylesheet=fedex_us&hl=en&site=us&headerFooterDir=us, 810-http://www.doublepimp.com/, 811-http://online.wsj.com/home-page, 812-http://online.wsj.com/article/SB10001424127887324436104578582082787214660.html?mod=WSJ_hpp_LEFTTopStories, 813-http://online.wsj.com/article/SB10001424127887324436104578580864003593342.html?mod=WSJ_hppMIDDLENexttoWhatsNewsSecond, 814-http://online.wsj.com/article/SB10001424127887324251504578581842366843974.html?mod=WSJ_hp_LEFTWhatsNewsCollection, 815-http://online.wsj.com/article/SB10001424127887324251504578581931036691650.html?mod=WSJ_hpp_LEFTTopStories, 816-http://online.wsj.com/search/term.html?KEYWORDS=golden&mod=DNH_S, 817-http://online.wsj.com/search/term.html?KEYWORDS=ray+lewis&mod=DNH_S, 818-http://online.wsj.com/search/term.html?KEYWORDS=stone+mountain&mod=DNH_S, 819-http://online.wsj.com/search/term.html?KEYWORDS=mcdonalds&mod=DNH_S, 820-https://vimeo.com/, 821-http://vimeo.com/search?q=golden, 822-http://vimeo.com/search?q=ray+lewis, 823-http://vimeo.com/search?q=stone+mountain, 824-http://vimeo.com/search?q=mcdonalds, 825-http://www.verizonwireless.com/wcms/consumer/iphone-offers.html, 826-https://m.verizonwireless.com/, 827-http://search.verizonwireless.com/?q=golden, 828-http://search.verizonwireless.com/?q=ray+lewis, 829-http://search.verizonwireless.com/?q=stone+mountain, 830-http://search.verizonwireless.com/?q=mcdonalds, 831-http://www.verizonwireless.com/b2c/storelocator/index.jsp, 832-https://m.verizonwireless.com/storelocator, 833-http://www.verizonwireless.com/wcms/consumer/shop.html, 834-https://m.verizonwireless.com/shop, 835-http://www.pof.com/, 836-http://www.tripadvisor.com/, 837-http://www.tripadvisor.com/Search?q=golden&sub-search=SEARCH&geo=&pid=3826&returnTo=__2F__, 838-http://www.tripadvisor.com/Search?q=ray+lewis&sub-search=SEARCH&geo=&pid=3826&returnTo=__2F__, 839-http://www.tripadvisor.com/Search?q=stone+mountain&sub-search=SEARCH&geo=&pid=3826&returnTo=__2F__, 840-http://www.tripadvisor.com/Search?q=mcdonalds&sub-search=SEARCH&geo=&pid=3826&returnTo=__2F__, 841-http://www.tripadvisor.com/ShowUserReviews-g186424-d2072790-r160198596-Detroits-Worcester_Worcestershire_England.html, 842-http://www.tripadvisor.com/Hotel_Review-g186424-d514528-Reviews-Travelodge_Worcester-Worcester_Worcestershire_England.html, 843-http://www.tripadvisor.com/Hotel_Review-g186424-d617017-Reviews-Pear_Tree_Inn_and_Country_Hotel-Worcester_Worcestershire_England.html, 844-http://www.tripadvisor.com/Hotel_Review-g186424-d193926-Reviews-Ye_

Olde_Talbot-Worcester_Worcestershire_England.html, 845-https://hootsuite.com/, 846-https://m.hootsuite.com/login?redirect=\%2F, 847-http://www.pogo.com/, 848-http://www.salesforce.com/, 849-http://www.salesforce.com/solutions/, 850-http://www.salesforce.com/site-search.jsp?cx=007946504037312675699\%3Aneen5rrs2_a&cof=FORID\%3A11\%3BNB\%3A1&q=golden, 851-http://www.salesforce.com/site-search.jsp?cx=007946504037312675699\%3Aneen5rrs2_a&cof=FORID\%3A11\%3BNB\%3A1&q=ray+lewis, 852-http://www.salesforce.com/site-search.jsp?cx=007946504037312675699\%3Aneen5rrs2_a&cof=FORID\%3A11\%3BNB\%3A1&q=stone+mountain, 853-http://www.salesforce.com/site-search.jsp?cx=007946504037312675699\%3Aneen5rrs2_a&cof=FORID\%3A11\%3BNB\%3A1&q=mcdonalds, 854-http://www.salesforce.com/solutions/financial-services/?d=70130000000szz3&internal=true, 855-http://www.salesforce.com/customers/stories/commonwealth-bank.jsp, 856-http://www.salesforce.com/solutions/high-tech/, 857-https://www.americanexpress.com/, 858-https://online.americanexpress.com/myca/mobl/us/login.do, 859-https://search.americanexpress.com/app/answers/list/search/1/kw/golden, 860-https://search.americanexpress.com/app/answers/list/search/1/kw/ray+lewis, 861-https://search.americanexpress.com/app/answers/list/search/1/kw/stone+mountain, 862-https://search.americanexpress.com/app/answers/list/search/1/kw/mcdonalds, 863-https://www304.americanexpress.com/credit-card/?inav=menu_cards_pc_chargecreditcard, 864-https://www262.americanexpress.com/card-application/unauth/featuredCardsPage.do?businessUnit=CCSG&inav=usmbl_menu_cards_personal_pr_body, 865-https://www262.americanexpress.com/card-application/unauth/cardDetail.do?id=1, 866-https://www304.americanexpress.com/credit-card/compare/25330?linknav=us-CCSG-ProspectNav-ViewAllCardsFilterinav=menu_cards_pc_chargecreditcard, 867-https://online.americanexpress.com/myca/mobl/us/static.do?page=un_help&content=CntUs&inav=usmbl_foot_gen_contact_pr, 868-http://www.tube8.com/, 869-http://m.tube8.com/, 870-http://www.tube8.com/searches.html?q=golden, 871-http://www.tube8.com/searches.html?q=ray+lewis, 872-http://www.tube8.com/searches.html?q=stone+mountain, 873-http://www.tube8.com/searches.html?q=mcdonalds, 874-http://www.constantcontact.com/beginnow?s_tnt=47758:13:0, 875-http://www.constantcontact.com/search/index.jsp?q=golden, 876-http://www.constantcontact.com/search/index.jsp?q=ray+lewis, 877-http://www.constantcontact.com/search/index.jsp?q=stone+mountain, 878-http://www.constantcontact.com/search/index.jsp?q=mcdonalds, 879-http://www.constantcontact.com/email-marketing, 880-http://www.constantcontact.com/social-campaigns, 881-http://www.constantcontact.com/social-campaigns/features/create, 882-http://www.constantcontact.com/social-campaigns/why-us, 883-http://aws.amazon.com/, 884-http://aws.amazon.com/search?searchQuery=golden&searchPath=all&x=0&y=0, 885-http://aws.amazon.com/search?searchQuery=ray+lewis&searchPath=all&x=0&y=0, 886-http://aws.amazon.com/search?searchQuery=stone+mountain&searchPath=all&x=0&y=0, 887-http:

//aws.amazon.com/search?searchQuery=mcdonalds&searchPath=all&x=0&y=0, 888-http://aws.amazon.com/ec2/, 889-http://aws.amazon.com/ebs/, 890-https://aws.amazon.com/amis?_encoding=UTF8&jiveRedirect=1, 891-https://aws.amazon.com/amis/windows-server-2008-r2-sp1-english-64bit-windows-media-services-4-1-2012-03-15, 892-http://www.yellowpages.com/, 893-http://www.yellowpages.com/chapel-hill-nc/golden?g=chapel+hill\%2C+nc, 894-http://www.yellowpages.com/chapel-hill-nc/mip/golden-taxi-cab-467961038?lid=467961038, 895-http://www.yellowpages.com/chapel-hill-nc/mip/golden-eldercare-management-pc-466871437?lid=466871437, 896-http://www.yellowpages.com/chapel-hill-nc/mip/golden-transportation-service-470808221?lid=470808221, 897-http://www.yellowpages.com/durham-nc/mip/golden-pine-ventures-460396995?lid=460396995, 898-http://www.yellowpages.com/chapel-hill-nc/ray-lewis?g=Chapel+Hill\%2C+NC&q=ray+lewis, 899-http://www.yellowpages.com/chapel-hill-nc/ray-lewis?g=Chapel+Hill\%2C+NC&q=stone+mountain, 900-http://www.yellowpages.com/chapel-hill-nc/ray-lewis?g=Chapel+Hill\%2C+NC&q=mcdonalds, 901-http://m.monster.com/, 902-http://www.monster.com/, 903-http://jobsearch.monster.com/search/golden_5, 904-http://jobsearch.monster.com/search/ray\%20lewis_5, 905-http://jobsearch.monster.com/search/stone\%20mountain_5, 906-http://jobsearch.monster.com/search/mcdonalds_5, 907-http://m.monster.com/JobSearch/Search?jobtitle=golden&keywords=&where=, 908-http://m.monster.com/JobSearch/Search?jobtitle=ray\%20lewis&keywords=&where=, 909-http://m.monster.com/JobSearch/Search?jobtitle=stone\%20mountain&keywords=&where=, 910-http://m.monster.com/JobSearch/Search?jobtitle=mcdonalds&keywords=&where=, 911-http://m.monster.com/Oracle-Golden-Gate-DBA-Job-Cupertino-CA-123166382, 912-http://jobview.monster.com/Oracle-Golden-Gate-DBA-Job-Cupertino-CA-123166382.aspx, 913-http://m.monster.com/Claims-Customer-Contact-Center-Supervisor-Golden-CO-Job-Englewood-CO-123479277, 914-http://jobview.monster.com/Claims-Customer-Contact-Center-Supervisor-Golden-CO-Job-Englewood-CO-123479277.aspx, 915-http://m.monster.com/Casino-The-Golden-Gate-Hotel-Job-Las-Vegas-NV-123503423, 916-http://jobview.monster.com/Casino-The-Golden-Gate-Hotel-Job-Las-Vegas-NV-123503423.aspx, 917-http://jobview.monster.com/Final-Expense-Golden-Opportunity-Job-Grand-Rapids-MI-123410019.aspx, 918-http://m.monster.com/Final-Expense-Golden-Opportunity-Job-Grand-Rapids-MI-123410019, 919-http://stackoverflow.com/, 920-http://stackoverflow.com/search?q=golden, 921-http://stackoverflow.com/search?q=ray+lewis, 922-http://stackoverflow.com/search?q=stone+mountain, 923-http://stackoverflow.com/search?q=mcdonalds, 924-http://stackoverflow.com/questions/12109826/issue-with-benthic-golden-workspace, 925-http://stackoverflow.com/questions/8594209/golden-ratio-webpage-template, 926-http://stackoverflow.com/questions/16385353/golden-section-method-advantages, 927-http://stackoverflow.com/questions/5395653/css-design-with-the-golden-ratio, 928-

http://m.lowes.com/, 929-http://www.lowes.com/, 930-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=golden, 931-http://www.lowes.com/Search=golden?storeId=10151&langId=-1&catalogId=10051&N=0&newSearch=true&Ntt=golden#!, 932-http://www.lowes.com/Search=ray+lewis?storeId=10151&langId=-1&catalogId=10051&N=0&newSearch=true&Ntt=ray+lewis#!, 933-http://www.lowes.com/Search=stone+mountain?storeId=10151&langId=-1&catalogId=10051&N=0&newSearch=true&Ntt=stone+mountain#!, 934-http://www.lowes.com/Search=mcdonalds?storeId=10151&langId=-1&catalogId=10051&N=0&newSearch=true&Ntt=mcdonalds#!, 935-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=ray+lewis, 936-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=stone+mountain, 937-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=mcdonalds, 938-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=3819849&store=595&view=detail, 939-http://www.lowes.com/pd_383763-63094-SRFP11222_0__?productId=3819849&Ntt=golden&pl=1&currentURL=\%3FNtt\%3Dgolden&facetInfo=, 940-http://www.lowes.com/pd_392043-19871-AR157_0__?productId=3645846&Ntt=golden&pl=1&currentURL=\%3FNtt\%3Dgolden&facetInfo=, 941-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=3645846&store=595&view=detail, 942-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=3319292&store=595&view=detail, 943-http://www.lowes.com/pd_159087-66150-20G+VSDB36_0__?productId=3285420&Ntt=golden&pl=1&currentURL=\%3FNtt\%3Dgolden&facetInfo=, 944-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=1238357&store=595&view=detail, 945-http://www.lowes.com/pd_240815-2120-34332_0__?productId=1238357&Ntt=golden&pl=1&currentURL=\%3FNtt\%3Dgolden&facetInfo=, 946-http://m.mapquest.com/, 947-http://www.mapquest.com/, 948-http://www.photobucket.com/, 949-http://m.photobucket.com/, 950-http://m.photobucket.com/images/ray+lewis, 951-http://m.photobucket.com/images/golden, 952-http://m.photobucket.com/images/stone+mountain, 953-http://m.photobucket.com/images/mcdonalds, 954-http://m1246.photobucket.com/image/ray\%20lewis/yourmyboyblue/raylewisebay_zpsc280184d.jpg.html?o=0, 955-http://m1292.photobucket.com/image/gold/Ronald_mark_Johnson/AMBER_GOLD_zpse6cd7fa6.jpg.html?o=0, 956-http://m1351.photobucket.com/image/stone/vhlsigs/stone_zps8e96cdf9.jpg.html?o=0, 957-http://m1163.photobucket.com/image/mcdonalds/cristinab42/thailand/IMG_6072.jpg.html?o=0, 958-http://photobucket.com/images/golden?page=1, 959-http://photobucket.

com/images/ray+lewis?page=1, 960-http://media.photobucket.com/user/yourmyboyblue/media/raylewisebay_zpsc280184d.jpg.html?filters[term]=ray\%20lewis&filters[primary]=images&filters[secondary]=videos&sort=1&o=0, 961-http://photobucket.com/images/stone+mountain?page=1, 962-http://photobucket.com/images/mcdonalds?page=1, 963-http://media.photobucket.com/user/Ronald_mark_Johnson/media/AMBER_GOLD_zpse6cd7fa6.jpg.html?filters[term]=gold&filters[primary]=images&filters[secondary]=videos&sort=1&o=0, 964-http://media.photobucket.com/user/vhlsigs/media/stone_zps8e96cdf9.jpg.html?filters[term]=stone&filters[primary]=images&filters[secondary]=videos&sort=1&o=0, 965-http://media.photobucket.com/user/cristinab42/media/thailand/IMG_6072.jpg.html?filters[term]=mcdonalds&filters[primary]=images&filters[secondary]=videos&sort=1&o=0, 966-http://www.tmz.com/, 967-http://m.tmz.com/, 968-http://m.tmz.com/2013/07/07/dwight-howard-game-lakers-rockets/, 969-http://www.tmz.com/2013/07/07/dwight-howard-game-lakers-rockets/, 970-http://m.tmz.com/2013/07/07/lady-gaga-rob-fusari-lawsuit-wendy-starland/, 971-http://www.tmz.com/2013/07/07/lady-gaga-rob-fusari-lawsuit-wendy-starland/, 972-http://www.tmz.com/2013/07/07/anna-nicole-smith-movie-lifetime-car-damage-200/, 973-http://m.tmz.com/2013/07/07/anna-nicole-smith-movie-lifetime-car-damage-200/, 974-http://www.tmz.com/2013/07/07/tameka-raymond-landlord-lawsuit/, 975-http://m.tmz.com/2013/07/07/tameka-raymond-landlord-lawsuit/, 976-http://www.tmz.com/search/news/golden/1/, 977-http://www.tmz.com/search/news/ray+lewis/1/, 978-http://www.tmz.com/search/news/stone+mountain/1/, 979-http://www.tmz.com/search/news/mcdonalds/1/, 980-http://www.forbes.com/, 981-http://www.forbes.com/search/?q=golden, 982-http://www.forbes.com/search/?q=ray+lewis, 983-http://www.forbes.com/search/?q=stone+mountain, 984-http://www.forbes.com/search/?q=mcdonalds, 985-http://www.forbes.com/sites/prishe/2013/01/30/legend-or-liar-should-ray-lewis-play-in-super-bowl-xlvii/, 986-http://www.forbes.com/sites/jamesgruber/2013/07/06/bonds-to-bounce-back/, 987-http://www.forbes.com/sites/ashleaebeling/2013/07/05/how-to-get-multiple-offers-on-your-home/, 988-http://www.forbes.com/sites/learnvest/2013/07/03/9-job-mistakes-that-could-stall-your-entire-career/, 989-http://www22.verizon.com/home/verizonglobalhome/ghp_landing.aspx, 990-https://m.verizon.com/PAMMobile/presignin.aspx, 991-http://search.verizon.com/?tp=r&rv=r&q=golden, 992-http://search.verizon.com/?tp=r&rv=r&q=ray+lewis, 993-http://search.verizon.com/?tp=r&rv=r&q=stone+mountain, 994-http://search.verizon.com/?tp=r&rv=r&q=mcdonalds, 995-https://m.verizon.com/mforyourhome/services.aspx, 996-http://www22.verizon.com/home/services/, 997-http://www.adnxs.com/, 998 -http://www.wordpress.org, 999-http://wordpress.org/search/golden, 1000-http://wordpress.org/search/ray\%20lewis, 1001-http://wordpress.org/search/stone\%20mountain, 1002-http://wordpress.org/search/mcdonalds, 1003-http://wordpress.org/hosting/, 1004-

http://wordpress.org/download/, 1005-http://codex.wordpress.org/Main_Page, 1006-http://codex.wordpress.org/Getting_Started_with_WordPress, 1007-http://www.trulia.com/, 1008-http://www.trulia.com/CO/Golden/, 1009-http://www.trulia.com/CO/Ray_Lewis/, 1010-http://www.trulia.com/CO/stone_mountain/, 1011-http://www.trulia.com/CO/mcdonalds/, 1012-http://www.trulia.com/property/3123083916-16259-W-10th-Ave-Golden-CO-80401, 1013-http://www.trulia.com/property/1062272057-293-White-Ash-Dr-Golden-CO-80403, 1014-http://www.trulia.com/property/1096166618-15052-W-13th-Ave-Golden-CO-80401, 1015-http://www.trulia.com/property/3043392355-1553-Robinson-Hill-Rd-Golden-CO-80403, 1016-http://www.latimes.com/, 1017-http://mobile.latimes.com/s.p?sId=7&m=b, 1018-http://www.latimes.com/search/dispatcher.front?Query=golden&target=adv_all, 1019-http://www.latimes.com/search/dispatcher.front?Query=ray+lewis&target=adv_all, 1020-http://www.latimes.com/search/dispatcher.front?Query=stone+mountain&target=adv_all, 1021-http://www.latimes.com/search/dispatcher.front?Query=mcdonalds&target=adv_all, 1022-http://www.latimes.com/sports/sportsnow/la-sp-sn-rg3-married-in-lavish-style-2013070703346887.story, 1023-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858928&postId=3858928&postUserId=7&sessionToken=&catId=6907&curAbsIndex=0&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\%3DsectionId\%253A6907\%26DPS\%3D0\%26DPL\%3D3, 1024-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858875&postId=3858875&postUserId=7&sessionToken=&catId=6907&curAbsIndex=1&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\%3DsectionId\%253A6907\%26DPS\%3D0\%26DPL\%3D3, 1025-http://www.latimes.com/sports/sportsnow/la-sp-sn-marion-bartoli-not-a-looker-remark-2013070707627216.story, 1026-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858853&postId=3858853&postUserId=7&sessionToken=&catId=5224&curAbsIndex=0&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\%3DsectionId\%253A5224\%26DPS\%3D0\%26DPL\%3D3, 1027-http://www.latimes.com/local/lanow/la-me-ln-sf-plane-crash-asiana-airlines-president-apologizes-20130707,0,6390842.story, 1028-http://www.latimes.com/news/world/worldnow/la-fg-wn-qatada-jordan-20130707,0,2027104.story, 1029-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858615&postId=3858615&postUserId=7&sessionToken=&catId=5217&curAbsIndex=1&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\%3DsectionId\%253A5217\%26DPS\%3D0\%26DPL\%3D3, 1030-http://www.sears.com/, 1031-http://m.sears.com/, 1032-http://m.sears.com/keyword.do?keyword=golden&vertName=&vName=, 1033-http://www.sears.com/search=golden?vName=Movies+Music&cName=Music&autoRedirect=true&viewItems=50&redirectType=CAT_REC, 1034-http://www.sears.com/search=ray+lews, 1035-http://www.sears.com/search=stone+mountain, 1036-http://www.sears.com/search=mcdonalds, 1037-http://m.sears.com/keyword.do?keyword=ray+lewis&vertName=&vName=, 1038-http://m.sears.com/keyword.do?keyword=stone+mountain&vertName=&vName=, 1039-http://m.sears.com/keyword.do?keyword=mcdonalds&vertName=&vName=, 1040-http://m.sears.com/productdetails.do?partNumber=05757212000P&reviewCount=zero&itemSrc=Online&threshold=59.0&fullFillment=TW, 1041-http://www.sears.com/panasonic-smart-network-blu-ray-disc-8482-player-with/p-05757212000P?prdNo=6&blockNo=6&blockType=G6, 1042-http://m.sears.com/productdetails.do?partNumber=05729406000P, 1043-http://www.sears.com/panasonic-65-in-smart-viera-s60-series-plasma/p-05729406000P?prdNo=1&blockNo=1&blockType=G1, 1044-http://www.sears.com/front-porch-classics-skittle-pool/p-05238123000P?prdNo=1&blockNo=1&blockType=G1, 1045-http://m.sears.com/productdetails.do?partNumber=05238123000P&reviewCount=zero&itemSrc=Online&threshold=59.0&fullFillment=VD, 1046-http://www.sears.com/carrom-skittles/p-00623806000P?prdNo=2&blockNo=2&blockType=G2, 1047-http://m.sears.com/productdetails.do?partNumber=00623806000P&reviewCount=zero&itemSrc=Online&threshold=0.0&fullFillment=VD, 1048-http://www.webmd.com/, 1049-http://www.m.webmd.com/, 1050-http://www.webmd.com/search/search_results/default.aspx?query=golden, 1051-http://www.webmd.com/search/search_results/default.aspx?query=ray+lewis, 1052-http://www.webmd.com/search/search_results/default.aspx?query=stone+mountain, 1053-http://www.webmd.com/search/search_results/default.aspx?query=mcdonalds, 1054-http://www.m.webmd.com/mobile-search/default.htm?query=golden, 1055-http://www.m.webmd.com/mobile-search/default.htm?query=ray+lewis, 1056-http://www.m.webmd.com/mobile-search/default.htm?query=stone+mountain, 1057-http://www.m.webmd.com/mobile-search/default.htm?query=mcdonalds, 1058-http://www.webmd.com/diet/rm-quiz-best-worst-foods-belly-fat, 1059-http://www.m.webmd.com/diet/rm-quiz-best-worst-foods-belly-fat, 1060-http://www.m.webmd.com/a-to-z-guides/rm-quiz-science-love, 1061-http://www.webmd.com/sex-relationships/rm-quiz-science-love, 1062-http://www.expedia.com/, 1063-http://www.expedia.com/MobileHotel?rfrr=-1065&, 1064-http://www.macys.com/, 1065-http://m.macys.com/, 1066-http://www1.macys.com/shop/search?keyword=golden, 1067-http://www1.macys.com/shop/search?keyword=ray+lewis, 1068-http://m.macys.com/shop/search?keyword=golden, 1069-http://m.macys.com/shop/search?keyword=ray+lewis, 1070-http://m.macys.com/shop/search?keyword=stone+mountain, 1071-http://m.macys.com/shop/search?keyword=mcdonalds, 1072-http://www1.macys.com/shop/search?keyword=stone+mountain, 1073-http://www1.macys.com/shop/search?keyword=mcdonalds, 1074-http://www1.macys.com/shop/product/levis-jeans-569-loose-straight?ID=778920&CategoryID=11221#fn=sp\%3D1\%26spc\%3D2\%26kws\%3Dray\%20lewis\%26slotId\%3D1, 1075-http://m.macys.com/shop/product/levis-jeans-569-loose-straight?ID=778920&CategoryID=11221#fn=sp\%3D1\%26spc\%3D2\%26kws\

%3Dray\%20lewis\%26slotId\%3D1, 1076-http://www1.macys.com/shop/product/hello-kitty-kids-toy-girls-or-little-girls-coloring-book?ID=867470&CategoryID=5991#fn=sp\%3D1\%26spc\%3D15\%26kws\%3Dbook\%26slotId\%3D1, 1077-http://m.macys.com/shop/product/hello-kitty-kids-toy-girls-or-little-girls-coloring-book?ID=867470&CategoryID=5991#fn=sp\%3D1\%26spc\%3D15\%26kws\%3Dbook\%26slotId\%3D1, 1078-http://www1.macys.com/shop/product/elizabeth-arden-ceramide-capsules-daily-youth-restoring-serum-total-60-capsules-95-fl-oz?ID=253891&CategoryID=30078#fn=sp\%3D1\%26spc\%3D6574\%26kws\%3Dgold\%26slotId\%3D1, 1079-http://m.macys.com/shop/product/elizabeth-arden-ceramide-capsules-daily-youth-restoring-serum-total-60-capsules-95-fl-oz?ID=253891&CategoryID=30078#fn=sp\%3D1\%26spc\%3D6574\%26kws\%3Dgold\%26slotId\%3D1, 1080-http://www1.macys.com/shop/product/kidz-delight-kids-toy-arthur-little-tv?ID=717950&CategoryID=5991#fn=sp\%3D1\%26spc\%3D13\%26kws\%3Dtv\%26slotId\%3D1, 1081-http://m.macys.com/shop/product/kidz-delight-kids-toy-arthur-little-tv?ID=717950&CategoryID=5991#fn=sp\%3D1\%26spc\%3D13\%26kws\%3Dtv\%26slotId\%3D1, 1082-http://fiverr.com/, 1083-http://fiverr.com/gigs/search?utf8=?&order=latest&search_in=everywhere&query=golden&x=0&y=0, 1084-http://fiverr.com/gigs/search?utf8=?&order=latest&search_in=everywhere&query=ray+lewis&x=0&y=0, 1085-http://fiverr.com/gigs/search?utf8=?&order=latest&search_in=everywhere&query=stone+mountain&x=0&y=0, 1086-http://fiverr.com/gigs/search?utf8=?&order=latest&search_in=everywhere&query=mcdonalds&x=0&y=0, 1087-http://fiverr.com/babita943/make-you-half-egyptian-link-copper-wire-bracelet, 1088-http://fiverr.com/dani7770/make-a-cool-papercraft-of-you, 1089-http://fiverr.com/missbluebird/make-you-a-bunny-ring, 1090-http://fiverr.com/elfstacy/send-a-harry-potter-fan-a-hogwarts-acceptance-letter, 1091-http://www.directrev.com/, 1092-http://www.swagbucks.com/, 1093-https://www.dropbox.com/m/login?cont=https\%3A//www.dropbox.com/m, 1094-https://www.dropbox.com/, 1095-http://www.dailymotion.com/us, 1096-http://www.dailymotion.com/us/relevance/search/golden/1, 1097-http://www.dailymotion.com/us/relevance/search/ray+lewis/1, 1098-http://www.dailymotion.com/us/relevance/search/stone+mountain/1, 1099-http://www.dailymotion.com/us/relevance/search/mcdonalds/1, 1100-http://wigetmedia.com/, 1101-http://wigetmedia.com/about, 1102-http://wigetmedia.com/contact, 1103-http://wigetmedia.com/publishers, 1104-http://www.nba.com/, 1105-http://www.nba.com/search/?text=golden, 1106-http://www.nba.com/search/?text=ray+lewis, 1107-http://www.nba.com/search/?text=stone+mountain, 1108-http://www.nba.com/search/?text=mcdonalds, 1109-http://www.nba.com/video/channels/nba_tv/2013/07/07/20130706-gt-warriors-future.nba/index.html, 1110-http://www.nba.com/2013/news/07/06/jazz-signs-trey-burke-and-rudy-gobert.ap/index.html, 1111-http://www.nba.com/2013/news/07/06/dorell-wright-signs-with-blazers.ap/index.html, 1112-http://www.nba.com/2013/news/07/06/shaq-on-howard-move.ap/index.html, 1113-

http://m.newegg.com/, 1114-http://www.newegg.com/, 1115-http://www.newegg.com/Product/ProductList.aspx?Submit=ENE&DEPA=0&Order=BESTMATCH&Description=golden&N=-1&isNodeId=1, 1116-http://www.newegg.com/Product/ProductList.aspx?Submit=ENE&DEPA=0&Order=BESTMATCH&Description=ray+lewis&N=-1&isNodeId=1, 1117-http://www.newegg.com/Product/ProductList.aspx?Submit=ENE&DEPA=0&Order=BESTMATCH&Description=stone+mountain&N=-1&isNodeId=1, 1118-http://www.newegg.com/Product/ProductList.aspx?Submit=ENE&DEPA=0&Order=BESTMATCH&Description=mcdonalds&N=-1&isNodeId=1, 1119-http://m.newegg.com/ProductList?Keyword=golden, 1120-http://m.newegg.com/ProductList?Keyword=ray+lewis, 1121-http://m.newegg.com/ProductList?Keyword=stone+mountain, 1122-http://m.newegg.com/ProductList?Keyword=mcdonalds, 1123-http://m.newegg.com/Product/index?itemNumber=N82E16834312242, 1124-http://www.newegg.com/Product/Product.aspx?Item=N82E16834312242, 1125-http://m.newegg.com/Product/index?itemNumber=N82E16834230987, 1126-http://www.newegg.com/Product/Product.aspx?Item=N82E16834230987, 1127-http://m.newegg.com/Product/index?itemNumber=N82E16823126097, 1128-http://www.newegg.com/Product/Product.aspx?Item=N82E16823126097, 1129-http://www.newegg.com/Product/Product.aspx?Item=N82E16832416552, 1130-http://m.newegg.com/Product/index?itemNumber=N82E16832416552, 1131-https://www.yieldmanager.com/, 1132-https://online.citibank.com/US/Welcome.c, 1133-https://online.citibank.com/US/JRS/globalsearch/Search.do?qt=golden, 1134-https://online.citibank.com/US/JRS/globalsearch/Search.do?qt=ray+lewis, 1135-https://online.citibank.com/US/JRS/globalsearch/Search.do?qt=stone_mountain, 1136-https://online.citibank.com/US/JRS/globalsearch/Search.do?qt=mcdonalds, 1137-http://msn.foxsports.com/, 1138-http://sports.mobile.msn.com/en-us/, 1139-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906510&acid=2&afid=0, 1140-http://msn.foxsports.com/tennis/story/andy-murray-wins-wimbledon-title-defeats-novak-djokovic-will-last-forever-070713, 1141-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906568&acid=2&afid=0, 1142-http://msn.foxsports.com/nascar/story/kurt-busch-leaves-daytona-unscathed-in-chase-for-the-sprint-cup-070713, 1143-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906467&acid=2&afid=0, 1144-http://msn.foxsports.com/ufc/story/Anderson-Silva-costs-himself-title-against-Chris-Weidman-at-UFC-162-070613, 1145-http://msn.foxsports.com/mlb/story/adam-jones-home-run-off-mariano-rivera-lift-baltimore-orioles-over-new-york-yankees-070713, 1146-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906639&acid=2&afid=0, 1147-http://msn.foxsports.com/search?sp_q=golden, 1148-http://msn.foxsports.com/search?sp_q=ray+lewis, 1149-http://msn.foxsports.com/search?sp_q=stone+mountain, 1150-http://msn.foxsports.com/search?sp_q=mcdonalds, 1151-http://mobile.backpage.com/, 1152-http://www.backpage.com/, 1153-http://auburn.backpage.com/Events/reptile-super-show-july-6-7-2013-san-diego-ca-concourse-civic-center-downtown/7978335, 1154-http://mobile.backpage.com/Events/reptile-super-show-july-6-7-2013-san-diego-ca-concourse-civic-center-downtown/7978335, 1155-http://mobile.backpage.com/Events/nighttalker-radio-show-michael-hastings-car-crash-assassination/7924589, 1156-http://auburn.backpage.com/Events/nighttalker-radio-show-michael-hastings-car-crash-assassination/7924589, 1157-http://auburn.backpage.com/Events/hookup-party-this-weekend-saturday-july-13th/8050194, 1158-http://mobile.backpage.com/Events/hookup-party-this-weekend-saturday-july-13th/8050194, 1159-http://mobile.backpage.com/Events/?keyword=golden, 1160-http://auburn.backpage.com/Events/?keyword=golden, 1161-http://mobile.backpage.com/Events/?keyword=ray+lewis, 1162-http://mobile.backpage.com/Events/?keyword=stone+mountain, 1163-http://mobile.backpage.com/Events/?keyword=mcdonalds, 1164-http://auburn.backpage.com/Events/?keyword=ray+lewis, 1165-http://auburn.backpage.com/Events/?keyword=stone+mountain, 1166-http://auburn.backpage.com/Events/?keyword=mcdonalds, 1167-http://www.southwest.com/, 1168-https://mobile.southwest.com/p?stpp=true&formid=main, 1169-https://login.live.com/login.srf?wa=wsignin1.0&ct=1373235406&rver=6.1.6206.0&sa=1&ntprob=-1&wp=MBI_SSL_SHARED&wreply=https:\%2F\%2Fmail.live.com\%2F\%3Fowa\%3D1\%26owasuffix\%3Dowa\%252f&id=64855&snsc=1&cbcxt=mail, 1170 -http://www.xhamstercams.com, 1171-http://www.xhamstercams.com/search.php?q=golden&submit=Search, 1172-http://www.xhamstercams.com/search.php?q=ray+lewis&submit=Search, 1173-http://www.xhamstercams.com/search.php?q=stone+mountain&submit=Search, 1174-http://www.xhamstercams.com/search.php?q=mcdonalds&submit=Search, 1175-http://www.goodreads.com/, 1176-http://www.goodreads.com/book/show/33507.Twenty_Thousand_Leagues_Under_the_Sea, 1177-http://www.goodreads.com/book/show/22463.The_Origin_of_Species, 1178-http://www.goodreads.com/book/show/16131193-the-astronaut-wives-club, 1179-http://www.goodreads.com/book/show/16248046-the-alley-of-love-and-yellow-jasmines, 1180-http://www.cbslocal.com/, 1181-http://www.baidu.com/, 1182-http://m.baidu.com/, 1183-http://www.baidu.com/s?wd=golden&rsv_bp=0&ch=&tn=baidu&bar=&rsv_spt=3&ie=utf-8&rsv_sug3=4&inputT=855, 1184-http://www.baidu.com/s?wd=ray+lewis&rsv_bp=0&ch=&tn=baidu&bar=&rsv_spt=3&ie=utf-8&rsv_sug3=4&inputT=855, 1185-http://www.baidu.com/s?wd=stone+mountain&rsv_bp=0&ch=&tn=baidu&bar=&rsv_spt=3&ie=utf-8&rsv_sug3=4&inputT=855, 1186-http://www.baidu.com/s?wd=mcdonalds&rsv_bp=0&ch=&tn=baidu&bar=&rsv_spt=3&ie=utf-8&rsv_sug3=4&inputT=855, 1187-http://m.baidu.com/ssid=0/from=0/bd_page_type=1/uid=51D9DEB13A6F4949623FC7CA2BC3E04D/baiduid=E6642719F9F904A9AF5B1ECBD542168B/s?word=golden, 1188-http://m.baidu.com/ssid=0/from=0/bd_page_type=1/uid=51D9DEB13A6F4949623FC7CA2BC3E04D/baiduid=E6642719F9F904A9AF5B1ECBD542168B/s?word=ray+lewis, 1189-http://m.baidu.com/ssid=0/from=0/bd_page_type=1/uid=51D9DEB13A6F4949623FC7CA2BC3E04D/baiduid=

E6642719F9F904A9AF5B1ECBD542168B/s?word=stone+mountain, 1190-http://m.baidu.com/ssid=0/from=0/bd_page_type=1/uid=51D9DEB13A6F4949623FC7CA2BC3E04D/baiduid=E6642719F9F904A9AF5B1ECBD542168B/s?word=mcdonalds, 1191-http://kickass.to/, 1192-http://kickass.to/usearch/golden/, 1193-http://kickass.to/usearch/ray\%20lewis/, 1194-http://kickass.to/usearch/stone\%20mountain/, 1195-http://kickass.to/usearch/mcdonalds/, 1196-http://kickass.to/maxim-magazine-usa-july-2013-aft3rlif3-t7548028.html, 1197-http://kickass.to/ride-to-hell-retribution-dlc-steam-rip-multi5-rg-gameworks-t7548891.html, 1198-http://kickass.to/deadpool-pc-full-game-en-ru-nosteam-t7547512.html, 1199-http://kickass.to/the-sims-3-island-paradise-full-games4theworld-t7543868.html, 1200-http://www.intuit.com/, 1201-http://m.intuit.com/, 1202-http://www.java.com/en/, 1203-http://search.oracle.com/search/search?group=Java.com&search_p_main_operator=any&search_p_atname=url&search_p_op=contains&search_p_val=en&q=golden&submit.x=0&submit.y=0, 1204-http://search.oracle.com/search/search?group=Java.com&search_p_main_operator=any&search_p_atname=url&search_p_op=contains&search_p_val=en&q=ray+lewis&submit.x=0&submit.y=0, 1205-http://search.oracle.com/search/search?group=Java.com&search_p_main_operator=any&search_p_atname=url&search_p_op=contains&search_p_val=en&q=stone+mountain&submit.x=0&submit.y=0, 1206-http://search.oracle.com/search/search?group=Java.com&search_p_main_operator=any&search_p_atname=url&search_p_op=contains&search_p_val=en&q=mcdonalds&submit.x=0&submit.y=0, 1207-http://www.java.com/en/download/faq/develop.xml#javaconf, 1208-http://www.java.com/en/download/help/index_installing.xml, 1209-http://www.java.com/en/download/help/disable_browser.xml, 1210-http://www.java.com/en/download/faq/index_general.xml, 1211-http://www.oracle.com/index.html, 1212-http://search.oracle.com/search/search?start=1&search_p_main_operator=all&q=golden, 1213-http://search.oracle.com/search/search?start=1&search_p_main_operator=all&q=ray+lewis, 1214-http://search.oracle.com/search/search?start=1&search_p_main_operator=all&q=stone+mountain, 1215-http://search.oracle.com/search/search?start=1&search_p_main_operator=all&q=mcdonalds, 1216-http://www.oracle.com/us/technologies/big-data/index.html, 1217-http://www.oracle.com/us/products/database/overview/index.html, 1218-http://www.oracle.com/us/products/applications/human-capital-management/overview/index.htm, 1219-http://www.oracle.com/us/corporate/contact/global-070511.html, 1220-http://www.ca.gov/, 1221-http://m.ca.gov/, 1222-http://www.ca.gov/Apps/SearchNew.aspx?search=golden&cx=001779225245372747843\%3Amdsmtl_vi1a&cof=&ie=UTF-8&submit.x=0&submit.y=0, 1223-http://www.ca.gov/Apps/SearchNew.aspx?search=ray+lewis&cx=001779225245372747843\%3Amdsmtl_vi1a&cof=&ie=UTF-8&submit.x=0&submit.y=0, 1224-http://www.ca.gov/Apps/SearchNew.aspx?search=stone+mountain&cx=001779225245372747843\%3Amdsmtl_vi1a&cof=&ie=UTF-8&submit.x=0&submit.y=0, 1225-http://www.ca.gov/Apps/SearchNew.aspx?search=mcdonalds&cx=

0017792252453727478 43\%3Amdsmtl_vi1a&cof=&ie=UTF-8&submit.x=0&submit.y=0, 1226-http://www.empowernetwork.com/, 1227-http://www.empowernetwork.com/blog-system.php?id=, 1228-http://www.empowernetwork.com/inner-circle-mastermind.php, 1229-http://www.empowernetwork.com/costa-rica-intensive.php, 1230-http://www.empowernetwork.com/video-hosting.php, 1231-http://www.ancestry.com/, 1232-http://www.nydailynews.com/, 1233-http://www.nydailynews.com/search-results/search-results-7.113?q=golden&selecturl=site, 1234-http://www.nydailynews.com/search-results/search-results-7.113?q=ray+lewis&selecturl=site, 1235-http://www.nydailynews.com/search-results/search-results-7.113?q=stone+mountain&selecturl=site, 1236-http://www.nydailynews.com/search-results/search-results-7.113?q=mcdonalds&selecturl=site, 1237-http://www.nydailynews.com/news/election/love-gov-eliot-spitzer-run-city-controller-article-1.1392529, 1238-http://www.nydailynews.com/news/crime/1-teen-killed-iowa-crash-driver-arrested-article-1.1392385, 1239-http://www.nydailynews.com/new-york/brooklyn/mta-worker-shoots-bklyn-station-article-1.1392193, 1240-http://www.nydailynews.com/news/politics/president-obama-u-s-backing-egyptian-party-group-article-1.1392196, 1241-http://m.realtor.com/, 1242-http://www.realtor.com/, 1243-http://m.realtor.com/#results?loc=golden&type=single_family\%2Ccondo\%2Cland, 1244-http://m.realtor.com/#results?loc=ray+lewis&type=single_family\%2Ccondo\%2Cland, 1245-http://m.realtor.com/#results?loc=stone+mountain&type=single_family\%2Ccondo\%2Cland, 1246-http://m.realtor.com/#results?loc=mcdonalds&type=single_family\%2Ccondo\%2Cland, 1247-http://m.cbsnews.com/, 1248-http://www.cbsnews.com/, 1249-http://www.cbsnews.com/1770-5_162-0.html?query=golden&tag=srch&searchtype=cbsSearch, 1250-http://www.cbsnews.com/1770-5_162-0.html?query=ray+lewis&tag=srch&searchtype=cbsSearch, 1251-http://www.cbsnews.com/1770-5_162-0.html?query=stone+mountain&tag=srch&searchtype=cbsSearch, 1252-http://www.cbsnews.com/1770-5_162-0.html?query=mcdonalds&tag=srch&searchtype=cbsSearch, 1253-http://m.cbsnews.com/searchstory.rbml?query=golden&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 1254-http://m.cbsnews.com/searchstory.rbml?query=ray+lewis&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 1255-http://m.cbsnews.com/searchstory.rbml?query=stone+mountain&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 1256-http://m.cbsnews.com/searchstory.rbml?query=mcdonalds&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 1257-http://www.cbsnews.com/8301-201_162-57592558/asiana-airlines-flight-214-tried-to-abort-landing/, 1258-http://m.cbsnews.com/storysynopsis.rbml?catid=57592558&feed_id=0&videofeed=36, 1259-http://www.cbsnews.com/8301-250_162-57592575/teresa-heinz-kerry-in-critical-condition-at-hospital/, 1260-http://m.cbsnews.com/storysynopsis.rbml?catid=57592575&feed_id=0&videofeed=36, 1261-http://m.cbsnews.com/storysynopsis.rbml?catid=57592567&feed_id=0&videofeed=36, 1262-http://www.cbsnews.com/8301-202_162-57592567/russianofficial-venezuela-is-snowdens-last-chance/, 1263-http://www.cbsnews.com/8301-202_162-57592553/death-toll-in-quebec-train-explosion-rises-to-5/, 1264-http://m.cbsnews.com/storysynopsis.rbml?catid=57592553&feed_id=0&videofeed=36, 1265-http://www.priceline.com/l/home.htm?&rdr=p2seti&sv3=Y, 1266-https://www.priceline.com/smartphone/home.do?plf=PCLN, 1267-http://disney.com/, 1268-http://search.disney.com/search?o=home&q=golden, 1269-http://search.disney.com/search?o=home&q=ray+lewis, 1270-http://search.disney.com/search?o=home&q=stone+mountain, 1271-http://search.disney.com/search?o=home&q=mcdonalds, 1272-http://disney.go.com/monsters-university/#/characters/mike-wazowski, 1273-http://www.warriorforum.com/, 1274-http://www.warriorforum.com/main-internet-marketing-discussion-forum/741403-warrior-forum-rules.html, 1275-http://www.warriorforum.com/main-internet-marketing-discussion-forum/47454-being-better-member-moderator.html, 1276-http://www.warriorforum.com/tags/allinone.html, 1277-http://www.warriorforum.com/main-internet-marketing-discussion-forum/300-you-moderator.html, 1278-http://m.wikihow.com/Main-Page, 1279-http://www.wikihow.com/Main-Page, 1280-http://www.wikihow.com/Special:GoogSearch?cx=008953293426798287586\%3Amr-gwotjmbs&cof=FORID\%3A10&ie=UTF-8&q=golden&siteurl=www.wikihow.com\%2FMain-Page, 1281-http://www.wikihow.com/Special:GoogSearch?cx=008953293426798287586\%3Amr-gwotjmbs&cof=FORID\%3A10&ie=UTF-8&q=ray+lewis&siteurl=www.wikihow.com\%2FMain-Page, 1282-http://www.wikihow.com/Special:GoogSearch?cx=008953293426798287586\%3Amr-gwotjmbs&cof=FORID\%3A10&ie=UTF-8&q=stone+mountain&siteurl=www.wikihow.com\%2FMain-Page, 1283-http://www.wikihow.com/Special:GoogSearch?cx=008953293426798287586\%3Amr-gwotjmbs&cof=FORID\%3A10&ie=UTF-8&q=mcdonalds&siteurl=www.wikihow.com\%2FMain-Page, 1284-http://www.google.com/cse/m?cx=008953293426798287586\%3Amr-gwotjmbs&ie=UTF-8&q=golden, 1285-http://www.google.com/cse/m?cx=008953293426798287586\%3Amr-gwotjmbs&ie=UTF-8&q=ray+lewis, 1286-http://www.google.com/cse/m?cx=008953293426798287586\%3Amr-gwotjmbs&ie=UTF-8&q=stone+mountain, 1287-http://www.google.com/cse/m?cx=008953293426798287586\%3Amr-gwotjmbs&ie=UTF-8&q=mcdonalds, 1288-http://m.wikihow.com/Write-a-Novel, 1289-http://www.wikihow.com/Write-a-Novel, 1290-http://www.wikihow.com/Accessorize-a-Polka-Dot-Dress, 1291-http://m.wikihow.com/Accessorize-a-Polka-Dot-Dress, 1292-http://www.wikihow.com/Cook-a-Sweet-Potato-in-the-Microwave, 1293-http://m.wikihow.com/Cook-a-Sweet-Potato-in-the-Microwave, 1294-http://www.wikihow.com/Find-Water-in-an-Emergency, 1295-http://m.wikihow.com/Find-Water-in-an-Emergency, 1296-http://www.retailmenot.com/, 1297-http://www.retailmenot.com/s/golden, 1298-http://www.retailmenot.com/s/ray+lewis, 1299-http://www.retailmenot.com/s/stone+mountain, 1300-http://www.retailmenot.com/s/mcdonalds, 1301-http://www.retailmenot.com/view/kohls.com, 1302-http://www.retailmenot.com/view/godaddy.com, 1303-http://www.retailmenot.com/view/us.asos.com, 1304-http://www.retailmenot.

com/view/bestbuy.com, 1305-http://m.slickdeals.net/, 1306-http://slickdeals.net/, 1307-http://slickdeals.net/newsearch.php?forumchoice\%5B\%5D=4&forumchoice\%5B\%5D=9&forumchoice\%5B\%5D=10&forumchoice\%5B\%5D=13&forumchoice\%5B\%5D=25&forumchoice\%5B\%5D=30&forumchoice\%5B\%5D=38&forumchoice\%5B\%5D=39&forumchoice\%5B\%5D=41&forumchoice\%5B\%5D=44&forumchoice\%5B\%5D=53&forumchoice\%5B\%5D=54&q=golden&showposts=0&archive=0&firstonly=1, 1308-http://slickdeals.net/newsearch.php?forumchoice\%5B\%5D=4&forumchoice\%5B\%5D=9&forumchoice\%5B\%5D=10&forumchoice\%5B\%5D=13&forumchoice\%5B\%5D=25&forumchoice\%5B\%5D=30&forumchoice\%5B\%5D=38&forumchoice\%5B\%5D=39&forumchoice\%5B\%5D=41&forumchoice\%5B\%5D=44&forumchoice\%5B\%5D=53&forumchoice\%5B\%5D=54&q=ray+lewis&showposts=0&archive=0&firstonly=1, 1309-http://slickdeals.net/newsearch.php?forumchoice\%5B\%5D=4&forumchoice\%5B\%5D=9&forumchoice\%5B\%5D=10&forumchoice\%5B\%5D=13&forumchoice\%5B\%5D=25&forumchoice\%5B\%5D=30&forumchoice\%5B\%5D=38&forumchoice\%5B\%5D=39&forumchoice\%5B\%5D=41&forumchoice\%5B\%5D=44&forumchoice\%5B\%5D=53&forumchoice\%5B\%5D=54&q=stone+mountain&showposts=0&archive=0&firstonly=1, 1310-http://slickdeals.net/newsearch.php?forumchoice\%5B\%5D=4&forumchoice\%5B\%5D=9&forumchoice\%5B\%5D=10&forumchoice\%5B\%5D=13&forumchoice\%5B\%5D=25&forumchoice\%5B\%5D=30&forumchoice\%5B\%5D=38&forumchoice\%5B\%5D=39&forumchoice\%5B\%5D=41&forumchoice\%5B\%5D=44&forumchoice\%5B\%5D=53&forumchoice\%5B\%5D=54&q=mcdonalds&showposts=0&archive=0&firstonly=1, 1311-http://m.slickdeals.net/newsearch.php?q=golden, 1312-http://m.slickdeals.net/newsearch.php?q=ray+lewis, 1313-http://m.slickdeals.net/newsearch.php?q=stone+mountain, 1314-http://m.slickdeals.net/newsearch.php?q=mcdonalds, 1315-http://m.slickdeals.net/f/6136248-10-Rolls-of-Kleenex-Cottonelle-2-Ply-Toilet-Paper-1-Free-Ship-to-Store, 1316-http://slickdeals.net/permadeal/98476/staples-10-rolls-of-kleenex-cottonelle-2ply-toilet-paper, 1317-http://m.slickdeals.net/f/6135978-Olay-Products-2-pack-of-15-2-oz-Quench-In-Shower-Body-Lotion-5-60-2-pack-of-23-6-oz-Cleansing-Body-Wash-7-30-2-pack-of-12-oz-Cleansing-Body-Wash-4-50-amp-More-Free-Shipping, 1318-http://slickdeals.net/permadeal/98442/amazon-olay-products-2pack-of-15.2oz-quench-inshower-body-lotion-5.60-2pack-of-23.6oz-cleansing-body-wash-7.30-2pack-of-12oz-cleansing-body-wash, 1319-http://m.slickdeals.net/f/6136496-Magic-School-Bus-The-Complete-Series-DVD-29-Free-Shipping, 1320-http://slickdeals.net/permadeal/98462/amazon-magic-school-bus-the-complete-series-dvd, 1321-http://slickdeals.net/permadeal/98466/amazon-documentary-blurays-samsara-or-baraka, 1322-http://m.slickdeals.net/f/6136620-Documentary-Blu-rays-Samsara-or-Baraka-13-each, 1323-http://sweetpacks.com/, 1324-http://sweetpacks.com/?s=golden&lang=en, 1325-http://sweetpacks.com/?s=ray+lewis&lang=en, 1326-http://sweetpacks.com/?s=stone+mountain&lang=en, 1327-http://sweetpacks.com/?s=mcdonalds&lang=en, 1328-

http://sweetpacks.com/about-us/, 1329-http://sweetpacks.com/about-us/management/, 1330-http://sweetpacks.com/career/, 1331-http://sweetpacks.com/contact-us/, 1332-https://soundcloud.com/, 1333-http://m.soundcloud.com/, 1334-https://soundcloud.com/search?q=golden, 1335-http://m.soundcloud.com/tracks/search?q=golden, 1336-http://m.soundcloud.com/tracks/search?q=ray+lewis, 1337-http://m.soundcloud.com/tracks/search?q=stone+mountain, 1338-http://m.soundcloud.com/tracks/search?q=mcdonalds, 1339-https://soundcloud.com/search?q=ray+lewis, 1340-https://soundcloud.com/search?q=stone+mountain, 1341-https://soundcloud.com/search?q=mcdonalds, 1342-https://soundcloud.com/kaytranada/jill-scott-golden-kaytranadas, 1343-http://m.soundcloud.com/kaytranada/jill-scott-golden-kaytranadas, 1344-http://m.soundcloud.com/damon-albarn-official/the-golden-dawn-clip, 1345-https://soundcloud.com/damon-albarn-official, 1346-https://soundcloud.com/skrillex, 1347-http://m.soundcloud.com/skrillex/golden-mummy-golden-bird-htb-vs-skrillex, 1348-http://m.soundcloud.com/yoshiki/golden-globe-theme-1-minute, 1349-https://soundcloud.com/yoshiki/golden-globe-theme-1-minute, 1350-http://thepiratebay.sx/, 1351-http://thepiratebay.sx/s/?q=golden&page=0&orderby=99, 1352-http://thepiratebay.sx/search/ray+lewis/0/99/0, 1353-http://thepiratebay.sx/search/stone+mountain/0/99/0, 1354-http://thepiratebay.sx/search/mcdonalds/0/99/0, 1355-http://thepiratebay.sx/torrent/8495002/Fats_and_Friends_-_with_Ray_Charles_and_Jerry_Lee_Lewis, 1356-http://thepiratebay.sx/torrent/8481620/Macklemore__amp__Ryan_Lewis_-_Can_t_Hold_Us_feat._Ray_Dalton.wav, 1357-http://thepiratebay.sx/torrent/8432463/Mackelmore__amp__Ryan_Lewis_-_Cant_t_Hold_Us_feat._Ray_Dalton_48, 1358-http://thepiratebay.sx/torrent/7689571/A_Football_Life_-_Ray_Lewis, 1359-http://www.legacy.com/NS/, 1360-http://www.legacy.com/memorial-sites/2013/, 1361-http://www.legacy.com/memorial-sites/arizona-firefighters/, 1362-http://www.legacy.com/ns/obituary.aspx?n=oliver-red-cloud&pid=165683485, 1363-http://www.legacy.com/memorial-sites/wars-in-iraq-and-afghanistan/, 1364-http://www.ign.com/, 1365-http://m.ign.com/, 1366-http://www.ign.com/search?q=golden, 1367-http://www.ign.com/search?q=ray+lewis, 1368-http://www.ign.com/search?q=stone+mountain, 1369-http://www.ign.com/search?q=mcdonalds, 1370-http://m.ign.com/search/product?q=golden, 1371-http://m.ign.com/search/product?q=ray+lewis, 1372-http://m.ign.com/search/product?q=stone+mountain, 1373-http://m.ign.com/search/product?q=mcdonalds, 1374-http://www.ign.com/articles/2013/07/08/dualshock-4-light-bar-is-always-on, 1375-http://m.ign.com/articles/2013/07/08/dualshock-4-light-bar-is-always-on, 1376-http://www.ign.com/articles/2013/07/08/russell-crowe-keen-on-man-of-steel-prequel?abthid=51da9a5a8a6b983431000002, 1377-http://m.ign.com/articles/2013/07/08/russell-crowe-keen-on-man-of-steel-prequel, 1378-http://m.ign.com/articles/2013/07/08/you-dont-need-dlc-for-the-wonderful-101, 1379-http://www.ign.com/articles/2013/07/08/you-dont-need-dlc-for-the-wonderful-101?abthid=51daa4d1a329253631000006, 1380-http://m.ign.com/articles/2013/07/08/the-10-most-heartwarming-moments-in-comic-book-movies, 1381-http://www.ign.com/articles/2013/07/08/the-10-most-heartwarming-moments-in-comic-book-movies, 1382-http://office.microsoft.com/en-us/business/office-365-enterprise-e3-business-software-FX103030346.aspx, 1383-http://office.microsoft.com/en-us/results.aspx?qu=golden&ex=2, 1384-http://office.microsoft.com/en-us/results.aspx?qu=ray+lewis&ex=2, 1385-http://office.microsoft.com/en-us/results.aspx?qu=stone+mountain&ex=2, 1386-http://office.microsoft.com/en-us/results.aspx?qu=mcdonalds&ex=2, 1387-http://office.microsoft.com/en-us/office-365-small-business-premium-office-online-FX103037625.aspx?WT\%2Eintid1=ODC\%5FENUS\%5FFX101825692\%5FXT104041502, 1388-http://office.microsoft.com/en-us/business/office-365-customer-stories-office-testimonials-FX103045622.aspx, 1389-http://office.microsoft.com/en-us/business/what-is-office-365-for-business-FX102997580.aspx, 1390-http://www.kohls.com/, 1391-http://www.kohls.com/product/prd-1399431/artcom-ray-lewis-2010-action-framed-art-print.jsp, 1392-http://www.kohls.com/product/prd-1443964/chef-buddy-over-the-sink-cutting-board.jsp?crosssell=true, 1393-http://www.kohls.com/product/prd-c20181/500-thread-count-deep-fitted-egyptian-cotton-sheet-set.jsp, 1394-http://www.kohls.com/product/prd-1004134/homevance-window-back-swivel-bar-stool.jsp, 1395-http://www.irs.gov/, 1396-http://search.irs.gov/search?q=golden&output=xml_no_dtd&proxystylesheet=irs_portals_frontend&client=irs_portals_frontend&oe=UTF-8&ie=UTF-8&num=10&ud=1&exclude_apps=1&site=default_collection&numgm=5&requiredfields=-archive\%3A1, 1397-http://search.irs.gov/search?q=ray+lewis&output=xml_no_dtd&proxystylesheet=irs_portals_frontend&client=irs_portals_frontend&oe=UTF-8&ie=UTF-8&num=10&ud=1&exclude_apps=1&site=default_collection&numgm=5&requiredfields=-archive\%3A1, 1398-http://search.irs.gov/search?q=stone+mountain&output=xml_no_dtd&proxystylesheet=irs_portals_frontend&client=irs_portals_frontend&oe=UTF-8&ie=UTF-8&num=10&ud=1&exclude_apps=1&site=default_collection&numgm=5&requiredfields=-archive\%3A1, 1399-http://search.irs.gov/search?q=mcdonalds&output=xml_no_dtd&proxystylesheet=irs_portals_frontend&client=irs_portals_frontend&oe=UTF-8&ie=UTF-8&num=10&ud=1&exclude_apps=1&site=default_collection&numgm=5&requiredfields=-archive\%3A1, 1400-http://www.irs.gov/uac/Newsroom/IRS-Statement-on-the-Supreme-Court-Decision-on-the-Defense-of-Marriage-Act, 1401-http://www.irs.gov/uac/Newsroom/Questions-and-Answers-on-501(c)-Organizations, 1402-http://www.irs.gov/uac/Newsroom/Penalty-Relief-Available-to-Some-Storm-Victims-Unable-To-File-On-Time, 1403-http://www.irs.gov/uac/IRS-Guidance, 1404-http://www.surveymonkey.com/, 1405-http://www.examiner.com/, 1406-http://www.examiner.com/search/google?query=golden&cx=partner-pub-7479725245717969\%3A9ze01gmnpyp&cof=

FORID\%3A9&ie=ISO-8859-1&sa=Search, 1407-http://www.examiner.com/search/google?query=ray+lewis&cx=partner-pub-7479725245717969\%3A9ze01gmnpyp&cof=FORID\%3A9&ie=ISO-8859-1&sa=Search, 1408-http://www.examiner.com/search/google?query=mcdonalds&cx=partner-pub-7479725245717969\%3A9ze01gmnpyp&cof=FORID\%3A9&ie=ISO-8859-1&sa=Search, 1409-http://www.examiner.com/search/google?query=stone+mountain&cx=partner-pub-7479725245717969\%3A9ze01gmnpyp&cof=FORID\%3A9&ie=ISO-8859-1&sa=Search, 1410-http://www.examiner.com/article/asiana-airlines-boeing-777-crash-lands-at-san-francisco-airport, 1411-http://www.examiner.com/article/wimbledon-2013-schedule-live-stream-and-tv-coverage-for-men-s-tennis-final-7-7, 1412-http://www.examiner.com/list/nfl-jailhouse-rock?cid=PROG-List-HomepageFeatured2-NFLJail17-070513, 1413-http://www.examiner.com/article/teresa-heinz-kerry-critically-ill-secretary-of-state-john-kerry-at-her-side, 1414-http://m.allrecipes.com/, 1415-http://allrecipes.com/, 1416-http://allrecipes.com/search/default.aspx?qt=k&wt=golden&rt=r&origin=Home\%20Page, 1417-http://allrecipes.com/search/default.aspx?qt=k&wt=ray+lewis&rt=r&origin=Home\%20Page, 1418-http://allrecipes.com/search/default.aspx?qt=k&wt=stone+mountain&rt=r&origin=Home\%20Page, 1419-http://allrecipes.com/search/default.aspx?qt=k&wt=mcdonalds&rt=r&origin=Home\%20Page, 1420-http://m.allrecipes.com/search/recipes?wt=golden&sort=, 1421-http://m.allrecipes.com/search/recipes?wt=ray+lewis&sort=, 1422-http://m.allrecipes.com/search/recipes?wt=stone+mountain&sort=, 1423-http://m.allrecipes.com/search/recipes?wt=mcdonalds&sort=, 1424-http://allrecipes.com/Recipe/Golden-Knots/Detail.aspx?event8=1&prop24=SR_Thumb&e11=golden&e8=Quick\%20Search&event10=1&e7=Home\%20Page, 1425-http://m.allrecipes.com/recipe/22508/golden-knots, 1426-http://allrecipes.com/Recipe/Golden-Cakes/Detail.aspx?event8=1&prop24=SR_Title&e11=golden&e8=Quick\%20Search&event10=1&e7=Home\%20Page, 1427-http://m.allrecipes.com/recipe/23636/golden-cakes, 1428-http://m.allrecipes.com/recipe/8925/golden-lasagna, 1429-http://allrecipes.com/Recipe/Golden-Lasagna/Detail.aspx?event8=1&prop24=SR_Title&e11=golden&e8=Quick\%20Search&event10=1&e7=Home\%20Page, 1430-http://allrecipes.com/Recipe/Golden-Sweet-Cornbread/Detail.aspx?event8=1&prop24=SR_Title&e11=golden&e8=Quick\%20Search&event10=1&e7=Home\%20Page, 1431-http://m.allrecipes.com/recipe/17891/golden-sweet-cornbread, 1432-http://www.meetup.com/find/, 1433-http://www.meetup.com/find/?keywords=golden&radius=5&userFreeform=Chapel+Hill\%2C+North+Carolina\%2C+USA&mcId=z27514&mcName=Chapel+Hill\%2C+NC&sort=default, 1434-http://www.meetup.com/find/?keywords=ray+lewis&radius=5&userFreeform=Chapel+Hill\%2C+North+Carolina\%2C+USA&mcId=z27514&mcName=Chapel+Hill\%2C+NC&sort=default, 1435-http://www.meetup.com/find/?keywords=stone+mountain&radius=5&userFreeform=Chapel+Hill\%2C+North+Carolina\%2C+USA&mcId=z27514&mcName=Chapel+Hill\%2C+NC&sort=default, 1436-http://www.meetup.com/find/?keywords=mcdonalds&radius=5&userFreeform=Chapel+

Hill\%2C+North+Carolina\%2C+USA&mcId=z27514&mcName=Chapel+Hill\%2C+NC&sort=default, 1437-http://www.pornhublive.com/, 1438-http://www.pornhublive.com/search.php?q=golden&submit=Search, 1439-http://www.pornhublive.com/search.php?q=ray+lewis&submit=Search, 1440-http://www.pornhublive.com/search.php?q=stone+mountain&submit=Search, 1441-http://www.pornhublive.com/search.php?q=mcdonalds&submit=Search, 1442-https://www.stumbleupon.com/, 1443-http://www.stumbleupon.com/interest/mobile-website, 1444-http://www.mozilla.org/en-US/, 1445-http://www.mozilla.org/en-US/about/, 1446-http://www.mozilla.org/en-US/products/, 1447-http://www.mozilla.org/en-US/contribute/, 1448-http://www.mozilla.org/en-US/mission/, 1449-http://www.cbs.com/, 1450-http://www.cbs.com/sitesearch/results/?q=golden, 1451-http://www.cbs.com/sitesearch/results/?q=ray+lewis, 1452-http://www.cbs.com/sitesearch/results/?q=stone+mountain, 1453-http://www.cbs.com/sitesearch/results/?q=mcdonalds, 1454-http://www.cbs.com/shows/2_broke_girls/, 1455-http://www.cbs.com/shows/48_hours/, 1456-http://www.cbs.com/shows/amazing_race/, 1457-http://www.cbs.com/shows/big_brother/, 1458-http://www.theblaze.com/, 1459-http://www.theblaze.com/stories/2013/07/08/third-amendment-violated-nev-police-allegedly-invade-familys-home-to-use-during-swat-call-arrest-two-for-obstruction-when-owner-refuses/, 1460-http://www.theblaze.com/stories/2013/07/08/zimmerman-trial-trayvon-martins-dad-answered-no-when-asked-if-screams-on-911-call-belonged-to-his-son-detective-testifies/, 1461-http://www.theblaze.com/stories/2013/07/08/shock-report-female-prisoners-were-sterilized-in-calif-prisons-without-state-approval/, 1462-http://www.theblaze.com/stories/2013/07/08/secret-move-keeps-osama-bin-laden-records-hidden-from-public-view-but-why/, 1463-https://www.adcash.com/en/index.php, 1464-https://www.adcash.com/en/publishers.php, 1465-https://www.adcash.com/en/advertisers.php, 1466-https://www.adcash.com/en/contact.php, 1467-https://www.adcash.com/en/campaign.php, 1468-http://www.whitepages.com/, 1469-http://www.whitepages.com/business?key=golden&where=30319, 1470-http://www.whitepages.com/business?key=ray+lewis&where=30319, 1471-http://www.whitepages.com/business?key=stone+mountain&where=30319, 1472-http://www.whitepages.com/business?key=mcdonalds&where=30319, 1473-http://www.whitepages.com/business/golden-livingcenter-northside-atlanta-ga, 1474-http://www.whitepages.com/business/details?uid=C2303984, 1475-http://www.whitepages.com/business/details?uid=C454068395, 1476-http://www.whitepages.com/business/details?uid=C2982900, 1477-https://www.usbank.com/index.html, 1478-http://m.usbank.com/mobile-web/index.asp?msg=, 1479-https://wwws.usbank.com/search/default2.asp?ui_mode=question&charset=UTF-8&language=en&restriction.level.collections=Collections.FullUnsecuredUSbank&question_box=golden, 1480-https://wwws.usbank.com/search/default2.asp?ui_mode=question&charset=UTF-8&language=en&restriction.level.collections=Collections.FullUnsecuredUSbank&question_box=ray+

lewis, 1481-https://wwws.usbank.com/search/default2.asp?ui_mode=question&charset=UTF-8&language=en&restriction.level.collections=Collections.FullUnsecuredUSbank&question_box=stone+mountain, 1482-https://wwws.usbank.com/search/default2.asp?ui_mode=question&charset=UTF-8&language=en&restriction.level.collections=Collections.FullUnsecuredUSbank&question_box=mcdonalds, 1483-https://www.usbank.com/en/AboutHome.cfm, 1484-http://m.usbank.com/mobile-web/about-us-nt.html, 1485-https://mm.usbank.com/b/LocatorSearch/FindATMBranch.aspx, 1486-https://www.usbank.com/locations/, 1487-https://www.usbank.com/privacy/index.html, 1488-https://mm.usbank.com/b/More/PrivacyOverview.aspx, 1489-https://www.fidelity.com/, 1490-http://www.fidelity.mobi/fiw/FiwHome, 1491-https://search.fidelity.com/search/getSearchResults?question=golden, 1492-https://search.fidelity.com/search/getSearchResults?question=ray+lewis, 1493-https://search.fidelity.com/search/getSearchResults?question=stone+mountain, 1494-https://search.fidelity.com/search/getSearchResults?question=mcdonalds, 1495-http://activequote.fidelity.com/webxpress/get_quote?bar=p, 1496-http://www.fidelity.mobi/fiw/QuotesMain;jsessionid=0000Yi97gJMix33ID0RGHd1WNI9:-1?__JWTS__=0, 1497-https://login.fidelity.com/ftgw/Fidelity/RtlCust/Login/Init/df.chf.ra/Summary?AuthRedUrl=https://scs.fidelity.com/customeronly/portfolio.shtml?bar=p, 1498-http://www.fidelity.mobi/fiw/BrokerageHome?__JWTS__=7, 1499-http://www.addthis.com/, 1500-https://www.addthis.com/get/sharing, 1501-http://www.addthis.com/advertising, 1502-http://www.addthis.com/data#.UdtfLhztUcM, 1503-http://support.addthis.com/, 1504-http://support.addthis.com/customer/portal/articles/search?q=golden, 1505-http://support.addthis.com/customer/portal/articles/search?q=ray+lewis, 1506-http://support.addthis.com/customer/portal/articles/search?q=stone+mountain, 1507-http://support.addthis.com/customer/portal/articles/search?q=mcdonalds, 1508-http://www.woot.com/, 1509-http://www.woot.com/offers/vizio-29-720p-led-hdtv-9?utm_expid=31924516-17&utm_referrer=http\%3A\%2F\%2Fwww.woot.com\%2F, 1510-http://home.woot.com/offers/5-pc-comforter-set-3-sizes-6-colors, 1511-http://sellout.woot.com/offers/3m-led-mobile-pocket-projector-2, 1512-http://tools.woot.com/offers/worx-24v-trimmer-edger-and-blower-combo, 1513-http://www.rr.com/, 1514-http://m.rr.com/, 1515-http://search.rr.com/#web/golden/1/, 1516-http://search.rr.com/#web/ray+lewis/1/, 1517-http://search.rr.com/#web/stone+mountain/1/, 1518-http://search.rr.com/#web/mcdonalds/1/, 1519-http://search.rr.com/#rrimage/ray\%20lewis/1/, 1520-http://search.rr.com/#rrimage/stone\%20mountain/1/, 1521-http://search.rr.com/#rrimage/mcdonalds/1/, 1522-http://search.rr.com/#rrimage/golden/1/, 1523-http://m.rr.com/rss.jsp;jsessionid=7ECF9B4389D01EB69F62A7E03FA7EB64.sonny2?rssid=25524411&item=http\%3a\%2f\%2fwww.rr.com\%2fservices\%2fpublicapi\%2fcontent\%2f1.0\%2f\%3fmethod\%3dgetMobileHeadlinesRss\

%26ci\%3d87499350\%26csi\%3d55255142&cid=25372441, 1524-http://www.rr.com/news/topic/article/rr/55255142/87499350/Captain_of_wrecked_cruise_ship_on_trial_in_Italy, 1525-http://www.rr.com/news/topic/article/rr/55255142/87504037/Explosion_rocks_Hezbollah_stronghold_in_Lebanon, 1526-http://m.rr.com/rss.jsp?rssid=25524411&item=http\%3a\%2f\%2fwww.rr.com\%2fservices\%2fpublicapi\%2fcontent\%2f1.0\%2f\%3fmethod\%3dgetMobileHeadlinesRss\%26ci\%3d87504037\%26csi\%3d55255142&cid=25372441, 1527-http://www.rr.com/news/topic/article/rr/55255105/87504820/Pilot_interviews_key_to_answers_in_SFO_crash, 1528-http://m.rr.com/rss.jsp?rssid=25524411&item=http\%3a\%2f\%2fwww.rr.com\%2fservices\%2fpublicapi\%2fcontent\%2f1.0\%2f\%3fmethod\%3dgetMobileHeadlinesRss\%26ci\%3d87504820\%26csi\%3d55255142&cid=25391891, 1529-http://m.rr.com/rss.jsp?rssid=25524411&item=http\%3a\%2f\%2fwww.rr.com\%2fservices\%2fpublicapi\%2fcontent\%2f1.0\%2f\%3fmethod\%3dgetMobileHeadlinesRss\%26ci\%3d87503287\%26csi\%3d55255142&cid=25391891, 1530-http://www.rr.com/news/topic/article/rr/55255105/87503287/911_calls_becoming_heart_of_Zimmerman_trial, 1531-http://www.nfl.com/, 1532-http://search.nfl.com/search?query=golden, 1533-http://search.nfl.com/search?query=ray+lewis, 1534-http://search.nfl.com/search?query=stone+mountain, 1535-http://search.nfl.com/search?query=mcdonalds, 1536-http://www.nfl.com/news/story/0ap1000000216825/article/robert-kraft-patriots-duped-by-aaron-hernandez, 1537-http://www.nfl.com/news/story/0ap1000000216712/article/victor-cruz-new-york-giants-strike-45879m-contract, 1538-http://www.nfl.com/news/story/0ap1000000216674/article/six-players-eligible-for-2013-nfl-supplemental-draft, 1539-http://www.nfl.com/news/story/0ap1000000216823/article/denver-broncos-exec-matt-russell-charged-with-dui, 1540-http://www.accuweather.com/, 1541-http://m.accuweather.com/, 1542-http://m.accuweather.com/en/us/chapel-hill-nc/27517/weather-forecast/11338_pc, 1543-http://www.accuweather.com/en/us/chapel-hill-nc/27516/weather-forecast/329826, 1544-http://m.accuweather.com/en/us/hampton-va/23669/weather-forecast/331251, 1545-http://www.accuweather.com/en/us/hampton-va/23669/weather-forecast/331251, 1546-http://m.accuweather.com/en/us/miami-fl/33130/weather-forecast/347936, 1547-http://www.accuweather.com/en/us/miami-fl/33128/weather-forecast/347936, 1548-http://m.accuweather.com/en/us/hilton-head-island-sc/29926/weather-forecast/340557, 1549-http://www.accuweather.com/en/us/hilton-head-island-sc/29926/weather-forecast/340557, 1550-http://www.reuters.com/, 1551-http://us.mobile.reuters.com/, 1552-http://www.reuters.com/article/2013/07/10/us-tribune-division-idUSBRE9690C620130710, 1553-http://us.mobile.reuters.com/shortArticle/topNews/idUSBRE9690C620130710, 1554-http://us.mobile.reuters.com/shortArticle/topNews/idUSBRE95Q0NO20130710, 1555-http://www.reuters.com/article/2013/07/10/us-egypt-protests-idUSBRE95Q0NO20130710, 1556-http://us.mobile.reuters.com/shortArticle/politicsNews/idUSBRE96815620130709, 1557-http://www.reuters.com/article/2013/07/09/us-usa-healthcare-republicans-idUSBRE96815620130709, 1558-http://www.reuters.com/article/2013/07/09/us-afghanistan-usa-troops-idUSBRE96815C20130709, 1559-http://us.mobile.reuters.com/shortArticle/politicsNews/idUSBRE96815C20130709, 1560-http://www.today.com/, 1561-http://www.today.com/money/golden-corral-buffet-chain-responds-gross-out-footage-6C10574630, 1562-http://www.today.com/money/weak-economy-means-fewer-babies-least-now-6C10471907, 1563-http://www.today.com/money/dreams-delayed-or-denied-young-adults-put-parenthood-6C10528964, 1564-http://www.today.com/money/getting-sick-doesnt-pay-many-us-workers-6C10565447, 1565-http://www.coupons.com/, 1566-http://www.coupons.com/store-loyalty-card-coupons/, 1567-http://www.coupons.com/local-offers/, 1568-http://www.coupons.com/coupon-codes/, 1569-http://www.coupons.com/coupons/Food-Coupons-107/, 1570-http://m.okcupid.com/, 1571-http://www.okcupid.com/, 1572-https://m.livingsocial.com/, 1573-https://www.livingsocial.com/, 1574-https://m.livingsocial.com/cities/331/deals/750392, 1575-https://m.livingsocial.com/cities/331/deals/747620, 1576-https://m.livingsocial.com/cities/1919/deals/750876, 1577-https://m.livingsocial.com/cities/331/deals/758880, 1578-http://www.people.com/people/, 1579-http://www.people.com/people/mobile/home/, 1580-http://www.people.com/people/article/02071651200.html, 1581-http://www.people.com/people/mobile/article/02071651200.html, 1582-http://search.people.com/results.html?search=golden&bu=&searchSubmit=Go, 1583-http://search.people.com/results.html?search=ray+lewis&bu=&searchSubmit=Go, 1584-http://search.people.com/results.html?search=stone+mountain&bu=&searchSubmit=Go, 1585-http://search.people.com/results.html?search=mcdonalds&bu=&searchSubmit=Go, 1586-http://www.people.com/people/mobile/article/02071651300.html, 1587-http://www.people.com/people/article/02071651300.html, 1588-http://www.people.com/people/article/02071638800.html, 1589-http://www.people.com/people/mobile/article/02071638800.html, 1590-http://vube.com/, 1591-http://vube.com/Golden+Empress/0ScnwgY6CJ?t=s, 1592-http://vube.com/Golden\%20Empress/SZM3m6bkxG?t=s, 1593-http://vube.com/PASTORSEXWIFE/3R66d3gsey?t=s, 1594-http://vube.com/Free+Naked+Cam+Sex/I4vAEVpe0M?t=s, 1595-http://www.jackhenrybanking.com/products/InternetBanking/NetTeller, 1596-http://www.jackhenrybanking.com/?S=golden, 1597-http://www.jackhenrybanking.com/?S=ray\%20lewis, 1598-http://www.jackhenrybanking.com/?S=stone\%20mountain, 1599-http://www.jackhenrybanking.com/?S=mcdonalds, 1600-http://www.nih.gov/, 1601-http://search.nih.gov/search?utf8=?&affiliate=nih&query=golden&commit.x=0&commit.y=0&commit=Search, 1602-http://search.nih.gov/search?utf8=?&affiliate=nih&query=ray+lewis&commit.x=0&commit.y=0&commit=Search, 1603-http://search.nih.gov/search?utf8=?&affiliate=nih&query=stone+mountain&commit.x=0&commit.y=0&commit=Search, 1604-http://search.nih.gov/search?utf8=?&affiliate=nih&query=mcdonalds&commit.x=0&commit.y=0&commit=Search, 1605-http://www.nih.gov/about/director/

09272012_celebrationofscience.htm, 1606-http://www.nih.gov/news/health/jun2013/nichd-25.htm, 1607-http://BRAINfeedback.nih.gov/, 1608-http://www.nih.gov/about/impact/index.htm, 1609-http://att.yahoo.com/, 1610-http://us.yhs4.search.yahoo.com/yhs/search?p=ray+lewis&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1611-http://us.yhs4.search.yahoo.com/yhs/search?p=golden&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1612-http://us.yhs4.search.yahoo.com/yhs/search?p=mcdonalds&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1613-http://us.yhs4.search.yahoo.com/yhs/search?p=stone+mountain&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1614-http://images.search.yahoo.com/yhs/search;_ylt=A0oG7pAKbd9RokUA1Q8PxQt.?p=ray+lewis&fr=&fr2=piv-web&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1615-http://images.search.yahoo.com/yhs/search;_ylt=A0oG7pAKbd9RokUA1Q8PxQt.?p=golden&fr=&fr2=piv-web&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1616-http://images.search.yahoo.com/yhs/search;_ylt=A0oG7pAKbd9RokUA1Q8PxQt.?p=stone+mountain&fr=&fr2=piv-web&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1617-http://images.search.yahoo.com/yhs/search;_ylt=A0oG7pAKbd9RokUA1Q8PxQt.?p=mcdonalds&fr=&fr2=piv-web&hspart=att&hsimp=yhs-att_001&type=att_lego_portal_home, 1618-http://www.worldstarhiphop.com/videos/, 1619-http://m.gap.com/, 1620-http://www.gap.com/, 1621-http://m.gap.com/category.html?search=golden&pdn=, 1622-http://m.gap.com/category.html?search=ray+lewis&pdn=, 1623-http://m.gap.com/category.html?search=stone+mountain&pdn=, 1624-http://m.gap.com/category.html?search=mcdonalds&pdn=, 1625-http://www.gap.com/browse/search.do?searchText=ray+lewis, 1626-http://www.gap.com/browse/search.do?searchText=golden, 1627-http://www.gap.com/browse/search.do?searchText=stone+mountain, 1628-http://www.gap.com/browse/search.do?searchText=mcdonalds, 1629-http://m.gap.com/product.html?dn=gp289760352&pdn=search, 1630-http://www.gap.com/browse/product.do?vid=1&pid=289760352, 1631-http://m.gap.com/product.html?dn=gp425280042&pdn=search, 1632-http://www.gap.com/browse/product.do?vid=5&pid=425280042, 1633-http://m.gap.com/product.html?dn=gp432934042&pdn=search, 1634-http://www.gap.com/browse/product.do?vid=1&pid=432934042, 1635-http://www.gap.com/browse/product.do?vid=1&pid=579177012, 1636-http://m.gap.com/product.html?dn=gp579177012&pdn=search, 1637-http://www.overstock.com/, 1638-http://www.overstock.com/search?keywords=golden&SearchType=Header, 1639-http://www.overstock.com/search?keywords=ray+lewis&SearchType=Header, 1640-http://www.overstock.com/search?keywords=stone+mountain&SearchType=Header, 1641-http://www.overstock.com/search?keywords=mcdonalds&SearchType=Header, 1642-http://www.overstock.com/Sports-Toys/Baltimore-Ravens-Ray-Lewis-9x12-Photo-Plaque/3825592/product.html?refccid=EI7A4SZYIDPJFE2OMHZDOB36MI&searchidx=0, 1643-http://www.overstock.com/Sports-Toys/University-of-North-Carolina-2009-National-Champions-Plaque/

4488041/product.html?rcmndsrc=2, 1644-http://www.overstock.com/Sports-Toys/Encore-Select-2010-NBA-Champions-LA-Lakers-Stat-Plaque-12x15/5888783/product.html?rcmndsrc=2, 1645-http://www.overstock.com/Books-Movies-Music-Games/Wii-Black-Console-with-Wii-Sports-Wii-Sports-Resort/7389195/product.html?searchidx=0, 1646-http://www.wunderground.com/, 1647-http://m.wund.com/, 1648-http://m.wund.com/cgi-bin/findweather/getForecast?brand=mobile&query=30319, 1649-http://www.wunderground.com/cgi-bin/findweather/hdfForecast?query=30319, 1650-http://www.wunderground.com/cgi-bin/findweather/getForecast?query=25517, 1651-http://m.wund.com/cgi-bin/findweather/getForecast?brand=mobile&query=25517, 1652-http://www.wunderground.com/cgi-bin/findweather/getForecast?query=27516, 1653-http://m.wund.com/cgi-bin/findweather/getForecast?brand=mobile&query=27516, 1654-http://www.wunderground.com/cgi-bin/findweather/getForecast?query=45672, 1655-http://m.wund.com/cgi-bin/findweather/getForecast?brand=mobile&query=45672, 1656-http://www.jcpenney.com/, 1657-http://m.jcpenney.com/mobile/index.jsp, 1658-http://www.jcpenney.com/dotcom/jsp/search/results.jsp?fromSearch=true&Ntt=golden&ruleZoneName=XGNSZone&grView=&_requestid=59470, 1659-http://www.jcpenney.com/dotcom/jsp/search/results.jsp?fromSearch=true&Ntt=ray+lewis&ruleZoneName=XGNSZone&grView=&_requestid=59470, 1660-http://www.jcpenney.com/dotcom/jsp/search/results.jsp?fromSearch=true&Ntt=stone+mountain&ruleZoneName=XGNSZone&grView=&_requestid=59470, 1661-http://www.jcpenney.com/dotcom/jsp/search/results.jsp?fromSearch=true&Ntt=mcdonalds&ruleZoneName=XGNSZone&grView=&_requestid=59470, 1662-http://m.jcpenney.com/mobile/jsp/browse/searchResults.jsp?fromSearch=true&Ntt=golden&ruleZoneName=XGNSZone&grView=null&_requestid=330237, 1663-http://m.jcpenney.com/mobile/jsp/browse/searchResults.jsp?fromSearch=true&Ntt=ray+lewis&ruleZoneName=XGNSZone&grView=null&_requestid=330237, 1664-http://m.jcpenney.com/mobile/jsp/browse/searchResults.jsp?fromSearch=true&Ntt=stone+mountain&ruleZoneName=XGNSZone&grView=null&_requestid=330237, 1665-http://m.jcpenney.com/mobile/jsp/browse/searchResults.jsp?fromSearch=true&Ntt=mcdonalds&ruleZoneName=XGNSZone&grView=null&_requestid=330237, 1666-http://m.jcpenney.com/mobile/jsp/browse/product.jsp?currIndx=1&Ntt=&subcat_Id=&dimCombo=&deprtId=&cateId=&ppId=pp5002860131&sub_catId=, 1667-http://www.jcpenney.com/dotcom/rachael-ray-porcelain-ii-12\%25c2\%25bd-open-skillet/prod.jump?ppId=pp5002860131&searchTerm=ray+lewis&dimCombo=null&dimComboVal=null&catId=SearchResults, 1668-http://www.jcpenney.com/dotcom/everyday-prices/premium-weight-yoga-mat/prod.jump?ppId=pp5002650111&searchTerm=weight&dimCombo=null&dimComboVal=null&catId=SearchResults, 1669-http://m.jcpenney.com/mobile/jsp/browse/product.jsp?currIndx=1&Ntt=weight&subcat_Id=&dimCombo=&deprtId=&cateId=&ppId=pp5002650111&sub_catId=, 1670-http://m.jcpenney.com/mobile/jsp/browse/product.jsp?currIndx=2&Ntt=weight&subcat_Id=&dimCombo=&deprtId=&cateId=&ppId=1bb8994&sub_catId=,

1671-http://www.jcpenney.com/dotcom/clearance/bed-bath/bath/weight-watchers-digital-scale/prod.jump?ppId=1bb8994&searchTerm=weight&dimCombo=null&dimComboVal=null&catId=SearchResults, 1672-http://m.jcpenney.com/mobile/jsp/browse/product.jsp?currIndx=8&Ntt=weight&subcat_Id=&dimCombo=&deprtId=&cateId=&ppId=pp5002691446&sub_catId=, 1673-http://www.jcpenney.com/dotcom/everyday-prices/body-by-jake-deluxe-weight-lifting-belt/prod.jump?ppId=pp5002691446&searchTerm=weight&dimCombo=null&dimComboVal=null&catId=SearchResults, 1674-http://www.thefreedictionary.com/, 1675-http://www.thefreedictionary.com/golden, 1676-http://encyclopedia.thefreedictionary.com/ray+lewis, 1677-http://encyclopedia.thefreedictionary.com/stone+mountain, 1678-http://encyclopedia.thefreedictionary.com/mcdonalds, 1679-http://www.concast.com/index_full.php, 1680-http://www.concast.com/site-search.php?cx=017040694683756441395\%3Avcobuueig5y&cof=FORID\%3A11&ie=UTF-8&q=golden&sa.x=0&sa.y=0, 1681-http://www.concast.com/site-search.php?cx=017040694683756441395\%3Avcobuueig5y&cof=FORID\%3A11&ie=UTF-8&q=ray+lewis&sa.x=0&sa.y=0, 1682-http://www.concast.com/site-search.php?cx=017040694683756441395\%3Avcobuueig5y&cof=FORID\%3A11&ie=UTF-8&q=stone+mountain&sa.x=0&sa.y=0, 1683-http://www.concast.com/site-search.php?cx=017040694683756441395\%3Avcobuueig5y&cof=FORID\%3A11&ie=UTF-8&q=mcdonalds&sa.x=0&sa.y=0, 1684-http://www.concast.com/c92200.php, 1685-http://www.concast.com/c83600.php, 1686-http://www.concast.com/c91100.php, 1687-http://www.concast.com/c95200.php, 1688-http://www.patch.com/, 1689-http://www.cbssports.com/, 1690-http://www.cbssports.com/info/search#q=golden, 1691-http://www.cbssports.com/info/search#q=ray\%20lewis, 1692-http://www.cbssports.com/info/search#q=stone\%20mountain, 1693-http://www.cbssports.com/info/search#q=mcdonalds, 1694-http://www.cbssports.com/nba/blog/eye-on-basketball/22744647/pat-riley-says-the-heat-wont-use-the-amnesty-provision, 1695-http://www.cbssports.com/nfl/blog/eye-on-football/22744066/police-aaron-hernandez-put-in-security-system-after-attempted-breakins, 1696-http://www.cbssports.com/nfl/blog/eye-on-football/22744066/police-aaron-hernandez-put-in-security-system-after-attempted-breakins, 1697-http://www.cbssports.com/nhl/blog/eye-on-hockey/22737447/nhls-second-tier-of-free-agents-still-offers-plenty-of-value, 1698-http://m.zappos.com/, 1699-http://www.zappos.com/, 1700-http://www.zappos.com/golden, 1701-http://m.zappos.com/golden, 1702-http://www.zappos.com/ray-lewis, 1703-http://m.zappos.com/ray-lewis, 1704-http://m.zappos.com/stone-mountain, 1705-http://m.zappos.com/mcdonalds, 1706-http://www.zappos.com/stone-mountain, 1707-http://www.zappos.com/mcdonalds, 1708-http://m.zappos.com/dkny-golden-delicious-gift-set-n-a, 1709-http://www.zappos.com/dkny-golden-delicious-gift-set-n-a?zfcTest=fcl\%3A0, 1710-http://www.zappos.com/dkny-dkny-golden-delicious-body-lotion-no-color?zfcTest=fcl\%3A0, 1711-http://m.zappos.com/dkny-dkny-golden-delicious-body-lotion-no-color, 1712-

http://www.zappos.com/bottega-veneta-silver-ring-argento-antico-lucid?zfcTest=fcl\%3A0, 1713-http://m.zappos.com/bottega-veneta-silver-ring-argento-antico-lucid, 1714-http://www.zappos.com/bottega-veneta-silver-bracelet-argento-antico-lucid?zfcTest=fcl\%3A0, 1715-http://m.zappos.com/bottega-veneta-silver-bracelet-argento-antico-lucid, 1716-http://www.roblox.com/Landing/Animated/, 1717-https://m.roblox.com/Login?ReturnUrl=\%2f, 1718 -http://www.download.com, 1719-http://www.incredibar.com/essentials/homepage, 1720-http://www.incredibar.com/music/homepage, 1721-http://www.incredibar.com/games/homepage, 1722-http://mobile.bloomberg.com/, 1723-http://www.bloomberg.com/, 1724-http://mobile.bloomberg.com/search/search?search=golden, 1725-http://search1.bloomberg.com/search/?content_type=all&page=1&q=golden, 1726-http://mobile.bloomberg.com/search/search?search=ray+lewis, 1727-http://search1.bloomberg.com/search/?content_type=all&page=1&q=ray+lewis, 1728-http://search1.bloomberg.com/search/?content_type=all&page=1&q=stone+mountain, 1729-http://mobile.bloomberg.com/search/search?search=stone+mountain, 1730-http://mobile.bloomberg.com/search/search?search=mcdonalds, 1731-http://search1.bloomberg.com/search/?content_type=all&page=1&q=mcdonalds, 1732-http://mobile.bloomberg.com/news/2013-07-11/flea-market-abortions-thrive-as-texas-may-close-clinics.html?cmpid=, 1733-http://www.bloomberg.com/news/2013-07-11/flea-market-abortions-thrive-as-texas-may-close-clinics.html, 1734-http://www.bloomberg.com/news/2013-07-12/u-s-stock-index-futures-little-changed-before-results.html, 1735-http://mobile.bloomberg.com/news/2013-07-12/u-s-stock-index-futures-little-changed-before-results.html?cmpid=, 1736-http://www.bloomberg.com/news/2013-07-12/asiana-pilots-raised-speed-concerns-seconds-before-crash.html, 1737-http://mobile.bloomberg.com/news/2013-07-12/asiana-pilots-raised-speed-concerns-seconds-before-crash.html?cmpid=, 1738-http://mobile.bloomberg.com/news/2013-07-12/at-t-to-acquire-leap-in-deal-that-values-carrier-at-1-2-billion.html?cmpid=, 1739-http://www.bloomberg.com/news/2013-07-12/at-t-to-acquire-leap-in-deal-that-values-carrier-at-1-2-billion.html, 1740-http://www.siteadvisor.com/, 1741-http://www.siteadvisor.com/websecurity/index.html, 1742-http://www.siteadvisor.com/howitworks/index.html, 1743-http://www.siteadvisor.com/analysis/, 1744-http://www.siteadvisor.com/webmasters/index.html, 1745-http://www.guardiannews.com/, 1746-http://m.guardian.co.uk/, 1747-http://www.guardian.co.uk/search?q=golden&section=, 1748-http://www.guardian.co.uk/search?q=ray+lewis&section=, 1749-http://www.guardian.co.uk/search?q=stone+mountain&section=, 1750-http://www.guardian.co.uk/search?q=mcdonalds&section=, 1751-http://m.guardian.co.uk/sport/2013/jul/13/the-ashes-england-australia-live-report, 1752-http://www.guardian.co.uk/sport/2013/jul/13/the-ashes-england-australia-live-report, 1753-http://m.guardian.co.uk/football/2013/jul/13/manchester-united-david-moyes-first-game-live, 1754-http://www.guardian.co.uk/football/2013/jul/13/manchester-united-david-moyes-first-game-live, 1755-http://m.guardian.co.uk/world/2013/jul/13/jesse-jackson-george-zimmerman-trayvon-martin, 1756-http://www.guardian.co.uk/world/2013/jul/13/jesse-jackson-george-zimmerman-trayvon-martin, 1757-http://www.guardian.co.uk/world/2013/jul/13/texas-passes-anti-abortion-law, 1758-http://m.guardian.co.uk/world/2013/jul/13/texas-passes-anti-abortion-law, 1759-http://m.tagged.com/, 1760-http://www.tagged.com/, 1761-https://myspace.com/, 1762-http://www.time.com/time/, 1763-http://nation.time.com/2013/07/13/in-move-to-university-of-california-napolitano-trades-one-challenging-bureaucracy-for-another/, 1764-http://swampland.time.com/2013/07/12/five-changes-to-justice-department-guidelines-designed-to-protect-reporters/?iid=sl-main-lead, 1765-http://business.time.com/2013/07/12/gas-prices-forecast-to-soar-during-peak-summer-vacation-period/, 1766-http://entertainment.time.com/2013/07/12/lady-gaga-arrested-development-and-james-bond-the-week-in-entertainment/, 1767-http://www.npr.org/, 1768-http://m.npr.org/, 1769-http://m.npr.org/story/search?searchTerm=golden, 1770-http://m.npr.org/story/search?searchTerm=ray+lewis, 1771-http://m.npr.org/story/search?searchTerm=stone+mountain, 1772-http://m.npr.org/story/search?searchTerm=mcdonalds, 1773-http://www.npr.org/search/index.php?searchinput=ray+lewis, 1774-http://www.npr.org/search/index.php?searchinput=golden, 1775-http://www.npr.org/search/index.php?searchinput=stone+mountain, 1776-http://www.npr.org/search/index.php?searchinput=mcdonalds, 1777-http://www.npr.org/blogs/thetwo-way/2013/07/14/202016100/angry-but-mostly-peaceful-protests-follow-zimmerman-acquittal, 1778-http://m.npr.org/news/front/202016100, 1779-http://m.npr.org/news/front/201925236, 1780-http://www.npr.org/2013/07/14/201925236/a-bipartisan-road-show-to-reform-the-tax-code, 1781-http://www.npr.org/blogs/thetwo-way/2013/07/14/201947778/actor-cory-monteith-who-played-finn-hudson-on-glee-found-dead, 1782-http://m.npr.org/news/front/201947778, 1783-http://www.npr.org/blogs/parallels/2013/07/14/201153551/The-Don-Whos-Taken-Charge-Of-Jordans-Biggest-Refugee-Camp, 1784-http://m.npr.org/news/front/201153551?start=5, 1785-http://mobile.boston.com/, 1786-http://www.boston.com/, 1787-http://www.boston.com/search/?q=golden, 1788-http://www.boston.com/search/?q=ray+lewis, 1789-http://www.boston.com/search/?q=stone+mountain, 1790-http://www.boston.com/search/?q=mcdonalds, 1791-http://www.boston.com/sports/other-sports/track-and-field/2013/07/14/tyson-gay-tests-positive-for-banned-substance/i8seJAftFs4H5nZjd3V25H/story.html, 1792-http://mobile.boston.com/art/35/sports/other-sports/track-and-field/2013/07/14/tyson-gay-tests-positive-for-banned-substance/i8seJAftFs4H5nZjd3V25H/story;jsessionid=89B8B743A139B3C0B188B7F6F7DE79A4, 1793-http://mobile.boston.com/art/35/news/nation/2013/07/14/zimmerman-cleared-attorney-says-safety-concern/IvZYJ6Zg1XmizasJ2YUECI/story, 1794-http://www.boston.com/news/nation/2013/07/14/zimmerman-cleared-attorney-says-safety-concern/

IvZYJ6Zg1XmizasJ2YUECI/story.html, 1795-http://www.boston.com/news/world/canada/2013/07/14/cory-monteith-star-hit-show-glee-found-dead/g7tw0wHC4voe7FfjFG1gWK/story.html, 1796-http://mobile.boston.com/art/35/news/world/canada/2013/07/14/cory-monteith-star-hit-show-glee-found-dead/g7tw0wHC4voe7FfjFG1gWK/story, 1797-http://m.foodnetwork.com/, 1798-http://www.foodnetwork.com/, 1799-http://m.foodnetwork.com/recipes/search?id=golden, 1800-http://www.foodnetwork.com/search/delegate.do?fnSearchString=golden&fnSearchType=site, 1801-http://m.foodnetwork.com/recipes/search?id=ray+lewis, 1802-http://www.foodnetwork.com/search/delegate.do?fnSearchString=ray+lewis&fnSearchType=site, 1803-http://www.foodnetwork.com/search/delegate.do?fnSearchString=stone+mountain&fnSearchType=site, 1804-http://www.foodnetwork.com/search/delegate.do?fnSearchString=mcdonalds&fnSearchType=site, 1805-http://m.foodnetwork.com/recipes/search?id=stone+mountain, 1806-http://m.foodnetwork.com/recipes/search?id=mcdonalds, 1807-http://www.foodnetwork.com/recipes/rachael-ray/pork-chops-with-golden-apple-sauce-recipe/index.html, 1808-http://m.foodnetwork.com/recipes/28806, 1809-http://m.foodnetwork.com/recipes/37273, 1810-http://www.foodnetwork.com/recipes/tyler-florence/ricotta-pancakes-with-roasted-golden-delicious-apples-and-roasted-prosciutto-recipe/index.html, 1811-http://m.foodnetwork.com/recipes/134915, 1812-http://www.foodnetwork.com/recipes/30-minute-meals/pork-chops-golden-apple-and-raisin-sauce-whole-wheat-pasta-mac-n-cheddar-recipe/index.html, 1813-http://m.foodnetwork.com/recipes/454772, 1814-http://www.foodnetwork.com/recipes/guy-fieri/golden-pound-cake-recipe/index.html, 1815-http://www.barnesandnoble.com/s/mobile-site, 1816-http://www.barnesandnoble.com/, 1817-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=golden, 1818-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=ray+lewis, 1819-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=stone+mountain, 1820-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=mcdonalds, 1821-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=ray+lewis, 1822-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=golden, 1823-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=stone+mountain, 1824-http://www.barnesandnoble.com/s/golden?store=allproducts&keyword=mcdonalds, 1825-http://www.barnesandnoble.com/w/golden-lady-antebellum/25982878?ean=5099997918721, 1826-http://www.barnesandnoble.com/p/toys-games-golden-retriever/25491917?ean=4005086163355, 1827-http://www.barnesandnoble.com/w/dvd-golden-girls-season-2-bea-arthur/9405771?ean=786936255935, 1828-http://www.barnesandnoble.com/w/golden-goblet-eloise-jarvis-mcgraw/1102157089?ean=9780140303353, 1829-http://www.wikimedia.org/, 1830-http://m.youjizz.com/, 1831-http://www.youjizz.com/, 1832-http://www.youjizz.com/search/golden-1.html, 1833-http://m.youjizz.com/search/golden/page1.html, 1834-

http://m.youjizz.com/search/ray+lewis/page1.html, 1835-http://www.youjizz.com/search/golden-1.html, 1836-http://www.youjizz.com/search/ray+lewis-1.html, 1837-http://www.youjizz.com/search/stone+mountain-1.html, 1838-http://www.youjizz.com/search/mcdonalds-1.html, 1839-http://m.ticketmaster.com/, 1840-http://www.ticketmaster.com/, 1841-http://m.ticketmaster.com/ticket/search.do?articles=tmus&query=golden&submit=, 1842-http://m.ticketmaster.com/ticket/search.do?articles=tmus&query=ray+lewis&submit=, 1843-http://m.ticketmaster.com/ticket/search.do?articles=tmus&query=stone+mountain&submit=, 1844-http://m.ticketmaster.com/ticket/search.do?articles=tmus&query=mcdonalds&submit=, 1845-http://www.ticketmaster.com/search?tm_link=tm_homeA_header_search&user_input=golden&q=golden, 1846-http://www.ticketmaster.com/search?tm_link=tm_homeA_header_search&user_input=ray+lewis&q=ray+lewis, 1847-http://www.ticketmaster.com/search?tm_link=tm_homeA_header_search&user_input=stone+mountain&q=stone=mountain, 1848-http://www.ticketmaster.com/search?tm_link=tm_homeA_header_search&user_input=mcdonalds&q=mcdonalds, 1849-http://en.cam4.co/, 1850-http://www.hp.com/, 1851-http://www8.hp.com/us/en/hp-search/search-results.html?ajaxpage=1#/page=1&/cc=us&/lang=en&/qt=golden, 1852-http://www8.hp.com/us/en/hp-search/search.html?nores=true&qt=ray\%20lewis, 1853-http://www8.hp.com/us/en/hp-search/search.html?nores=true&qt=stone\%20mountain, 1854-http://www8.hp.com/us/en/hp-search/search.html?nores=true&qt=mcdonalds, 1855-https://www.discovercard.com/cardmembersvcs/mobile/app/loginlogout/auth, 1856-https://www.discover.com/, 1857-http://m.kayak.com/p/front/?prevmode=front, 1858-http://www.kayak.com/, 1859-https://m.cox.net/mydvr/scheduling/splash.action, 1860-http://intercept.cox.com/dispatch/8905267496139349943/intercept.cox?lob=residential&s=filter&dest=http\%3A\%2F\%2Fww2.cox.com\%2Fmyconnection\%2Fhome.cox, 1861-http://www.search-results.com/, 1862-http://www.search-results.com/web?qsrc=0&o=15621&l=dir&fhp=1&q=golden&locale=en_US, 1863-http://www.search-results.com/web?qsrc=0&o=15621&l=dir&fhp=1&q=ray+lewis&locale=en_US, 1864-http://www.search-results.com/web?qsrc=0&o=15621&l=dir&fhp=1&q=stone+mountain&locale=en_US, 1865-http://www.search-results.com/web?qsrc=0&o=15621&l=dir&fhp=1&q=mcdonalds&locale=en_US, 1866-http://www.search-results.com/pictures?qsrc=167&o=15621&l=dir&q=ray+lewis&locale=en_US, 1867-http://www.search-results.com/pictures?qsrc=167&o=15621&l=dir&q=golden&locale=en_US, 1868-http://www.search-results.com/pictures?qsrc=167&o=15621&l=dir&q=stone+mountain&locale=en_US, 1869-http://www.search-results.com/pictures?qsrc=167&o=15621&l=dir&q=mcdonalds&locale=en_US, 1870-http://m.costco.com/, 1871-http://www.costco.com/, 1872-http://www.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=golden, 1873-http://m.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=golden, 1874-http://www.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=

ray+lewis, 1875-http://m.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=ray+lewis, 1876-http://m.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=stone+mountain, 1877-http://m.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=mcdonalds, 1878-http://www.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=stone+mountain, 1879-http://www.costco.com/CatalogSearch?storeId=10301&catalogId=10701&langId=-1&keyword=mcdonalds, 1880-http://m.costco.com/\%22Golden-Skylines-Change-in-Spirits\%22-by-Hilary-Williams.product.100049066.html, 1881-http://www.costco.com/\%22Golden-Skylines-Change-in-Spirits\%22-by-Hilary-Williams.product.100049066.html, 1882-http://m.costco.com/Laurel-Designs-Golden-Bronze-Finish-6-Light-Chandelier.product.100010112.html, 1883-http://www.costco.com/Laurel-Designs-Golden-Bronze-Finish-6-Light-Chandelier.product.100010112.html, 1884-http://www.costco.com/Pentax-K-30-2-Lens-Weatherproof-DSLR-Bundle-Blue.product.100034661.html, 1885-http://m.costco.com/Pentax-K-30-2-Lens-Weatherproof-DSLR-Bundle-Blue.product.100034661.html, 1886-http://www.costco.com/Safco-Veer-Series-Stacking-Chair-Blue-\%2526-Chrome-Color-4ct-SAF-4286BU.product.11612499.html, 1887-http://m.costco.com/Safco-Veer-Series-Stacking-Chair-Blue-\%2526-Chrome-Color-4ct-SAF-4286BU.product.11612499.html, 1888-https://mobile.usaa.com/inet/ent_logon/Logon, 1889-https://www.usaa.com/inet/ent_logon/Logon, 1890-https://www.usaa.com/inet/ent_search/CpPrvtSearch?SearchPhrase=golden&maac_page_ref=ent_login_member, 1891-https://www.usaa.com/inet/ent_search/CpPrvtSearch?SearchPhrase=ray+lewis&maac_page_ref=ent_login_member, 1892-https://www.usaa.com/inet/ent_search/CpPrvtSearch?SearchPhrase=stone+mountain&maac_page_ref=ent_login_member, 1893-https://www.usaa.com/inet/ent_search/CpPrvtSearch?SearchPhrase=mcdonalds&maac_page_ref=ent_login_member, 1894-https://www.usaa.com/inet/pages/auto_insurance_main?offerName=pubHomePro_PrdBckt_1_030512_PnC_AutoIns_learnmore, 1895-https://mobile.usaa.com/inet/ent_mobile_storefront/StoreFrontApp/SubProductDetailPage?key=insurance-mobile-auto-product, 1896-https://www.usaa.com/inet/pages/insurance_home_main?wa_ref=lf_product_ins_home_property, 1897-https://mobile.usaa.com/inet/ent_mobile_storefront/StoreFrontApp/SubProductDetailPage?key=insurance-mobile-homeowners-product, 1898-https://mobile.usaa.com/inet/ent_mobile_storefront/StoreFrontApp/SubProductDetailPage?key=banking-mobile-checking-product, 1899-https://www.usaa.com/inet/pages/bank_main?wa_ref=lf_product_bank, 1900-https://www.usaa.com/inet/pages/investments_iras_main?wa_ref=lf_product_invest_iras, 1901-https://mobile.usaa.com/inet/ent_mobile_storefront/StoreFrontApp/SubProductDetailPage?key=investments-mobile-ira-product, 1902-http://mobile.weather.gov/#typeLocation, 1903-http://www.weather.gov/, 1904-http://forecast.weather.gov/MapClick.php?lat=33.8730946&lon=-84.33842900000002&site=all&smap=1&searchresult=Atlanta\%2C\%20GA\%2030319\%2C\

%20USA#.UeNQuBy4wcM, 1905-http://mobile.weather.gov/index.php?lat=33.8730946&lon=-84.33842900000002, 1906-http://forecast.weather.gov/MapClick.php?lat=35.9722081&lon=-79.04755590000002&site=all&smap=1&searchresult=Chapel\%20Hill\%2C\%20NC\%2027514\%2C\%20USA#.UeNQ_Ry4wcM, 1907-http://mobile.weather.gov/index.php?lat=35.9722081&lon=-79.04755590000002, 1908-http://mobile.weather.gov/index.php?lat=46.6414266&lon=-94.8835345, 1909-http://forecast.weather.gov/MapClick.php?lat=46.6414266&lon=-94.8835345&site=all&smap=1&searchresult=Nimrod\%2C\%20MN\%2056478\%2C\%20USA#.UeNRORy4wcM, 1910-http://mobile.weather.gov/index.php?lat=33.9799999&lon=-118.38999999999999, 1911-http://forecast.weather.gov/MapClick.php?lat=33.9799999&lon=-118.38999999999999&site=all&smap=1&searchresult=Culver\%20City\%2C\%20CA\%2090233\%2C\%20USA#.UeNRgxy4wcM, 1912-http://www.cnbc.com/, 1913-http://search.cnbc.com/main.do?target=all&categories=exclude&partnerId=2000&keywords=golden, 1914-http://search.cnbc.com/main.do?target=all&categories=exclude&partnerId=2000&keywords=ray+lewis, 1915-http://search.cnbc.com/main.do?target=all&categories=exclude&partnerId=2000&keywords=stone+mountain, 1916-http://search.cnbc.com/main.do?target=all&categories=exclude&partnerId=2000&keywords=mcdonalds, 1917-http://www.cnbc.com/id/100882699, 1918-http://www.cnbc.com/id/100873221, 1919-http://www.cnbc.com/id/100884648, 1920-http://www.cnbc.com/id/100880818, 1921-https://m.webmail.earthlink.net/login, 1922-http://www.earthlink.net/, 1923-http://www.slideshare.net/, 1924-http://www.slideshare.net/search/slideshow?searchfrom=header&q=golden, 1925-http://www.slideshare.net/search/slideshow?searchfrom=header&q=ray+lewis, 1926-http://www.slideshare.net/search/slideshow?searchfrom=header&q=stone+mountain, 1927-http://www.slideshare.net/search/slideshow?searchfrom=header&q=mcdonalds, 1928-http://www.slideshare.net/mattthemathman/moz-2013-ranking-factors-matt-peters-mozcon-24036187, 1929-http://www.slideshare.net/infonote/5-rhea-mozcon2012drysdale, 1930-http://www.slideshare.net/philipbuckley/everything-you-know-about-seo-is-wrong, 1931-http://www.slideshare.net/jennyhalasz/seo-meetup-the-search-year-in-review, 1932-http://www.usmagazine.com/, 1933-http://www.usmagazine.com/search?cx=007258338474195852663\%3Asc01_1qj3z0&cof=FORID\%3A10&ie=UTF-8&q=golden&sa=\%C2\%A0, 1934-http://www.usmagazine.com/search?cx=007258338474195852663\%3Asc01_1qj3z0&cof=FORID\%3A10&ie=UTF-8&q=ray+lewis&sa=\%C2\%A0, 1935-http://www.usmagazine.com/search?cx=007258338474195852663\%3Asc01_1qj3z0&cof=FORID\%3A10&ie=UTF-8&q=stone+mountain&sa=\%C2\%A0, 1936-http://www.usmagazine.com/search?cx=007258338474195852663\%3Asc01_1qj3z0&cof=FORID\%3A10&ie=UTF-8&q=mcdonalds&sa=\%C2\%A0, 1937-http://www.usmagazine.com/celebrity-body/news/sofia-vergara-wears-revealing-black-cutout-swimsuit-hugs-fiance-in-greece-2013157, 1938-http://www.usmagazine.com/celebrity-news/news/cory-monteiths-death-3-charities-to-make-donations-in-glee-stars-memory-2013157, 1939-

http://www.usmagazine.com/celebrity-news/news/fergie-to-change-legal-name-to-fergie-duhamel-2013157, 1940-http://www.usmagazine.com/celebrity-body/news/uma-thurman-and-arpad-busson-in-swimsuits-engage-in-pda-on-yacht-2013157, 1941-http://www.fool.com/, 1942-http://m.fool.com/, 1943-http://m.fool.com/search?sort=date&query=golden&thisform=Submit, 1944-http://m.fool.com/search?sort=date&query=ray+lewis&thisform=Submit, 1945-http://m.fool.com/search?sort=date&query=stone+mountain&thisform=Submit, 1946-http://m.fool.com/search?sort=date&query=mcdonalds&thisform=Submit, 1947-http://www.fool.com/search/solr.aspx?exchange-input=&q=golden&source=ignsittn0000001, 1948-http://www.fool.com/search/solr.aspx?exchange-input=&q=ray+lewis&source=ignsittn0000001, 1949-http://www.fool.com/search/solr.aspx?exchange-input=&q=stone+mountain&source=ignsittn0000001, 1950-http://www.fool.com/search/solr.aspx?exchange-input=&q=mcdonalds&source=ignsittn0000001, 1951-http://m.fool.com/investing/general/2013/07/15/an-unexpected-but-welcome-consequence-of-rising-mo?source=izmmblmfa0000001, 1952-http://www.fool.com/investing/general/2013/07/15/an-unexpected-but-welcome-consequence-of-rising-mo.aspx?source=ihpsitth0000001, 1953-http://m.fool.com/investing/general/2013/07/15/this-is-what-poor-retirement-planning-looks-like?source=izmmblmta0000001&amp;lidx=1, 1954-http://www.fool.com/investing/general/2013/07/15/this-is-what-poor-retirement-planning-looks-like.aspx?source=ihpsitota0000001&lidx=1, 1955-http://m.fool.com/investing/general/2013/07/15/presentation-slides-what-makes-us-bad-investors?source=iaasitlnk0000003?source=izmmblmta0000001&amp;lidx=2, 1956-http://www.fool.com/investing/general/2013/07/15/presentation-slides-what-makes-us-bad-investors.aspx?source=ihpsitota0000001&lidx=2, 1957-http://www.fool.com/investing/general/2013/07/15/3-reasons-to-ignore-the-coming-flood-of-hedge-fund.aspx?source=ihpsitota0000001&lidx=3, 1958-http://m.fool.com/investing/general/2013/07/15/3-reasons-to-ignore-the-coming-flood-of-hedge-fund?source=izmmblmta0000001&amp;lidx=3, 1959-http://www.staples.com/, 1960-http://m.staples.com/, 1961-http://m.staples.com/golden/directory_golden?autocompletesearchkey=golden, 1962-http://m.staples.com/golden/directory_golden?autocompletesearchkey=ray+lewis, 1963-http://m.staples.com/golden/directory_golden?autocompletesearchkey=stone+mountain, 1964-http://m.staples.com/golden/directory_golden?autocompletesearchkey=mcdonalds, 1965-http://www.staples.com/golden/directory_golden?, 1966-http://www.staples.com/ray+lewis/directory_ray+lewis?, 1967-http://www.staples.com/stone+mountain/directory_stone+mountain?, 1968-http://www.staples.com/mcdonalds/directory_mcdonalds?, 1969-http://m.staples.com/Golden-Grahams-Treats-Cereal-Bars-21-oz-12-Bars-Box/product_865332, 1970-http://www.staples.com/Golden-Grahams-Treats-Cereal-Bars-21-oz-12-Bars-Box/product_865332, 1971-http://m.staples.com/Golden-Nugget-Gift-Card-25/product_142343, 1972-http://www.staples.com/Golden-Nugget-Gift-Card-25/product_142343, 1973-http://www.staples.com/BelVita-Breakfast-Biscuits-Golden-Oat-8-Packs-Box/product_177562, 1974-http://m.staples.com/BelVita-Breakfast-Biscuits-Golden-Oat-8-Packs-Box/product_177562, 1975-http://m.staples.com/Kenroy-Home-Cromwell-Floor-Lamp-Golden-Flecked-Bronze-Finish/product_149329, 1976-http://www.staples.com/Kenroy-Home-Cromwell-Floor-Lamp-Golden-Flecked-Bronze-Finish/product_149329, 1977-http://m.t-mobile.com/, 1978-http://www.t-mobile.com/, 1979-http://find.t-mobile.com/controller?N=0&Ntk=primary&Ntx=mode\%2Bmatchallpartial&Ntt=golden, 1980-http://find.t-mobile.com/controller?N=0&Ntk=primary&Ntx=mode\%2Bmatchallpartial&Ntt=ray+lewis, 1981-http://find.t-mobile.com/controller?N=0&Ntk=primary&Ntx=mode\%2Bmatchallpartial&Ntt=stone+mountain, 1982-http://find.t-mobile.com/controller?N=0&Ntk=primary&Ntx=mode\%2Bmatchallpartial&Ntt=mcdonalds, 1983-http://explore.t-mobile.com/phone-upgrade?link=jump, 1984-http://m.t-mobile.com/phone-upgrade?cm_mmc_o=FB_bkwCjC-czywEwllCjC3J4VCjC3J4VFB_bkwFzy6Aww, 1985-http://www.google.com, 1986-http://www.google.com/#q=lewis+james, 1987-http://www.google.com/#q=hard+rock+place, 1988-http://www.google.com/#q=holy+youth, 1989-http://www.google.com/#q=a+hello+berry, 1990-http://www.google.com/#q=man+results, 1991-http://www.google.com/search?um=1&hl=en&tbo=d&biw=1280&bih=596&tbm=isch&sa=1&q=lewis+james, 1992-http://www.google.com/search?um=1&hl=en&tbo=d&biw=1280&bih=596&tbm=isch&sa=1&q=hard+rock+place, 1993-http://www.google.com/search?um=1&hl=en&tbo=d&biw=1280&bih=596&tbm=isch&sa=1&q=holy+youth, 1994-http://www.google.com/search?um=1&hl=en&tbo=d&biw=1280&bih=596&tbm=isch&sa=1&q=a+hello+berry, 1995-http://www.google.com/search?um=1&hl=en&tbo=d&biw=1280&bih=596&tbm=isch&sa=1&q=man+results, 1996-http://www.google.com/search?hl=en&gl=us&tbm=nws&q=lewis+james, 1997-http://www.google.com/search?hl=en&gl=us&tbm=nws&q=hard+rock+place, 1998-http://www.google.com/search?hl=en&gl=us&tbm=nws&q=holy+youth, 1999-http://www.google.com/search?hl=en&gl=us&tbm=nws&q=a+hello+berry, 2000-http://www.google.com/search?hl=en&gl=us&tbm=nws&q=man+results, 2001-https://www.facebook.com, 2002-https://www.facebook.com/officialraylewis, 2003-https://www.facebook.com/ladygaga, 2004-https://www.facebook.com/DonaldTrump, 2005-https://www.facebook.com/barackobama, 2006-https://www.facebook.com/McDonalds, 2007-http://www.facebook.com/search.php?q=lewis+james, 2008-http://www.facebook.com/search.php?q=hard+rock+place, 2009-http://www.facebook.com/search.php?q=holy+youth, 2010-http://www.facebook.com/search.php?q=a+hello+berry, 2011-http://www.facebook.com/search.php?q=man+results, 2012-http://www.youtube.com, 2013-http://www.youtube.com/results?search_query=lewis+james, 2014-http://www.youtube.com/results?search_query=hard+rock+place, 2015-http://www.youtube.com/results?search_query=holy+youth, 2016-http://www.youtube.com/results?search_query=a+hello+berry, 2017-http://www.youtube.com/results?search_query=man+results, 2018-http://www.yahoo.com, 2019-http://news.yahoo.com/, 2020-http://news.yahoo.com/blogs/lookout/gay-marriage-mississippi-newspaper-owner-140311568.html, 2021-http://news.yahoo.com/britain-india-diamond-royal-crown-ours-000950576.html, 2022-http://news.yahoo.com/internet-advertisers-kill-text-based-captcha-205416291.html, 2023-http://news.yahoo.com/former-senator-admits-fathering-child-other-senators-daughter-200530146--abc-news-politics.html, 2024-http://news.yahoo.com/drones-large-small-coming-us-010537668.html, 2025-http://finance.yahoo.com, 2026-http://finance.yahoo.com/q?s=INTC, 2027-http://finance.yahoo.com/q?s=MSFT, 2028-http://finance.yahoo.com/q?s=AAPL, 2029-http://finance.yahoo.com/q?s=BAC, 2030-http://finance.yahoo.com/q?s=CSCO, 2031-http://news.yahoo.com, 2032-http://weather.yahoo.com, 2033-http://sports.yahoo.com, 2034-http://sports.yahoo.com/nba, 2035-http://sports.yahoo.com/nfl, 2036-http://sports.yahoo.com/mlb, 2037-http://sports.yahoo.com/nhl, 2038-http://sports.yahoo.com/golf, 2039-http://sports.yahoo.com/boxing, 2040-http://sports.yahoo.com/soccer, 2041-http://sports.yahoo.com/college-basketball, 2042-http://sports.yahoo.com/college-football, 2043-http://omg.yahoo.com, 2044-http://omg.yahoo.com/blogs/celeb-news/, 2045-http://omg.yahoo.com/photos, 2046-http://omg.yahoo.com/videos, 2047-http://omg.yahoo.com/top-celebrities, 2048-http://www.amazon.com, 2049-http://www.amazon.com/s/ref=nb_sb_noss_2?url=search-alias\%3Daps&field-keywords=lewis+james, 2050-http://www.amazon.com/s/ref=nb_sb_noss_2?url=search-alias\%3Daps&field-keywords=hard+rock+place, 2051-http://www.amazon.com/s/ref=nb_sb_noss_2?url=search-alias\%3Daps&field-keywords=holy+youth, 2052-http://www.amazon.com/s/ref=nb_sb_noss_2?url=search-alias\%3Daps&field-keywords=a+hello+berry, 2053-http://www.amazon.com/s/ref=nb_sb_noss_2?url=search-alias\%3Daps&field-keywords=man+results, 2054-http://www.amazon.com/Mediabridge-High-Speed-Cable-Ethernet/dp/B0019EHU8G/ref=sr_1_1?ie=UTF8&qid=1361416525&sr=8-1&keywords=product, 2055-http://www.amazon.com/Mediabridge-Toslink-Cable-Optical-Digital/dp/B005M4IWNQ/ref=pd_sim_e_4, 2056-http://www.amazon.com/ZVOX-4003201-Z-Base-Low-Profile-Cabinet/dp/B006O711V0/ref=pd_sim_e_3, 2057-http://www.amazon.com/Deluxe-chrome-Rubberized-Snap--Iphone/dp/B005LUBUT4/ref=sr_1_1?s=electronics&ie=UTF8&qid=1361474021&sr=1-1&keywords=hot, 2058-http://www.amazon.com/Combo-Polka-Flex-Case-Iphone/dp/B008EU7HRM/ref=pd_sim_e_5, 2059-http://www.ebay.com, 2060-http://www.ebay.com/sch/i.html?_trksid=p5197.m570.l1313&_nkw=lewis+james&_sacat=0&_from=R40, 2061-http://www.ebay.com/sch/i.html?_trksid=p5197.m570.l1313&_nkw=hard+rock+place&_sacat=0&_from=R40, 2062-http://www.ebay.com/sch/i.html?_trksid=p5197.m570.l1313&_nkw=holy+youth&_sacat=0&_from=R40, 2063-http://www.ebay.com/sch/i.html?_trksid=p5197.m570.l1313&_nkw=a+hello+berry&_sacat=0&_from=R40, 2064-http://www.ebay.com/sch/i.html?_trksid=p5197.m570.l1313&_nkw=man+results&_sacat=0&_from=R40, 2065-http://www.ebay.com/itm/DISNEY-PIXAR-CARS-2-DELUXE-IVAN-MATER-SINGLE-GREAT-CARD-

AWESOME-NEW-CAR-RARE-/321072787434?pt=TV_Movie_Character_Toys_US&hash=item4ac16defea, 2066-http://cgi.ebay.com/ebaymotors/Ford-Mustang-Coupe-Ford-Mustang-1966-Poppy-Red-Head-Turner-/251230916774?pt=US_Cars_Trucks&hash=item3a7e8790a6#ht_500wt_1182, 2067-http://www.ebay.com/itm/Nintendo-Wii-Sports-Sports-Resort-Pack-Black-Console-NTSC-RVKSKAAU-/360597210459?pt=Video_Games&hash=item53f544c55b, 2068-http://www.ebay.com/itm/NINTENDO-WII-BLACK-CONSOLE-100-WORKING-REPLACE-YOUR-DAMAGE-UNIT-WIIWARE-GAMES-/261173008235?pt=Video_Games&hash=item3ccf1fd76b, 2069-http://www.ebay.com/itm/Black-Nintendo-Wii-Video-Game-System-/370764009184?pt=Video_Games&hash=item565341cee0, 2070-http://www.wikipedia.org/, 2071-http://en.wikipedia.org/w/index.php?search=man+results&title=Special\%3ASearch, 2072-http://en.wikipedia.org/w/index.php?search=lewis+james&title=Special\%3ASearch, 2073-http://en.wikipedia.org/w/index.php?search=hard+rock+place&title=Special\%3ASearch, 2074-http://en.wikipedia.org/w/index.php?search=a+hello+berry&title=Special\%3ASearch, 2075-http://en.wikipedia.org/w/index.php?search=holy+youth&title=Special\%3ASearch, 2076-http://en.wikipedia.org/wiki/Physics, 2077-http://en.wikipedia.org/wiki/Adam_Morrison, 2078-http://en.wikipedia.org/wiki/Michael_Jordan, 2079-http://en.wikipedia.org/wiki/University_of_North_Carolina_at_Chapel_Hill, 2080-http://en.wikipedia.org/wiki/North_Carolina, 2081-http://www.craigslist.org/about/sites, 2082-http://charlotte.craigslist.org, 2083-http://charlotte.craigslist.org/search/?areaID=41&subAreaID=&query=lewis+james&catAbb=sss, 2084-http://charlotte.craigslist.org/search/?areaID=41&subAreaID=&query=hard+rock+place&catAbb=sss, 2085-http://charlotte.craigslist.org/search/?areaID=41&subAreaID=&query=holy+youth&catAbb=sss, 2086-http://charlotte.craigslist.org/search/?areaID=41&subAreaID=&query=a+hello+berry&catAbb=sss, 2087-http://charlotte.craigslist.org/search/?areaID=41&subAreaID=&query=man+results&catAbb=sss, 2088-http://charlotte.craigslist.org/mcd/3580319747.html, 2089-http://charlotte.craigslist.org/ptd/3576143018.html, 2090-http://charlotte.craigslist.org/zip/3635517954.html, 2091-http://live.com, 2092-http://twitter.com, 2093-http://twitter.com/search, 2094-https://twitter.com/search?q=lewis\%20james&src=typd, 2095-https://twitter.com/search?q=hard\%20rock\%20place&src=typd, 2096-https://twitter.com/search?q=holy\%20youth&src=typd, 2097-https://twitter.com/search?q=a\%20hello\%20berry&src=typd, 2098-https://twitter.com/search?q=man\%20results&src=typd, 2099-https://twitter.com/raylewis, 2100-https://twitter.com/ladygaga, 2101-https://twitter.com/realDonaldTrump, 2102-https://twitter.com/barackobama, 2103-https://twitter.com/McDonalds, 2104-http://www.linkedin.com/, 2105-http://www.linkedin.com/pub/dir/lewis/james, 2106-http://www.linkedin.com/pub/dir/hard/rock, 2107-http://www.linkedin.com/pub/dir/holy/youth, 2108-http://www.linkedin.com/pub/dir/hello/berry, 2109-http://www.linkedin.com/pub/dir/man/results, 2110-http://www.linkedin.com/in/barackobama, 2111-http://www.linkedin.com/pub/ray-lewis/5/461/815, 2112-http://www.linkedin.com/pub/lady-gaga/61/357/130, 2113-http://www.linkedin.com/pub/donald-trump/21/737/441, 2114-http://www.linkedin.com/company/mcdonald’s-corporation, 2115-http://www.bing.com, 2116-http://www.bing.com/search?q=lewis+james, 2117-http://www.bing.com/search?q=hard+rock+place, 2118-http://www.bing.com/search?q=holy+youth, 2119-http://www.bing.com/search?q=a+hello&berry, 2120-http://www.bing.com/search?q=man+results, 2121-http://www.bing.com/images/search?q=lewis+james&FORM=HDRSC2, 2122-http://www.bing.com/images/search?q=hard+rock+place&FORM=HDRSC2, 2123-http://www.bing.com/images/search?q=holy+youth&FORM=HDRSC2, 2124-http://www.bing.com/images/search?q=a+hello+berry&FORM=HDRSC2, 2125-http://www.bing.com/images/search?q=man+results&FORM=HDRSC2, 2126-http://www.bing.com/videos/search?q=lewis+james&FORM=HDRSC3, 2127-http://www.bing.com/videos/search?q=hard+rock+place&FORM=HDRSC3, 2128-http://www.bing.com/videos/search?q=holy+youth&FORM=HDRSC3, 2129-http://www.bing.com/videos/search?q=a+hello+berry&FORM=HDRSC3, 2130-http://www.bing.com/videos/search?q=man+results&FORM=HDRSC3, 2131-http://www.bing.com/news?q=lewis+james&FORM=HDRSC6, 2132-http://www.bing.com/news?q=hard+rock+place&FORM=HDRSC6, 2133-http://www.bing.com/news?q=holy+youth&FORM=HDRSC6, 2134-http://www.bing.com/news?q=a+hello+berry&FORM=HDRSC6, 2135-http://www.bing.com/news?q=man+results&FORM=HDRSC6, 2136-http://www.blogspot.com, 2137-http://obamabarack.blogspot.com/, 2138-http://john-nevarez.blogspot.com/, 2139-http://gagacheat.blogspot.com/, 2140-http://gagajournal.blogspot.com/, 2141-http://mymannypacquiao.blogspot.com/, 2142-http://pinterest.com, 2143-http://pinterest.com/search/pins/?q=lewis+james, 2144-http://pinterest.com/search/pins/?q=hard+rock+place, 2145-http://pinterest.com/search/pins/?q=holy+youth, 2146-http://pinterest.com/search/pins/?q=a+hello+berry, 2147-http://pinterest.com/search/pins/?q=man+results, 2148-http://pinterest.com/pin/53972895505405246/, 2149-http://pinterest.com/pin/125819383312005050/, 2150-http://pinterest.com/pin/51369251971746854/, 2151-http://pinterest.com/pin/94294185918187201/, 2152-http://pinterest.com/pin/48223707256954728 4/, 2153-http://go.com, 2154-http://www.msn.com/, 2155-http://investing.money.msn.com/investments/stock-price?symbol=INTC&x=0&y=0, 2156-http://investing.money.msn.com/investments/stock-price?symbol=MSFT&x=0&y=0, 2157-http://investing.money.msn.com/investments/stock-price?symbol=BAC&x=0&y=0, 2158-http://investing.money.msn.com/investments/stock-price?symbol=AAPL&x=0&y=0, 2159-http://investing.money.msn.com/investments/stock-price?symbol=CSCO&x=0&y=0, 2160-http://news.msn.com, 2161-http://entertainment.msn.com, 2162-http://msn.foxsports.com, 2163-http://msn.foxsports.com/nascar/story/danica-patrick-eases-through-budweiser-duel-ready-for-daytona-500-racing-022113, 2164-http://msn.foxsports.com/nfl/story/Scouting-Combine-New-York-Jets-Rex-Ryan-John-Idzik-dispel-Darrelle-Revis-rumors-Day-1-022113, 2165-http://msn.foxsports.com/golf/story/Tiger-Woods-Rory-McIlroygone-after-one-match-at-Accenture-match-play-022113, 2166-http://msn.foxsports.com/nfl, 2167-http://msn.foxsports.com/nba, 2168-http://msn.foxsports.com/mlb, 2169-http://msn.foxsports.com/nhl, 2170-http://msn.foxsports.com/collegefootball, 2171-http://msn.foxsports.com/collegebasketball, 2172-http://msn.foxsports.com/golf, 2173-http://msn.foxsports.com/nascar, 2174-http://msn.foxsports.com/foxsoccer, 2175-http://msn.foxsports.com/ufc, 2176-http://money.msn.com, 2177-http://living.msn.com, 2178-http://living.msn.com/family-parenting, 2179-http://living.msn.com/style-beauty, 2180-http://living.msn.com/home-decor, 2181-http://local.msn.com, 2182-http://local.msn.com/weather.aspx?zip=27517, 2183-http://local.msn.com/hourly.aspx?zip=27517, 2184-http://local.msn.com/weather.aspx?zip=27510, 2185-http://local.msn.com/hourly.aspx?zip=27510, 2186-http://local.msn.com/weather.aspx?zip=27514, 2187-http://local.msn.com/hourly.aspx?zip=27514, 2188-http://local.msn.com/weather.aspx?zip=30319, 2189-http://local.msn.com/hourly.aspx?zip=30319, 2190-http://local.msn.com/weather.aspx?zip=30332, 2191-http://local.msn.com/hourly.aspx?zip=30332, 2192-http://www.aol.com, 2193-http://search.aol.com/aol/search?enabled_terms=&s_it=comsearch51&q=give+me, 2194-http://search.aol.com/aol/image?q=give+me&v_t=comsearch51&s_it=searchtabs, 2195-http://search.aol.com/aol/video?q=give+me&v_t=comsearch51&s_it=searchtabs, 2196-http://www.dailyfinance.com/?icis=navbar_rootfinance_main5, 2197-http://www.dailyfinance.com/quote/nasdaq/INTC, 2198-http://www.dailyfinance.com/quote/nasdaq/MSFT, 2199-http://www.dailyfinance.com/quote/nasdaq/BAC, 2200-http://www.dailyfinance.com/quote/nasdaq/AAPL, 2201-http://www.dailyfinance.com/quote/nasdaq/CSCO, 2202-http://weather.aol.com, 2203-http://weather.aol.com/forecast/todays/us/nc/chapel-hill/id/27599, 2204-http://www.huffingtonpost.com/entertainment/?icid=navbar_rootentertainment_main5, 2205-http://www.huffingtonpost.com, 2206-http://www.huffingtonpost.com/paul-raushenbush/tim-tebow-first-baptist-dallas_b_2734677.html, 2207-http://www.huffingtonpost.com/bianca-jagger/violence-against-women_b_2733708.html, 2208-http://www.huffingtonpost.com/2013/02/21/laura-bush-gay-marriage_n_2733619.html?utm_hp_ref=mostpopular, 2209-http://www.huffingtonpost.com/politics, 2210-http://www.huffingtonpost.com/business, 2211-http://www.huffingtonpost.com/sports, 2212-http://www.huffingtonpost.com/news/nfl, 2213-http://www.huffingtonpost.com/news/nba, 2214-http://www.huffingtonpost.com/news/college-football, 2215-http://www.huffingtonpost.com/news/college-basketball, 2216-http://www.huffingtonpost.com/news/mlb, 2217-http://www.stylelist.com/?icid=navbar_rootstyle_main5, 2218-https://www.tumblr.com, 2219-http://www.tumblr.com/tagged/lewis+james, 2220-http://www.tumblr.com/tagged/hard+rock+place, 2221-http://www.tumblr.com/tagged/holy+youth, 2222-http://www.tumblr.com/tagged/a+hello+berry, 2223-http://www.tumblr.com/tagged/man+results, 2224-https://www.paypal.com/home, 2225-http://www.cnn.com, 2226-

http://www.cnn.com/2013/02/22/us/weather-winter-storm/index.html?hpt=hp_c2, 2227-http://www.cnn.com/2013/02/21/us/lincoln-babysitter/index.html?hpt=hp_c2, 2228-http://www.cnn.com/2013/02/21/us/cnnheroes-carter-strive-for-college/index.html?hpt=hp_c2, 2229-http://www.cnn.com/2013/02/21/us/fbi-misbehavior/index.html?hpt=hp_c2, 2230-http://www.cnn.com/video, 2231-http://www.cnn.com/search/?query=lewis\%20james&sortBy=date, 2232-http://www.cnn.com/search/?query=hard\%20rock\%20place&sortBy=date, 2233-http://www.cnn.com/search/?query=holy\%20youth&sortBy=date, 2234-http://www.cnn.com/search/?query=a\%20hello\%20berry&sortBy=date, 2235-http://www.cnn.com/search/?query=man\%20results&sortBy=date, 2236-http://www.cnn.com/search/?query=lewis\%20james&primaryType=video&sortBy=date&intl=false, 2237-http://www.cnn.com/search/?query=hard\%20rock\%20place&primaryType=video&sortBy=date&intl=false, 2238-http://www.cnn.com/search/?query=holy\%20youth&primaryType=video&sortBy=date&intl=false, 2239-http://www.cnn.com/search/?query=a\%20hello\%20berry&primaryType=video&sortBy=date&intl=false, 2240-http://www.cnn.com/search/?query=man\%20results&primaryType=video&sortBy=date&intl=false, 2241-http://www.cnn.com/SHOWBIZ/, 2242-http://www.cnn.com/POLITICS, 2243-http://www.cnn.com/LIVING, 2244-http://money.cnn.com/, 2245-http://money.cnn.com/quote/quote.html?symb=INTC, 2246-http://money.cnn.com/quote/quote.html?symb=MSFT, 2247-http://money.cnn.com/quote/quote.html?symb=AAPL, 2248-http://money.cnn.com/quote/quote.html?symb=BAC, 2249-http://money.cnn.com/quote/quote.html?symb=CSCO, 2250-http://espn.go.com, 2251-http://espn.go.com/nfl/story/_/id/8973985/atlanta-falcons-likely-release-michael-turner-sources, 2252-http://espn.go.com/nfl/story/_/id/8973973/philadelphia-eagles-ask-cb-nnamdi-asomugha-restructure-contract-released-sources, 2253-http://espn.go.com/college-sports/story/_/id/8973498/bobby-valentine-athletics-director-sacred-heart-pioneers, 2254-http://espn.go.com/nfl, 2255-http://espn.go.com/nba, 2256-http://espn.go.com/mlb, 2257-http://espn.go.com/nhl, 2258-http://espn.go.com/college-football, 2259-http://espn.go.com/mens-college-basketball, 2260-http://soccernet.espn.espn.go.com, 2261-http://www.ask.com, 2262-http://www.ask.com/answers/301602801/what-s-the-best-way-to-lower-your-cholesterol-with-out-any-medication?qsrc=4034, 2263-http://www.ask.com/answers/301719201/how-can-i-make-money-from-my-blog?qsrc=4034, 2264-http://www.ask.com/answers/301625281/what-s-the-best-way-to-prepare-for-the-act-test?qsrc=4034, 2265-http://www.weather.com/weather/right-now/27514, 2266-http://www.weather.com/weather/right-now/27517, 2267-http://www.weather.com/weather/right-now/27510, 2268-http://www.weather.com/weather/right-now/30319, 2269-http://www.weather.com/weather/right-now/30332, 2270-http://www.weather.com/weather/5-day/27514, 2271-http://www.weather.com/weather/5-day/27517, 2272-http://www.weather.com/weather/5-day/27510, 2273-http://www.weather.com/weather/5-day/30319, 2274-http://www.weather.com/weather/5-day/30332, 2275-http://www.weather.com/, 2276-http://www.bankofamerica.com, 2277-http://www.bankofamerica.com/smallbusiness, 2278-http://www.bankofamerica.com/planning/investment.go, 2279-http://corp.bankofamerica.com/business/bi/home, 2280-http://wordpress.com, 2281-http://jxpaton.wordpress.com/, 2282-https://www.chase.com, 2283-https://www.chase.com/business-banking, 2284-https://www.chase.com/online/commercial-bank/commercial-bank.htm, 2285-http://www.imdb.com, 2286-http://www.imdb.com/movies-in-theaters/, 2287-http://www.imdb.com/title/tt2024432/, 2288-http://www.imdb.com/title/tt0882977/, 2289-http://www.imdb.com/title/tt2387433/, 2290-http://www.imdb.com/title/tt2024432/?ref_=inth_ov_vi#lb, 2291-http://www.imdb.com/tv, 2292-http://www.imdb.com/media/rm3340280320/tt2269550?slideshow=1, 2293-http://www.imdb.com/features/video/tv/, 2294-http://www.imdb.com/boards/, 2295-http://www.imdb.com/tvgrid/2013-02-09/, 2296-http://www.imdb.com/tv/blog, 2297-http://www.imdb.com/news, 2298-http://www.microsoft.com, 2299-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=lewis\%20james, 2300-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=hard\%20rock\%20place, 2301-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=holy\%20youth, 2302-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=a\%20hello\%20berry, 2303-http://search.microsoft.com/en-us/results.aspx?form=MSHOME&setlang=en-us&q=man\%20results, 2304-http://www.avg.com/us-en/homepage, 2305-http://www.apple.com, 2306-http://store.apple.com/us, 2307-http://store.apple.com/us/seach?find=ipad, 2308-http://store.apple.com/us/buy/home/shop_ipad/family/ipad_mini?product=MD528LL/A, 2309-http://www.apple.com/search/?q=users\%20story, 2310-http://support.apple.com/kb/index?page=search&product=&q=lewis\%20james&src=support_site, 2311-http://support.apple.com/kb/index?page=search&product=&q=hard\%20rock\%20place&src=support_site, 2312-http://support.apple.com/kb/index?page=search&product=&q=holy\%20youth&src=support_site, 2313-http://support.apple.com/kb/index?page=search&product=&q=a\%20hello\%20berry&src=support_site, 2314-http://support.apple.com/kb/index?page=search&product=&q=man\%20results&src=support_site, 2315-http://www.apple.com/mac, 2316-http://www.apple.com/ipod, 2317-http://www.apple.com/iphone, 2318-http://www.apple.com/ipad, 2319-http://www.about.com/, 2320-http://search.about.com/?q=man+results, 2321-http://search.about.com/?q=hard+rock+place, 2322-http://search.about.com/?q=lewis+james, 2323-http://search.about.com/?q=holy+youth, 2324-http://search.about.com/?q=a+hello+berry, 2325-http://www.about.com/compute/, 2326-http://www.about.com/education, 2327-http://www.about.com/autos, 2328-http://www.about.com/food, 2329-http://www.about.com/money, 2330-http://www.about.com/careers, 2331-https://www.wellsfargo.com, 2332-https://www.wellsfargo.com/biz, 2333-https://www.wellsfargo.com/com, 2334-http://www.foxnews.com, 2335-http://www.foxnews.com/us/2013/02/22/body-found-in-los-angeles-hotel-water-tank-needs-more-tests/?test=latestnews, 2336-http:

//www.foxnews.com/world/2013/02/22/japan-identifies-spate-boeing-787-jet-problems-but-still-investigating/?test=latestnews, 2337-http://www.foxnews.com/us/2013/02/22/woman-admits-leaving-baby-to-die-on-us-road/?test=latestnews, 2338-http://video.foxnews.com/, 2339-http://www.foxnews.com/search-results/search?q=lewis+james&submit=Search, 2340-http://www.foxnews.com/search-results/search?q=hard+rock+place&submit=Search, 2341-http://www.foxnews.com/search-results/search?q=holy+youth&submit=Search, 2342-http://www.foxnews.com/search-results/search?q=a+hello+berry&submit=Search, 2343-http://www.foxnews.com/search-results/search?q=man+results&submit=Search, 2344-http://www.foxnews.com/search-results/search?&submit=Search&q=lewis+james&mc_Text=192245&mc_Video=85334&mc_Blog=3834&mc_Slideshow=779&mediatype=Text, 2345-http://www.foxnews.com/search-results/search?&submit=Search&q=hard+rock+place&mc_Text=192245&mc_Video=85334&mc_Blog=3834&mc_Slideshow=779&mediatype=Text, 2346-http://www.foxnews.com/search-results/search?&submit=Search&q=holy+youth&mc_Text=192245&mc_Video=85334&mc_Blog=3834&mc_Slideshow=779&mediatype=Text, 2347-http://www.foxnews.com/search-results/search?&submit=Search&q=a+hello+berry&mc_Text=192245&mc_Video=85334&mc_Blog=3834&mc_Slideshow=779&mediatype=Text, 2348-http://www.foxnews.com/search-results/search?&submit=Search&q=man+results&mc_Text=192245&mc_Video=85334&mc_Blog=3834&mc_Slideshow=779&mediatype=Text, 2349-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=lewis+james&mc_Blog=3834&mc_Video=85334&mediatype=Video, 2350-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=hard+rock+place&mc_Blog=3834&mc_Video=85334&mediatype=Video, 2351-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=holy+youth&mc_Blog=3834&mc_Video=85334&mediatype=Video, 2352-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=a+hello+berry&mc_Blog=3834&mc_Video=85334&mediatype=Video, 2353-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=man+results&mc_Blog=3834&mc_Video=85334&mediatype=Video, 2354-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=hard+rock+place&mc_Blog=3834&mc_Video=85334&mediatype=Blog, 2355-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=lewis+james&mc_Blog=3834&mc_Video=85334&mediatype=Blog, 2356-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=holy+youth&mc_Blog=3834&mc_Video=85334&mediatype=Blog, 2357-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=192245&q=a+hello+berry&mc_Blog=3834&mc_Video=85334&mediatype=Blog, 2358-http://www.foxnews.com/search-results/search?&mc_Slideshow=779&submit=Search&mc_Text=

192245&q=man+results&mc_Blog=3834&mc_Video=85334&mediatype=Blog, 2359-http://www.foxnews.com/politics/index.html, 2360-http://www.foxnews.com/entertainment/index.html, 2361-http://www.foxnews.com/leisure/index.html, 2362-http://www.foxnews.com/sports/index.html, 2363-http://www.foxnews.com/sports/football/index.html, 2364-http://www.foxnews.com/sports/basketball/index.html, 2365-http://www.foxnews.com/sports/hockey/index.html, 2366-http://www.foxnews.com/sports/baseball/index.html, 2367-http://www.foxnews.com/sports/college/index.html, 2368-http://www.foxnews.com/sports/tennis/index.html, 2369-http://www.foxnews.com/sports/nascar/index.html, 2370-http://www.foxnews.com/sports/golf/index.html, 2371-http://www.walmart.com, 2372-http://www.walmart.com/search/search-ng.do?search_query=lewis+james&ic=16_0&Find=Find&search_constraint=0, 2373-http://www.walmart.com/search/search-ng.do?search_query=hard+rock+place&ic=16_0&Find=Find&search_constraint=0, 2374-http://www.walmart.com/search/search-ng.do?search_query=a+hello+berry&ic=16_0&Find=Find&search_constraint=0, 2375-http://www.walmart.com/search/search-ng.do?search_query=holy+youth&ic=16_0&Find=Find&search_constraint=0, 2376-http://www.walmart.com/search/search-ng.do?search_query=man+results&ic=16_0&Find=Find&search_constraint=0, 2377-http://www.walmart.com/ip/Rabbids-Go-Home-A-Comedy-Adventure-Wii/12165824, 2378-http://www.walmart.com/ip/Your-Shape-Wii/12311695, 2379-http://www.walmart.com/ip/Secret-Of-The-Wings-Widescreen/20606771, 2380-http://www.walmart.com/ip/Sceptre-32-Class-LCD-720p-60Hz-HDTV-X322BV-HD/15739136, 2381-http://www.walmart.com/ip/Wall-Mount-Adjustable-DVD-Shelf-Black/17376632?findingMethod=rr, 2382-http://mywebsearch.com, 2383-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=lewis+james, 2384-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=hard+rock+place, 2385-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=holy+youth, 2386-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=a+hello+berry, 2387-http://search.mywebsearch.com/mywebsearch/GGmain.jhtml?searchfor=man+results, 2388-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=lewis+james, 2389-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=hard+rock+place, 2390-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=holy+youth, 2391-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=a+hello+berry, 2392-http://search.mywebsearch.com/mywebsearch/AJimage.jhtml?searchfor=man+results, 2393-http://search.mywebsearch.com/mywebsearch/news.jhtml?searchfor=lewis+james, 2394-http://search.mywebsearch.com/mywebsearch/news.jhtml?searchfor=hard+rock+place, 2395-http://search.mywebsearch.com/mywebsearch/news.jhtml?searchfor=holy+youth, 2396-http://search.mywebsearch.com/mywebsearch/news.jhtml?searchfor=a+hello+berry, 2397-http://search.mywebsearch.com/mywebsearch/news.jhtml?searchfor=man+results, 2398-http://search.mywebsearch.com/mywebsearch/video.jhtml?searchfor=lewis+james, 2399-

http://search.mywebsearch.com/mywebsearch/video.jhtml?searchfor=hard+rock+place, 2400-http://search.mywebsearch.com/mywebsearch/video.jhtml?searchfor=holy+youth, 2401-http://search.mywebsearch.com/mywebsearch/video.jhtml?searchfor=a+hello+berry, 2402-http://search.mywebsearch.com/mywebsearch/video.jhtml?searchfor=man+results, 2403-http://xfinity.comcast.net, 2404-http://search.comcast.net/?cat=web&con=betac&q=lewis+james, 2405-http://search.comcast.net/?cat=web&con=betac&q=hard+rock+place, 2406-http://search.comcast.net/?cat=web&con=betac&q=holy+youth, 2407-http://search.comcast.net/?cat=web&con=betac&q=a+hello+berry, 2408-http://search.comcast.net/?cat=web&con=betac&q=man+results, 2409-http://search.comcast.net/?q=lewis+james&cat=images, 2410-http://search.comcast.net/?q=hard+rock+place&cat=images, 2411-http://search.comcast.net/?q=holy+youth&cat=images, 2412-http://search.comcast.net/?q=a+hello+berry&cat=images, 2413-http://search.comcast.net/?q=man+results&cat=images, 2414-http://search.comcast.net/?cat=news&con=net&form_submit=1&q=lewis+james, 2415-http://search.comcast.net/?cat=news&con=net&form_submit=1&q=hard+rock+place, 2416-http://search.comcast.net/?cat=news&con=net&form_submit=1&q=holy+youth, 2417-http://search.comcast.net/?cat=news&con=net&form_submit=1&q=a+hello+berry, 2418-http://search.comcast.net/?cat=news&con=net&form_submit=1&q=man+results, 2419-http://www.nytimes.com/, 2420-http://www.nytimes.com/2013/02/23/business/global/daily-euro-zone-watch.html?hpw&_r=0, 2421-http://www.nytimes.com/reuters/2013/02/22/sports/olympics/22reuters-safrica-pistorius-ancwomen.html?hp, 2422-http://www.nytimes.com/reuters/2013/02/22/business/22reuters-volkswagen-results-pay.html?hp, 2423-http://query.nytimes.com/search/sitesearch/#/lewis+james, 2424-http://query.nytimes.com/search/sitesearch/#/hard+rock+place, 2425-http://query.nytimes.com/search/sitesearch/#/holy+youth, 2426-http://query.nytimes.com/search/sitesearch/#/a+hello+berry, 2427-http://query.nytimes.com/search/sitesearch/#/man+results, 2428-http://www.nytimes.com/most-popular, 2429-http://well.blogs.nytimes.com/2013/02/13/why-four-workouts-a-week-may-be-better-than-six/, 2430-http://www.nytimes.com/2013/02/10/opinion/sunday/relax-youll-be-more-productive.html?_r=0, 2431-http://www.nytimes.com/2013/02/17/arts/design/flood-control-in-the-netherlands-now-allows-sea-water-in.html, 2432-http://www.nytimes.com/pages/todayspaper/index.html, 2433-http://www.nytimes.com/video/, 2434-http://www.nytimes.com/pages/world/index.html, 2435-http://www.nytimes.com/pages/national/index.html, 2436-http://www.nytimes.com/pages/sports/index.html, 2437-http://www.nytimes.com/pages/sports/baseball/index.html, 2438-http://www.nytimes.com/pages/sports/basketball/index.html, 2439-http://www.nytimes.com/pages/sports/ncaafootball/index.html, 2440-http://www.nytimes.com/pages/sports/football/index.html, 2441-http://www.nytimes.com/pages/sports/hockey/index.html, 2442-http://www.nytimes.com/pages/sports/soccer/index.html, 2443-http://www.nytimes.com/

pages/sports/golf/index.html, 2444-http://www.nytimes.com/pages/sports/tennis/index.html, 2445-http://imgur.com/, 2446-http://imgur.com/gallery/6BTbkl0, 2447-http://imgur.com/gallery/ZNIAU, 2448-http://imgur.com/gallery/42l4Ytj, 2449-http://www.yelp.com, 2450-http://www.yelp.com/search?find_desc=lewis+james&find_loc=Cary\%2C+NC&ns=1, 2451-http://www.yelp.com/search?find_desc=hard+rock+place&find_loc=Cary\%2C+NC&ns=1, 2452-http://www.yelp.com/search?find_desc=holy+youth&find_loc=Cary\%2C+NC&ns=1, 2453-http://www.yelp.com/search?find_desc=a+hello+berry&find_loc=Cary\%2C+NC&ns=1, 2454-http://www.yelp.com/search?find_desc=man+results&find_loc=Cary\%2C+NC&ns=1, 2455-http://www.yelp.com/biz/taipei-101-cary#query:chinese\%20stew, 2456-http://www.yelp.com/biz/grand-asia-market-raleigh#query:chinese\%20stew, 2457-http://www.yelp.com/biz/waraji-japanese-restaurant-raleigh#query:chinese\%20stew, 2458-http://www.ehow.com, 2459-http://www.ehow.com/search.html?s=lewis+james&skin=corporate&t=all, 2460-http://www.ehow.com/search.html?s=hard+rock+place&skin=corporate&t=all, 2461-http://www.ehow.com/search.html?s=holy+youth&skin=corporate&t=all, 2462-http://www.ehow.com/search.html?s=a+hello+berry&skin=corporate&t=all, 2463-http://www.ehow.com/search.html?s=man+results&skin=corporate&t=all, 2464-http://www.ehow.com/search.html?s=lewis+james&skin=corporate&t=article, 2465-http://www.ehow.com/search.html?s=hard+rock+place&skin=corporate&t=article, 2466-http://www.ehow.com/search.html?s=holy+youth&skin=corporate&t=article, 2467-http://www.ehow.com/search.html?s=a+hello+berry&skin=corporate&t=article, 2468-http://www.ehow.com/search.html?s=man+results&skin=corporate&t=article, 2469-http://www.ehow.com/search.html?s=lewis+james&skin=corporate&t=video, 2470-http://www.ehow.com/search.html?s=hard+rock+place&skin=corporate&t=video, 2471-http://www.ehow.com/search.html?s=holy+youth&skin=corporate&t=video, 2472-http://www.ehow.com/search.html?s=a+hello+berry&skin=corporate&t=video, 2473-http://www.ehow.com/search.html?s=man+results&skin=corporate&t=video, 2474-http://www.ehow.com/video_4975639_ride-bicycle.html, 2475-http://www.ehow.com/video_4974888_is-riding-bike-good-exercise.html, 2476-http://www.ehow.com/how_2002837_ride-bicycle.html, 2477-http://www.ehow.com/how_2001409_yourself-ride-bicycle-ride-bike.html, 2478-http://instagram.com, 2479-http://blog.instagram.com/, 2480-http://instagram.com/seanwes, 2481-http://instagram.com/seanmmadden, 2482-http://www.babylon.com/, 2483-http://www.conduit.com/, 2484-http://blog.conduit.com/, 2485-http://www.conduit.com/Search.aspx?cx=010301873083402539744:nxaq5wgrtuo&cof=FORID:11&ie=UTF-8&sa=Search&q=lewis\%20james, 2486-http://www.conduit.com/Search.aspx?cx=010301873083402539744:nxaq5wgrtuo&cof=FORID:11&ie=UTF-8&sa=Search&q=hard\%20rock\%20place, 2487-http://www.conduit.com/Search.aspx?cx=010301873083402539744:nxaq5wgrtuo&cof=FORID:11&ie=UTF-8&sa=Search&q=holy\%20youth, 2488-http://www.conduit.com/Search.aspx?cx=010301873083402539744:nxaq5wgrtuo&cof=FORID:11&ie=UTF-8&sa=Search&q=

a\%20hello\%20berry, 2489-http://www.conduit.com/Search.aspx?cx=010301873083402539744:nxaq5wgrtuo&cof=FORID:11&ie=UTF-8&sa=Search&q=man\%20results, 2490-http://www.etsy.com/, 2491-http://www.etsy.com/search?q=lewis\%20james&view_type=gallery&ship_to=US, 2492-http://www.etsy.com/search?q=hard\%20rock\%20place&view_type=gallery&ship_to=US, 2493-http://www.etsy.com/search?q=holy\%20youth&view_type=gallery&ship_to=US, 2494-http://www.etsy.com/search?q=a\%20hello\%20berry&view_type=gallery&ship_to=US, 2495-http://www.etsy.com/search?q=man\%20results&view_type=gallery&ship_to=US, 2496-http://www.etsy.com/listing/102976362/30-rock-i-want-to-go-to-there-liz-lemon?ref=sr_gallery_1&ga_search_query=go+there&ga_view_type=gallery&ga_ship_to=US&ga_search_type=all, 2497-http://www.etsy.com/listing/84346654/im-going-to-be-a-big-brother-shirt-big?ref=sr_gallery_2&ga_search_query=go+there&ga_view_type=gallery&ga_ship_to=US&ga_search_type=all, 2498-http://www.zedo.com/, 2499-http://www.zedo.com/news/, 2500-http://www.zedo.com/blog/, 2501-http://www.zedo.com/?s=lewis+james, 2502-http://www.zedo.com/?s=hard+rock+place, 2503-http://www.zedo.com/?s=holy+youth, 2504-http://www.zedo.com/?s=a+hello+berry, 2505-http://www.zedo.com/?s=man+results, 2506-http://www.cnet.com/, 2507-http://reviews.cnet.com/1770-5_7-0.html?query=lewis+james&tag=srch&searchtype=products, 2508-http://reviews.cnet.com/1770-5_7-0.html?query=hard+rock+place&tag=srch&searchtype=products, 2509-http://reviews.cnet.com/1770-5_7-0.html?query=holy+youth&tag=srch&searchtype=products, 2510-http://reviews.cnet.com/1770-5_7-0.html?query=a+hello+berry&tag=srch&searchtype=products, 2511-http://reviews.cnet.com/1770-5_7-0.html?query=man+results&tag=srch&searchtype=products, 2512-http://reviews.cnet.com/sedan/2013-honda-accord/4505-10865_7-35426697.html, 2513-http://reviews.cnet.com/suv/2013-bmw-x1-xdrive28i/4505-10868_7-35602742.html, 2514-http://news.cnet.com/, 2515-http://news.cnet.com/8301-1023_3-57569310-93/judge-tosses-some-shareholder-suits-over-facebooks-ipo-flop/, 2516-http://www.nbc.com, 2517-http://www.nbc.com/search?q=lewis+james, 2518-http://www.nbc.com/search?q=hard+rock+place, 2519-http://www.nbc.com/search?q=holy+youth, 2520-http://www.nbc.com/search?q=a+hello+berry, 2521-http://www.nbc.com/search?q=man+results, 2522-http://www.nbc.com/search?&q=lewis\%20james&mediatype=Video, 2523-http://www.nbc.com/search?&q=hard\%20rock\%20place&mediatype=Video, 2524-http://www.nbc.com/search?&q=holy\%20youth&mediatype=Video, 2525-http://www.nbc.com/search?&q=a\%20hello\%20berry&mediatype=Video, 2526-http://www.nbc.com/search?&q=man\%20results&mediatype=Video, 2527-http://www.nbc.com/schedule/, 2528-http://www.nbc.com/video/, 2529-http://www.nbc.com/shows/, 2530-http://www.flickr.com, 2531-http://www.flickr.com/search/?q=lewis\%20james, 2532-http://www.flickr.com/search/?q=hard\%20rock\%20place, 2533-http://www.flickr.com/search/?q=holy\%20youth, 2534-http://www.flickr.com/search/?q=a\%20hello\%20berry, 2535-http://www.flickr.com/

search/?q=man\%20results, 2536-http://www.flickr.com/photos/teche/385202303/, 2537-http://www.flickr.com/photos/silviosousacabral/2761392392/, 2538-http://www.outbrain.com/, 2539-http://www.outbrain.com/blog/, 2540-http://www.outbrain.com/blog/2013/02/google-attempts-to-redefine-mobile-market.html, 2541-http://www.outbrain.com/blog/2012/12/how-great-content-can-help-increase-your-holiday-traffic-and-sales.html, 2542-http://www.outbrain.com/amplify/, 2543-http://www.outbrain.com/engage/, 2544-http://www.hulu.com/, 2545-http://www.hulu.com/search?q=lewis+james, 2546-http://www.hulu.com/search?q=hard+rock+place, 2547-http://www.hulu.com/search?q=holy+youth, 2548-http://www.hulu.com/search?q=a+hello+berry, 2549-http://www.hulu.com/search?q=man+results, 2550-http://www.hulu.com/watch/229341, 2551-http://optmd.com/, 2552-http://optmd.com/about.html, 2553-http://optmd.com/optout.html, 2554-http://www.pandora.com/, 2555-http://www.pandora.com/station/play/1279370045779007469, 2556-http://www.pandora.com/station/play/2054685791586280609, 2557-http://www.pandora.com/station/play/2054686715004249249, 2558-http://www.pandora.com/station/play/2054687307709736097, 2559-http://www.pandora.com/station/play/2054687664192021665, 2560-http://www.pandora.com/station/play/2054688501710644385, 2561-http://www.pandora.com/station/play/2054689107301033121, 2562-http://www.pandora.com/station/play/2054689742956192929, 2563-http://www.pandora.com/search/lewis\%20james\%20, 2564-http://www.pandora.com/search/hard\%20rock\%20place\%20, 2565-http://www.pandora.com/search/holy\%20youth\%20, 2566-http://www.pandora.com/search/a\%20hello\%20berry\%20, 2567-http://www.pandora.com/search/man\%20results\%20, 2568-http://www.intuit.com/, 2569-http://quickbooks.intuit.com/, 2570-http://payroll.intuit.com/payroll_services/, 2571-http://payroll.intuit.com/payroll_resources/payroll_101/, 2572-http://payments.intuit.com/, 2573-http://quickbooks.intuit.com/search/small-business.jsp?searchTerm=lewis+james, 2574-http://quickbooks.intuit.com/search/small-business.jsp?searchTerm=hard+rock+place, 2575-http://quickbooks.intuit.com/search/small-business.jsp?searchTerm=holy+youth, 2576-http://quickbooks.intuit.com/search/small-business.jsp?searchTerm=a+hello+berry, 2577-http://quickbooks.intuit.com/search/small-business.jsp?searchTerm=man+results, 2578-http://www.reddit.com/, 2579-http://t.co/, 2580-http://thepiratebay.se/, 2581-http://thepiratebay.se/search/lewis\%20james, 2582-http://thepiratebay.se/search/hard\%20rock\%20place, 2583-http://thepiratebay.se/search/holy\%20youth, 2584-http://thepiratebay.se/search/a\%20hello\%20berry, 2585-http://thepiratebay.se/search/man\%20results, 2586-http://thepiratebay.se/torrent/7359487/How_Math_Can_Save_Your_Life_-_And_Make_You_Rich__Help_You_Find_T, 2587-http://www.target.com/, 2588-http://www.target.com/s?searchTerm=lewis+james, 2589-http://www.target.com/s?searchTerm=hard+rock+place, 2590-http://www.target.com/s?searchTerm=holy+youth, 2591-http://www.target.com/s?searchTerm=a+hello+berry, 2592-http://www.target.com/s?searchTerm=man+results, 2593-http://www.target.

com/p/cryin-won-t-help-you/-/A-12017970#prodSlot=medium_1_2, 2594-http://www.target.com/p/believe/-/A-14161776?reco=Rec\%7Cpdp\%7C14161776\%7CPopularProductsInCategory\%7Citem_page.vertical_1&lnk=Rec\%7Cpdp\%7CPopularProductsInCategory\%7Citem_page.vertical_1, 2595-http://www.pch.com/unrecognized, 2596-http://www.bestbuy.com/, 2597-http://www.bestbuy.com/site/Won't+You+Help+Me+to+Raise+-+CD/10117946.p?id=91438&skuId=10117946&st=help\%20me\%20and\%20you&lp=1&cp=1, 2598-http://www.bestbuy.com/site/Greatest+Hits\%3A+If+You+Can't+Help+Me\%2C+Please...+-+CD/6457041.p?id=1359836&skuId=6457041&st=help\%20me\%20and\%20you&lp=2&cp=1, 2599 -http://www.blogger.com, 2600-http://www.adobe.com/, 2601-http://www.adobe.com/products/catalog.html?promoid=KAWQI, 2602-http://www.adobe.com/products/creativecloud.html, 2603-http://www.adobe.com/products/creativesuite/design-web-premium.html?promoid=KCHGT, 2604-http://www.adobe.com/solutions.html?promoid=KAWQJ, 2605-http://helpx.adobe.com/support.html?promoid=KAWQK, 2606-http://www.adobe.com/cfusion/search/index.cfm?term=lewis+james&loc=en_us&siteSection=support.html\%3Fpromoid\%3DKAWQK, 2607-http://www.adobe.com/cfusion/search/index.cfm?term=hard+rock+place&loc=en_us&siteSection=support.html\%3Fpromoid\%3DKAWQK, 2608-http://www.adobe.com/cfusion/search/index.cfm?term=holy+youth&loc=en_us&siteSection=support.html\%3Fpromoid\%3DKAWQK, 2609-http://www.adobe.com/cfusion/search/index.cfm?term=a+hello+berry&loc=en_us&siteSection=support.html\%3Fpromoid\%3DKAWQK, 2610-http://www.adobe.com/cfusion/search/index.cfm?term=man+results&loc=en_us&siteSection=support.html\%3Fpromoid\%3DKAWQK, 2611-http://www.indeed.com/, 2612-http://www.indeed.com/jobs?q=lewis+james&l=Chapel+Hill\%2C+NC, 2613-http://www.indeed.com/jobs?q=hard+rock+place&l=Chapel+Hill\%2C+NC, 2614-http://www.indeed.com/jobs?q=holy+youth&l=Chapel+Hill\%2C+NC, 2615-http://www.indeed.com/jobs?q=a+hello+berry&l=Chapel+Hill\%2C+NC, 2616-http://www.indeed.com/jobs?q=man+results&l=Chapel+Hill\%2C+NC, 2617-http://www.indeed.com/cmp/The-Dharma-House/jobs/Sales-Representative-Sales-Manager-28773a922407f358, 2618-https://www.pcrecruiter.net/pcrbin/reg5.exe?db=odT2cGE2KBfpQpYkM\%2fS47ef\%2bBr00QQ\%3d\%3d&id=102339118452252&src=Indeed&rid=www\%2Eindeed\%2Ecom, 2619-https://www.pcrecruiter.net/pcrbin/reg5.exe?db=odT2cGE2KBfpQpYkM\%2fS47ef\%2bBr00QQ\%3d\%3d&id=130685852797461&src=Indeed&rid=www\%2Eindeed\%2Ecom, 2620-http://www.indeed.com/p/viewjob.php?pid=1963660370408977&id=1640946, 2621-https://www.usps.com/, 2622-https://www.usps.com/ship/ship-a-package.htm, 2623-https://www.usps.com/send/send-mail.htm, 2624-https://www.usps.com/manage/manage-your-mail.htm, 2625-https://store.usps.com/store/, 2626-https://store.usps.com/store/browse/productDetailSingleSku.jsp?categoryNav=false&navAction=jump&navCount=0&productId=S_470404&categoryId=, 2627-https://store.usps.com/store/browse/uspsProductDetailMultiSkuDropDown.jsp?categoryNav=false&navAction=jump&navCount=2&productId=S_788704&categoryId=, 2628-http://www.answers.com/, 2629-

http://www.answers.com/topic/go-diego-go-animated-tv-series-2005-children-tv-series, 2630-http://wiki.answers.com/Q/Does_choclate_help_you_go_to_bed, 2631-http://wiki.answers.com/Q/Where_can_drug_abusers_go_to_get_help, 2632-http://www.answers.com/T/Sports, 2633-http://www.answers.com/T/Technology, 2634-http://www.answers.com/T/Entertainment_and_Arts, 2635-http://www.att.com, 2636-http://www.att.com/gen/general?pid=11627, 2637-http://www.att.com/shop/, 2638-http://www.irs.gov, 2639-http://search.irs.gov/search?q=lewis+james, 2640-http://search.irs.gov/search?q=hard+rock+place, 2641-http://search.irs.gov/search?q=holy+youth, 2642-http://search.irs.gov/search?q=a+hello+berry, 2643-http://search.irs.gov/search?q=man+results, 2644-http://www.irs.gov/Filing, 2645-http://www.irs.gov/Payments, 2646-http://www.irs.gov/Refunds, 2647-http://www.irs.gov/Credits-&-Deductions, 2648-http://www.irs.gov/uac/Small-Business-Health-Care-Tax-Credit-for-Small-Employers, 2649-http://www.irs.gov/Individuals/EITC-Home-Page--Its-easier-than-ever-to-find-out-if-you-qualify-for-EITC, 2650-http://www.irs.gov/uac/Tax-Benefits-for-Education:-Information-Center, 2651-http://www.reference.com/, 2652-http://thesaurus.com/, 2653-http://dictionary.reference.com/, 2654-http://quotes.dictionary.com/, 2655-http://dynamo.dictionary.com/, 2656-http://dictionary.reference.com/browse/smart?s=t, 2657-http://dictionary.reference.com/browse/help?s=t, 2658-http://dictionary.reference.com/browse/rude?s=t, 2659-http://thesaurus.com/browse/rude?s=t, 2660-http://thesaurus.com/browse/help?s=t, 2661-http://thesaurus.com/browse/smart?s=t, 2662-http://www.ups.com/, 2663-http://www.ups.com/content/us/en/index.jsx, 2664-https://www.ups.com/one-to-one/login?returnto=https\%3a//www.ups.com/myWorkspace/home\%3floc\%3den_US\%26WT.svl\%3dPriNav&reasonCode=-1, 2665-http://www.ups.com/content/us/en/shipping/index.html?WT.svl=PriNav, 2666-http://www.ups.com/WebTracking/track?loc=en_US&WT.svl=PriNav, 2667-http://www.ups.com/content/us/en/freight/index.html?WT.svl=PriNav, 2668-http://www.ups.com/search/quick?loc=en_US&results=25&view=both&query=help+you&searchButton=, 2669-http://www.godaddy.com/, 2670-http://support.godaddy.com/search/all/lewis+james/, 2671-http://support.godaddy.com/search/all/hard+rock+place/, 2672-http://support.godaddy.com/search/all/holy+youth/, 2673-http://support.godaddy.com/search/all/a+hello+berry/, 2674-http://support.godaddy.com/search/all/man+results/, 2675-http://www.groupon.com/, 2676-http://www.groupon.com/browse/raleigh-durham, 2677-http://www.groupon.com/browse/atlanta, 2678-http://www.zillow.com/, 2679-http://www.zillow.com/homes/University-of-North-Carolina_rb/, 2680-http://www.zillow.com/homedetails/111-Borden-Ave-Wilmington-NC-28403/54302349_zpid/, 2681-http://www.zillow.com/homedetails/2206-Gibson-Ave-Wilmington-NC-28403/54310299_zpid/, 2682-http://www.deviantart.com/, 2683-http://browse.deviantart.com/?qh=&section=&global=1&q=lewis+james, 2684-http://browse.deviantart.com/?qh=&section=&global=1&q=a+hello+berry, 2685-http://browse.

deviantart.com/?qh=&section=&global=1&q=hard+rock+place, 2686-http://browse.deviantart.com/?qh=&section=&global=1&q=holy+youth, 2687-http://browse.deviantart.com/?qh=&section=&global=1&q=man+results, 2688-http://avotius.deviantart.com/art/Go-40657251, 2689-http://spammishrice.deviantart.com/art/Go-For-It-93434234, 2690-http://www.wikia.com/Wikia, 2691-http://metalgear.wikia.com/wiki/Metal_Gear_Wiki, 2692-http://narutoanw.wikia.com/wiki/Alternate_Naruto_World_Wiki, 2693-http://naruto.wikia.com/wiki/Kakashi_Hatake, 2694-http://www.pof.com/, 2695-http://www.bbc.com/, 2696-http://www.bbc.com/news/, 2697-http://www.bbc.co.uk/news/business-21544328, 2698-http://www.bbc.co.uk/news/technology-21547947, 2699-http://www.bbc.co.uk/news/world-europe-21544991, 2700-http://www.bbc.co.uk/weather/, 2701-http://www.bbc.co.uk/weather/6057856, 2702-http://www.match.com, 2703-https://www.capitalone.com/, 2704-http://www.capitalone.com/credit-cards/?Log=1&EventType=Link&ComponentType=T&LOB=MTS::L0RT6ME8Z&SubLob=&PageName=Home\%20Page\%20C&PortletLocation=2&ComponentName=primary_nav&ComponentStrategy=&ContentElement=5\%3BCredit+Cards&TargetLob=MTS\%3A\%3ALCTMMQC4S&TargetPageName=Credit+Cards+Home&linkid=&email_delivery_id=&referer=http://www.capitalone.com/homepage&external_id=, 2705-http://www.capitalone.com/directbanking/?Log=1&EventType=Link&ComponentType=T&LOB=MTS::LCTMMQC4S&SubLob=&PageName=Credit\%20Cards\%20Home&PortletLocation=2&ComponentName=primary_nav&ComponentStrategy=&ContentElement=73\%3BBanking&TargetLob=MTS\%3A\%3ALCTMNE8UU&TargetPageName=Personal+Banking&linkid=&email_delivery_id=&referer=http://www.capitalone.com/homepage&external_id=, 2706-http://www.capitalone.com/loans/?Log=1&EventType=Link&ComponentType=T&LOB=MTS::LCTMNE8UU&SubLob=MTS::KV0LVIE8Z&PageName=Personal\%20Banking&PortletLocation=2&ComponentName=primary_nav&ComponentStrategy=&ContentElement=130\%3BLoans&TargetLob=MTS\%3A\%3ALCTMNE8UU&TargetPageName=Loans+Home&linkid=&email_delivery_id=&referer=&external_id=, 2707-http://www.directrev.com/, 2708-http://www.directrev.com/technology/, 2709-http://www.aweber.com/, 2710-http://www.aweber.com/email-marketing-features.htm, 2711-http://www.aweber.com/pricing.htm, 2712-http://www.dailymail.co.uk/ushome/index.html, 2713-http://www.dailymail.co.uk/money/index.html, 2714-http://www.dailymail.co.uk/money/news/article-2282806/Sterling-claws-ground-euro-EU-admits-eurozone-recession-2014.html, 2715-http://www.dailymail.co.uk/money/pensions/article-2282180/Retirees-lose-3600-year-claiming-pension-credit.html, 2716-http://www.dailymail.co.uk/money/markets/article-2282506/Global-stock-markets-hammered-French-fuel-eurozone-recession-fears.html, 2717-http://www.dailymail.co.uk/news/index.html, 2718-http://www.dailymail.co.uk/sport/index.html, 2719-http://www.dailymail.co.uk/sport/headlines/index.html, 2720-http://www.dailymail.co.uk/sport/football/index.html, 2721-http://www.dailymail.co.uk/sport/cricket/index.html, 2722-http://www.dailymail.co.uk/sport/boxing/index.html, 2723-http://www.dailymail.co.uk/sport/tennis/index.html, 2724-http://www.dailymail.co.uk/sport/golf/index.html, 2725-http://www.dailymail.co.uk/sport/racing/index.html, 2726-http://www.dailymail.co.uk/health/index.html, 2727-http://drudgereport.com/, 2728-http://www.drudgereportArchives.com/dsp/search.htm?searchFor=lewis+james, 2729-http://www.drudgereportArchives.com/dsp/search.htm?searchFor=hard+rock+place, 2730-http://www.drudgereportArchives.com/dsp/search.htm?searchFor=holy+youth, 2731-http://www.drudgereportArchives.com/dsp/search.htm?searchFor=a+hello+berry, 2732-http://www.drudgereportArchives.com/dsp/search.htm?searchFor=man+results, 2733-http://www.verizonwireless.com/b2c/index.html, 2734-http://www22.verizon.com/?lid=//global//residential, 2735-http://www22.verizon.com/home/verizonglobalhome/ghp_business.aspx, 2736-http://www.rr.com/, 2737-http://www.rr.com/news/topic/article/rr/9000/81864000/Storm_slows_Midwest_commute_buries_Plains_in_snow, 2738-http://www.rr.com/news/topic/article/rr/9000/81872359/Olive_Garden_owner_Darden_warns_on_3rd_quarter, 2739-http://www.rr.com/news/topic/article/rr/9000/81860193/Guatemala_Probing_reports_drug_lord_may_be_dead, 2740-http://www.rr.com/weather/weatherchannel/, 2741-http://www.rr.com/weather/weatherchannel/USGA0028, 2742-http://www.rr.com/weather/weatherchannel/USNC0120, 2743-http://www.rr.com/weather/weatherchannel/USNC0105, 2744-http://search.rr.com/#web/lewis\%20james/1/, 2745-http://search.rr.com/#web/hard\%20rock\%20place/1/, 2746-http://search.rr.com/#web/holy\%20youth/1/, 2747-http://search.rr.com/#web/a\%20hello\%20berry/1/, 2748-http://search.rr.com/#web/man\%20results/1/, 2749-http://search.rr.com/#rrimage/lewis\%20james/1/, 2750-http://search.rr.com/#rrimage/hard\%20rock\%20place/1/, 2751-http://search.rr.com/#rrimage/man\%20results/1/, 2752-http://search.rr.com/#rrimage/holy\%20youth/1/, 2753-http://search.rr.com/#rrimage/a\%20hello\%20berry/1/, 2754-http://search.rr.com/#rrvideo/lewis\%20james/1/, 2755-http://search.rr.com/#rrvideo/hard\%20rock\%20place/1/, 2756-http://search.rr.com/#rrvideo/holy\%20youth/1/, 2757-http://search.rr.com/#rrvideo/a\%20hello\%20berry/1/, 2758-http://search.rr.com/#rrvideo/man\%20results/1/, 2759-http://search.rr.com/#rrnews/hard\%20rock\%20place/1/, 2760-http://search.rr.com/#rrnews/lewis\%20james/1/, 2761-http://search.rr.com/#rrnews/a\%20hello\%20berry/1/, 2762-http://search.rr.com/#rrnews/holy\%20youth/1/, 2763-http://search.rr.com/#rrnews/man\%20results/1/, 2764-http://www.rr.com/news/news, 2765-http://www.rr.com/sports/sports, 2766-http://scoreboard.rr.com/sports.asp?sport=NFL, 2767-http://scoreboard.rr.com/sports.asp?sport=NBA, 2768-http://scoreboard.rr.com/sports.asp?sport=MLB, 2769-http://scoreboard.rr.com/sports.asp?sport=CBK, 2770-http://scoreboard.rr.com/sports.asp?sport=CFB, 2771-http://scoreboard.rr.com/sports.asp?sport=MLS, 2772-http://scoreboard.rr.com/sports.asp?sport=GOLF, 2773-http://scoreboard.rr.com/sports.asp?sport=NHL, 2774-http://scoreboard.rr.com/sports.asp?sport=TENNIS, 2775-

http://www.bankrate.com/partners/rdr/personal-finance.aspx, 2776-http://rr.websol.barchart.com/?module=stockDetail&selected=overview&symbol=\%24DOWI&lang=EN, 2777-http://rr.websol.barchart.com/?redirect=&module=stockDetail&redirect=&selected=overview&lang=&1_symbol=INTX&lang=&uniqueid=&1_symbol=INTC&x=0&y=0, 2778-http://rr.websol.barchart.com/?redirect=&module=stockDetail&redirect=&selected=overview&lang=&1_symbol=INTC&lang=&uniqueid=&1_symbol=MSFT&x=0&y=0, 2779-http://rr.websol.barchart.com/?redirect=&module=stockDetail&redirect=&selected=overview&lang=&1_symbol=MSFT&lang=&uniqueid=&1_symbol=CSCO&x=0&y=0, 2780-http://rr.websol.barchart.com/?redirect=&module=stockDetail&redirect=&selected=overview&lang=&1_symbol=MSFT&lang=&uniqueid=&1_symbol=AAPL&x=0&y=0, 2781-http://rr.websol.barchart.com/?redirect=&module=stockDetail&redirect=&selected=overview&lang=&1_symbol=MSFT&lang=&uniqueid=&1_symbol=BAC&x=0&y=0, 2782-http://www.rr.com/entertainment/entertainment, 2783-http://beta.photobucket.com/?fromLegacy=true, 2784-http://beta.photobucket.com/browse, 2785-http://beta.photobucket.com/images/help\%20me, 2786-http://media.beta.photobucket.com/user/schroder11/media/helpmeeat_zps5f461c53.png.html?filters[term]=help\%20me&filters[primary]=images&filters[secondary]=videos&sort=1&o=0, 2787-http://media.beta.photobucket.com/user/Tokis-Phoenix/media/Morning\%20Musume/Sayumi\%20Michishige/SHelpme19_zps3bdcd0b3.jpg.html?filters[term]=help\%20me&filters[primary]=images&filters[secondary]=videos&sort=1&o=13, 2788-http://incredibar.com/essentials/homepage, 2789-http://incredibar.com/music/homepage, 2790-http://incredibar.com/games/homepage, 2791-http://bleacherreport.com/, 2792-http://bleacherreport.com/articles/1538747-orlando-magic-reportedly-trade-jj-redick-to-milwaukee-bucks, 2793-http://bleacherreport.com/articles/1539624-atlanta-falcons-will-reportedly-release-rb-michael-turner, 2794-http://bleacherreport.com/articles/1539230-rory-mcilroy-and-tiger-woods-eliminated-at-wgc-accenture-match-play-championship, 2795-http://bleacherreport.com/nfl, 2796-http://bleacherreport.com/nba, 2797-http://bleacherreport.com/college-football, 2798-http://bleacherreport.com/mlb, 2799-http://bleacherreport.com/nhl, 2800-http://bleacherreport.com/college-basketball, 2801-http://bleacherreport.com/world-football, 2802-http://bleacherreport.com/nascar, 2803-http://bleacherreport.com/search?q=lewis+james, 2804-http://bleacherreport.com/search?q=hard+rock+place, 2805-http://bleacherreport.com/search?q=holy+youth, 2806-http://bleacherreport.com/search?q=a+hello+berry, 2807-http://bleacherreport.com/search?q=man+results, 2808-http://bleacherreport.com/articles/1520702-pain-and-self-loathing-pushed-me-to-brink-says-craig-spearman?search_query=help\%20me, 2809-http://bleacherreport.com/articles/1520078-how-well-know-when-andrew-bynum-is-back-to-full-strength-with-sixers?search_query=help\%20me, 2810-http://www.washingtonpost.com/, 2811-http://www.washingtonpost.com/newssearch/search.html?st=lewis+james&submit=Submit, 2812-http:

//www.washingtonpost.com/newssearch/search.html?st=hard+rock+place&submit=Submit, 2813-http://www.washingtonpost.com/newssearch/search.html?st=holy+youth&submit=Submit, 2814-http://www.washingtonpost.com/newssearch/search.html?st=a+hello+berry&submit=Submit, 2815-http://www.washingtonpost.com/newssearch/search.html?st=man+results&submit=Submit, 2816-http://www.washingtonpost.com/politics, 2817-http://www.washingtonpost.com/opinions, 2818-http://www.washingtonpost.com/local, 2819-http://www.washingtonpost.com/sports, 2820-http://www.washingtonpost.com/national, 2821-http://www.washingtonpost.com/world, 2822-http://www.washingtonpost.com/business, 2823-http://www.washingtonpost.com/business/economy/north-carolinas-jobless-face-a-double-whammy-of-aid-reductions/2013/02/14/6e32fa2c-7601-11e2-95e4-6148e45d7adb_story.html, 2824-http://washpost.bloomberg.com/Story?docId=1376-MI7NYH07SXKX01-5DKO2LGO4I1GQV00RCFT58EJB6, 2825-http://www.usatoday.com/, 2826-http://www.usatoday.com/news/, 2827-http://www.usatoday.com/sports/, 2828-http://www.usatoday.com/sports/nfl/, 2829-http://www.usatoday.com/sports/mlb/, 2830-http://www.usatoday.com/sports/nba/, 2831-http://www.usatoday.com/sports/nhl/, 2832-http://www.usatoday.com/sports/ncaaf/, 2833-http://www.usatoday.com/sports/ncaab/, 2834-http://www.usatoday.com/sports/nascar/, 2835-http://www.usatoday.com/life/, 2836-http://www.usatoday.com/money/, 2837-http://www.usatoday.com/money/lookup/stocks/INTC/?, 2838-http://www.usatoday.com/money/lookup/stocks/MSFT/?, 2839-http://www.usatoday.com/money/lookup/stocks/AAPL/?, 2840-http://www.usatoday.com/money/lookup/stocks/BAC/?, 2841-http://www.usatoday.com/money/lookup/stocks/CSCO/?, 2842-http://www.usatoday.com/story/money/business/2013/02/14/john-kerry-portfolio-heinz-deal/1920009/, 2843-http://www.usatoday.com/story/tech/columnist/talkingtech/2013/02/13/atlanta-for-tech-startups/1911353/, 2844-http://www.usatoday.com/story/tech/gaming/2013/02/14/after-burner-climax/1919707/, 2845-http://www.fedex.com/, 2846-http://www.fedex.com/us/, 2847-http://www.fedex.com/us/ship/, 2848-http://www.fedex.com/us/track/, 2849-http://www.fedex.com/us/manage/, 2850-http://m.facebook.com/, 2851-https://m.facebook.com/officialraylewis, 2852-https://m.facebook.com/ladygaga, 2853-https://m.facebook.com/DonaldTrump, 2854-https://m.facebook.com/barackobama, 2855-https://m.facebook.com/McDonalds, 2856-http://m.youtube.com/, 2857-http://m.youtube.com/results?client=mv-google&hl=en&gl=US&q=search+me&submit=Search, 2858-http://m.yahoo.com/, 2859-http://m.yahoo.com/w/search\%3B_ylt=A2KLt8oAfx5RoygArSEp89w4?submit=oneSearch&.ysid=.kN6oSWLKFD889zpx1Kbp_Be&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=lewis+james&x=0&y=0, 2860-http://m.yahoo.com/w/search\%3B_ylt=A2KLt8oAfx5RoygArSEp89w4?submit=oneSearch&.ysid=.kN6oSWLKFD889zpx1Kbp_Be&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=hard+rock+place&x=0&y=0, 2861-http://m.yahoo.com/w/search\%3B_ylt=A2KLt8oAfx5RoygArSEp89w4?submit=oneSearch&.ysid=.kN6oSWLKFD889zpx1Kbp_Be&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=holy+youth&x=0&y=0,

2862-http://m.yahoo.com/w/search\%3B_ylt=A2KLt8oAfx5RoygArSEp89w4?submit=oneSearch&.ysid=.kN6oSWLKFD889zpx1Kbp_Be&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=a+hello+berry&x=0&y=0, 2863-http://m.yahoo.com/w/search\%3B_ylt=A2KLt8oAfx5RoygArSEp89w4?submit=oneSearch&.ysid=.kN6oSWLKFD889zpx1Kbp_Be&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=man+results&x=0&y=0, 2864-http://m.yahoo.com/w/legobpengine/finance/details/\%3B_ylt=A2KL8wrYgB5RzisAhgop89w4?symbol=\%5EDJI&.intl=us&.lang=en, 2865-http://m.yahoo.com/w/legobpengine/finance/details/?.b=quotessearch\%2F&symbol=INTC&.ss=INTC&.ts=1360953610&.intl=us&.lang=en, 2866-http://m.yahoo.com/w/legobpengine/finance/details/?.b=quotessearch\%2F&symbol=MSFT&.ss=MSFT&.ts=1360953610&.intl=us&.lang=en, 2867-http://m.yahoo.com/w/legobpengine/finance/details/?.b=quotessearch\%2F&symbol=AAPL&.ss=AAPL&.ts=1360953610&.intl=us&.lang=en, 2868-http://m.yahoo.com/w/legobpengine/finance/details/?.b=quotessearch\%2F&symbol=BAC&.ss=BAC&.ts=1360953610&.intl=us&.lang=en, 2869-http://m.yahoo.com/w/legobpengine/finance/details/?.b=quotessearch\%2F&symbol=CSCO&.ss=CSCO&.ts=1360953610&.intl=us&.lang=en, 2870-http://m.yahoo.com/w/ygo-weather/forecast.bp\%3B_ylt=A2KL8x1PgR5RGx4Ahwwp89w4?l=USCA1116\%7C2502265&.intl=US&.lang=en, 2871-http://m.yahoo.com/w/ygo-weather/forecast.bp?l=2424766&.intl=US&.lang=en, 2872-http://m.yahoo.com/w/legobpengine/news/?.intl=us&.lang=en-US, 2873-http://m.yahoo.com/w/sports?.intl=us&.lang=en, 2874-http://m.yahoo.com/w/sports/all/?.ts=1360953905&.intl=us&.lang=en, 2875-http://m.yahoo.com/w/sports/nba/?.ts=1360953935&.intl=us&.lang=en, 2876-http://m.yahoo.com/w/sports/nfl/?.ts=1360953994&.intl=us&.lang=en, 2877-http://m.yahoo.com/w/sports/ncaaf/?.ts=1360954018&.intl=us&.lang=en, 2878-http://m.yahoo.com/w/sports/nhl/?.ts=1360954041&.intl=us&.lang=en, 2879-http://m.yahoo.com/w/sports/mlb/?.ts=1360954064&.intl=us&.lang=en, 2880-http://m.ebay.com/, 2881-http://m.ebay.com/Pages/SearchResults.aspx?sv=lewis\%20james&emvcc=0&nbcol=0\%7Cnull, 2882-http://m.ebay.com/Pages/SearchResults.aspx?sv=hard\%20rock\%20place&emvcc=0&nbcol=0\%7Cnull, 2883-http://m.ebay.com/Pages/SearchResults.aspx?sv=holy\%20youth&emvcc=0&nbcol=0\%7Cnull, 2884-http://m.ebay.com/Pages/SearchResults.aspx?sv=a\%20hello\%20berry&emvcc=0&nbcol=0\%7Cnull, 2885-http://m.ebay.com/Pages/SearchResults.aspx?sv=man\%20results&emvcc=0&nbcol=0\%7Cnull, 2886-http://m.ebay.com/Pages/ViewItem.aspx?aid=270832717214&sv=help\%20me\%20there&emvcc=0&nbcol=0\%7Cnull, 2887-http://en.m.wikipedia.org/?useformat=mobile, 2888-http://en.m.wikipedia.org/wiki/Physics, 2889-http://en.m.wikipedia.org/wiki/Adam_Morrison, 2890-http://en.m.wikipedia.org/wiki/Michael_Jordan, 2891-http://en.m.wikipedia.org/wiki/University_of_North_Carolina_at_Chapel_Hill, 2892-http://en.m.wikipedia.org/wiki/North_Carolina, 2893-http://www.amazon.com/gp/aw, 2894-http://www.amazon.com/gp/aw/s/ref=is_box_?k=lewis+james, 2895-http://www.amazon.com/gp/aw/s/ref=is_box_?k=hard+rock+place, 2896-http://www.amazon.com/gp/aw/s/ref=is_box_?k=holy+youth, 2897-http://www.amazon.com/gp/aw/s/ref=is_box_?k=a+hello+berry, 2898-http://www.amazon.com/gp/aw/s/ref=is_box_?k=man+results, 2899-http://www.amazon.com/gp/aw/d/0385739869/ref=mp_s_a_1?qid=1360954565&sr=8-2, 2900-http://www.amazon.com/gp/aw/d/1455512796/ref=mr_books_bs_p1_, 2901-http://www.amazon.com/gp/aw/d/1451695195/ref=mr_books_bs_p1_, 2902-https://mobile.twitter.com/signup, 2903-https://mobile.twitter.com/search?q=\%23mobile, 2904-https://mobile.twitter.com/raylewis, 2905-https://mobile.twitter.com/ladygaga, 2906-https://mobile.twitter.com/realDonaldTrump, 2907-https://mobile.twitter.com/barackobama, 2908-https://mobile.twitter.com/McDonalds, 2909-http://m.bing.com/, 2910-http://m.bing.com/search?q=lewis+james&FORM=BLXBSS&btsrc=internal, 2911-http://m.bing.com/search?q=hard+rock+place&FORM=BLXBSS&btsrc=internal, 2912-http://m.bing.com/search?q=holy+youth&FORM=BLXBSS&btsrc=internal, 2913-http://m.bing.com/search?q=a+hello+berry&FORM=BLXBSS&btsrc=internal, 2914-http://m.bing.com/search?q=man+results&FORM=BLXBSS&btsrc=internal, 2915-http://m.bing.com/images/search?q=lewis+james&FORM=ILXBSS&btsrc=internal, 2916-http://m.bing.com/images/search?q=hard+rock+place&FORM=ILXBSS&btsrc=internal, 2917-http://m.bing.com/images/search?q=holy+youth&FORM=ILXBSS&btsrc=internal, 2918-http://m.bing.com/images/search?q=a+hello+berry&FORM=ILXBSS&btsrc=internal, 2919-http://m.bing.com/images/search?q=man+results&FORM=ILXBSS&btsrc=internal, 2920-http://m.bing.com/news/search?q=lewis+james&form=IRXBSS&IIG=aff146eb3dd04ea4b66ac721d648f378&kval=4.1&AppNs=mSERP, 2921-http://m.bing.com/news/search?q=hard+rock+place&form=IRXBSS&IIG=aff146eb3dd04ea4b66ac721d648f378&kval=4.1&AppNs=mSERP, 2922-http://m.bing.com/news/search?q=a+hello+berry&form=IRXBSS&IIG=aff146eb3dd04ea4b66ac721d648f378&kval=4.1&AppNs=mSERP, 2923-http://m.bing.com/news/search?q=holy+youth&form=IRXBSS&IIG=aff146eb3dd04ea4b66ac721d648f378&kval=4.1&AppNs=mSERP, 2924-http://m.bing.com/news/search?q=man+results&form=IRXBSS&IIG=aff146eb3dd04ea4b66ac721d648f378&kval=4.1&AppNs=mSERP, 2925-https://touch.www.linkedin.com/#login, 2926-http://m.blogspot.com/, 2927-http://m.blogspot.com/search?q=lewis+james, 2928-http://m.blogspot.com/search?q=hard+rock+place, 2929-http://m.blogspot.com/search?q=holy+youth, 2930-http://m.blogspot.com/search?q=a+hello+berry, 2931-http://m.blogspot.com/search?q=man+results, 2932-http://m.pinterest.com/, 2933-http://m.pinterest.com/search/pins/?q=lewis+james, 2934-http://m.pinterest.com/search/pins/?q=hard+rock+place, 2935-http://m.pinterest.com/search/pins/?q=holy+youth, 2936-http://m.pinterest.com/search/pins/?q=a+hello+berry, 2937-http://m.pinterest.com/search/pins/?q=man+results, 2938-http://m.pinterest.com/pin/171207223306514510/, 2939-http://m.pinterest.com/pin/171207223306558969/, 2940-http://m.pinterest.com/pin/171207223306316751/, 2941-http://m.pinterest.com/pin/171207223306255087/, 2942-http://m.pinterest.com/pin/171207223306391098/, 2943-http://onmobile.msn.com/, 2944-http://m.aol.com/portal/, 2945-http://m.aol.com/portal/directory.do?tab=Directory&icid=tb_dir,

2946-http://m.aol.com/dailyfinance/default/home.do?icid=dr_dailyfin, 2947-http://m.aol.com/dailyfinance/default/basicStock.do?symbol=INTC&exchange=NAS&icid=df_getquote, 2948-http://m.aol.com/dailyfinance/default/basicStock.do?symbol=MSFT&exchange=NAS&icid=df_getquote, 2949-http://m.aol.com/dailyfinance/default/basicStock.do?symbol=AAPL&exchange=NAS&icid=df_getquote, 2950-http://m.aol.com/dailyfinance/default/basicStock.do?symbol=BAC&exchange=NAS&icid=df_getquote, 2951-http://m.aol.com/dailyfinance/default/basicStock.do?symbol=CSCO&exchange=NAS&icid=df_getquote, 2952-http://m.mapquest.com/, 2953-http://m.aol.com/search/aol/search?q=lewis+james&invocationType=srch_entr, 2954-http://m.aol.com/search/aol/search?q=hard+rock+place&invocationType=srch_entr, 2955-http://m.aol.com/search/aol/search?q=holy+youth&invocationType=srch_entr, 2956-http://m.aol.com/search/aol/search?q=a+hello+berry&invocationType=srch_entr, 2957-http://m.aol.com/search/aol/search?q=man+results&invocationType=srch_entr, 2958-http://m.aol.com/moviefone/default/home.do?icid=dr_mf, 2959-http://m.aol.com/weather/default/home.do?icid=dr_weather, 2960-http://m.autoblog.com/?icid=dr_autoblog, 2961-http://www.huffingtonpost.com/blackberry/, 2962-http://www.huffingtonpost.com/blackberry/p.html?id=2741494, 2963-http://www.huffingtonpost.com/blackberry/p.html?id=2739652, 2964-http://www.huffingtonpost.com/blackberry/p.html?id=2740960, 2965-http://www.huffingtonpost.com/blackberry/p.html?id=2696325, 2966-http://m.sportingnews.com/, 2967-http://m.sportingnews.com/sport/story/2013-02-22/oscar-pistorius-case-bail-hearing-shooting-death-murder-reeva-steenkamp, 2968-http://m.sportingnews.com/nfl/story/2013-02-21/wr-nate-burleson-restructures-contract-with-lions, 2969-http://m.sportingnews.com/mlb/story/2013-02-22/rockies-michael-cuddyer-tougher-penalties-ped-steroids-biogenesis-ryan-braun, 2970-http://m.sportingnews.com/nfl, 2971-http://m.sportingnews.com/mlb, 2972-http://m.sportingnews.com/nba, 2973-http://m.sportingnews.com/nhl, 2974-http://m.sportingnews.com/ncaa-football, 2975-http://m.tumblr.com/, 2976-http://m.tumblr.com/image/43047343328, 2977-http://m.tumblr.com/image/42497881029, 2978-http://m.tumblr.com/search/lewis+james, 2979-http://m.tumblr.com/search/hard+rock+place, 2980-http://m.tumblr.com/search/holy+youth, 2981-http://m.tumblr.com/search/a+hello+berry, 2982-http://m.tumblr.com/search/man+results, 2983-https://mobile.paypal.com/us/cgi-bin/wapapp?cmd=_wapapp-homepage, 2984-http://www.weather.com/mobile/wap.html, 2985-http://m.espn.go.com/wireless/index?w=1cm8g&i=MCOM, 2986-http://m.espn.go.com/nfl/story?storyId=8973985, 2987-http://m.espn.go.com/wireless/story?storyId=8973498, 2988-http://m.espn.go.com/nfl/, 2989-http://m.espn.go.com/mlb/, 2990-http://m.espn.go.com/nba/, 2991-http://m.espn.go.com/nhl/, 2992-http://m.espn.go.com/ncf/, 2993-http://m.espn.go.com/ncb/, 2994-https://www.bankofamerica.com/mobile/, 2995-http://m.wordpress.com/, 2996-https://mobilebanking.chase.com/, 2997-http://m.imdb.com/, 2998-http://m.imdb.com/title/tt2024432/, 2999-http://m.imdb.com/title/tt0882977/, 3000-http://m.imdb.com/title/tt2387433/, 3001-http://m.imdb.com/find?q=lewis+james&button.x=0&button.y=0&button=Search, 3002-http://m.imdb.com/find?q=hard+rock+place&button.x=0&button.y=0&button=Search, 3003-http://m.imdb.com/find?q=holy+youth&button.x=0&button.y=0&button=Search, 3004-http://m.imdb.com/find?q=a+hello+berry&button.x=0&button.y=0&button=Search, 3005-http://m.imdb.com/find?q=man+results&button.x=0&button.y=0&button=Search, 3006-http://m.imdb.com/title/tt0104410/, 3007-http://m.avg.com/, 3008-http://m.microsoft.com/en-us/default.mspx, 3009-http://m.microsoft.com/en-us/Products/default.mspx?prodtype=windows, 3010-http://www.foxnews.mobi/, 3011-http://www.foxnews.mobi/quickPage.html?page=38321&content=89358298, 3012-http://www.foxnews.mobi/quickPage.html?page=38321&content=89359227, 3013-http://www.foxnews.mobi/quickPage.html?page=38321&content=89358299, 3014-http://www.foxnews.mobi/quickPage.html?page=18573&cc=location.html, 3015-http://www.foxnews.mobi/quickPage.html?page=38321&content=88928914, 3016-http://www.foxnews.mobi/quickPage.html?page=38321&content=89358298, 3017-http://www.foxnews.mobi/quickPage.html?page=38321&content=89332450, 3018-http://world.foxnews.mobi/, 3019-http://sports.foxnews.mobi/, 3020-https://m.wellsfargo.com/, 3021-http://mobile.walmart.com/, 3022-http://mobile.walmart.com/ip/16621480, 3023-http://mobile.walmart.com/ip/14322438, 3024-http://m.comcast.net/m/, 3025-http://m.comcast.net/m/weather/, 3026-http://m.comcast.net/m/weather/30319/, 3027-http://m.comcast.net/m/weather/30332/, 3028-http://m.comcast.net/m/weather/27517/, 3029-http://m.comcast.net/m/weather/27510/, 3030-http://m.comcast.net/m/weather/27514/, 3031-http://m.comcast.net/m/news/, 3032-http://m.comcast.net/m/news/news/, 3033-http://m.comcast.net/m/articles/news-general/20130215/US.Jesse.Jackson.Jr/, 3034-http://m.comcast.net/m/articles/news-general/20130222/ML.Egypt/, 3035-http://m.comcast.net/m/news/news-sports/, 3036-http://m.imgur.com/, 3037-http://news.mobile.msn.com/en-us/default.aspx, 3038-http://news.mobile.msn.com/en-us/sports.aspx, 3039-http://news.mobile.msn.com/en-us/article_spt.aspx?aid=50829742&afid=19, 3040-http://news.mobile.msn.com/en-us/article_spt.aspx?aid=50829421&afid=19, 3041-http://news.mobile.msn.com/en-us/business.aspx, 3042-http://news.mobile.msn.com/en-us/article_biz.aspx?aid=16130963&afid=16, 3043-http://news.mobile.msn.com/en-us/article_biz.aspx?aid=16130394&afid=16, 3044-http://m.flickr.com/#/home, 3045-http://m.flickr.com/#/search/advanced/_QM_q_IS_lewis+james, 3046-http://m.flickr.com/#/search/advanced/_QM_q_IS_hard+rock+place, 3047-http://m.flickr.com/#/search/advanced/_QM_q_IS_holy+youth, 3048-http://m.flickr.com/#/search/advanced/_QM_q_IS_a+hello+berry, 3049-http://m.flickr.com/#/search/advanced/_QM_q_IS_man+results, 3050-http://m.flickr.com/#/photos/irestylianou/8475115153/in/search_QM_q_IS_help+me, 3051-http://m.flickr.com/#/photos/ronsombilongallery/5249681492/in/search_QM_q_IS_help+me, 3052-http://m.intuit.com/, 3053-http://m.intuit.com/#quickbooks, 3054-http://m.intuit.com/#payroll, 3055-http://www.reddit.com/.compact, 3056-http://m.target.com/, 3057-http://m.target.com/mcategories, 3058-http://m.target.com/store-locator/find-stores, 3059-http://m.target.com/s?category=0\%7CAll\%7Cmatchallany\%7Call+categories&searchTerm=lewis+james&x=0&y=0, 3060-http://m.target.com/s?category=0\%7CAll\%7Cmatchallany\%7Call+categories&searchTerm=hard+rock+place&x=0&y=0, 3061-http://m.target.com/s?category=0\%7CAll\%7Cmatchallany\%7Call+categories&searchTerm=holy+youth&x=0&y=0, 3062-http://m.target.com/s?category=0\%7CAll\%7Cmatchallany\%7Call+categories&searchTerm=a+hello+berry&x=0&y=0, 3063-http://m.target.com/s?category=0\%7CAll\%7Cmatchallany\%7Call+categories&searchTerm=man+results&x=0&y=0, 3064-http://m.target.com/p/awake-my-soul-help-me-to-sing/-/A-11374583, 3065-http://m.target.com/p/help-me-mr-mutt-hardcover/-/A-12704446, 3066-http://www.blogger.com/mobile-start.g, 3067-http://m.bestbuy.com/m/b/, 3068-http://m.usps.com/, 3069-http://m.usps.com/MobileTrackPackage.aspx, 3070-http://m.irs.gov/mt/www.irs.gov, 3071-http://www.answers.com/, 3072-http://www.answers.com/topic/love, 3073-https://m.ups.com/mobile/home, 3074-http://m.att.com/, 3075-http://www.att.com/gen/general?pid=11627, 3076-http://m.att.com/shopmobile/find-a-store.html, 3077-http://m.dictionary.com/r/, 3078-http://m.dictionary.com/d/?q=help, 3079-http://m.dictionary.com/d/?q=rude, 3080-http://m.dictionary.com/d/?q=smart, 3081-http://www.godaddymobile.com/, 3082-http://m.bbc.co.uk/sport, 3083-http://m.bbc.co.uk/weather, 3084-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www, 3085-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www/products, 3086-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www/atm, 3087-http://touch.match.com/, 3088-http://www.idrudgereport.com/, 3089-https://m.verizonwireless.com/, 3090-https://m.verizonwireless.com/explore, 3091-https://m.verizonwireless.com/shop, 3092-https://m.verizonwireless.com/myverizon, 3093-http://m.rr.com/, 3094-http://m.rr.com/c.jsp?cid=25390571, 3095-http://m.rr.com/rss.jsp?rssid=25390791&item=http\%3a\%2f\%2fwww.rr.com\%2fservices\%2fpublicapi\%2fcontent\%2f1.0\%2f\%3fmethod\%3dgetMobileHeadlinesRss\%26ci\%3d81605210\%26csi\%3d55254959&cid=25372441, 3096-http://m.rr.com/rss.jsp?rssid=25390791&item=http\%3a\%2f\%2fwww.rr.com\%2fservices\%2fpublicapi\%2fcontent\%2f1.0\%2f\%3fmethod\%3dgetMobileHeadlinesRss\%26ci\%3d81622876\%26csi\%3d55254959&cid=25372441, 3097-http://m.usatoday.com/, 3098-http://m.usatoday.com/news, 3099-http://m.usatoday.com/sports, 3100-http://m.usatoday.com/sports/mlb, 3101-http://m.usatoday.com/sports/nba, 3102-http://m.usatoday.com/sports/nfl, 3103-

http://m.usatoday.com/sports/nhl, 3104-http://m.usatoday.com/sports/collfootball, 3105-http://m.usatoday.com/sports/collbasketball, 3106-http://m.usatoday.com/sports/golf, 3107-http://m.usatoday.com/sports/tennis, 3108-http://m.usatoday.com/sports/boxing, 3109-http://m.usatoday.com/sports/soccer, 3110-http://m.usatoday.com/money, 3111-http://m.usatoday.com/stocksSearchResults?ticker=INTC&submitButton=search, 3112-http://m.usatoday.com/stocksSearchResults?ticker=MSFT&submitButton=search, 3113-http://m.usatoday.com/stocksSearchResults?ticker=BAC&submitButton=search, 3114-http://m.usatoday.com/stocksSearchResults?ticker=AAPL&submitButton=search, 3115-http://m.usatoday.com/stocksSearchResults?ticker=CSCO&submitButton=search, 3116-http://m.usatoday.com/article/news/1923257, 3117-http://m.usatoday.com/article/news/1917367, 3118-http://m.fedex.com/mt/www.fedex.com, 3119-http://m.fedex.com/mt/www.fedex.com/us/?un_zip_uat=&un_jtt_redirect, 3120-http://m.fedex.com/mt/www.fedex.com/Tracking?cntry_code=us&un_zip_uat=, 3121-http://m.fedex.com/mt/www.fedex.com/us/?un_zip_uat=&un_jtt_v_target=login, 3122-http://www.amazon.com/gp/product/B00KFVCQ7Y/ref=atv_terms_dp, 3123-http://www.amazon.com/gp/product/B006GLLTL6/ref=dv_web_u_TH_s_l_1?pf_rd_p=1811853042&pf_rd_s=center-2&pf_rd_t=101&pf_rd_i=293883011&pf_rd_m=ATVPDKIKX0DER&pf_rd_r=0DVKQ123FWHGAW6JKRFS#, 3124-http://www.amazon.com/gp/product/B006VREHRS/ref=dv_web_u_TH_s_l_2?pf_rd_p=1811853042&pf_rd_s=center-2&pf_rd_t=101&pf_rd_i=293883011&pf_rd_m=ATVPDKIKX0DER&pf_rd_r=0DVKQ123FWHGAW6JKRFS#, 3125-http://www.amazon.com/gp/product/B006GLM5EQ/ref=s9_al_bw_g318_i3_a_l?pf_rd_p=1814658222&pf_rd_s=center-4&pf_rd_t=101&pf_rd_i=2858778011&pf_rd_m=ATVPDKIKX0DER&pf_rd_r=0Y08GTDMVJBS5TWGNHQH#, 3126-http://www.amazon.com/gp/product/B00688628M/ref=s9_al_bw_g318_i5_a_l?pf_rd_p=1814658222&pf_rd_s=center-4&pf_rd_t=101&pf_rd_i=2858778011&pf_rd_m=ATVPDKIKX0DER&pf_rd_r=0Y08GTDMVJBS5TWGNHQH, 3127-http://dropbox.com, 3128-http://mp3.com/top-downloads/genre/classical/, 3129-http://soundowl.com/track/6rco, 3130-http://mp3skull.com/, 3131-http://thepiratebay.se/torrent/10198961/Dangerous.Mind.of.a.Hooligan.2014.BDRip.X264-SONiDO, 3132-http://thepiratebay.se/torrent/10198903/Jagadeka_Veeruni_Katha_\%281961\%29_Telugu_Xvid_2cd_-_Eng_Subs_-_NTR__, 3133-http://thepiratebay.se/torrent/10198608/Pompeii, 3134-http://thepiratebay.se/torrent/10198235/Los_ojos_de_Julia, 3135-http://thepiratebay.se/torrent/10198963/\%5BSmooth_Jazz\%5D_Acoustic_Alchemy_-_Live_in_London_2014__320_\%28By_Ja, 3136-http://thepiratebay.se/torrent/10198832/Kitty_Cleveland_-_The_Miracle_of_Divine_Mercy_\%28Chaplet_in_Song\%29, 3137-http://thepiratebay.se/torrent/10198777/Jon_Foreman__curren__Discography_\%5B2007_-_2008\%5D, 3138-http://thepiratebay.se/torrent/10198758/Attack_on_Religious_Liberty__The_Battle_for_the_Faith_in_Mexico, 3139-http://www.youtube.com/watch?client=mv-google&hl=en&gl=US&v=SBpmuRmSgeg&nomobile=1, 3140-http://www.youtube.com/watch?gl=US&client=mv-google&hl=en&v=GcSBw76_tpk&nomobile=1,

3141-http://www.youtube.com/watch?hl=en&client=mv-google&gl=US&v=-Sa7nVpfxvE&nomobile=1, 3142-http://www.youtube.com/watch?hl=en&gl=US&client=mv-google&v=4QCXr79Rkcw&nomobile=1, 3143-http://www.youtube.com/watch?v=SyCDgj9MtYc, 3144-http://m.pornhub.com/video/show/title/sarah_lewis_interracial_gangbang/vkey/2131985194, 3145-http://www.pornhub.com/view_video.php?viewkey=2131985194, 3146-http://m.pornhub.com/video/show/title/aleska_dp_pt1/vkey/282934724, 3147-http://www.pornhub.com/view_video.php?viewkey=282934724, 3148-http://m.pornhub.com/video/show/title/two_for_a_slut/vkey/1045490591, 3149-http://www.pornhub.com/view_video.php?viewkey=1045490591, 3150-http://m.pornhub.com/video/show/title/glamour_prostitute_threesome_group_sex/vkey/50325729, 3151-http://www.pornhub.com/view_video.php?viewkey=50325729, 3152-http://xhamster.com/movies/625607/ginni_lewis_anal.html, 3153-http://xhamster.com/movies/643252/breezy_aka_sabrina_lewis.html, 3154-http://xhamster.com/movies/444220/hairy_sabrina.html, 3155-http://xhamster.com/movies/610760/sabrina_starr_ask_for.html, 3156-http://www.netflix.com/WiPlayer?movieid=70197057&trkid=13462050&tctx=1\%2C1\%2C9cfac4a1-c260-42ee-b068-b7c12fca297e-5116201, 3157-http://www.netflix.com/WiPlayer?movieid=70248290&trkid=13462100&tctx=-99\%2C-99\%2C9cfac4a1-c260-42ee-b068-b7c12fca297e-5116201, 3158-http://www.netflix.com/WiPlayer?movieid=70262640&trkid=13462050&tctx=1\%2C0\%2C9cfac4a1-c260-42ee-b068-b7c12fca297e-5116201, 3159-http://www.netflix.com/WiPlayer?movieid=70136117&trkid=13462293&tctx=2\%2C3\%2C9cfac4a1-c260-42ee-b068-b7c12fca297e-5116201, 3160-http://www.netflix.com/WiPlayer?movieid=70136120&trkid=13462293&tctx=2\%2C4\%2C9cfac4a1-c260-42ee-b068-b7c12fca297e-5116201, 3161-http://www.hulu.com/watch/638995, 3162-http://www.hulu.com/watch/637084, 3163-http://www.hulu.com/watch/639053, 3164-http://www.hulu.com/watch/639038, 3165-http://www.hulu.com/watch/638457, 3166-http://www.hulu.com/watch/640097, 3167-http://www.hulu.com/watch/640098, 3168-http://www.hulu.com/watch/470578, 3169-http://www.hulu.com/watch/528131, 3170-http://www.hulu.com/watch/621444, 3171-http://www.hulu.com/watch/634412, 3172-http://www.hulu.com/watch/631072, 3173-http://www.hulu.com/watch/635920, 3174-http://www.hulu.com/watch/631702, 3175-http://www.hulu.com/watch/635448, 3176-http://www.hulu.com/watch/289301, 3177-http://www.hulu.com/watch/300181, 3178-http://www.hulu.com/watch/181431, 3179-http://www.hulu.com/watch/183504, 3180-http://www.hulu.com/watch/186428, 3181-http://www.hulu.com/watch/583728, 3182-http://www.hulu.com/watch/626395, 3183-http://www.hulu.com/watch/630183, 3184-http://www.hulu.com/watch/633065, 3185-http://www.hulu.com/watch/91506, 3186-http://www.hulu.com/watch/91508, 3187-http://www.hulu.com/watch/207044, 3188-http://www.hulu.com/watch/207042, 3189-http://www.hulu.com/watch/498419, 3190-http://www.hulu.com/watch/207895, 3191-http://www.redtube.com/36982, 3192-http://www.redtube.com/47565, 3193-http://www.redtube.com/16479, 3194-http://www.redtube.com/11878, 3195-http://www.youporn.com/watch/

578172/backroom-cumshot-compilation-ii/?from=related3&al=2&from_id=578172&pos=3, 3196-http://www.youporn.com/watch/523631/erica-and-brittney-sorry-girls/?from=related3&al=2&from_id=523631&pos=1, 3197-http://www.youporn.com/watch/334250/she-s-not-buying-my-lie-fucks-me-anyhow/?from=related3&al=2&from_id=334250&pos=5, 3198-http://www.dailymotion.com/video/xz4f7k_macklemore-ryan-lewis-ft-ray-can-t-hold-us-official-music-video_music?search_algo=2, 3199-http://www.dailymotion.com/video/xz4f7k_macklemore-ryan-lewis-ft-ray-can-t-hold-us-official-music-video_music?search_algo=2, 3200-http://www.dailymotion.com/video/xzo6n0_ray-donovan-liev-schreiber_shortfilms, 3201-http://www.dailymotion.com/video/xxlcr4_the-reluctant-fundamentalist-trailer-hd-2012-mira-nair-kate-hudson-kiefer-sutherland-liev-schreiber_shortfilms#.UdnLnRy4wcM, 3202-http://www.worldstarhiphop.com/videos/video.php?v=wshhW7X2dkn9jnNoB8g8, 3203-http://www.worldstarhiphop.com/videos/video.php?v=wshhKBuRq3v50a9GOk7F, 3204-http://www.worldstarhiphop.com/videos/video.php?v=wshh2h4ig9zFaKz8A2jA, 3205-http://www.worldstarhiphop.com/videos/video.php?v=wshh2h4ig9zFaKz8A2jA, 3206-http://www.youtube.com/watch?v=W-UQsVW4-MU, 3207-http://www.youtube.com/watch?v=qpunQZ4cUyI, 3208-http://www.youtube.com/user/theinternshipmovie/featured?v=ehJFc1W0VKE, 3209-http://www.youtube.com/watch?v=3Wj8Yxa309E, 3210-http://www.youtube.com/user/personalliberty?v=P-Dky4zWkko, 3211-http://www.youtube.com/watch?v=ATG6oHSNSm0, 3212-http://www.youtube.com/watch?v=PZErG2QOv9c, 3213-http://www.youtube.com/watch?v=n-B_kmAebbQ, 3214-http://www.youtube.com/watch?v=z9P6fPXFht0, 3215-http://www.youtube.com/watch?v=p5ijbo-XeH4, 3216-http://www.youtube.com/watch?v=-rMMTv7XLYw, 3217-http://www.youtube.com/watch?v=Fae0j1WN1zA&list=PLzVF1nAqI9VmKRcgzZX0L0diFoApovY88, 3218-http://www.youtube.com/watch?v=hrIad1RVFV0, 3219-http://www.youtube.com/watch?v=voN8omBe2r4, 3220-http://www.youtube.com/watch?v=Ih5Mr93E-2c, 3221-http://www.youtube.com/watch?v=HkB9VJdu27M, 3222-http://www.youtube.com/watch?v=ml-a2HbtAWs, 3223-http://www.youtube.com/watch?v=eUxtmELaTss, 3224-http://www.youtube.com/watch?v=sm7bkc1REUI, 3225-http://www.youtube.com/watch?v=ZzAgktRXlPY, 3226-http://www.youtube.com/watch?v=5u550YOIfT0, 3227-http://www.youtube.com/watch?v=GMOgsuW15gM, 3228-http://www.youtube.com/watch?v=-lRjl1oiG-M, 3229-http://www.youtube.com/watch?v=eLlTcdp839E, 3230-http://www.youtube.com/watch?v=jjbRL0HdVwQ, 3231-http://www.youtube.com/watch?v=Nz8AXNi4wnI, 3232-http://www.youtube.com/watch?v=gr7KCp4Eui0, 3233-http://www.youtube.com/watch?v=blFgHPVOx6k, 3234-http://www.youtube.com/watch?v=sjYBJfAV__A, 3235-http://www.youtube.com/watch?v=S37NGQQadNg, 3236-http://www.youtube.com/watch?v=LYRhzYREk90, 3237-http://www.youtube.com/watch?v=LAr20-YOFgw, 3238-http://www.youtube.com/watch?v=lptVexxVmps, 3239-http://www.youtube.com/watch?v=vNWByUk22sI, 3240-http://www.youtube.com/watch?v=EAYcAEnq0vc, 3241-http://www.youtube.com/watch?v=jjbRL0HdVwQ, 3242-http://www.youtube.com/watch?v=

uECtg1-nnTA, 3243-http://www.youtube.com/watch?v=wg_cx741sBA, 3244-http://www.youtube.com/watch?v=FPFll_ResuA, 3245-http://www.youtube.com/watch?v=a8R43Yy6k7A, 3246-http://www.youtube.com/watch?v=smqnFHTVr2U, 3247-http://www.youtube.com/watch?v=fr3YMGNpFCc, 3248-http://www.youtube.com/watch?v=RarGG4did6Y, 3249-http://www.youtube.com/watch?v=P-2y1Ap31nc, 3250-http://www.youtube.com/watch?v=rs1g6uCYSz8, 3251-http://www.youtube.com/watch?v=GpUbS6IyEWg, 3252-http://www.youtube.com/watch?v=nBvBN17OFS8, 3253-http://www.youtube.com/watch?v=T0SZZYt_Wak, 3254-http://www.youtube.com/watch?v=1yF25q4OL_4, 3255-http://www.youtube.com/watch?v=INrz9w9ZIjU, 3256-http://www.netflix.com/WiPlayer?movieid=70288438&trkid=13462067&tctx=5\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3257-http://www.netflix.com/WiPlayer?movieid=70304979&trkid=13462073&tctx=7\%2C3\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3258-http://www.netflix.com/WiPlayer?movieid=60004481&trkid=13462055&tctx=8\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3259-http://www.netflix.com/WiPlayer?movieid=1171557&trkid=13462277&tctx=11\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3260-http://www.netflix.com/WiPlayer?movieid=70230151&trkid=13462075&tctx=13\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3261-http://www.netflix.com/WiPlayer?movieid=60028097&trkid=13462062&tctx=14\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3262-http://www.netflix.com/WiPlayer?movieid=70153391&trkid=13462055&tctx=16\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3263-http://www.netflix.com/WiPlayer?movieid=70166091&trkid=13462275&tctx=18\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3264-http://www.netflix.com/WiPlayer?movieid=70197037&trkid=13462275&tctx=18\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3265-http://www.netflix.com/WiPlayer?movieid=70299454&trkid=13462541&tctx=23\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3266-http://www.netflix.com/WiPlayer?movieid=70300066&trkid=13462064&tctx=27\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3267-http://www.netflix.com/WiPlayer?movieid=70108783&trkid=13462682&tctx=29\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3268-http://www.netflix.com/WiPlayer?movieid=70098333&trkid=13462577&tctx=31\%2C3\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3269-http://www.netflix.com/WiPlayer?movieid=60021957&trkid=13462577&tctx=31\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3270-http://www.netflix.com/WiPlayer?movieid=26198935&trkid=13462656&tctx=34\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3271-http://www.netflix.com/WiPlayer?movieid=70301595&trkid=13462055&tctx=36\%2C3\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3272-http://www.netflix.com/WiPlayer?movieid=70120143&trkid=13462292&tctx=-99\%2C-99\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3273-http://www.netflix.com/WiPlayer?movieid=70222627&trkid=13462292&tctx=-99\%2C-99\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3274-http:

//www.netflix.com/WiPlayer?movieid=60002777&trkid=13462274&tctx=33\%2C4\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3275-http://www.netflix.com/WiPlayer?movieid=70302491&trkid=13462656&tctx=34\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3276-http://www.netflix.com/WiPlayer?movieid=70266998&trkid=13462063&tctx=35\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3277-http://www.netflix.com/WiPlayer?movieid=70122321&trkid=13462055&tctx=36\%2C9\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3278-http://www.netflix.com/WiPlayer?movieid=70200744&trkid=13462055&tctx=36\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3279-http://www.netflix.com/WiPlayer?movieid=70266228&trkid=13462055&tctx=36\%2C4\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3280-http://www.netflix.com/WiPlayer?movieid=70285977&trkid=13462071&tctx=32\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3281-http://www.netflix.com/WiPlayer?movieid=70142822&trkid=13462071&tctx=32\%2C3\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3282-http://www.netflix.com/WiPlayer?movieid=70242803&trkid=13462071&tctx=32\%2C6\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3283-http://www.netflix.com/WiPlayer?movieid=70140403&trkid=13462286&tctx=30\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3284-http://www.netflix.com/WiPlayer?movieid=70181716&trkid=13462682&tctx=29\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3285-http://www.netflix.com/WiPlayer?movieid=269880&trkid=13462068&tctx=28\%2C1\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3286-http://www.netflix.com/WiPlayer?movieid=70241754&trkid=13462068&tctx=28\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3287-http://www.netflix.com/WiPlayer?movieid=70180057&trkid=13462286&tctx=30\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3288-http://www.netflix.com/WiPlayer?movieid=60021957&trkid=13462577&tctx=31\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3289-http://www.netflix.com/WiPlayer?movieid=70098333&trkid=13462577&tctx=31\%2C3\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3290-http://www.netflix.com/WiPlayer?movieid=70176656&trkid=13462577&tctx=31\%2C0\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3291-http://www.netflix.com/WiPlayer?movieid=70044883&trkid=13462577&tctx=31\%2C4\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3292-http://www.netflix.com/WiPlayer?movieid=70267269&trkid=13462656&tctx=34\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3293-http://www.netflix.com/WiPlayer?movieid=60011552&trkid=13462656&tctx=34\%2C3\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3294-http://www.netflix.com/WiPlayer?movieid=70039530&trkid=13462063&tctx=35\%2C4\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3295-http://www.netflix.com/WiPlayer?movieid=445386&trkid=13462656&tctx=34\%2C5\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311, 3296-http://www.netflix.com/WiPlayer?movieid=26656173&trkid=13462274&tctx=33\%2C2\%2C14ca88ce-4166-4500-8d01-e25f3144faa7-844311,

3297-https://www.sandvine.com/downloads/general/global-internet-phenomena/2013/sandvine-global-internet-phenomena-report-1h-2013.pdf, 3298-https://www.sandvine.com/downloads/general/global-internet-phenomena/2013/2h-2013-global-internet-phenomena-snapshot-na-fixed.pdf, 3299-http://www.google.com/patents/US7292531, 3300-http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.7975&rep=rep1&type=pdf, 3301-http://www.cs.jhu.edu/~cwright/hmm-extd-abstract.pdf, 3302-http://www.itk.ilstu.edu/faculty/ytang/traffic/QoS-TC-IMC04.pdf, 3303-http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.1482&rep=rep1&type=pdf, 3304-http://security.riit.tsinghua.edu.cn/~bhyang/paper_read/Sigmetrics2005_Internet\%20traffic\%20classification\%20using\%20bayesian\%20analysis\%20techniques.pdf, 3305-http://escholarship.org/uc/item/1wn9n8kt#page-2, 3306-https://www.google.com/?gws_rd=ssl, 3307-https://www.google.com/?gws_rd=ssl#q=yellowstone, 3308-https://www.google.com/?gws_rd=ssl#q=bose, 3309-https://www.google.com/?gws_rd=ssl#q=taylor+swift, 3310-https://www.facebook.com, 3311-https://www.facebook.com/YellowstoneNPS?ref=stream, 3312-https://www.facebook.com/TaylorSwift, 3313-https://www.facebook.com/Bose/info?tab=page_info, 3314-https://www.youtube.com, 3315-https://www.youtube.com/watch?v=e-ORhEE9VVg, 3316-https://www.youtube.com/watch?v=-XkewnRxv-A, 3317-https://www.youtube.com/watch?v=TKRUYnxqib4, 3318-https://www.yahoo.com, 3319-http://news.yahoo.com/portlands-ferguson-protest-boy-hugs-officer-174417767.html, 3320-https://www.yahoo.com/tech/confessions-of-a-smartphone-thief-101677365014.html, 3321-http://sports.yahoo.com/blogs/ncaaf-dr-saturday/ohio-state-qb-j-t--barrett-leaves-game-with-injury-200742837.html, 3322-http://www.baidu.com, 3323-http://www.baidu.com/s?ie=utf-8&f=8&rsv_bp=1&rsv_idx=1&tn=baidu&wd=yellowstone&rsv_pq=cece6992003bd5db&rsv_t=e335EAaF0pKFIlLtiHcWaMKO8tJWZV98cSUFfJrzBHT\%2FBVnXpUJfYWR8W50&rsv_enter=1&rsv_sug3=16&rsv_sug4=544&rsv_sug2=0&inputT=2674, 3324-http://www.baidu.com/s?ie=utf-8&f=8&rsv_bp=1&rsv_idx=1&tn=baidu&wd=bose&rsv_pq=abbf395a00380a42&rsv_t=d8700mptKf4Bk1o13v7Hf017CRlrPiQR7zyZgjrhrB3HXwWRANYbvbeKSmo&rsv_enter=1&rsv_sug3=4&rsv_sug2=0&inputT=592, 3325-http://www.baidu.com/s?ie=utf-8&f=8&rsv_bp=1&rsv_idx=1&tn=baidu&wd=taylor\%20swift&rsv_pq=99847bc1003605cb&rsv_t=b6780VvCNMinAngRp7JhXlm7zNFVWhLpyqR0HoAq1Q2XXUi8Wj5CFLMoDOU&rsv_enter=1&rsv_sug3=12&rsv_sug4=595&rsv_sug1=2&rsv_sug2=0&inputT=2176, 3326-http://www.amazon.com, 3327-http://www.amazon.com/gp/product/B00GLXKDPC/ref=s9_spu_gw_g121_i4?pf_rd_m=ATVPDKIKX0DER&pf_rd_s=desktop-2&pf_rd_r=07ST8FZAF98YX6BDB1HB&pf_rd_t=36701&pf_rd_p=1976974222&pf_rd_i=desktop, 3328-http://www.amazon.com/dp/B00HKH0I3A?psc=1, 3329-http://www.amazon.com/Foods-Psyllium-Husk-500mg-Capsules/dp/B0013OW2KS/ref=pd_sim_hpc_3?ie=UTF8&refRID=1FSJ3B7XADMEY39WKYSP, 3330-http://www.wikipedia.org, 3331-http://en.wikipedia.org/wiki/Yellowstone_National_Park, 3332-http://en.wikipedia.org/wiki/

Bose_Corporation, 3333-http://en.wikipedia.org/wiki/Taylor_Swift, 3334-http://www.taobao.com/market/global/index_new.php, 3335-http://detail.tmall.com/item.htm?spm=a230r.1.14.1.h8Hy7z&id=39129574563&ad_id=&am_id=&cm_id=140105335569ed55e27b&pm_id=&abbucket=14, 3336-http://detail.tmall.com/item.htm?spm=a220o.1000855.0.0.tMAu7D&id=36135055513, 3337-http://detail.tmall.com/item.htm?spm=a220o.1000855.1998099587.2.5GcDG5&id=41442682216&bi_from=tm_comb, 3338-https://twitter.com, 3339-https://twitter.com/bose, 3340-https://twitter.com/taylorswift13, 3341-https://twitter.com/yellowstonenps, 3342-http://www.qq.com, 3343-http://news.qq.com/a/20141130/001667.htm?tu_biz=1.114.1.0, 3344-http://sports.qq.com/nba/, 3345-http://ent.qq.com/star/, 3346-http://games.qq.com, 3347-http://dy.qq.com/article.htm?id=20141130A0000B00&tu_biz=v1, 3348-http://news.qq.com/photo.shtml, 3349-http://edu.qq.com/abroad/, 3350-https://www.google.co.in/?gws_rd=ssl, 3351-https://www.google.co.in/?gws_rd=ssl#q=yellowstone, 3352-https://www.google.co.in/?gws_rd=ssl#q=taylor+swift, 3353-https://www.google.co.in/?gws_rd=ssl#q=bose, 3354-https://login.live.com/, 3355-https://www.linkedin.com, 3356-https://www.linkedin.com/company/bose-corporation, 3357-https://www.linkedin.com/company/yellowstone-capital-llc, 3358-https://www.linkedin.com/pub/dir/Taylor/Swift, 3359-http://www.sina.com.cn, 3360-http://news.sina.com.cn/c/2014-11-29/201931222521.shtml, 3361-http://news.sina.com.cn/c/2014-11-29/174231222321.shtml, 3362-http://news.sina.com.cn/c/2014-11-29/184931222411.shtml, 3363-http://news.sina.com.cn/c/2014-11-29/194331222496.shtml, 3364-http://news.sina.com.cn/c/2014-11-29/221631222640.shtml, 3365-http://news.sina.com.cn/c/p/2014-11-29/201731222538.shtml, 3366-http://news.sina.com.cn/c/z/twllbt2014/, 3367-http://mil.news.sina.com.cn/2014-11-29/0942813430.html, 3368-http://www.tmall.com, 3369-http://brand.tmall.com/?spm=3.7396704.20000005.d1.vdejCk&abbucket=&acm=tt-1138874-37187.1003.8.74460&uuid=74460&abtest=&scm=1003.8.tt-1138874-37187.OTHER_1416855521756_74460&pos=2, 3370-http://www.tmall.hk/?spm=3.7396704.20000005.d8.vdejCk&abbucket=&acm=tt-1138874-37187.1003.8.74460&uuid=74460&abtest=&scm=1003.8.tt-1138874-37187.OTHER_1416359774024_74460&pos=8, 3371-http://nvzhuang.tmall.com/?spm=3.7396704.20000008.3.vdejCk&abbucket=&acm=tt-1142055-39052.1003.8.86935&uuid=86935&abtest=&scm=1003.8.tt-1142055-39052.OTHER_1418855675811_86935&pos=9, 3372-http://elegantprosper.tmall.com/?spm=a221t.7270053.1997467905.101.MGTbsa, 3373-http://detail.tmall.com/item.htm?spm=a1z10.1.w8860581-8809821628.3.a7V7Co&id=41856444812&scene=taobao_shop, 3374-http://detail.tmall.com/item.htm?spm=a220o.1000855.1998025129.2.CBDE4E&id=41777095499&abbucket=_AB-M32_B5&rn=&acm=03054.1003.1.115927&uuid=id3VVlDa_fyoBDbcwIk0CAUuxjsjCvwBr&abtest=_AB-LR32-PR32&scm=1003.1.03054.ITEM_41777095499_115927&pos=2, 3375-http://detail.tmall.com/item.htm?spm=a220o.1000855.1998025129.3.fmOWQg&id=41777571371&abbucket=_AB-M32_B5&rn=&acm=03054.1003.1.115927&uuid=XCoMVDJN_fyoBDbcwIk0CAUuxjsjCvwBr&abtest=_AB-

LR32-PR32&scm=1003.1.03054.ITEM_41777571371_115927&pos=3, 3376-http://us.weibo.com/gb, 3377-http://weibo.com/pingwest, 3378-http://weibo.com/2132734472/ByBMg7uHN, 3379-http://weibo.com/chinesewsj, 3380-http://weibo.com/p/1001603782194387653740?from=page_100206_profile&wvr=6&mod=wenzhangmod, 3381-http://weibo.com/p/1001603782188658234894?from=page_100206_profile&wvr=6&mod=wenzhangmod, 3382-http://weibo.com/p/1001603782185839625055?from=page_100206_profile&wvr=6&mod=wenzhangmod, 3383-http://www.yahoo.co.jp, 3384-http://news.yahoo.co.jp/pickup/6140452, 3385-http://news.yahoo.co.jp/pickup/6140458, 3386-http://news.yahoo.co.jp/pickup/6140460, 3387-http://news.yahoo.co.jp/pickup/6140456, 3388-http://news.yahoo.co.jp/pickup/6140444, 3389-http://www.hao123.com, 3390-http://xyx.hao123.com, 3391-http://www.hao123.com/rili, 3392-http://v.hao123.com/zongyi/, 3393-http://hao123.hunantv.com/video/short?id=2913, 3394-http://www.yandex.ru, 3395-http://news.yandex.ru/yandsearch?cl4url=www.interfax-russia.ru\%2FMoscow\%2Fspecial.asp\%3Fid\%3D563063\%26sec\%3D1725&lang=ru&lr=109906, 3396-http://market.yandex.ru/index?clid=506, 3397-http://auto.yandex.ru/?from=morda&_openstat=yandex_c_b;title;vned5y;earalr5y_c_b_2, 3398-http://rabota.yandex.ru/?from=morda&_openstat=yandex;title;filter;filter3all, 3399-http://rabota.yandex.ru/search.xml/?job_industry=310&rid=213, 3400-http://vk.com, 3401-https://vk.com/taylorswift, 3402-http://vk.com/bose_life, 3403-http://vk.com/yellowstoneusa, 3404-https://www.google.de/?gws_rd=ssl, 3405-https://www.google.de/?gws_rd=ssl#q=yellowstone, 3406-https://www.google.de/?gws_rd=ssl#q=taylor+swift, 3407-https://www.google.de/?gws_rd=ssl#q=bose, 3408-http://www.sohu.com, 3409-http://news.sohu.com/20141129/n406504112.shtml, 3410-http://news.sohu.com/20141129/n406502976.shtml, 3411-http://mil.sohu.com/20141129/n406496994.shtml, 3412-http://news.sohu.com/20141129/n406503474.shtml, 3413-http://news.sohu.com/20141129/n406502564.shtml, 3414-https://www.google.co.jp/?gws_rd=ssl, 3415-https://www.google.co.jp/?gws_rd=ssl#q=bose, 3416-https://www.google.co.jp/?gws_rd=ssl#q=yellowstone, 3417-https://www.google.co.jp/?gws_rd=ssl#q=taylor+swift, 3418-https://www.google.co.uk/?gws_rd=ssl, 3419-https://www.google.co.uk/?gws_rd=ssl#q=taylor+swift, 3420-https://www.google.co.uk/?gws_rd=ssl#q=bose, 3421-https://www.google.co.uk/?gws_rd=ssl#q=yellowstone, 3422-http://www.alibaba.com, 3423-http://www.alibaba.com/product-detail/bra-set-bra-panty-set-bra_140753102.html?spm=5386.7328861.1998097322.6, 3424-http://www.alibaba.com/product-detail/plastic-bee-hive-for-beekeeping_630267515.html?spm=5386.7374605.1998159253.31&tracelog=agnysb06, 3425-http://www.alibaba.com/product-detail/Fresh-Apple_60033353329.html?spm=5386.7374605.1998151692.25&tracelog=agjj01, 3426-http://www.alibaba.com/product-detail/metronome-education-equipment_50904225.html?s=p, 3427-http://www.alibaba.com/product-detail/toy-musical-instrument-cheap-hot-selling_317584061.html, 3428-https://www.google.fr/?gws_rd=ssl, 3429-https://www.google.fr/?gws_rd=ssl#q=taylor+swift, 3430-https://www.google.fr/?gws_rd=ssl#q=bose, 3431-https://www.google.fr/?gws_rd=ssl#q=yellowstone, 3432-https://mail.ru, 3433-http://news.mail.ru/politics/20303407/?frommail=1, 3434-http://news.mail.ru/incident/20305841/?frommail=1, 3435-http://news.mail.ru/economics/20305444/?frommail=1, 3436-http://sport.mail.ru/news/football/20305137/?frommail=1, 3437-http://auto.mail.ru/article/53041-razbiraem_vazhnye_obnovleniya_v_pdd/, 3438-http://www.aliexpress.com, 3439-http://www.aliexpress.com/item/Sexy-Modal-Boxers-Underwear-and-Cotton-Men-Underwear-and-Boxer-Shorts-Mens-High-quality-mini-order/1990015113.html?spm=5261.7132366.1998156808.1, 3440-http://www.aliexpress.com/item/For-iphone5-5s-4-4s-cases-Transparent-Simpson-Hand-grasp-the-logo-cell-phone-cases-covers/1870580330.html?spm=5261.7132366.1998156808.9, 3441-http://www.aliexpress.com/item/Min-order-is-8-mix-order-sexy-tattoo-girl-hot-sale-case-for-apple-iphone-4/1916913056.html, 3442-http://www.aliexpress.com/item/Min-order-is-8-mix-order-The-king-od-the-wood-3D-pattern-tiger-lion-cover/1911372575.html, 3443-http://www.aliexpress.com/item/Min-order-is-8-mix-order-hard-case-for-iphone-4-4S-5-5S-design-proctective/1910966200.html, 3444-https://www.google.com.br/?gws_rd=ssl, 3445-https://www.google.com.br/?gws_rd=ssl#q=taylor+swift, 3446-https://www.google.com.br/?gws_rd=ssl#q=bose, 3447-https://www.google.com.br/?gws_rd=ssl#q=yellowstone, 3448-https://www.google.ru/?gws_rd=ssl, 3449-https://www.google.ru/?gws_rd=ssl#newwindow=1&q=taylor+swift, 3450-https://www.google.ru/?gws_rd=ssl#newwindow=1&q=bose, 3451-https://www.google.ru/?gws_rd=ssl#newwindow=1&q=yellowstone, 3452-http://360.cn, 3453-http://bbs.360safe.com/thread-5067998-1-1.html, 3454-http://bbs.360safe.com/thread-4999272-1-1.html, 3455-http://bbs.360safe.com/thread-4884796-1-1.html, 3456-http://www.amazon.co.jp, 3457-http://www.amazon.co.jp/?????-??????-??????-ATBC-PVC-???????????/dp/B00P7LBFKG/ref=sr_1_1?s=hobby&ie=UTF8&qid=1417304898&sr=1-1, 3458-http://www.amazon.co.jp/selector-infected-WIXOSS-??????-ATBC-PVC???????????/dp/B00PRR6I5C/ref=pd_sim_hb_6?ie=UTF8&refRID=19Z2TRD58EPXAM4T9VRN, 3459-http://www.amazon.co.jp/???????-??????-??????-ATBC-PVC-???????????/dp/B00Q48AIEK/ref=pd_sim_hb_3?ie=UTF8&refRID=1T93M5J0KPFVCJKJ5T8F, 3460-http://www.amazon.co.jp/???????-??????-??????-ATBC-PVC-???????????/dp/B00PHY6MGU/ref=pd_sim_hb_4?ie=UTF8&refRID=02HRN11ZJJ2075CK8WGD, 3461-http://www.amazon.co.jp/??????????????-?????-??????-???????-???/dp/B00K05VNGK/ref=sr_1_2?m=AN1VRQENFRJN5&s=hobby&srs=3206718051&ie=UTF8&qid=1417305062&sr=1-2, 3462-http://www.163.com/, 3463-http://news.163.com/14/1130/01/AC8T6PBI00014AED.html, 3464-http://news.163.com/14/1130/00/AC8P30UD000146BE.html, 3465-http://news.163.com/14/1130/02/AC92AF1500014AED.html, 3466-http://auto.163.com/photoview/5BD20008/170817.html#p=AC78NEQA5BD20008, 3467-http://auto.163.com/photoview/5BD20008/170649.html?from=tj_day, 3468-https://www.google.it/?gws_rd=ssl, 3469-https://www.google.it/?gws_rd=ssl#q=yellowstone, 3470-https://www.google.it/?gws_rd=ssl#q=taylor+swift, 3471-https:

//www.google.it/?gws_rd=ssl#q=bose, 3472-https://www.google.es/?gws_rd=ssl, 3473-https://www.google.es/?gws_rd=ssl#q=taylor+swift, 3474-https://www.google.es/?gws_rd=ssl#q=bose, 3475-https://www.google.es/?gws_rd=ssl#q=yellowstone, 3476-http://www.amazon.de, 3477-http://www.amazon.de/Samsung-UE48H6270-LED-Backlight-Fernseher-schwarz-silber/dp/B00IVX814K/ref=sr_1_1?m=A3JWKAKR8XB7XF&s=ce-de&ie=UTF8&qid=1417307590&sr=1-1, 3478-http://www.amazon.de/dp/B00E9WPQTU?psc=1, 3479-http://www.amazon.de/dp/B00LUROVDE?psc=1, 3480-http://soso.com, 3481-http://www.soso.com/q?ie=utf8&pid=s.idx&cid=s.idx.se&unc=&query=yellowstone&w=&sut=2113&sst0=1417307704721&lkt=11\%2C1417307702608\%2C1417307704535, 3482-http://www.soso.com/q?ie=utf8&pid=s.idx&cid=s.idx.se&unc=&query=bose, 3483-http://www.soso.com/q?ie=utf8&pid=s.idx&cid=s.idx.se&unc=&query=taylor+swift, 3484-http://fc2.com, 3485-http://video.fc2.com, 3486-http://video.fc2.com/en/content/20140910JEyFAP4N, 3487-http://video.fc2.com/en/content/201409171QW5LZgd&suggest, 3488-http://video.fc2.com/en/content/20140807xM7UndY5&suggest, 3489-http://gmw.cn, 3490-http://politics.gmw.cn/2014-11/29/content_14004947.htm, 3491-http://politics.gmw.cn/2014-11/29/content_14005036.htm, 3492-http://politics.gmw.cn/2014-11/29/content_14002932.htm, 3493-http://politics.gmw.cn/2014-11/29/content_14005463.htm, 3494-https://www.google.ca/?gws_rd=ssl, 3495-https://www.google.ca/?gws_rd=ssl#q=bose, 3496-https://www.google.ca/?gws_rd=ssl#q=taylor\%20swift, 3497-https://www.google.ca/?gws_rd=ssl#q=yellowstone, 3498-https://www.google.com.mx/?gws_rd=ssl, 3499-https://www.google.com.mx/?gws_rd=ssl#q=taylor+swift, 3500-https://www.google.com.mx/?gws_rd=ssl#q=bose, 3501-https://www.google.com.mx/?gws_rd=ssl#q=yellowstone, 3502-https://www.google.com.hk/?gws_rd=ssl, 3503-https://www.google.com.hk/?gws_rd=ssl#q=yellowstone, 3504-https://www.google.com.hk/?gws_rd=ssl#q=bose, 3505-https://www.google.com.hk/?gws_rd=ssl#q=taylor+swift, 3506-http://www.amazon.co.uk, 3507-http://www.amazon.co.uk/gp/product/B00KC6KMWI/ref=s9_pop_gw_g424_i3/280-2846582-0617240?pf_rd_m=A3P5ROKL5A1OLE&pf_rd_s=center-2&pf_rd_r=09A3BMSNBY4NJDKXRFC3&pf_rd_t=101&pf_rd_p=358550247&pf_rd_i=468294, 3508-http://www.amazon.co.uk/gp/product/B005N8W1MO/ref=s9_pop_gw_g23_i5/280-2846582-0617240?pf_rd_m=A3P5ROKL5A1OLE&pf_rd_s=center-2&pf_rd_r=09A3BMSNBY4NJDKXRFC3&pf_rd_t=101&pf_rd_p=358550247&pf_rd_i=468294, 3509-http://www.amazon.co.uk/dp/B00JLFC0KI?psc=1, 3510-http://www.amazon.co.uk/dp/B00CD3IDFG?psc=1, 3511-http://www.amazon.co.uk/dp/B00BBYJCEO?psc=1, 3512-http://www.cntv.cn, 3513-http://english.cntv.cn, 3514-http://english.cntv.cn/2014/11/28/VIDE1417178162604135.shtml, 3515-http://english.cntv.cn/2014/11/29/VIDE1417232520686377.shtml, 3516-http://english.cntv.cn/2014/11/29/ARTI1417245893034403.shtml, 3517-http://english.cntv.cn/2014/11/29/ARTI1417227633551434.shtml, 3518-http://youradexchange.com, 3519-https://www.google.com.tr/?gws_rd=ssl, 3520-https://www.google.com.tr/?gws_rd=ssl#q=taylor+swift, 3521-https://www.google.com.tr/?gws_rd=ssl#q=bose, 3522-https://www.

google.com.tr/?gws_rd=ssl#q=yellowstone, 3523-http://www.ebay.de, 3524-http://ok.ru, 3525-https://www.alipay.com/?src=alipay.com, 3526-https://www.google.pl/?gws_rd=ssl, 3527-https://www.google.pl/?gws_rd=ssl#q=yellowstone, 3528-https://www.google.pl/?gws_rd=ssl#q=bose, 3529-https://www.google.pl/?gws_rd=ssl#q=taylor+swift, 3530-http://www.rakuten.co.jp, 3531-http://event.rakuten.co.jp/campaign/supersale/20141130/genre/mensfashion/?l-id=top_normal_flashbnr_10_523&l-id=ppf_pc_s1_pc_web_t1_7206, 3532-http://item.rakuten.co.jp/ueno-yayoi/501-original-sp/, 3533-http://item.rakuten.co.jp/ueno-yayoi/gloverall-monty-special/, 3534-http://item.rakuten.co.jp/trident/20000mah-01/, 3535-http://www.naver.com, 3536-http://newsstand.naver.com/?list=ct1&pcode=109, 3537-http://newsstand.naver.com/?list=ct1&pcode=016, 3538-http://newsstand.naver.com/?list=ct1&pcode=052, 3539-http://happybean.naver.com/donations/H000000107658, 3540-http://happybean.naver.com/donations/H000000107667, 3541-https://www.google.com.au/?gws_rd=ssl, 3542-https://www.google.com.au/?gws_rd=ssl#q=yellowstone, 3543-https://www.google.com.au/?gws_rd=ssl#q=bose, 3544-https://www.google.com.au/?gws_rd=ssl#q=taylor+swift, 3545-http://people.com.cn, 3546-http://politics.people.com.cn/n/2014/1129/c1024-26118450.html, 3547-http://opinion.people.com.cn/n/2014/1130/c1003-26118982.html, 3548-http://world.people.com.cn/n/2014/1130/c1002-26118804.html, 3549-http://politics.people.com.cn/n/2014/1129/c1001-26118487.html, 3550-http://finance.people.com.cn/n/2014/1130/c1004-26118977.html, 3551-http://www.flipkart.com, 3552-http://www.flipkart.com/huggies-wonder-pants-medium/p/itmdhts6gagwmr3n?pid=DPRDHTS6QYUVESVR&srno=b_1&offer=DOTDOnDiaper_Nov29.&ref=c3df162c-48f6-427a-b187-0b53e3d0226a, 3553-http://www.flipkart.com/huggies-dry-diaper-medium/p/itmdv5gmudkqkdye?pid=DPRDAHHEUBKAFFUU&icmpid=reco_pp_recobundle_babycare_diaper_1_2&ppid=DPRDHTS6QYUVESVR, 3554-http://www.flipkart.com/spice-double-bubble-2429-flats/p/itmefhy3dhxpgggr?pid=SNDEFHY3HKYQFUZ5&srno=b_2&ref=c3185a6a-6097-4e7d-9662-b5758f39d2ea, 3555-http://www.flipkart.com/spice-double-bubble-2429-flats/p/itmefhy3gka6kdjp?pid=SNDEFHY3YCTMCEUY&icmpid=reco_pp_same_footwear_sandal_3&ppid=SNDEFHY3HKYQFUZ5, 3556-http://www.flipkart.com/ocean-patio-tumbler-set-5b1831006g0000/p/itmdyp5hrapyqfgh?pid=GLSDYP5H9US6ZEF3&icmpid=reco_pp_historyFooter_footwear_na_2&ppid=SNDEFHY3HKYQFUZ5, 3557-http://xinhuanet.com, 3558-http://news.xinhuanet.com/politics/2014-11/29/c_1113457654.htm, 3559-http://news.xinhuanet.com/politics/2014-11/29/c_1113456363.htm, 3560-http://news.xinhuanet.com/local/2014-11/30/c_127262525.htm, 3561-http://news.xinhuanet.com/politics/2014-11/29/c_1113455511.htm, 3562-http://news.xinhuanet.com/ziliao/2014-11/30/c_127262658.htm, 3563-http://www.amazon.cn, 3564-http://www.amazon.cn/??-it-??/dp/B00MFEI7RW/ref=cngwv1_cefloor_ha_3_B00MFEI7RW/476-6793934-2865314?pf_rd_m=A1AJ19PSB66TGU&pf_rd_s=center-4&pf_rd_r=0X885VXWY7M22WP8PKXV&pf_rd_t=101&pf_rd_p=241224932&pf_rd_i=899254051, 3565-http://www.amazon.cn/dp/B00H990J1U?psc=1, 3566-http://www.amazon.cn/????/dp/B00MOAG9AO/ref=pd_sim_sbs_pc_18?ie=UTF8&refRID=12J1JK8YWMPB6SMVRZDX, 3567-http://www.amazon.cn/????/dp/B00MO9CREC/ref=pd_sim_sa_5?ie=UTF8&refRID=1238GQ7XNEJA7DZDQGPQ, 3568-http://www.amazon.cn/????/dp/B00HC5YKWU/ref=pd_sim_sbs_luggage_2?ie=UTF8&refRID=18R8R95DGCEK81SGRF90, 3569-https://www.pixnet.net, 3570-https://www.pixnet.net/blog/profile/pf77501, 3571-https://www.pixnet.net/blog/profile/jnee, 3572-http://angelpomelo.pixnet.net/blog/post/188672283, 3573-https://www.pixnet.net/blog, 3574-http://pigx3.pixnet.net/blog/post/41776549, 3575-http://www.ebay.co.uk, 3576-http://www.ebay.co.uk/cln/gift-curation/Tech-Treats/135776598014, 3577-http://www.sogou.com, 3578-http://www.sogou.com/web?query=taylor+swift, 3579-http://www.sogou.com/web?query=bose, 3580-http://www.sogou.com/web?query=yellowstone, 3581-http://www.indiatimes.com, 3582-http://www.indiatimes.com/lifestyle/technology/top-bosses-and-their-top-mobile-apps-revealed-228671.html, 3583-http://www.indiatimes.com/lifestyle/10-annoying-things-people-do-when-you-are-travelling-alone-228665.html, 3584-http://www.indiatimes.com/lifestyle/these-7word-stories-will-stir-your-imagination-228658.html, 3585-http://www.indiatimes.com/culture/who-we-are/20-indian-unesco-world-heritage-sites-you-need-to-visit-228634.html, 3586-http://www.indiatimes.com/entertainment/celebs/, 3587-http://www.tudou.com, 3588-http://www.tudou.com/listplay/V180IZeKqqg/X5pH0cTb5To.html, 3589-http://imake.tudou.com, 3590-http://www.tudou.com/albumplay/rCZxzemAmMs.html, 3591-http://v.tudou.com/choufengdeyu/, 3592-http://www.tudou.com/listplay/e7BGSm86eU8/cKHJI6wXj1M.html, 3593-http://www.uol.com.br, 3594-http://esporte.uol.com.br/futebol/campeonatos/brasileiro/serie-b/ultimas-noticias/2014/11/29/boa-tropeca-contra-icasa-america-mg-vence-em-bh-mas-4-vaga-e-do-avai.htm, 3595-http://www1.folha.uol.com.br/saopaulo/2014/11/1554807-blogueiras-cobram-ate-r-40-mil-por-postagem-de-marcas.shtml, 3596-http://esporte.uol.com.br/futebol-americano/ultimas-noticias/2014/11/29/milionario-deixou-esporte-para-plantar-e-doar-batata-internet-ensinou-tudo.htm, 3597-http://televisao.uol.com.br/noticias/redacao/2014/11/29/maria-antonieta-de-las-nieves-diz-que-atritos-com-bolanos-nao-eram-pessoais.htm, 3598-http://www1.folha.uol.com.br/educacao/2014/11/1555428-nota-de-matematica-recua-na-rede-publica.shtml, 3599-https://www.google.com.tw/?gws_rd=ssl, 3600-https://www.google.com.tw/?gws_rd=ssl#q=yellowstone, 3601-https://www.google.com.tw/?gws_rd=ssl#q=bose, 3602-https://www.google.com.tw/?gws_rd=ssl#q=taylor+swift, 3603-https://www.google.com.eg/?gws_rd=ssl, 3604-https://www.google.com.eg/?gws_rd=ssl#q=bose, 3605-https://www.google.com.eg/?gws_rd=ssl#q=taylor+swift, 3606-https://www.google.com.eg/?gws_rd=ssl#q=yellowstone, 3607-https://www.google.com.sa/?gws_rd=ssl, 3608-https://www.google.com.sa/search?site=&source=hp&q=bose, 3609-https://www.google.com.sa/search?site=&source=hp&q=yellowstone, 3610-https://www.google.com.sa/search?q=taylor+swift, 3611-http://www.amazon.fr, 3612-

http://www.amazon.fr/Pentax-K-S1-Appareil-numrique-Objectif/dp/B00N3XBBMC/ref=br_lf_m_1000835833_1_7_ttl?ie=UTF8&m=A1X6FK5RDHNB96&s=photo&pf_rd_p=550440487&pf_rd_s=center-3&pf_rd_t=1401&pf_rd_i=1000835833&pf_rd_m=A1X6FK5RDHNB96&pf_rd_r=13C5YJBARGEMHM4MGHTM, 3613-http://www.amazon.fr/Pentax-Appareil-numrique-Reflex-Objectif/dp/B00O1VX36M/ref=pd_sim_sbs_ph_4?ie=UTF8&refRID=0SHM0KWTZBE7DCTA80HN, 3614-http://www.amazon.fr/Pentax-K-S1-Appareil-numrique-Objectif/dp/B00N3XBGI6/ref=pd_sim_sbs_ph_6?ie=UTF8&refRID=06E1P2PXXPD55BW5ANQ5, 3615-http://www.amazon.fr/Apple-gnration--Tablette-tactile-Retina/dp/B00G55JTBA/ref=sr_1_1?s=electronics&ie=UTF8&qid=1417318023&sr=1-1&keywords=ipad, 3616-http://www.amazon.fr/Rotatif-fonction-SmartCover-STYLET-OFFERTS/dp/B00G424MAM/ref=pd_sim_ce_5?ie=UTF8&refRID=1NYS9C0RGCVRFJCQ5NPG, 3617-https://www.google.co.kr/?gws_rd=ssl, 3618-https://www.google.co.kr/?gws_rd=ssl#newwindow=1&q=yellowstone, 3619-https://www.google.co.kr/?gws_rd=ssl#newwindow=1&q=bose, 3620-https://www.google.co.kr/?gws_rd=ssl#newwindow=1&q=taylor+swift, 3621-https://www.google.com.pk/?gws_rd=ssl, 3622-https://www.google.com.pk/?gws_rd=ssl#q=bose, 3623-https://www.google.com.pk/?gws_rd=ssl#q=yellowstone, 3624-https://www.google.com.pk/?gws_rd=ssl#q=taylor+swift, 3625-http://www.amazon.in/, 3626-http://www.amazon.in/gp/product/B00J60MFTO/ref=s9_pop_gw_g23_i3/277-6350587-2410146?pf_rd_m=A1VBAL9TL5WCBF&pf_rd_s=center-3&pf_rd_r=00T6EREX8100E8C54W8T&pf_rd_t=101&pf_rd_p=518027687&pf_rd_i=1320006031, 3627-http://www.amazon.in/gp/product/0143423002/ref=s9_ri_gw_g14_i3/277-6350587-2410146?pf_rd_m=A1VBAL9TL5WCBF&pf_rd_s=center-6&pf_rd_r=00T6EREX8100E8C54W8T&pf_rd_t=101&pf_rd_p=518028667&pf_rd_i=1320006031, 3628-http://www.amazon.in/INTEX-CAPACITY-POWER-4000mAh-mobile/dp/B00IUGMQEE/ref=sr_1_8?s=electronics&ie=UTF8&qid=1417439824&sr=1-8, 3629-http://www.amazon.in/Like-Happened-Yesterday-Ravinder-Singh/dp/0143418807/ref=pd_rhf_se_s_cp_4_GR0V?ie=UTF8&refRID=0JTCA6H769RNH6EARWYC, 3630-http://www.amazon.in/Half-Girlfriend-Chetan-Bhagat/dp/8129135728/ref=pd_cp_b_0, 3631-http://www.ettoday.net/, 3632-http://www.ettoday.net/news/20141201/433344.htm, 3633-http://www.ettoday.net/news/20141201/433347.htm, 3634-http://www.ettoday.net/news/20141201/433342.htm, 3635-http://www.ettoday.net/news/20141201/433272.htm, 3636-http://www.ettoday.net/news/20141125/429986.htm, 3637-http://www.pconline.com.cn/, 3638-http://mobile.pconline.com.cn/579/5795376.html, 3639-http://mobile.pconline.com.cn/580/5807678.html, 3640-http://www1.pconline.com.cn/act/thanksgiving/index.html#area1, 3641-http://ivy.pconline.com.cn/adpuba/click?adid=346546&id=pc.sy.wzld.syzt., 3642-http://dc.pconline.com.cn/579/5795216.html, 3643-http://ameblo.jp/, 3644-http://ameblo.jp/lilicom/entry-11959360143.html, 3645-http://ameblo.jp/kasumiarimura/entry-11959024321.html, 3646-http://ameblo.jp/matoba-koji/entry-11959292301.html, 3647-http://ameblo.jp/ninpu-hitomi/entry-11959039785.html, 3648-http://ameblo.jp/sunsuntaiyo/, 3649-http://www.life.com.tw/, 3650-http://www.life.com.tw/?app=view&no=200929, 3651-http://www.life.com.tw/?app=view&no=200913, 3652-http://www.life.com.tw/?app=view&no=200911, 3653-http://www.life.com.tw/?app=view&no=201787, 3654-http://www.life.com.tw/?app=view&no=201831, 3655-http://www.globo.com/, 3656-http://extra.globo.com/noticias/rio/jacare-flagrado-atravessando-rua-no-rio-de-janeiro-14706703.html#ixzz3KeKjOus1, 3657-http://globoesporte.globo.com/futebol/times/botafogo/noticia/2014/12/botafogo-antecipa-viagem-de-volta-apos-queda-e-despista-protestos.html, 3658-http://gshow.globo.com/novelas/imperio/vem-por-ai/noticia/2014/12/racha-na-familia-maria-isis-desperta-desejo-de-vinganca-em-magnolia.html, 3659-http://oglobo.globo.com/economia/emprego/concurso-vai-escolher-um-testador-de-casas-de-ferias-que-vai-ganhar-para-viajar-brasil-afora-14705788, 3660-http://ego.globo.com/famosos/noticia/2014/12/angelina-jolie-sofre-acidente-de-carro-em-los-angeles-diz-site.html, 3661-https://www.google.co.th/, 3662-https://www.google.co.th/#q=taylor+swift, 3663-https://www.google.co.th/#q=bose, 3664-https://www.google.co.th/#q=yellowstone, 3665-http://bongacams.com/, 3666-https://www.google.nl/, 3667-https://www.google.nl/#q=taylor+swift, 3668-https://www.google.nl/#q=bose, 3669-https://www.google.nl/#q=yellowstone, 3670-https://www.google.com.ar/, 3671-https://www.google.com.ar/#q=taylor+swift, 3672-https://www.google.com.ar/#q=bose, 3673-https://www.google.com.ar/#q=yellowstone, 3674-http://www.jd.com/, 3675-http://sale.jd.com/act/rDfC40zGRHdY3cap.html, 3676-http://sale.jd.com/act/SW0BJoR614mZ.html, 3677-http://sale.jd.com/act/Fi8bkCnmpy.html, 3678-http://item.jd.com/1253580930.html, 3679-http://item.jd.com/1351505239.html, 3680-http://themeforest.net/, 3681-https://www.google.co.za/, 3682-https://www.google.co.za/#q=taylor+swift, 3683-https://www.google.co.za/#q=bose, 3684-https://www.google.co.za/#q=yellowstone, 3685-http://www.zol.com.cn/, 3686-http://4g.zol.com.cn/493/4936464.html, 3687-http://news.zol.com.cn/493/4936713.html, 3688-http://tupian.zol.com.cn/tushuo/4938241.html, 3689-http://news.zol.com.cn/493/4936742.html, 3690-http://power.zol.com.cn/493/4936621.html, 3691-http://www.snapdeal.com/, 3692-http://www.snapdeal.com/products/home-kitchen-bed-linen, 3693-http://www.snapdeal.com/product/home-candy-pink-floral-cotton/2077484786#bcrumbLabelId:211, 3694-http://www.snapdeal.com/product/coirfit-daydream-45-inches-pocket/1759816080, 3695-http://www.snapdeal.com/product/maroon-single-mattress-waterproof-cover/228285877, 3696-http://www.snapdeal.com/product/blue-eyes-attractive-jacquard-weaved/1741085571, 3697-http://coccoc.com/, 3698-http://adf.ly/, 3699-http://www.amazon.it/, 3700-http://www.amazon.it/gp/product/B00H72FKZY/ref=s9_pop_gw_g23_i1/276-6748803-5279137?pf_rd_m=A11IL2PNWYJU7H&pf_rd_s=center-2&pf_rd_r=0YMC6QNCR0CEQSMVJWD7&pf_rd_t=101&pf_rd_p=312233767&pf_rd_i=426865031, 3701-http://www.amazon.it/CSL-Auricolari-Reduction-trasporto-Hardcover/dp/B00JQ9J76O/ref=pd_sim_ce_2?ie=UTF8&refRID=0D6K484Z76VXV5J2KSN7,

3702-http://www.amazon.it/Casio-MQ-24-7BLL-MQ247BLL-Orologio-unisex/dp/B000JNKABW/ref=pd_sim_ce_40?ie=UTF8&refRID=1SHFYJ6EDH2BTT130680, 3703-http://www.amazon.it/Casio-MQ24-7B2-Orologio-da-Uomo/dp/B000GB0G7A/ref=pd_sim_w_5?ie=UTF8&refRID=0XQS38C2QM16R3791M9V, 3704-https://www.amazon.it/gp/product/B00BL3R7FQ/gcrnsts?ie=UTF8&qid=1417441776&ref_=sr_1_8&s=gift-cards&sr=1-8, 3705-http://diply.com/, 3706-http://diply.com/creativeideas/24-rare-historical-photos-that-will-leave-you-speechless/67089, 3707-http://diply.com/weird-facts/she-melts-a-red-crayon-on-stove-end/67098, 3708-http://diply.com/trendyjoe/these-pop-art-illustrations-reveal-secret-lives-fictional/67084, 3709-http://diply.com/just-a-dream/21-life-hacks-every-woman-needs-know/66984, 3710-http://diply.com/bigtheory/22-hilarious-times-weve-all-experienced-that-awkward-moment-when/66980, 3711-http://www.leboncoin.fr/, 3712-http://www.leboncoin.fr/annonces/offres/alsace/, 3713-http://www.leboncoin.fr/vetements/399185816.htm?ca=1_s, 3714-http://www.leboncoin.fr/chaussures/740434558.htm?ca=1_s, 3715-http://www.leboncoin.fr/motos/713507431.htm?ca=1_s, 3716-http://www.leboncoin.fr/ameublement/740434321.htm?ca=1_s, 3717-http://www.livejournal.com/, 3718-https://www.google.co.id/?gws_rd=ssl, 3719-https://www.google.co.id/?gws_rd=ssl#q=taylor+swift, 3720-https://www.google.co.id/?gws_rd=ssl#q=bose, 3721-https://www.google.co.id/?gws_rd=ssl#q=yellowstone, 3722-http://www.bycontext.com/, 3723-http://www.youku.com/, 3724-http://i.youku.com/u/UMTQ0NTg2NTUwOA==?from=y1.3-idx-grid-1519-9909.86850.2-3, 3725-http://i.youku.com/u/UNTk5MTE5MzQ0, 3726-http://i.youku.com/u/UMTU4MTY2MTI3Ng==, 3727-http://v.youku.com/v_show/id_XODM5MDQ2NDc2.html?f=22450459&from=y1.3-music-grid-140-9922.87061.1-1, 3728-http://v.youku.com/v_show/id_XODA4OTA5MzEy.html?from=y1.2-3-95.3.2-2.1-4-1-1, 3729-http://v.youku.com/v_show/id_XODM2Mjg5NTU2.html?from=y1.2-1-95.3.9-2.1-1-1-8, 3730-http://v.youku.com/v_show/id_XODI4OTc1ODI0.html?from=y1.2-1-95.3.17-2.1-1-1-16, 3731-http://v.youku.com/v_show/id_XODI3Mzc0OTE2.html?from=y1.2-1-95.3.16-2.1-1-1-15, 3732-http://www.nicovideo.jp/, 3733-http://live.nicovideo.jp/watch/lv198414275?cc_referrer=ustop, 3734-http://www.nicovideo.jp/watch/sm25007348?cc_referrer=ustop, 3735-http://www.nicovideo.jp/watch/1406279621, 3736-http://www.nicovideo.jp/watch/1409143185, 3737-http://www.nicovideo.jp/watch/1409904011, 3738-http://blogfa.com/, 3739-http://dargaheeshgh.blogfa.com/, 3740-http://saneyanm.blogfa.com/, 3741-http://rosukh84.blogfa.com/, 3742-http://asheghanehbaallah.blogfa.com/, 3743-http://meyomeykhaneh.blogfa.com/, 3744-http://www.tubecup.com/, 3745-http://www.tubecup.com/videos/90656/crystal-gunns-mambos/, 3746-http://www.tubecup.com/videos/90041/hawt-mother-i-d-like-to-fuck-strict-jerkoff-instruction/, 3747-http://www.tubecup.com/videos/132854/japanese-momteach-about-sex-xlx/, 3748-http://www.tubecup.com/videos/130111/julia-great-girlfriend-4-by-packmans-cen/, 3749-http://www.tubecup.com/videos/130469/julia-great-girlfriend-5-by-packmans-cen/, 3750-https://www.google.gr/, 3751-https://www.google.gr/#q=taylor+swift, 3752-https://www.google.gr/#q=bose, 3753-https://www.google.gr/#q=yellowstone, 3754-http://naverland.naver.jp/?p=7471, 3755-http://www.douban.com/, 3756-http://www.douban.com/note/450392083/, 3757-http://www.douban.com/note/460474481/, 3758-http://www.douban.com/photos/photo/717637981/, 3759-http://www.douban.com/doulist/2406267/, 3760-http://www.douban.com/photos/album/138417243/, 3761-http://www.twitch.tv/, 3762-http://www.twitch.tv/flosd, 3763-http://www.twitch.tv/skumbagkrepo, 3764-http://www.twitch.tv/directory/game/World\%20of\%20Warcraft\%3A\%20Warlords\%20of\%20Draenor, 3765-http://www.twitch.tv/directory/game/Grand\%20Theft\%20Auto\%20V, 3766-http://www.twitch.tv/ellohime, 3767-http://www.chinadaily.com.cn/, 3768-http://usa.chinadaily.com.cn/china/2014-12/01/content_19002348.htm, 3769-http://usa.chinadaily.com.cn/world/2014-12/01/content_19002955.htm, 3770-http://usa.chinadaily.com.cn/world/2014-12/01/content_19002343.htm, 3771-http://usa.chinadaily.com.cn/china/2014-12/01/content_19004494.htm, 3772-http://www.chinadaily.com.cn/culture/2014-11/25/content_18970468.htm, 3773-http://www.daum.net/, 3774-http://media.daum.net/issue/854/?newsId=20141201212129204, 3775-http://go.shopping.daum.net/link/go.daum?dataseq=9byy7&do=0, 3776-http://webtoon.daum.net/webtoon/viewer/28165, 3777-http://webtoon.daum.net/webtoon/viewer/28165, 3778-http://shopping.daum.net/go.daum?url=VKA00CqTSjHD9wY4IxZ.j_iXfpiE5B.EwibIsnmKoVwLsPchTdfPbZ_Tx5Lzf8bpJRYPlMKLUb7CRcWwaEnAJ7t7IjdijpX2Qx.kgPjyGqaYGPQprWjg5VWdyFZxeDSut4.GVR4Jv.wM2Nqn.D2xfJfl6Cf9O-yovOxEf, 3779-https://www.google.co.ve/?gws_rd=ssl, 3780-https://www.google.co.ve/?gws_rd=ssl#q=taylor+swift, 3781-https://www.google.co.ve/?gws_rd=ssl#q=bose, 3782-https://www.google.co.ve/?gws_rd=ssl#q=yellowstone, 3783-http://allegro.pl/, 3784-http://allegro.pl/dzial/strefa-okazji/furby-sweet-by-hasbro-caly-w-grochy-22243.html?ref=mainpage-bargain, 3785-http://allegro.pl/show_item.php?item=4847316876&sh_dwh_token=d0b6dd74a8394345128e20d296cb1c5e, 3786-http://allegro.pl/listing/user/listing.php?id=13393&us_id=3765198, 3787-http://allegro.pl/pilka-nozna-puma-evoforce-5-nowosc-2014-3-kolory-i4773869216.html, 3788-http://allegro.pl/4f-bluza-bielizna-polarowa-meska-bimp001-nowosc-s-i4754848462.html, 3789-https://www.google.com.my/?gws_rd=ssl, 3790-https://www.google.com.my/?gws_rd=ssl#q=taylor+swift, 3791-https://www.google.com.my/?gws_rd=ssl#q=bose, 3792-https://www.google.com.my/?gws_rd=ssl#q=yellowstone, 3793-http://ok.ru/, 3794-http://ask.fm/, 3795-https://www.popads.net/, 3796-https://www.google.com.ua/?gws_rd=ssl, 3797-https://www.google.com.ua/?gws_rd=ssl#q=taylor+swift, 3798-https://www.google.com.ua/?gws_rd=ssl#q=bose, 3799-https://www.google.com.ua/?gws_rd=ssl#q=yellowstone, 3800-http://www.dmm.co.jp/, 3801-http://www.dmm.co.jp/en/top/?_ga=1.166312573.2131691940.1417449138/, 3802-http://www.jabong.com/, 3803-http://www.jabong.com/men/?icn=home-new-UI&ici=r1_b2_men,

3804-http://www.jabong.com/poe-Stripes-Dark-Grey-Casual-Shirt-964689.html?pos=1, 3805-http://www.jabong.com/poe-Solid-Navy-Blue-Slim-Fit-Casual-Shirt-764682.html, 3806-http://www.jabong.com/poe-Solid-Black-Casual-Shirt-964625.html, 3807-http://www.jabong.com/silver-streak-Solid-Black-Slim-Fit-Casual-Shirt-831663.html, 3808-http://www.onet.pl/, 3809-http://wiadomosci.onet.pl/tylko-w-onecie/wybory-samorzadowe-2014-dr-flis-charakter-relacji-prezydenta-z-rzadem-to-wielka/wf52l, 3810-http://wiadomosci.onet.pl/krakow/krakow-lekarze-wyprowadzili-dwulatka-z-glebokiej-hipotermii/c9kjee, 3811-http://wiadomosci.onet.pl/kraj/wyznania-zbigniewa-lubienieckiego-lowcy-sowietow/bf12hp, 3812-http://wiadomosci.onet.pl/swiat/ukrainska-armia-rosyjskie-sily-specjalne-atakuja-lotnisko-w-doniecku/8zj2f, 3813-http://wiadomosci.onet.pl/swiat/andriej-illarionow-o-katastrofie-smolenskiej-zadna-brzoza-nie-jest-w-stanie-zniszczyc/z6p81, 3814-https://www.google.com.ng/?gws_rd=ssl, 3815-https://www.google.com.ng/?gws_rd=ssl#q=taylor+swift, 3816-https://www.google.com.ng/?gws_rd=ssl#q=bose, 3817-https://www.google.com.ng/?gws_rd=ssl#q=yellowstone, 3818-http://mailchimp.com/, 3819-https://www.google.com.vn/, 3820-https://www.google.com.vn/#q=taylor+swift, 3821-https://www.google.com.vn/#q=bose, 3822-https://www.google.com.vn/#q=yellowstone, 3823-http://badoo.com/, 3824-http://badoo.com/vi/01293829326/, 3825-http://badoo.com/01289803882/, 3826-http://badoo.com/01304999488/, 3827-http://badoo.com/01291775751/, 3828-http://badoo.com/01287181015/, 3829-https://archive.org/, 3830-https://archive.org/details/canadianrose1967cana, 3831-https://archive.org/details/SivaBhaktaCharithamu, 3832-https://blog.archive.org/2014/11/05/inviting-the-internet-over-to-play/, 3833-https://blog.archive.org/2014/11/11/lost-landscapes-of-san-francisco-fundraiser-benefitting-internet-archive-friday-december-19-2014/, 3834-http://blog.archive.org/2014/10/28/building-libraries-together/, 3835-http://feedly.com/#welcome, 3836-http://feedly.com/#explore\%2F\%23tech, 3837-http://feedly.com/#explore\%2F\%23entrepreneurship, 3838-http://feedly.com/#explore\%2F\%23business, 3839-http://feedly.com/#explore\%2F\%23marketing, 3840-http://feedly.com/#explore\%2F\%23vimeo, 3841-http://9gag.com/, 3842-http://9gag.com/gag/aGVA0OG, 3843-http://9gag.com/gag/aBQPewQ?ref=fsidebar, 3844-http://9gag.com/gag/anXxBKL?ref=fsidebar, 3845-http://9gag.com/gag/aqZx2qM?ref=fsidebar, 3846-http://9gag.com/gag/aj6BNqw?ref=fsidebar, 3847-https://www.quora.com/, 3848-http://torrentz.eu/, 3849-http://torrentz.eu/search?q=taylor+swift, 3850-http://torrentz.eu/search?q=bose, 3851-http://torrentz.eu/search?q=yellowstone, 3852-http://mystart.com/, 3853-http://www.orange.fr/, 3854-http://boutique.orange.fr/mobile/choisir-forfait, 3855-http://boutique.orange.fr/mobile/forfait-m6-mobile, 3856-http://boutique.orange.fr/mobile/forfaits-origami-jet, 3857-http://boutique.orange.fr/tablette-et-cle, 3858-http://travel.orange.fr/, 3859-http://www.mama.cn/, 3860-http://www.mama.cn/baby/art/

20141201/774994.html, 3861-http://www.mama.cn/baby/art/20141128/774979.html, 3862-http://www.mama.cn/baby/art/20141129/774987.html, 3863-http://www.mama.cn/baby/art/20141121/774935.html, 3864-http://www.mama.cn/ask/jingxuan/17552/, 3865-http://www.gmx.net/, 3866-http://www.gmx.net/magazine/services/singletreff/#msct_cid=29953;lpos=banner;ltype=pic;mtype=wsite;bid=377659, 3867-http://www.gmx.net/magazine/panorama/tugce-strafe-taeter-erwartet-30249254, 3868-http://www.gmx.net/magazine/unterhaltung/tv-film/weihnachten-minions-30249252, 3869-http://www.gmx.net/magazine/shopping/trentino/, 3870-http://www.gmx.net/magazine/shopping/weihnachtsmarkt/, 3871-https://www.google.com.co/, 3872-https://www.google.com.co/#q=taylor+swift, 3873-https://www.google.com.co/#q=bose, 3874-https://www.google.com.co/#q=yellowstone, 3875-http://www.xcar.com.cn/, 3876-http://newcar.xcar.com.cn/99/?zoneclick=101402, 3877-http://newcar.xcar.com.cn/m23561/, 3878-http://newcar.xcar.com.cn/m23561/, 3879-http://dealer.xcar.com.cn/92824/price_m6658682.htm?zoneclick=101188, 3880-http://dealer.xcar.com.cn/64902/price_s99_1_1.htm, 3881-http://www.amazon.es/, 3882-http://www.amazon.es/gp/product/B00KD7UZH8/ref=s9_ri_gw_g23_i2/279-4600046-3260246?pf_rd_m=A1AT7YVPFBWXBL&pf_rd_s=center-2&pf_rd_r=0JQ76FESX2XS91KXPNMZ&pf_rd_t=101&pf_rd_p=454936587&pf_rd_i=602357031, 3883-http://www.amazon.es/Cubierta-cuero-artificial-Motorola-Moto/dp/B00L3C52EI/ref=pd_sim_e_1?ie=UTF8&refRID=1956AZZV5JZWQ2SVGXP7, 3884-http://www.amazon.es/Motorola-Moto-Smartphone-pantalla-Dual-Core/dp/B00KD7UV8G/ref=pd_cp_e_1, 3885-http://www.amazon.es/Motorola-Moto-Smartphone-pantalla-Dual-Core/dp/B00KD7UV8G/ref=pd_cp_e_1, 3886-http://www.amazon.es/Bluedio-S2-auriculares-inal\%C3\%A1mbricos-incorporado/dp/B00J9A13VC/ref=sr_1_1?s=electronics-accessories&ie=UTF8&qid=1417457271&sr=1-1, 3887-http://www.reimageplus.com/, 3888-http://www.espncricinfo.com/, 3889-http://www.espncricinfo.com/australia-v-india-2014-15/content/current/story/805993.html, 3890-http://www.espncricinfo.com/india/content/current/story/806191.html, 3891-http://www.espncricinfo.com/australia/content/current/story/805747.html, 3892-http://www.espncricinfo.com/ci/content/video_audio/modern_masters/index.html, 3893-http://www.espncricinfo.com/ci/content/video_audio/805397.html, 3894-http://www.yaolan.com/, 3895-http://www.yaolan.com/talk/congronghehu/, 3896-http://tiny.yaolan.com/topic/kaichi/, 3897-http://bbs.yaolan.com/thread-52607447-1-2.html, 3898-http://www.yaolan.com/news/201412011857660.shtml, 3899-http://bbs.yaolan.com/thread-52607036-1-1.html, 3900-https://www.avito.ru/, 3901-https://www.avito.ru/rossiya/rabota, 3902-https://www.avito.ru/voronezhskaya_oblast/rabota, 3903-https://www.avito.ru/voronezh/vakansii/prodavets-konsultant_istore_g._voronezh_180142158, 3904-https://www.avito.ru/moskva/vakansii/administrator_v_saunu_152752623, 3905-https://www.avito.ru/leningradskaya_oblast/rabota, 3906-http://www.goo.ne.jp/, 3907-

http://news.goo.ne.jp/topstories/politics/79/b15f3af6e2ec87c66c4d9b50a133ac8a.html, 3908-http://news.goo.ne.jp/article/fminyu/region/fminyu-14846304.html, 3909-http://news.goo.ne.jp/topstories/nation/31/076d2ff1db83913747cf299dd30516a1.html, 3910-http://news.goo.ne.jp/topstories/nation/31/3632c287a37f1afa79c42f470ba6af66.html, 3911-http://news.goo.ne.jp/topstories/world/366/c03d7ed4caac589ca843afed778050d6.html, 3912-https://www.google.be/, 3913-https://www.google.be/#q=taylor+swift, 3914-https://www.google.be/#q=bose, 3915-https://www.google.be/#q=yellowstone, 3916-https://www.google.se/, 3917-https://www.google.se/#q=taylor+swift, 3918-https://www.google.se/#q=yellowstone, 3919-https://www.google.se/#q=bose, 3920-http://www.china.com.cn/, 3921-http://news.china.com.cn/node_7206134.htm, 3922-http://news.china.com.cn/politics/2014-12/01/content_34197014.htm, 3923-http://news.china.com.cn/2014-11/06/content_33989711.htm, 3924-http://news.china.com.cn/2014-12/01/content_34195875.htm, 3925-http://finance.china.com.cn/industry/hotnews/20141201/2821019.shtml, 3926-http://www.ndtv.com/, 3927-http://www.ndtv.com/article/cities/church-gutted-in-east-delhi-police-say-arson-628449?utm_source=ndtv&utm_medium=top-stories-widget&utm_campaign=story-2-http\%3a\%2f\%2fwww.ndtv.com\%2farticle\%2fcities\%2fchurch-gutted-in-east-delhi-police-say-arson-628449, 3928-http://www.ndtv.com/article/cheat-sheet/rohtak-sisters-who-took-on-harassers-to-be-honoured-on-republic-day-10-developments-628360?utm_source=ndtv&utm_medium=top-stories-widget&utm_campaign=story-5-http\%3a\%2f\%2fwww.ndtv.com\%2farticle\%2fcheat-sheet\%2frohtak-sisters-who-took-on-harassers-to-be-honoured-on-republic-day-10-developments-628360, 3929-http://profit.ndtv.com/news/corporates/article-radio-cab-business-set-to-zoom-on-the-fast-lane-704947?utm_source=ndtv&utm_medium=top-stories-widget&utm_campaign=story-10-http\%3a\%2f\%2fprofit.ndtv.com\%2fnews\%2fcorporates\%2farticle-radio-cab-business-set-to-zoom-on-the-fast-lane-704947, 3930-http://gadgets.ndtv.com/apps/news/google-unveils-best-apps-of-2014-section-on-play-store-628302?utm_source=ndtv&utm_medium=top-stories-widget&utm_campaign=story-16-http\%3a\%2f\%2fgadgets.ndtv.com\%2fapps\%2fnews\%2fgoogle-unveils-best-apps-of-2014-section-on-play-store-628302, 3931-http://cooks.ndtv.com/article/show/how-to-lose-weight-according-to-ayurveda-627471?utm_source=ndtv&utm_medium=top-stories-widget&utm_campaign=story-17-http\%3a\%2f\%2fcooks.ndtv.com\%2farticle\%2fshow\%2fhow-to-lose-weight-according-to-ayurveda-627471, 3932-http://www.aili.com/, 3933-http://celeb.aili.com/2239/2612722p.html, 3934-http://luxury.aili.com/2356/2612641.html, 3935-http://digi.aili.com/2258/2612645p.html, 3936-http://fashion.aili.com/233/2612511.html, 3937-http://beauty.aili.com/4/2612701.html, 3938-http://web.de/, 3939-http://web.de/magazine/unterhaltung/topde/#.homepage.news_spotlight.Comeback\%20mit\%20XXL-Muckis.0, 3940-http://web.de/magazine/unterhaltung/tv-film/paare-hochzeit-blick-

30248244, 3941-http://web.de/magazine/politik/us-luftwaffe-bombardiert-strategie-30249676, 3942-http://web.de/magazine/reise/pkw-winterurlaub-30249440, 3943-http://suche.web.de/web?q=Schlag+den+Raab&rq=true&origin=hotspots,

APPENDIX 3: LIST OF WEB PAGE FEATURES

This appendix includes a list of the web page features analyzed in Chapter 3 and 4.

Primary Traffic Features

1 - Client sends SYN segments, 2 - server sends SYN segments, 3 - Bidirectional SYN segments, 4 - Client sends RESET segments, 5 - server sends SYNACK segments, 6 - server sends RESET segments, 7 - Bidirectional SYNACK segments, 8 - Bidirectional RESET segments, 9 - Client sends PUSH segments, 10 - Client sends FIN segments, 11 - server sends PUSH segments, 12 - server sends FIN segments, 13 - Bidirectional PUSH segments, 14 - Bidirectional FIN segments, 15 - Client sends segments, 16 - Client sends Bytes, 17 - server sends segments, 18 - server sends Bytes, 19 - Bidirectional segments, 20 - Bidirectional Bytes, 21 - number of IPPairs, 22 - number of TCP connections, 23- number of DNS requests, 24- Sum Web Object Requests, 25- number of unused TCP connections, 26 - number of HTTPS connections

TCP/IP header-based features: 1 - Client sends SYN segments, 2 - server sends SYN segments, 3 - Bidirectional SYN segments, 4 - Client sends RESET segments, 5 - server sends SYNACK segments, 6 - server sends RESET segments, 7 - Bidirectional SYNACK segments, 8 - Bidirectional RESET segments, 9 - Client sends PUSH segments, 10 - Client sends FIN segments, 11 - server sends PUSH segments, 12 - server sends FIN segments, 13 - Bidirectional PUSH segments, 14 - Bidirectional FIN segments, 15 - Client sends segments, 16 - Client sends Bytes, 17 - server sends segments, 18 - server sends Bytes, 19 - Bidirectional segments, 20 - Bidirectional Bytes, 21 - number of IPPairs, 22 - number of TCP connections, 23 - Mean number of RESET segments per TCP connection server sends, 24 - Minimum RESET per TCP connection server, 25 - 10 percentile RESET per TCP connection server, 26 - 25 percentile RESET per TCP connection server, 27 - 50 percentile RESET per TCP connection server, 28 - 75 percentile RESET per TCP connection server, 29 - 90 percentile RESET per TCP connection server, 30 - Maximum percentile RESET per TCP connection server, 31 - Mean number of RESET segments per TCP connection Client sends, 32 - Minimum RESET per TCP connection server, 33 - 10 percentile RESET per TCP connection client, 34 - 25 percentile RESET per TCP connection client, 35 - 50 percentile RESET per TCP connection client, 36 - 75 percentile RESET per TCP connection client, 37 - 90 percentile RESET per TCP connection client, 38 - Maximum percentile RESET per TCP connection client, 39 - Mean number of PUSH segments per TCP connection Client sends, 40 - Minimum PUSH per TCP connection client, 41 - 10 percentile PUSH per TCP connection client, 42 - 25 percentile PUSH per TCP connection client, 43 - 50 percentile PUSH per TCP connection client, 44 - 75 percentile PUSH per TCP connection client, 45 - 90 percentile PUSH per TCP connection client, 46 - Maximum percentile PUSH per TCP connection client, 47 - Mean number of PUSH segments per TCP connection server sends, 48 - Minimum PUSH per TCP connection server, 49 - 10 percentile PUSH per TCP connection server, 50 - 25 percentile PUSH per TCP connection server, 51 - 50 percentile PUSH per TCP connection server, 52 - 75 percentile PUSH per TCP connection server, 53 - 90 percentile PUSH per TCP connection server, 54 - Maximum percentile PUSH per TCP connection server, 55 - Mean number of bytes per TCP connection Client sends, 56 - Minimum bytes per TCP connection client, 57 - 10 percentile bytes per TCP connection client, 58 - 25 percentile bytes per TCP connection client, 59 - 50 percentile bytes per TCP connection client, 60 - 75 percentile bytes per TCP connection client, 61 - 90 percentile bytes per TCP connection client, 62 - Maximum percentile bytes per TCP connection client, 63 - number of unused TCP connections client client, 64 - Mean number of bytes per TCP connection server sends, 65 - Minimum bytes per TCP connection server, 66 - 10 percentile bytes per TCP connection server, 67 - 25 percentile bytes per TCP connection server, 68 - 50 percentile bytes per TCP connection server, 69 - 75 percentile bytes per TCP connection server, 70 - 90 percentile bytes per

TCP connection server, 71 - Maximum percentile bytes per TCP connection server, 72 - number of unused TCP connections client server, 73 - number of unused TCP connections server, 74 - Mean number of bytes per IP Pair Client sends, 75 - Minimum bytes per IP Pair client, 76 - 10 percentile bytes per IP Pair client, 77 - 25 percentile bytes per IP Pair client, 78 - 50 percentile bytes per IP Pair client, 79 - 75 percentile bytes per IP Pair client, 80 - 90 percentile bytes per IP Pair client, 81 - Maximum percentile bytes per IP Pair client, 82 - number of unused IP Pairs client , 83 - Mean number of bytes per IP Pair server sends, 84 - Minimum bytes per IP Pair client, 85 - 10 percentile bytes per IP Pair server, 86 - 25 percentile bytes per IP Pair server, 87 - 50 percentile bytes per IP Pair server, 88 - 75 percentile bytes per IP Pair server, 89 - 90 percentile bytes per IP Pair server, 90 - Maximum percentile bytes per IP Pair server, 91 - number of unused IP Pairs server, 92 - number of port 80 TCP connections, 93 - Mean RTT port 80 TCP connection, 94 - Minimum RTT port 80 TCP connection, 95 - 10 percentile RTT port 80 TCP connection, 96 - 25 percentile RTT port 80 TCP connection, 97 - 50 percentile RTT port 80 TCP connection, 98 - 75 percentile RTT port 80 TCP connection, 99 - 90 percentile RTT port 80 TCP connection, 100 - Maximum percentile RTT port 80 TCP connection, 101 - number of port 443 TCP connections, 102 - Mean RTT port 443 TCP connection, 103 - Minimum RTT port 443 TCP connection, 104 - 10 percentile RTT port 443 TCP connection, 105 - 25 percentile RTT port 443 TCP connection, 106 - 50 percentile RTT port 443 TCP connection, 107 - 75 percentile RTT port 443 TCP connection, 108 - 90 percentile RTT port 443 TCP connection, 109 - Maximum percentile RTT port 443 TCP connection, 110 - number of Unused Secure connections, 111 - number of port 443 and 80 TCP connections, 112 - Mean RTT port 443 and 80 TCP connection, 113 - Minimum RTT port 443 and 80 TCP connection, 114 - 10 percentile RTT port 443 and 80 TCP connection, 115 - 25 percentile RTT port 443 and 80 TCP connection, 116 - 50 percentile RTT port 443 and 80 TCP connection, 117 - 75 percentile RTT port 443 and 80 TCP connection, 118 - 90 percentile RTT port 443 and 80 TCP connection, 119 - Maximum percentile RTT port 443 and 80 TCP connection, 120 - duration TCP connections, 121 - Mean duration TCP connection, 122 - Minimum duration TCP connection, 123 - 10 percentile duration TCP connection, 124 - 25 percentile duration TCP connection, 125 - 50 percentile duration TCP connection, 126 - 75 percentile duration TCP connection, 127 - 90 percentile duration TCP connection, 128 - Maximum percentile duration TCP connection, 129 - Sum duration TCP connections, 130 - Mean bytes port 443 TCP connection, 131 - Minimum bytes per 443 TCP connection , 132 - 10 percentile 443 bytes per TCP Connection, 133 - 25 percentile 443 bytes per TCP Connection, 134 - 50 percentile 443 bytes per TCP Connection, 135 - 75 percentile 443 bytes per TCP Connection, 136 - 90 percentile 443 bytes per TCP Connection, 137 - Maximum percentile 443 bytes per TCP Connection, 138 - Sum bytes port 443 TCP connection, 139 - Mean bytes port 80 TCP connection, 140 - Minimum bytes per 80 TCP connection server, 141 - 10 percentile bytes per 80 TCP connection server, 142 - 25 percentile bytes per 80 TCP connection server, 143 - 50 percentile bytes per 80 TCP connection server, 144 - 75 percentile bytes per 80 TCP connection server, 145 - 90 percentile bytes per 80 TCP connection server, 146 - Maximum percentile bytes per 80 TCP connection server, 147 - Sum bytes port 80 TCP connection, 148 - Mean duration port 80 TCP connection, 149 - Minimum duration port 80, 150 - 10 percentile duration port 80, 151 - 25 percentile duration port 80, 152 - 50 percentile duration port 80, 153 - 75 percentile duration port 80, 154 - 90 percentile duration port 80, 155 - Maximum percentile duration port 80, 156 - Mean duration port 443 TCP connection, 157 - Minimum duration port 443, 158 - 10 percentile duration port 443, 159 - 25 percentile duration port 443, 160 - 50 percentile duration port 443, 161 - 75 percentile duration port 443, 162 - 90 percentile duration port 443, 163 - Maximum percentile duration port 443, 164 - Mean inter TCP connection arrival time, 165 - Minimum inter TCP connection arrival time, 166 - 10 percentile inter TCP connection arrival time, 167 - 25 percentile inter TCP connection arrival time, 168 - 50 percentile inter TCP connection arrival time, 169 - 75 percentile inter TCP connection arrival time, 170 - 90 percentile inter TCP connection arrival time, 171 - Maximum percentile inter TCP connection arrival time, 172 - Mean number of epochs per TCP connection, 173 - Minimum number of epochs per TCP connection, 174 - 10 percentile number of epochs per TCP connection, 175 - 25 percentile number of epochs per TCP connection, 176 - 50 percentile number of epochs per TCP connection, 177 - 75 percentile number of epochs per TCP connection, 178 - 90 percentile number of epochs per TCP connection, 179 - Maximum percentile number of epochs per TCP connection, 180 - Minimum bytes sent client per epoch, 181 - number of epochs, 182 - Mean number of bytes sent by client per epoch, 183 - Minimum bytes sent client per epoch, 184 - 10 percentile number of bytes sent by client per epoch, 185 - 25 percentile number of bytes sent by client per epoch, 186 - 50 percentile number of bytes sent by client per epoch, 187 - 75 percentile number of bytes sent by client per epoch, 188 - 90 percentile number of bytes sent by client per epoch, 189 - Maximum percentile number of bytes sent by client per epoch, 190 - Sum number of bytes sent by client epoch, 191 - Minimum bytes sent client per epoch, 192 - Mean interarrival epoch, 193 - Minimum interarrival epoch, 194 - 10 percentile interarrival epoch, 195 - 25 percentile interarrival epoch, 196 - 50 percentile interarrival epoch, 197 - 75 percentile interarrival epoch, 198 - 90 percentile interarrival epoch, 199 - Maximum percentile interarrival epoch, 200 - Time until 5 percent bytes load, 201 - Time until 10 percent bytes load, 202 - Time until 25 percent bytes load, 203 - Time until 50 percent bytes load, 204 - Time until 75 percent bytes load, 205 - Time until 90 percent bytes load, 206 - Time until 95 percent Bytes, 207 - 235 , 208 - Mean number of bytes sent by server per epoch, 209 - Minimum bytes sent server per epoch, 210 - 10 percentile number of bytes sent by server per epoch, 211 - 25 percentile number of bytes sent by server per epoch, 212 - 50 percentile number of bytes sent by server per epoch, 213 - 75 percentile number of bytes sent by server per epoch, 214 - 90 percentile number of bytes sent by server per epoch, 215 - Maximum percentile number of bytes sent by server per epoch, 216 - Sum number of bytes sent by server epoch

HTTP-based features:

217 - Mean Web Object Length (HTTP level), 218 - 10 percentile Web Object Length (HTTP level) , 219 - 25 percentile Web Object Length (HTTP level), 220 - 50 percentile Web Object Length (HTTP level), 221 - 75 percentile Web Object Length (HTTP level), 222 - 90 percentile Web Object Length (HTTP level), 223 - Maximum percentile Web Object Length (HTTP level), 224 - Sum Web Object Requests , 225 - Mean Total Web Object Length per TCP connections, 226 - Minimum Total Web Object Length per TCP connections, 227 - 10 Total Web Object Length per TCP connections, 228 - 25 Total Web Object Length per TCP connections, 229 - 50 Total Web Object Length per TCP connections, 230 - 75 Total Web Object Length per TCP connections, 231 - 90 Total Web Object Length per TCP connections, 232 - Maximum Total Web Object Length per TCP connections, 233 - Sum Total Web Object Length across TCP connections, 234 - Mean GET request per TCP connections, 235 - Minimum GET request per TCP connections, 236 - 10 GET request per TCP connections, 237 - 25 GET request per TCP connections, 238 - 50 GET request per TCP connections, 239 - 75 GET request per TCP connections, 240 - 90 GET request per TCP connections, 241 - Maximum GET request per TCP connections, 242 - Sum GET Requests, 243 - Mean POST request per TCP connections, 244 - Minimum POST request per TCP connections, 245 - 10 POST request per TCP connections, 246 - 25 POST request per TCP connections, 247 - 50 POST request per TCP connections, 248 - 75 POST request per TCP connections, 249 - 90 POST request per TCP connections, 250 - Maximum POST request per TCP connections, 251 - Sum POST Requests, 252 - Mean HEAD request per TCP connections, 253 - Minimum HEAD request per TCP connections, 254 - 10 HEAD request per TCP connections, 255 - 25 HEAD request per TCP connections, 256 - 50 HEAD request per TCP connections, 257 - 75 HEAD request per TCP connections, 258 - 90 HEAD request

per TCP connections, 259 - Maximum HEAD request per TCP connections, 260 - Sum HEAD Requests, 261 - Mean TRACE request per TCP connections, 262 - Minimum TRACE request per TCP connections, 263 - 10 TRACE request per TCP connections, 264 - 25 TRACE request per TCP connections, 265 - 50 TRACE request per TCP connections, 266 - 75 TRACE request per TCP connections, 267 - 90 TRACE request per TCP connections, 268 - Maximum TRACE request per TCP connections, 269 - Sum TRACE Requests, 270 - Mean PUT request per TCP connections, 271 - Minimum PUT request per TCP connections, 272 - 10 PUT request per TCP connections , 273 - 25 PUT request per TCP connections, 274 - 50 PUT request per TCP connections, 275 - 75 PUT request per TCP connections, 276 - 90 PUT request per TCP connections , 277 - Maximum PUT request per TCP connections , 278 - Sum PUT Requests , 279 - Mean DELETE request per TCP connections, 280 - Minimum DELETE request per TCP connections, 281 - 10 DELETE request per TCP connections, 282 - 25 DELETE request per TCP connections, 283 - 50 DELETE request per TCP connections, 284 - 75 DELETE request per TCP connections, 285 - 90 DELETE request per TCP connections, 286 - Maximum DELETE request per TCP connections, 287 - Sum DELETE Requests, 288 - Mean OPTIONS request per TCP connections, 289 - Minimum OPTIONS request per TCP connections, 290 - 10 OPTIONS request per TCP connections, 291 - 25 OPTIONS request per TCP connections, 292 - 50 OPTIONS request per TCP connections, 293 - 75 OPTIONS request per TCP connections, 294 - 90 OPTIONS request per TCP connections, 295 - Maximum OPTIONS request per TCP connections, 296 - Sum OPTIONS Requests, 297 - Mean PATCH request per TCP connections, 298 - Minimum PATCH request per TCP connections, 299 - 10 PATCH request per TCP connections, 300 - 25 PATCH request per TCP connections, 301 - 50 PATCH request per TCP connections, 302 - 75 PATCH request per TCP connections, 303 - 90 PATCH request per TCP connections, 304 - Maximum PATCH request per TCP connections, 305 - Sum PATCH Requests, 306 - Mean HTTP request per TCP connections, 307 - Minimum HTTP request per TCP connections, 308 - 10 HTTP request per TCP connections, 309 - 25 HTTP request per TCP connections, 310 - 50 HTTP request per TCP connections, 311 - 75 HTTP request per TCP connections, 312 - 90 HTTP request per TCP connections, 313 - Maximum HTTP request per TCP connections, 314 - Sum HTTP Requests, 315 - number of HTTP hostnames contacted, 316 - Mean number of HTTP objects per hostname, 317 - Minimum number of HTTP objects per hostname, 318 - 10 number of HTTP objects per hostname, 319 - 25 number of HTTP objects per hostname, 320 - 50 number of HTTP objects per hostname, 321 - 75 number of HTTP objects per hostname, 322 - 90 number of HTTP objects per hostname, 323 - Maximum number of HTTP objects per hostname, 324 - Total number of HTTP objects for hostname factor, 325 - Mean number Duplication of an Object, 326 - Minimum number Duplication of an Object, 327 - 10 number Duplication of an Object, 328 - 25 number Duplication of an Object, 329 - 50 number Duplication of an Object, 330 - 75 number Duplication of an Object, 331 - 90 number Duplication of an Object, 332 - Maximum number Duplication of an Object, 333 - number Objects Object, 334 - Mean number of IP addresses Contacted for each Object, 335 - Minimum number of IP addresses Contacted for each Object, 336 - 10 number of IP addresses Contacted for each Object, 337 - 25 number of IP addresses Contacted for each Object, 338 - 50 number of IP addresses Contacted for each Object, 339 - 75 number of IP addresses Contacted for each Object, 340 - 90 number of IP addresses Contacted for each Object, 341 - Maximum number of IP addresses Contacted for each Object, 342 - number of IP addresses Contacted for each Object, 343 - Mean number Duplication of an Object, 344 - Minimum number Duplication of an Object, 345 - 10 number Duplication of an Object, 346 - 25 number Duplication of an Object, 347 - 50 number Duplication of an Object, 348 - 75 number Duplication of an Object, 349 - 90 number Duplication of an Object, 350 - Maximum number Duplication of an Object, 351 - number Objects Object, 352 - Mean number of IP addresses Contacted for each Object, 353 - Minimum number of IP addresses Contacted for each

Object, 354 - 10 number of IP addresses Contacted for each Object, 355 - 25 number of IP addresses Contacted for each Object, 356 - 50 number of IP addresses Contacted for each Object, 357 - 75 number of IP addresses Contacted for each Object, 358 - 90 number of IP addresses Contacted for each Object, 359 - Maximum number of IP addresses Contacted for each Object, 360 - number of IP addresses Contacted for each Web Object Overall, 361 - Mean number of Hosts Contacted for each Object, 362 - Minimum number of Hosts Contacted for each Object, 363 - 10 number of Hosts Contacted for each Object, 364 - 25 number of Hosts Contacted for each Object, 365 - 50 number of Hosts Contacted for each Object, 366 - 75 number of Hosts Contacted for each Object, 367 - 90 number of Hosts Contacted for each Object, 368 - Maximum number of Hosts Contacted for each Object, 369 - Status code 200, 370 - Status code 204 , 371 - Status code 300 , 372 - Status code 301 , 373 - Status code 302 , 374 - Status code 304 , 375 - Status code 307 , 376 - Status code 400 , 377 - Status code 401 , 378 - Status code 403 , 379 - Status code 404 , 380 - Status code 410 , 381 - Status code 500 , 382 - Status code 501 , 383 - Status code 503 , 384 - Status code 550 , 385 - Total Web Object Length (HTTP level), 386 - MIME type application, 387 - MIME type text, 388 - MIME type image, 389 - MIME type video, 390 - MIME type audio, 391 - MIME type font, 392 - MIME type content, 393 - MIME type binary, 394 - MIME type application/js, 395 - MIME type application/swf, 396 - MIME type application/xhtmlxml, 397 - MIME type application/xjavascript, 398 - MIME type application/pkixcrl, 399 - MIME type application/javascript, 400 - MIME type application/xml, 401 - MIME type application/fontwoff, 402 - MIME type application/rssxml, 403 - MIME type application/pdf, 404 - MIME type application/fontttf, 405 - MIME type application/xpnacl, 406 - MIME type application/vndgoogle, 407 - MIME type application/xfontwoff, 408 - MIME type application/xmsdod, 409 - MIME type application/opensearch, 410 - MIME type application/msfont, 411 - MIME type application/xjson, 412 - MIME type application/xshockwaveflash, 413 - MIME type application/xfontttf, 414 - MIME type application/json, 415 - MIME type application/xjavascript, 416 - MIME type application/unknown, 417 - MIME type application/fonteot, 418 - MIME type application/xmlxhtml, 419 - MIME type application/vndgooglesafebrowsingupdate, 420 - MIME type application/xfcs, 421 - MIME type application/smil, 422 - MIME type application/octet_stream, 423 - MIME type application/xamf, 424 - MIME type application/xwoff, 425 - MIME type application/atomxml, 426 - MIME type application/xchromeextension, 427 - MIME type application/xpkcs7crl, 428 - MIME type application/ocspresponse, 429 - MIME type binary/octetstream, 430 - MIME type image/webp, 431 - MIME type image/jpeg, 432 - MIME type image/png, 433 - MIME type image/bmp, 434 - MIME type image/xicon, 435 - MIME type image/vndmicrosoft, 436 - MIME type image/svgxml, 437 - MIME type image/gif, 438 - MIME type multipartmixed, 439 - MIME type video/realgravity, 440 - MIME type video/xflv, 441 - MIME type video/mp4, 442 - MIME type video/xm4v, 443 - MIME type video/webm, 444 - MIME type video/xmsasf, 445 - MIME type audio/mp4, 446 - MIME type audio/ogg, 447 - MIME type audio/mpeg, 448 - MIME type contentunknown, 449 - MIME type font/xwoff, 450 - MIME type font/eot, 451 - MIME type font/woff, 452 - MIME type font/ttf, 453 - MIME type text/css, 454 - MIME type text/crossdomain, 455 - MIME type text/html, 456 - MIME type text/javascript, 457 - MIME type text/json, 458 - MIME type text/cachemanifest, 459 - MIME type text/plain, 460 - MIME type text/xml, 461 - MIME type text/xcomponent, 462 - MIME type text/ecmascript, 463 - MIME type text/xcrossdomain, 464 - MIME type text/xjson

DNS-based features: 465 - Mean number of DNS responses per Query, 466 - Minimum number of DNS responses per Query, 467 - 10 number of DNS responses per Query, 468 - 25 number of DNS responses per Query, 469 - 50 number of DNS responses per Query, 470 - 75 number of DNS responses per Query, 471 - 90 number of DNS responses per Query, 472 - Maximum number of DNS responses per Query, 473 - Sum number of DNS responses ALL Queries, 474 - number of DNS responses ALL Queries, 475 - Mean number of DNS A responses per Query, 476 - Minimum number of DNS A responses per Query, 477 - 10 number of DNS A responses per Query, 478 - 25 number of DNS A responses per Query, 479 - 50 number of DNS A responses per Query, 480 - 75 number of DNS A responses per Query, 481 - 90 number of DNS A responses per Query, 482 - Maximum number of DNS A responses per Query, 483 - Sum number of DNS A responses ALL Queries, 484 - Mean number of DNS PTR responses per Query, 485 - Minimum number of DNS PTR responses per Query, 486 - 10 number of DNS PTR responses per Query, 487 - 25 number of DNS PTR responses per Query, 488 - 50 number of DNS PTR responses per Query, 489 - 75 number of DNS PTR responses per Query, 490 - 90 number of DNS PTR responses per Query, 491 - Maximum number of DNS PTR responses per Query, 492 - Sum number of DNS PTR responses ALL Queries, 493 - Mean number of DNS CNAME responses per Query, 494 - Minimum number of DNS CNAME responses per Query, 495 - 10 number of DNS CNAME responses per Query, 496 - 25 number of DNS CNAME responses per Query, 497 - 50 number of DNS CNAME responses per Query, 498 - 75 number of DNS CNAME responses per Query, 499 - 90 number of DNS CNAME responses per Query, 500 - Maximum number of DNS CNAME responses per Query, 501 - Sum number of DNS CNAME responses ALL Queries, 502 - Sum PTR overall, 503 - Sum A overall, 504 - Sum CNAME overall, 505 - number of DNS queries with TTL responses, 506 - Mean DNS TTL, 507 - Minimum DNS TTL, 508 - 10 percentile DNS TTL, 509 - 25 percentile DNS TTL, 510 - 50 percentile DNS TTL, 511 - 75 percentile DNS TTL, 512 - 90 percentile DNS TTL, 513 - Maximum percentile DNS TTL, 514 - number of DNS queries with A TTL responses, 515 - Mean DNS A TTL, 516 - Minimum DNS A TTL, 517 - 10 percentile DNS A TTL, 518 - 25 percentile DNS A TTL, 519 - 50 percentile DNS A TTL, 520 - 75 percentile DNS A TTL, 521 - 90 percentile DNS A TTL, 522 - Maximum percentile DNS A TTL, 523 - number of DNS queries with CNAME TTL responses, 524 - Mean DNS CNAME TTL, 525 - Minimum DNS CNAME TTL, 526 - 10 percentile DNS CNAME TTL, 527 - 25 percentile DNS CNAME TTL, 528 - 50 percentile DNS CNAME TTL, 529 - 75 percentile DNS CNAME TTL, 530 - 90 percentile DNS CNAME TTL, 531 - Maximum percentile DNS CNAME TTL, 532 - number of DNS queries with responses, 533 - number of DNS queries without responses, 534 - Mean DNS Response Time, 535 - Minimum DNS Response Time, 536 - 10 percentile DNS Response Time, 537 - 25 percentile DNS Response Time, 538 - 50 percentile DNS Response Time, 539 - 75 percentile DNS Response Time, 540 - 90 percentile DNS Response Time, 541 - Maximum percentile DNS Response Time, 542 - Total DNS queries, 543 - number of NXDomain responses, 544 - Mean DNS bytes Client, 545 - Minimum DNS bytes Client, 546 - 10 percentile DNS bytes Client, 547 - 25 percentile DNS bytes Client, 548 - 50 percentile DNS bytes Client, 549 - 75 percentile DNS bytes Client, 550 - 90 percentile DNS bytes Client, 551 - Maximum percentile DNS bytes Client, 552 - Mean DNS bytes Server, 553 - Minimum DNS bytes Server, 554 - 10 percentile DNS bytes Server, 555 - 25 percentile DNS bytes Server, 556 - 50 percentile DNS bytes Server, 557 - 75 percentile DNS bytes Server, 558 - 90 percentile DNS bytes Server, 559 - Maximum percentile DNS bytes Server, 560 - Mean DNS bytes Bidirectional, 561 - Minimum DNS bytes Bidirectional, 562 - 10 percentile DNS bytes Bidirectional, 563 - 25 percentile DNS bytes Bidirectional, 564 - 50 percentile DNS bytes Bidirectional, 565 - 75 percentile DNS bytes Bidirectional, 566 - 90 percentile DNS bytes Bidirectional, 567 - Maximum percentile DNS bytes Bidirectional, 568 - Mean DNS interarrival, 569 - Minimum DNS interarrival, 570 - 10 percentile DNS interarrival, 571 - 25 percentile DNS interarrival, 572 - 50 percentile DNS interarrival, 573 - 75 percentile DNS interarrival, 574 - 90 percentile DNS interarrival, 575 - Maximum percentile DNS interarrival

HTML-based features: 576 - number of code tags, 577 - number of kbd tags, 578 - number of tbody tags, 579 - number of font tags, 580 - number of noscript tags, 581 - number of style tags, 582 - number of img tags, 583 - number of title tags, 584 -

number of menu tags, 585 - number of tt tags, 586 - number of tr tags, 587 - number of param tags, 588 - number of li tags, 589 - number of source tags, 590 - number of tfoot tags, 591 - number of th tags, 592 - number of input tags, 593 - number of td tags, 594 - number of main tags, 595 - number of dl tags, 596 - number of blockquote tags, 597 - number of fieldset tags, 598 - number of extensions of type image, 599 - number of dd tags, 600 - number of meter tags, 601 - number of optgroup tags, 602 - number of dt tags, 603 - number of wbr tags, 604 - number of button tags, 605 - number of summary tags, 606 - number of p tags, 607 - number of menuitem tags, 608 - number of output tags, 609 - number of div tags, 610 - number of dir tags, 611 - number of em tags, 612 - number of datalist tags, 613 - number of frame tags, 614 - number of hgroup tags, 615 - number of meta tags, 616 - number of video tags, 617 - number of characters within script tags, 618 - number of .jpeg extension, 619 - number of rt tags, 620 - number of canvas tags, 621 - number of rp tags, 622 - number of sub tags, 623 - number of section tags, 624 - number of bdi tags, 625 - number of label tags, 626 - number of acronym tags, 627 - number of progress tags, 628 - number of body tags, 629 - number of HTML5 tags, 630 - number of basefont tags, 631 - number of small tags, 632 - number of base tags, 633 - number of br tags, 634 - number of address tags, 635 - number of article tags, 636 - number of strong tags, 637 - number of legend tags, 638 - number of ol tags, 639 - number of caption tags, 640 - number of s tags, 641 - number of dialog tags, 642 - number of col tags, 643 - number of a tags, 644 - number of h1 tags, 645 - number of header tags, 646 - number of table tags, 647 - number of select tags, 648 - number of noframes tags, 649 - number of span tags, 650 - number of area tags, 651 - number of .gif extension, 652 - number of mark tags, 653 - number of dfn tags, 654 - number of strike tags, 655 - number of cite tags, 656 - number of thead tags, 657 - number of head tags, 658 - number of option tags, 659 - number of form tags, 660 - number of var tags, 661 - Percent of HTML5 tags, 662 - number of ruby tags, 663 - number of b tags, 664 - number of colgroup tags, 665 - number of link tags, 666 - number of keygen tags, 667 - number of ul tags, 668 - number of applet tags, 669 - number of del tags, 670 - number of iframe tags, 671 - number of embed tags, 672 - number of pre tags, 673 - number of frameset tags, 674 - number of Included Elements, 675 - number of figure tags, 676 - number of ins tags, 677 - number of nonjavascript scripts, 678 - number of javascript scripts, 679 - number of isResponsive tags, 680 - number of aside tags, 681 - number of html tags, 682 - number of nav tags, 683 - number of details tags, 684 - number of samp tags, 685 - number of map tags, 686 - number of track tags, 687 - number of object tags, 688 - number of style tags, 689 - number of figcaption tags, 690 - number of script tags, 691 - number of .png extensions, 692 - number of center tags, 693 - number of textarea tags, 694 - number of footer tags, 695 - number of i tags, 696 - number of q tags, 697 - number of u tags, 698 - number of time tags, 699 - number of audio tags, 700 - number of abbr tags, 701 - number of Words, 702 - number of Different Words

APPENDIX 4: LIST OF STATISTICALLY SIGNIFICANT HTML-BASED MOBILE WEB PAGE FEATURES

This appendix includes a list of the statistically significant HTML-based mobile web page features.

1-number of tbody tags, 1.040251e-05, 2-number of font tags, 8.189950e-03, 3-number of img tags, 2.221504e-21, 4-number of noscript tags, 1.848504e-13, 5-number of style tags, 8.029846e-06, 6-number of title tags, 2.985116e-05, 7-number of tr tags, 3.388648e-14, 8-number of param tags, 9.754458e-06, 9-number of li tags, 7.207601e-45, 10-number of th tags, 4.937559e-06, 11-number of input tags, 1.922421e-25, 12-number of td tags, 6.939682e-14, 13-number of dl tags, 4.268245e-03, 14-number of fieldset tags, 2.361830e-09, 15-number of extensions of type image, 5.173600e-25, 16-number of dd tags, 5.369374e-03, 17-number of dt tags, 3.042719e-03, 18-number of wbr tags, 3.926414e-03, 19-number of button tags, 7.107152e-06, 20-number of p tags, 9.265052e-19, 21-number of div tags, 1.304351e-27, 22-number of em tags, 8.101569e-04, 23-number of meta tags, 1.406754e-12, 24-number of .jpeg extension, 2.867897e-02, 25-number of section tags, 3.130847e-07, 26-number of label tags, 7.414462e-31, 27-number of body tags, 3.598524e-08, 28-number of HTML5 tags, 5.245314e-05, 29-number of br tags, 1.291483e-09, 30-number of strong tags, 1.488312e-10, 31-number of legend tags, 3.048210e-05, 32-number of ol tags, 2.911072e-02, 33-number of nonjavascript scripts, 4.849987e-13, 34-number of s tags, 7.361247e-05, 35-number of a tags, 8.850201e-34, 36-number of h1 tags, 5.385250e-06, 37-number of header tags, 2.114162e-03, 38-number of table tags, 1.562030e-14, 39-number of select tags, 8.698680e-06, 40-number of span tags, 1.281466e-25, 41-number of area tags, 1.168378e-12, 42-number of .gif extension, 2.657471e-29, 43-number of strike tags, 8.593438e-03, 44-number of option tags, 4.112953e-06, 45-number of form tags, 5.795814e-24, 46-number of link tags, 4.637175e-16, 47-Percent of HTML5 tags, 5.245314e-05, 48-number of b tags, 2.650473e-11, 49-number of link tags, 1.213222e-17, 50-number of ul tags, 7.655419e-46, 51-number of iframe tags, 8.557930e-15, 52-number of embed tags, 3.585427e-02, 53-number of Included Elements, 1.330106e-19, 54-number of ins tags, 2.807726e-06, 55-number of script tags, 4.849987e-13, 56-number of isResponsive tags, 1.144375e-152, 57-number of map tags, 2.094166e-12, 58-number of object tags, 3.023122e-04, 59-number of javascript scripts, 2.488999e-12, 60-number of .png extensions, 5.148947e-17, 61-number of center tags, 4.900131e-02, 62-number of textarea tags, 3.586174e-03, 63-number of footer tags, 6.930330e-18, 64-number of i tags, 4.357242e-03, 65-number of q tags, 3.460826e-02, 66-number of Different Words, 7.580967e-65

APPENDIX 5: LIST OF MOBILE WEB PAGES STUDIED

This appendix includes a list of the mobile web pages studied in Section 1.5.2.

1-http://m.facebook.com/, 2-https://m.facebook.com/officialraylewis, 3-https://m.facebook.com/McDonalds, 4-https://m.facebook.com/stonemountainpark, 5-http://m.youtube.com/, 6-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=golden, 7-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=ray+lewis, 8-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=mcdonalds, 9-http://m.youtube.com/results?hl=en&gl=US&client=mv-google&q=stone+mountain, 10-http://m.yahoo.com/, 11-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=golden&x=0&y=0, 12-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=mcdonalds&x=0&y=0, 13-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=ray+lewis&x=0&y=0, 14-http://m.yahoo.com/w/search\%3B_ylt=A2KL8yAj1LxRZ1AA.AIp89w4?submit=oneSearch&.intl=US&.lang=en&.tsrc=yahoo&.sep=fp&p=stone+mountain&x=0&y=0, 15-http://m.yahoo.com/w/ygo-frontpage/lp/story/us/3348478/coke.bp\%3B_ylt=A2KL8xzg1bxR2UwAAggp89w4\%3B_ylu=X3oDMTFybGdnODRoBGNwb3MDMQRjc2VjA21vYmlsZS10ZARpbnRsA3VzBHBrZwNpZC=0zMzQ4NDc4BHBvcwMzBHNsawNtb3Jl?ref_w=frontdoors&view=today&.intl=US&.lang=en, 16-http://m.yahoo.com/w/ygo-frontpage/lp/story/us/3346216/coke.bp\%3B_ylt=A2KL8xv.1bxRYTMAMg4p89w4\%3B_ylu=X3oDMTFzc21sYmhiBGNwb3MDMwRjc2VjA21vYmlsZS10ZARpbnRsA3VzBHBrZwNpZC0zM=zQ2MjE2BHBvcwMxBHNsawN0aHVtYg--?ref_w=frontdoors&view=today&.intl=US&.lang=en, 17-http://m.yahoo.com/w/ygo-frontpage/lp/story/us/3345618/coke.bp\%3B_ylt=A2KL8yBo17xRO3UA9QIp89w4\%3B_ylu=X3oDMTFzMjZpYWIwBGNwb3MDNARjc2VjA21vYmlsZS10ZARpbnRsA3VzBHBrZwNp=ZC0zMzQ1NjE4BHBvcwMxBHNsawN0aHVtYg--?ref_w=frontdoors&view=today&.intl=US&.lang=en, 18-http://www.amazon.com/gp/aw, 19-http://www.amazon.com/gp/aw/s/ref=is_box_?k=golden, 20-http://www.amazon.com/gp/aw/d/1442452161/ref=mp_s_a_1_1?qid=1371330635&sr=8-1, 21-http://www.amazon.com/gp/aw/s/ref=is_s_?ie=UTF8&k=ray+lewis&url=i\%3Daps, 22-http://www.amazon.com/gp/aw/d/B00BEZQJ5A/ref=mp_s_a_1_10?qid=1371330819&sr=8-10, 23-http://www.amazon.com/gp/aw/s/ref=is_s_?ie=UTF8&k=stone+mountain&url=i\%3Daps, 24-http://www.amazon.com/gp/aw/d/1596296828/ref=mp_s_a_1_1?qid=1371331323&sr=8-1, 25-http://m.bing.com/, 26-http://m.bing.com/search?q=golden, 27-http://m.bing.com/search?q=ray+lewis, 28-http://m.bing.com/search?q=stone+mountain, 29-http://m.bing.com/search?q=mcdonalds, 30-http://m.bing.com/images/search?q=golden, 31-http://m.bing.com/images/search?q=ray+lewis, 32-http://m.bing.com/images/search?q=stone+mountain, 33-http://m.bing.com/images/

search?q=mcdonalds, 34-http://en.m.wikipedia.org/wiki/Main_Page, 35-http://en.m.wikipedia.org/wiki/Golden, 36-http://en.m.wikipedia.org/wiki/Ray_Lewis, 37-http://en.m.wikipedia.org/wiki/Stone_Mountain, 38-http://en.m.wikipedia.org/wiki/Mcdonalds, 39-https://touch.www.linkedin.com/login.html, 40-https://login.live.com/?wa=wsignin1.0&rpsnv=11&ct=1368921021&rver=6.1.6206.0&wp=MBI_SSL_SHARED&wreply=https\%3a\%2f\%2fmail.live.com\%2fm\%2f\%3ffl\%3d635045178212091183&lc=1033&id=64855&mspco=1&pcexp=false, 41-https://mobile.twitter.com/signup, 42-https://mobile.twitter.com/search?q=mcdonalds&src=typd, 43-https://mobile.twitter.com/search?q=ray+lewis&src=typd, 44-https://mobile.twitter.com/search?q=golden&src=typd, 45-https://mobile.twitter.com/search?q=stone+mountain&src=typd, 46-https://mobile.twitter.com/raylewis, 47-https://mobile.twitter.com/McDonalds, 48-https://mobile.twitter.com/StoneMtnPark, 49-https://mobile.twitter.com/Ashton5SOS, 50-http://www.blogger.com/mobile-start.g, 51-http://m.aol.com/portal/, 52-http://m.aol.com/search/aol/search?q=mcdonalds&s_it=srch_entr, 53-http://m.aol.com/search/aol/search?q=ray+lewis&s_it=srch_entr, 54-http://m.aol.com/search/aol/search?q=stone+mountain&s_it=srch_entr, 55-http://m.aol.com/search/aol/search?q=mcdonalds&s_it=srch_entr, 56-http://m.aol.com/search/aol/images?q=ray+lewis&v_t=srch_entr, 57-http://m.aol.com/search/aol/images?q=mcdonalds&v_t=srch_entr, 58-http://m.aol.com/search/aol/images?q=stone+mountain&v_t=srch_entr, 59-http://m.aol.com/search/aol/images?q=golden&v_t=srch_entr, 60-http://www.huffingtonpost.com/2013/06/16/pope-francis-blessing, 61-http://www.huffingtonpost.com/2013/06/14/the-best-summer-ice-cream_n_3417468.html?icid=maing-grid7\%7Cmain5\%7Cdl2\%7Csec1_lnk2\%26pLid\%3D329265, 62-http://www.sportingnews.com/mlb/story/2013-06-15/alex-cobb-tampa-bay-rays-hit-by-line-drive-eric-hosmer-royal-skull-stretcher?icid=maing-grid7\%7Cmain5\%7Cdl3\%7Csec1_lnk2\%26pLid\%3D330099, 63-http://m.pinterest.com/, 64-http://m.pinterest.com/search/pins/?q=ray+lewis, 65-http://m.pinterest.com/search/pins/?q=mcdonalds, 66-http://m.pinterest.com/search/pins/?q=stone+mountain, 67-http://m.pinterest.com/search/pins/?q=golden, 68-http://m.pinterest.com/pin/9922061651802862/, 69-http://m.pinterest.com/pin/510103095264410678/, 70-http://m.pinterest.com/pin/34199278393284055/, 71-http://m.pinterest.com/pin/29766047510066706/, 72-http://m.now.msn.com/, 73-http://now.msn.com/SiteSearch?q=golden&x=0&y=0&form=MSNTRE, 74-http://m.now.msn.com/SiteSearch?q=ray+lewis&x=0&y=0&form=MSNTRE, 75-http://m.now.msn.com/SiteSearch?q=stone+mountain&x=0&y=0&form=MSNTRE, 76-http://m.now.msn.com/SiteSearch?q=mcdonalds&x=0&y=0&form=MSNTRE, 77-https://mobile.paypal.com/us/cgi-bin/wapapp?cmd=_wapapp-homepage, 78-http://m.espn.go.com/wireless/, 79-http://m.espn.go.com/wireless/search/results?q=mcdonalds&fromForm=true, 80-http://m.espn.go.com/wireless/search/results?q=ray+lewis&fromForm=true, 81-http://m.espn.go.com/wireless/search/results?q=stone+mountain&fromForm=true, 82-http://m.espn.

go.com/wireless/search/results?q=golden&fromForm=true, 83-https://www.bankofamerica.com/mobile/banking.go, 84-http://m.microsoft.com/en-us/default.mspx, 85-http://m.microsoft.com/en-us/Products/default.mspx?prodtype=office, 86-https://mobilebanking.chase.com/, 87-http://www.foxnews.mobi/, 88-http://www.foxnews.mobi/quickPage.html?page=38321&content=94097423, 89-http://www.foxnews.mobi/quickPage.html?page=38321&content=94098112, 90-http://politics.foxnews.mobi/quickPage.html?page=23888&external=2194212.proteus.fma, 91-http://m.imdb.com/, 92-http://m.imdb.com/news/ni55804425, 93-http://m.imdb.com/title/tt0944947/, 94-http://m.imdb.com/name/nm0227759/?ref_=tt_cl_t1, 95-http://m.imdb.com/title/tt0340377/, 96-http://m.imdb.com/find?q=golden&button.x=0&button.y=0&button=Search, 97-http://m.imdb.com/find?q=ray+lewis&button.x=0&button.y=0&button=Search, 98-http://m.imdb.com/find?q=stone+mountain&button.x=0&button.y=0&button=Search, 99-http://m.imdb.com/find?q=mcdonalds&button.x=0&button.y=0&button=Search, 100-http://m.pornhub.com/, 101-http://m.pornhub.com/video/search?query=mcdonalds, 102-http://m.pornhub.com/video/search?query=stone+mountain, 103-http://m.pornhub.com/video/search?query=ray+lewis, 104-http://m.pornhub.com/video/search?query=golden, 105-http://m.comcast.net/m/, 106-http://m.comcast.net/m/articles/news-general/20130616/US-The-Secret-Government/, 107-http://m.comcast.net/m/articles/news-general/20130616/GLF--US.Open-Stefani.Ace/, 108-http://m.comcast.net/m/articles/news-politics/20130616/US-Cheney/, 109-http://mobile.nytimes.com/, 110-http://mobile.nytimes.com/2013/06/17/world/europe/turkey.html?from=homepage, 111-http://mobile.nytimes.com/2013/06/17/business/economy/for-g-8-meeting-talk-of-economy-but-syria-looms-large.html?from=homepage, 112-http://mobile.nytimes.com/travel/2013/06/16/travel/looking-for-clementine-hunters-louisiana.html?from=homepage, 113-http://mobile.nytimes.com/search?query=golden&sort=rel, 114-http://mobile.nytimes.com/search?query=ray+lewis&sort=rel, 115-http://mobile.nytimes.com/search?query=stone+mountain&sort=rel, 116-http://mobile.nytimes.com/search?query=mcdonalds&sort=rel, 117-http://news.mobile.msn.com/en-us/default.aspx, 118-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=golden&go=Go, 119-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=ray+lewis&go=Go, 120-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=stone+mountain&go=Go, 121-http://m.bing.com/search/?MID=3001&LC=en-us&PC=msn.msnbc.msn&h=4msd029m&q=mcdonalds&go=Go, 122-http://news.mobile.msn.com/en-us/articles.aspx?aid=18987063&afid=11, 123-http://news.mobile.msn.com/en-us/articles.aspx?aid=6C10313219&afid=11, 124-http://news.mobile.msn.com/en-us/articles.aspx?aid=18987359&afid=11, 125-http://news.mobile.msn.com/en-us/articles.aspx?aid=18987395&afid=11, 126-http://m.xhamster.com/, 127-http://m.xhamster.com/search.html?search=golden, 128-http://m.xhamster.com/search.html?search=ray+lewis, 129-http://m.xhamster.com/search.html?search=stone+mountain, 130-http://m.xhamster.com/search.html?search=mcdonalds, 131-http://mobile.walmart.com/, 132-http://mobile.walmart.com/m/searchr;jsessionid=75DCE9533EB1E2432ADBE096151DAA6B?search_query=golden, 133-http://mobile.walmart.com/ip/23914822, 134-http://mobile.walmart.com/m/searchr?search_query=ray+lewis, 135-http://mobile.walmart.com/ip/Inspector-Lewis-Series-4-Blu-ray-Widescreen/17126113, 136-http://mobile.walmart.com/m/searchr;jsessionid=0D4DCE2A266C7F64307D26911AE25830?search_query=stone+mountain, 137-http://mobile.walmart.com/ip/Mountain-Jewels-Gravel-Aquariums-25-lb/10449968, 138-http://mobile.walmart.com/m/searchr;jsessionid=C9B61EC47010A7B9F8988B0BA7F10003?search_query=mcdonalds, 139-http://mobile.walmart.com/ip/Mary-McDonald-Interiors-The-Allure-of-Style/13430747, 140-https://m.wellsfargo.com/, 141-http://m.reddit.com/, 142-http://i.imgur.com/7deC2lV.png, 143-http://m.flickr.com/#/home, 144-http://m.flickr.com/#/search/advanced/_QM_q_IS_golden, 145-http://m.flickr.com/#/search/advanced/_QM_q_IS_ray+lewis, 146-http://m.flickr.com/#/search/advanced/_QM_q_IS_stone+mountain, 147-http://m.flickr.com/#/search/advanced/_QM_q_IS_mcdonalds, 148-http://m.flickr.com/#/photos/luis_villablanca/9059399766/in/explore-1371382723/, 149-http://m.flickr.com/#/photos/snowyturner/9055322469/in/explore-1371382723/, 150-http://m.flickr.com/#/photos/vineetradhakrishnan/9058440177/in/explore-1371382723/, 151-http://m.flickr.com/#/photos/petezelewski/9058318873/in/explore-1371382723/, 152-http://m.livejasmin.com/en/, 153-http://m.target.com/, 154-http://m.target.com/s/golden#keywords=golden, 155-http://m.target.com/s/ray+lewis, 156-http://m.target.com/s/stone+mountain, 157-http://m.target.com/s/mcdonalds, 158-http://m.target.com/p/nfl-player-throw-ray-lewis/-/A-10148918, 159-http://m.target.com/p/sun-maid-golden-raisins-bag-10oz/-/A-13207041, 160-http://m.target.com/p/funky-monkey-bananamon-1oz/-/A-12935987, 161-http://m.target.com/p/ortega-yellow-corn-taco-shells-12-ct/-/A-13388903, 162-http://m.shopathome.com/, 163-http://m.shopathome.com/Search?sf=golden, 164-http://m.shopathome.com/Search?sf=ray+lewis, 165-http://m.shopathome.com/Search?sf=stone+mountain, 166-http://m.shopathome.com/Search?sf=mcdonalds, 167-http://m.redtube.com.brazzersmobile.com/, 168-http://m.homedepot.com/p/TrafficMaster-6-in-x-36-in-Golden-Maple-Resilient-Vinyl-Plank-Flooring-24-sq-ft-case-161215/100595231/, 169-http://m.homedepot.com/p/Delray-Plants-8-in-Golden-Pothos-HB-in-Plastic-Pot-8POTHB/202204580/, 170-http://m.att.com/, 171-http://m.usps.com/, 172-http://m.usps.com/MobileTrackPackage.aspx, 173-http://m.usps.com/MobilePOLocator.aspx, 174-http://m.usps.com/about.aspx, 175-https://m.ups.com/mobile/home, 176-https://m.ups.com/mobile/trackhome?loc=en_US, 177-https://m.ups.com/mobile/locator?loc=en_US, 178-https://m.ups.com/one-to-one/mdotlogin?returnto=https\%3a//m.ups.com/ums/m.ship\%3floc\%3den_US&reasonCode=-1&appid=UIS, 179-http://m.usatoday.com/, 180-http://m.usatoday.com/article/news/2464119, 181-http://m.usatoday.com/article/news/2481207, 182-http://m.usatoday.com/article/news/2479185, 183-http://m.dictionary.com/r/, 184-http://m.dictionary.com/r/?q=golden&submit-result-SEARCHR=Search, 185-http://m.dictionary.com/r/?q=ray+lewis&submit-result-SEARCHR=Search, 186-http://m.dictionary.com/r/?q=stone+mountain&submit-result-SEARCHR=Search, 187-http://m.dictionary.com/r/?q=mcdonalds&submit-result-SEARCHR=Search, 188-http://m.dictionary.com/r/?q=computerAD, 189-http://m.dictionary.com/r/?q=Guise, 190-http://m.dictionary.com/r/?q=family&submit-result-SEARCHR=Search, 191-http://m.godaddy.com/, 192-http://m.bestbuy.com/m/e/digital/, 193-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=9615177&pid=&ev=prodView, 194-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=9406171&pid=&ev=prodView, 195-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=6836663&pid=&ev=prodView, 196-http://m.bestbuy.com/m/e/product/detail.jsp?skuId=9136379&pid=&ev=prodView, 197-http://m.groupon.com/raleigh-durham?z=skip, 198-http://m.groupon.com/deals/shutterfly-334-raleigh-durham, 199-http://m.groupon.com/deals/the-golf-warriors, 200-http://m.groupon.com/deals/cinellis-3, 201-http://m.groupon.com/deals/amf-bowling-centers-nat-5-raleigh-durham, 202-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www, 203-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www/cards/login?redirect=www/cards/accounts, 204-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www/atm, 205-https://moblprod.capitalone.com/worklight/apps/services/www/EnterpriseMobileBanking/mobilewebapp/default/EnterpriseMobileBanking.html#www/contact, 206-http://www.idrudgereport.com/, 207-http://m.mlb.com/, 208-http://m.mlb.com/scores/, 209-http://m.mlb.com/news/article/2013070252487424/, 210-http://m.mlb.com/news/article/2013070252452294/, 211-http://m.mlb.com/news/article/2013070252456948/, 212-http://www.match.com/home/mymatch.aspx?lid=2, 213-http://m.fedex.com/mt/www.fedex.com/us/?un_zip_uat=&un_jtt_redirect, 214-https://m.verizonwireless.com/, 215-https://m.verizonwireless.com/storelocator, 216-https://m.verizonwireless.com/shop, 217-https://m.hootsuite.com/login?redirect=\%2F, 218-https://online.americanexpress.com/myca/mobl/us/login.do, 219-https://www262.americanexpress.com/card-application/unauth/featuredCardsPage.do?businessUnit=CCSG&inav=usmbl_menu_cards_personal_pr_body, 220-https://www262.americanexpress.com/card-application/unauth/cardDetail.do?id=1, 221-https://online.americanexpress.com/myca/mobl/us/static.do?page=un_help&content=CntUs&inav=usmbl_foot_gen_contact_pr, 222-http://m.tube8.com/, 223-http://m.monster.com/, 224-http://m.monster.com/JobSearch/Search?jobtitle=golden&keywords=&where=, 225-http://m.monster.com/JobSearch/Search?jobtitle=ray\%20lewis&keywords=&where=, 226-

http://m.monster.com/JobSearch/Search?jobtitle=stone\%20mountain&keywords=&where=, 227-http://m.monster.com/JobSearch/Search?jobtitle=mcdonalds&keywords=&where=, 228-http://m.monster.com/Oracle-Golden-Gate-DBA-Job-Cupertino-CA-123166382, 229-http://m.monster.com/Claims-Customer-Contact-Center-Supervisor-Golden-CO-Job-Englewood-CO-123479277, 230-http://m.monster.com/Casino-The-Golden-Gate-Hotel-Job-Las-Vegas-NV-123503423, 231-http://m.monster.com/Final-Expense-Golden-Opportunity-Job-Grand-Rapids-MI-123410019, 232-http://m.lowes.com/, 233-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=golden, 234-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=ray+lewis, 235-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=stone+mountain, 236-http://m.lowes.com/productlist?langId=-1&storeId=10702&catalogId=10051&nValue=productsearch&store=0595&view=list&pageSize=20&firstRecord=0&sort=ts&searchTerm=mcdonalds, 237-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=3819849&store=595&view=detail, 238-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=3645846&store=595&view=detail, 239-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=3319292&store=595&view=detail, 240-http://m.lowes.com/product?langId=-1&storeId=10702&catalogId=10051&productId=1238357&store=595&view=detail, 241-http://m.mapquest.com/, 242-http://m.photobucket.com/, 243-http://m.photobucket.com/images/ray+lewis, 244-http://m.photobucket.com/images/golden, 245-http://m.photobucket.com/images/stone+mountain, 246-http://m.photobucket.com/images/mcdonalds, 247-http://m1246.photobucket.com/image/ray\%20lewis/yourmyboyblue/raylewisebay_zpsc280184d.jpg.html?o=0, 248-http://m1292.photobucket.com/image/gold/Ronald_mark_Johnson/AMBER_GOLD_zpse6cd7fa6.jpg.html?o=0, 249-http://m1351.photobucket.com/image/stone/vhlsigs/stone_zps8e96cdf9.jpg.html?o=0, 250-http://m1163.photobucket.com/image/mcdonalds/cristinab42/thailand/IMG_6072.jpg.html?o=0, 251-http://m.tmz.com/, 252-http://m.tmz.com/2013/07/07/dwight-howard-game-lakers-rockets/, 253-http://m.tmz.com/2013/07/07/lady-gaga-rob-fusari-lawsuit-wendy-starland/, 254-http://m.tmz.com/2013/07/07/anna-nicole-smith-movie-lifetime-car-damage-200/, 255-http://m.tmz.com/2013/07/07/tameka-raymond-landlord-lawsuit/, 256-https://m.verizon.com/PAMMobile/presignin.aspx, 257-https://m.verizon.com/mforyourhome/services.aspx, 258-http://mobile.latimes.com/s.p?sId=7&m=b, 259-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858928&postId=3858928&postUserId=7&sessionToken=&catId=6907&curAbsIndex=0&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\

%3DsectionId\%253A6907\%26DPS\%3D0\%26DPL\%3D3, 260-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858875&postId=3858875&postUserId=7&sessionToken=&catId=6907&curAbsIndex=1&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\%3DsectionId\%253A6907\%26DPS\%3D0\%26DPL\%3D3, 261-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858853&postId=3858853&postUserId=7&sessionToken=&catId=5224&curAbsIndex=0&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\%3DsectionId\%253A5224\%26DPS\%3D0\%26DPL\%3D3, 262-http://mobile.latimes.com/p.p?m=b&a=rp&id=3858615&postId=3858615&postUserId=7&sessionToken=&catId=5217&curAbsIndex=1&resultsUrl=DID\%3D9\%26DFCL\%3D1000\%26DSB\%3Drank\%2523desc\%26DBFQ\%3DuserId\%253A7\%26DL.w\%3D\%26DL.d\%3D10\%26DQ\%3DsectionId\%253A5217\%26DPS\%3D0\%26DPL\%3D3, 263-http://m.sears.com/, 264-http://m.sears.com/keyword.do?keyword=golden&vertName=&vName=, 265-http://m.sears.com/keyword.do?keyword=ray+lewis&vertName=&vName=, 266-http://m.sears.com/keyword.do?keyword=stone+mountain&vertName=&vName=, 267-http://m.sears.com/keyword.do?keyword=mcdonalds&vertName=&vName=, 268-http://m.sears.com/productdetails.do?partNumber=05757212000P&reviewCount=zero&itemSrc=Online&threshold=59.0&fullFillment=TW, 269-http://m.sears.com/productdetails.do?partNumber=05729406000P, 270-http://m.sears.com/productdetails.do?partNumber=05238123000P&reviewCount=zero&itemSrc=Online&threshold=59.0&fullFillment=VD, 271-http://m.sears.com/productdetails.do?partNumber=00623806000P&reviewCount=zero&itemSrc=Online&threshold=0.0&fullFillment=VD, 272-http://www.m.webmd.com/, 273-http://www.m.webmd.com/mobile-search/default.htm?query=golden, 274-http://www.m.webmd.com/mobile-search/default.htm?query=ray+lewis, 275-http://www.m.webmd.com/mobile-search/default.htm?query=stone+mountain, 276-http://www.m.webmd.com/mobile-search/default.htm?query=mcdonalds, 277-http://www.m.webmd.com/diet/rm-quiz-best-worst-foods-belly-fat, 278-http://www.m.webmd.com/a-to-z-guides/rm-quiz-science-love, 279-http://www.expedia.com/MobileHotel?rfrr=-1065&, 280-http://m.macys.com/, 281-http://m.macys.com/shop/search?keyword=golden, 282-http://m.macys.com/shop/search?keyword=ray+lewis, 283-http://m.macys.com/shop/search?keyword=stone+mountain, 284-http://m.macys.com/shop/search?keyword=mcdonalds, 285-http://m.macys.com/shop/product/levis-jeans-569-loose-straight?ID=778920&CategoryID=11221#fn=sp\%3D1\%26spc\%3D2\%26kws\%3Dray\%20lewis\%26slotId\%3D1, 286-http://m.macys.com/shop/product/hello-kitty-kids-toy-girls-or-little-girls-coloring-book?ID=867470&CategoryID=5991#fn=sp\%3D1\%26spc\%3D15\%26kws\%3Dbook\%26slotId\%3D1, 287-http://m.macys.com/shop/product/elizabeth-arden-ceramide-capsules-daily-youth-restoring-serum-total-60-capsules-95-fl-oz?ID=253891&CategoryID=30078#fn=sp\%3D1\%26spc\%3D6574\%26kws\%3Dgold\%26slotId\%3D1, 288-http://m.macys.com/shop/product/kidz-delight-kids-toy-arthur-little-tv?ID=

717950&CategoryID=5991#fn=sp\%3D1\%26spc\%3D13\%26kws\%3Dtv\%26slotId\%3D1, 289-https://www.dropbox.com/m/login?cont=https\%3A//www.dropbox.com/m, 290-http://m.newegg.com/, 291-http://m.newegg.com/ProductList?Keyword=golden, 292-http://m.newegg.com/ProductList?Keyword=ray+lewis, 293-http://m.newegg.com/ProductList?Keyword=stone+mountain, 294-http://m.newegg.com/ProductList?Keyword=mcdonalds, 295-http://m.newegg.com/Product/index?itemNumber=N82E16834312242, 296-http://m.newegg.com/Product/index?itemNumber=N82E16834230987, 297-http://m.newegg.com/Product/index?itemNumber=N82E16823126097, 298-http://m.newegg.com/Product/index?itemNumber=N82E16832416552, 299-http://m.now.msn.com/, 300-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906510&acid=2&afid=0, 301-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906568&acid=2&afid=0, 302-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906467&acid=2&afid=0, 303-http://sports.mobile.msn.com/en-us/articles.aspx?aid=1906639&acid=2&afid=0, 304-http://mobile.backpage.com/, 305-http://mobile.backpage.com/Events/reptile-super-show-july-6-7-2013-san-diego-ca-concourse-civic-center-downtown/7978335, 306-http://mobile.backpage.com/Events/nighttalker-radio-show-michael-hastings-car-crash-assassination/7924589, 307-http://mobile.backpage.com/Events/hookup-party-this-weekend-saturday-july-13th/8050194, 308-http://mobile.backpage.com/Events/?keyword=golden, 309-http://mobile.backpage.com/Events/?keyword=ray+lewis, 310-http://mobile.backpage.com/Events/?keyword=stone+mountain, 311-http://mobile.backpage.com/Events/?keyword=mcdonalds, 312-https://mobile.southwest.com/p?stpp=true&formid=main, 313-http://m.baidu.com/, 314-http://m.baidu.com/ssid=0/from=0/bd_page_type=1/uid=51D9DEB13A6F4949623FC7CA2BC3E04D/baiduid=E6642719F9F904A9AF5B1ECBD542168B/s?word=golden, 315-http://m.baidu.com/ssid=0/from=0/bd_page_type=1/uid=51D9DEB13A6F4949623FC7CA2BC3E04D/baiduid=E6642719F9F904A9AF5B1ECBD542168B/s?word=ray+lewis, 316-http://m.baidu.com/ssid=0/from=0/bd_page_type=1/uid=51D9DEB13A6F4949623FC7CA2BC3E04D/baiduid=E6642719F9F904A9AF5B1ECBD542168B/s?word=mcdonalds, 317-http://m.intuit.com/, 318-http://m.ca.gov/, 319-http://m.realtor.com/, 320-http://m.realtor.com/#results?loc=golden&type=single_family\%2Ccondo\%2Cland, 321-http://m.realtor.com/#results?loc=ray+lewis&type=single_family\%2Ccondo\%2Cland, 322-http://m.realtor.com/#results?loc=stone+mountain&type=single_family\%2Ccondo\%2Cland, 323-http://m.realtor.com/#results?loc=mcdonalds&type=single_family\%2Ccondo\%2Cland, 324-http://m.cbsnews.com/, 325-http://m.cbsnews.com/searchstory.rbml?query=golden&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 326-http://m.cbsnews.com/searchstory.rbml?query=ray+lewis&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 327-http://m.cbsnews.com/searchstory.rbml?query=stone+mountain&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 328-http://m.cbsnews.com/searchstory.

rbml?query=mcdonalds&btnSearch.x=0&btnSearch.y=0&nbActionFormEncoding=UTF-8, 329-
http://m.cbsnews.com/storysynopsis.rbml?catid=57592558&feed_id=0&videofeed=36, 330-
http://m.cbsnews.com/storysynopsis.rbml?catid=57592575&feed_id=0&videofeed=36, 331-
http://m.cbsnews.com/storysynopsis.rbml?catid=57592567&feed_id=0&videofeed=36, 332-
http://m.cbsnews.com/storysynopsis.rbml?catid=57592553&feed_id=0&videofeed=36

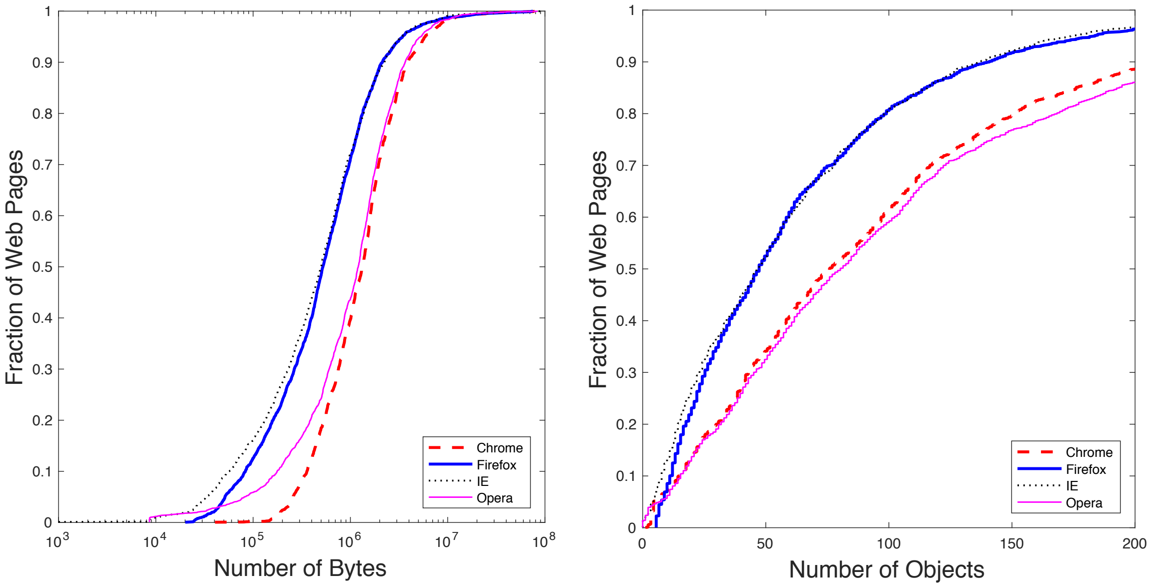
APPENDIX 6: LIST OF UPDATED BROWSER VERSIONS DURING COLLECTION

This appendix includes a list of the different browser updates that occurred for Firefox and Chrome during data collection. This list is provided below:

- Chrome/38.0.2125.122 - Updated on December 13, 2014
- Chrome/39.0.2171.95 - Updated on December 13, 2014
- Chrome/39.0.2171.99 - Updated on January 20, 2015
- Chrome/40.0.2214.91 - Updated on January 23, 2015
- Chrome/40.0.2214.93 - Updated on January 27, 2015
- Chrome/40.0.2214.94 - Updated on February 5, 2015
- Chrome/40.0.2214.111 - Updated on February 7, 2015
- Chrome/40.0.2214.115 - Updated on February 22, 2015
- Chrome/41.0.2272.76 - Updated on March 11, 2015
- Chrome/41.0.2272.89 - Updated on March 16, 2015
- Firefox/33.0 - Updated on December 13, 2014
- Firefox/34.0 - Updated on January 20, 2015
- Firefox/35.0 - Updated on February 7, 2015
- Firefox/36.0 - Updated on March 19, 2015

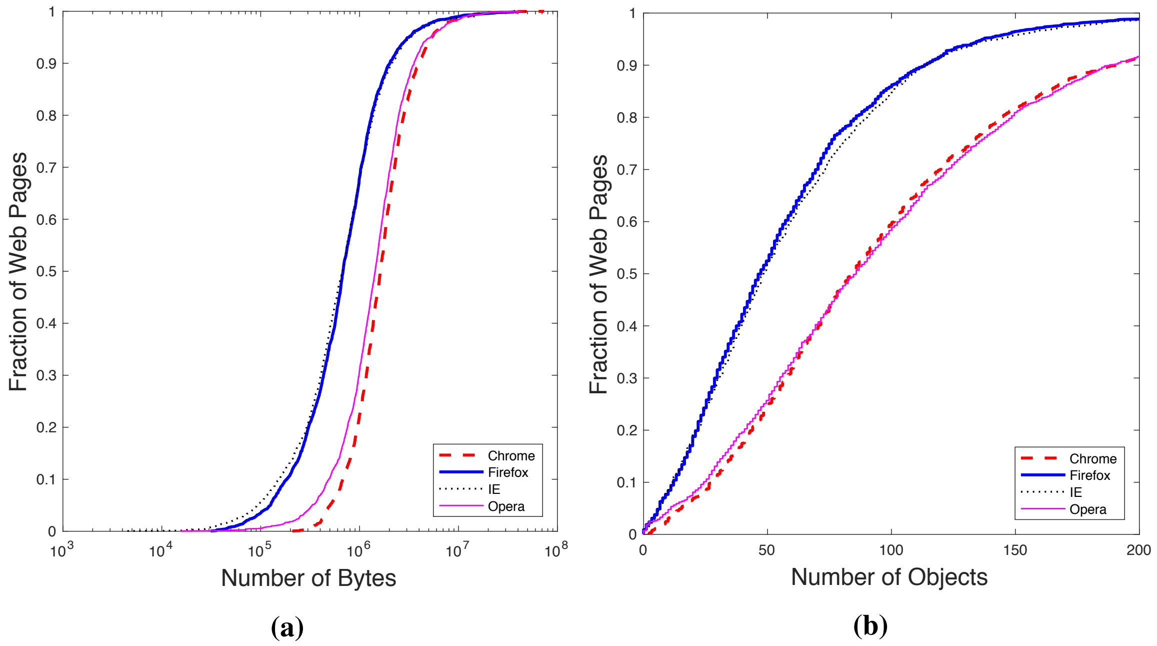
APPENDIX 7: ADDITIONAL CLIENT PLATFORM DIVERSITY PLOTS

This appendix includes plots which show that the most prominent differences across client platforms (browsers and operating systems) were repeatable in this dataset.



(a) **(b)**

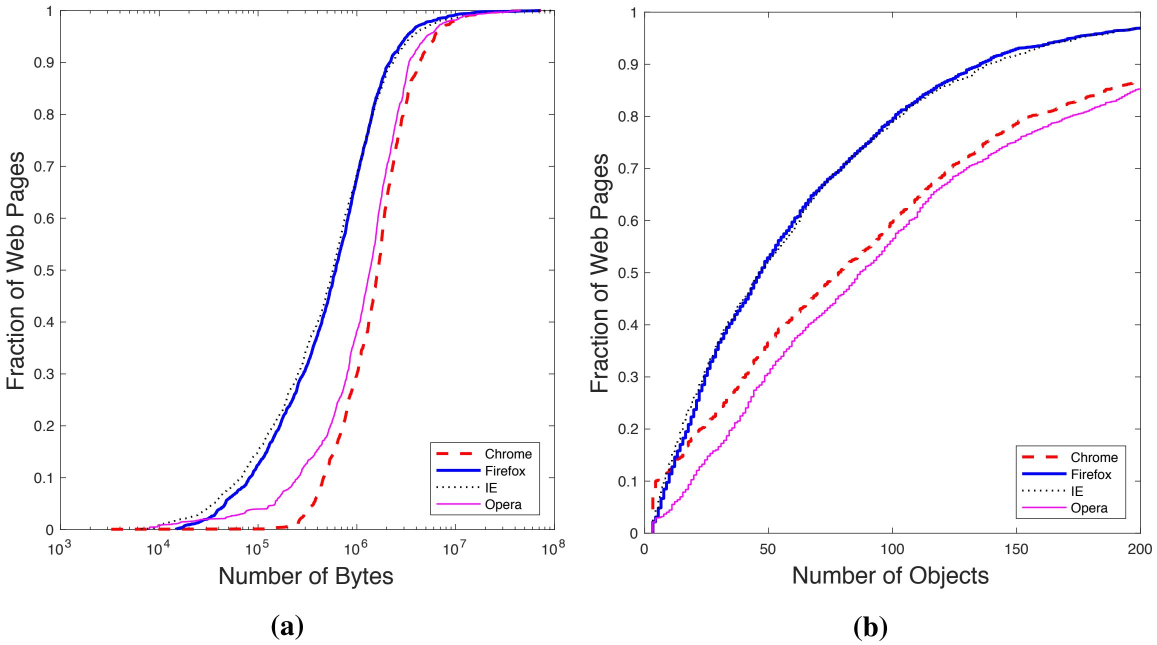
Figure 6.1: Differences in the number of bytes (a) and objects (b) observed across browser — Sample 2.



(a)

(b)

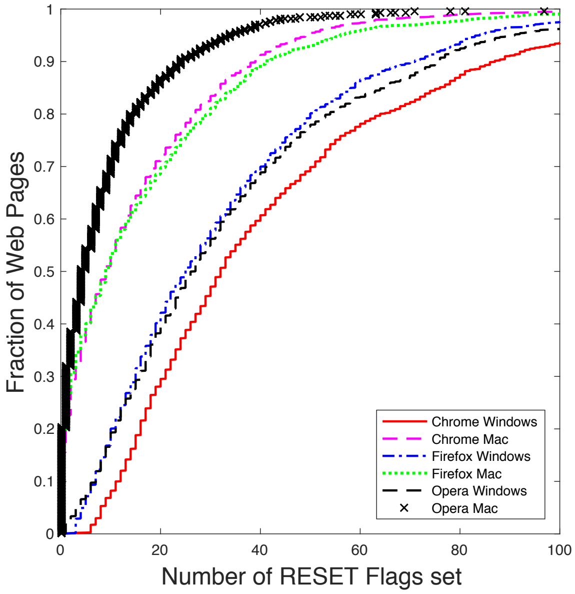
Figure 6.2: Differences in the number of bytes (a) and objects (b) observed across browser — Sample 3.



(a)

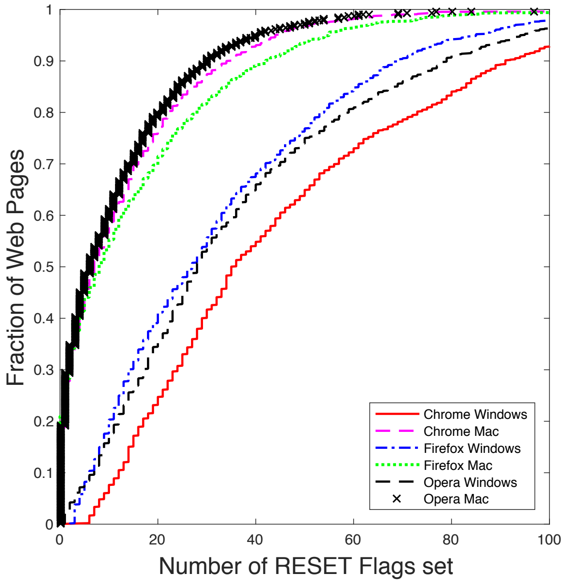
(b)

Figure 6.3: Differences in the number of bytes (a) and objects (b) observed across browser — Sample 4.



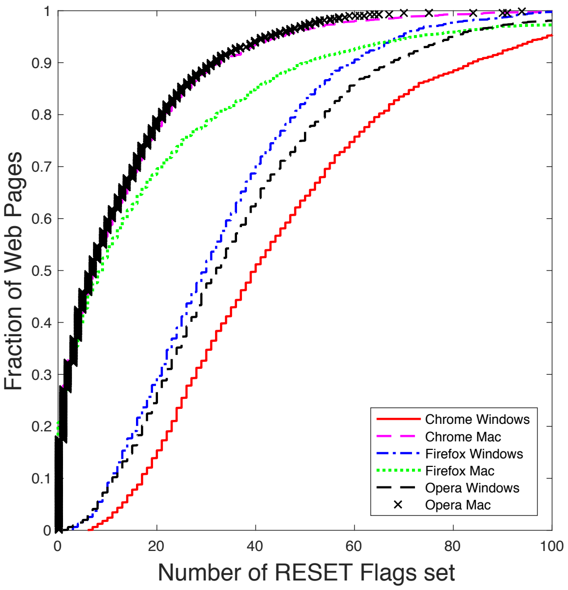
(a)

Figure 6.4: Differences in the number of RESET packets across operating system — Sample 2.



(a)

Figure 6.5: Differences in the number of RESET packets across operating system — Sample 3.



(a)

Figure 6.6: Differences in the number of RESET packets across operating system — Sample 4.

APPENDIX 8: TCP/IP TRAFFIC FEATURES USED FOR WEB PAGE CLASSIFICATION

This appendix includes a list of the TCP/IP-based features that are used for web page classification in Chapter 4.

1 - Client sends SYN segments, 2 - server sends SYN segments, 3 - Bidirectional SYN segments, 4 - Client sends RESET segments, 5 - server sends SYNACK segments, 6 - server sends RESET segments, 7 - Bidirectional SYNACK segments, 8 - Client sends PUSH segments, 9 - server sends PUSH segments, 10 - Bidirectional PUSH segments, 11 - Client sends segments, 12 - Client sends Bytes, 13 - server sends segments, 14 - server sends Bytes, 15 - Bidirectional segments, 16 - Bidirectional Bytes, 17 - number of IPPairs, 18 - number of TCP connections, 19 - Mean number of PUSH segments per TCP connection Client sends, 20 - Minimum PUSH per TCP connection client, 21 - 10 percentile PUSH per TCP connection client, 22 - 25 percentile PUSH per TCP connection client, 23 - 50 percentile PUSH per TCP connection client, 24 - 75 percentile PUSH per TCP connection client, 25 - 90 percentile PUSH per TCP connection client, 26 - Maximum percentile PUSH per TCP connection client, 27 - Mean number of PUSH segments per TCP connection server sends, 28 - Minimum PUSH per TCP connection server, 29 - 10 percentile PUSH per TCP connection server, 30 - 25 percentile PUSH per TCP connection server, 31 - 50 percentile PUSH per TCP connection server, 32 - 75 percentile PUSH per TCP connection server, 33 - 90 percentile PUSH per TCP connection server, 34 - Maximum percentile PUSH per TCP connection server, 35 - Mean number of bytes per TCP connection Client sends, 36 - Minimum bytes per TCP connection client, 37 - 10 percentile bytes per TCP connection client, 38 - 25 percentile bytes per TCP connection client, 39 - 50 percentile bytes per TCP connection client, 40 - 75 percentile bytes per TCP connection client, 41 - 90 percentile bytes per TCP connection client, 42 - Maximum percentile bytes per TCP connection client, 43 - number of unused TCP connections client client, 44 - Mean number of bytes per TCP connection server sends, 45 - Minimum bytes per TCP connection server, 46 - 10 percentile bytes per TCP connection server, 47 - 25 percentile bytes per TCP connection server, 48 - 50 percentile bytes per TCP connection server, 49 - 75 percentile bytes per TCP connection server, 50 - 90 percentile bytes per TCP connection server, 51 - Maximum percentile bytes per TCP connection server, 52 - number of unused TCP connections client server, 53 - number of unused TCP connections server, 54 - Mean number of bytes per IP Pair Client sends, 55 - Minimum bytes per IP Pair client, 56 - 10 percentile bytes per IP Pair client, 57 - 25 percentile bytes per IP Pair client, 58 - 50 percentile bytes per IP Pair client, 59 - 75 percentile bytes per IP Pair client, 60 - 90 percentile bytes per IP Pair

client, 61 - Maximum percentile bytes per IP Pair client, 62 - number of unused IP Pairs client , 63 - Mean number of bytes per IP Pair server sends, 64 - Minimum bytes per IP Pair client, 65 - 10 percentile bytes per IP Pair server, 66 - 25 percentile bytes per IP Pair server, 67 - 50 percentile bytes per IP Pair server, 68 - 75 percentile bytes per IP Pair server, 69 - 90 percentile bytes per IP Pair server, 70 - Maximum percentile bytes per IP Pair server, 71 - number of unused IP Pairs server, 72 - number of port 80 TCP connections, 73 - Mean bytes port 443 TCP connection, 74 - Minimum bytes per 443 TCP connection , 75 - 10 percentile 443 bytes per TCP Connection, 76- 25 percentile 443 bytes per TCP Connection, 77 - 50 percentile 443 bytes per TCP Connection, 78 - 75 percentile 443 bytes per TCP Connection, 79 - 90 percentile 443 bytes per TCP Connection, 80 - Maximum percentile 443 bytes per TCP Connection, 81 - Sum bytes port 443 TCP connection, 82 - Mean bytes port 80 TCP connection, 83 - Minimum bytes per 80 TCP connection server, 84 - 10 percentile bytes per 80 TCP connection server, 85 - 25 percentile bytes per 80 TCP connection server, 86 - 50 percentile bytes per 80 TCP connection server, 87 - 75 percentile bytes per 80 TCP connection server, 88 - 90 percentile bytes per 80 TCP connection server, 89 - Maximum percentile bytes per 80 TCP connection server, 90 - Sum bytes port 80 TCP connection, 91 - 75 percentile inter TCP connection arrival time, 92 - 90 percentile inter TCP connection arrival time, 93 - Maximum percentile inter TCP connection arrival time, 94 - Mean number of epochs per TCP connection, 95 - Minimum number of epochs per TCP connection, 96 - 10 percentile number of epochs per TCP connection, 97 - 25 percentile number of epochs per TCP connection, 98 - 50 percentile number of epochs per TCP connection, 99 - 75 percentile number of epochs per TCP connection, 100 - 90 percentile number of epochs per TCP connection, 101 - Maximum percentile number of epochs per TCP connection, 102 - Minimum bytes sent client per epoch, 103 - number of epochs, 104 - Mean number of bytes sent by client per epoch, 105 - Minimum bytes sent client per epoch, 106 - 10 percentile number of bytes sent by client per epoch, 107 - 25 percentile number of bytes sent by client per epoch, 108 - 50 percentile number of bytes sent by client per epoch, 109 - 75 percentile number of bytes sent by client per epoch, 110 - 90 percentile number of bytes sent by client per epoch, 111 - Maximum percentile number of bytes sent by client per epoch, 112 - Sum number of bytes sent by client epoch, 113 - Minimum bytes sent client per epoch, 114 - Mean number of bytes sent by server per epoch, 115 - Minimum bytes sent server per epoch, 116 - 10 percentile number of bytes sent by server per epoch, 117 - 25 percentile number of bytes sent by server per epoch, 118 - 50 percentile number of bytes sent by server per epoch, 119 - 75 percentile number of bytes sent by server per epoch, 120 - 90 percentile number of bytes sent by server per epoch, 121 - Maximum percentile number of bytes sent by server per epoch, 122 - Sum number of bytes sent by server epoch, 123 - The number of bytes in the first segment in the first TCP connection, 124 - The number of bytes in the second segment in the first TCP connection, 125 - The number of bytes in the third segment in the first TCP connection, 126 - The number of bytes in the fourth segment in the first TCP connection, 127 - The number of bytes in the fifth segment in the first TCP connection, 128 - The number of bytes in the sixth segment in the first TCP connection, 129 - The number of bytes in the seventh segment in the first TCP, 130 - The number of bytes in the eighth segment in the first TCP connection connection, 131 - The number of bytes in the ninth segment in the first TCP connection, 132 - The number of bytes in the tenth segment in the first TCP connection, 133 - The number of bytes in the first segment in the second TCP connection, 134 - The number of bytes in the second segment in the second TCP connection, 135 - The number of bytes in the third segment in the second TCP connection, 136 - The number of bytes in the fourth segment in the second TCP connection, 137 - The number of bytes in the fifth segment in the second TCP connection, 138 - The number of bytes in the sixth segment in the second TCP connection, 139 - The number of bytes in the seventh segment in the second TCP, 140 - The number of bytes in the eighth segment in the second TCP connection connection, 141 - The number of bytes in the ninth segment in the second TCP connection, 142 - The number of bytes in the tenth segment in the second TCP connection, 143 - The number of bytes in the first segment in the third TCP connection, 144 - The number of bytes in the second segment in the third TCP connection, 145 - The number of bytes in the third segment in the third TCP connection, 146 - The number of bytes in the fourth segment in the third TCP connection, 147 - The number of bytes in the fifth segment in the third TCP connection, 148 - The number of bytes in the sixth segment in the third TCP connection, 149 - The number of bytes in the seventh segment in the third TCP, 150 - The number of bytes in the eighth segment in the third TCP connection connection, 151 - The number of bytes in the ninth segment in the third TCP connection, 152 - The number of bytes in the tenth segment in the third TCP connection, 153 - The number of bytes in the first segment in the fourth TCP connection, 154 - The number of bytes in the second segment in the fourth TCP connection, 155 - The number of bytes in the third segment in the fourth TCP connection, 156 - The number of bytes in the fourth segment in the fourth TCP connection, 157 - The number of bytes in the fifth segment in the fourth TCP connection, 158 - The number of bytes in the sixth segment in the fourth TCP connection, 159 - The number of bytes in the seventh segment in the fourth TCP, 160 - The number of bytes in the eighth segment in the fourth TCP connection connection, 161 - The number of bytes in the ninth segment in the fourth TCP connection, 162 - The number of bytes in the tenth segment in the fourth TCP connection

APPENDIX 9: LIST OF INFORMATIVE TCP/IP FEATURES

This appendix includes a list of the informative TCP/IP features used in Chapter 4. Each feature has the labeling schemes in which they were identified as being informative.

Group 1:

Client Sends Push Packets - video, mobile, genre, content Number of IPPairs - video, mobile, genre, content Bidirectional Push Packets - video, mobile, genre, content Server Sends Push Packets - video, mobile, genre, content Number of TCP connections - video, mobile, genre, content Number of Port 80 TCP connections - video, mobile, genre, content Client Sends Syn Packets - video, mobile, genre, content Bidirectional Syn Packets - video, mobile, genre, content Server Sends Syn Packets - video, mobile, genre, content Server Sends Synack Packets - video, mobile, genre, content Bidirectional Synack Packets - video, mobile, genre, content Number of epochs/objects - video, mobile, genre, content

Group 2:

90 percentile Number of epochs per TCP connection - mobile, genre, content 90 percentile Number of Bytes sent by client per epoch - genre,mobile,content,video 75 percentile push per TCP Connection server - content,mobile,genre 75 percentile bytes per TCP Connection client - mobile, content,genre 75 percentile push per TCP Connection client - genre,content,mobile 90 percentile push per TCP Connection client - genre, mobile, content 90 percentile push per TCP Connection server - genre, content, mobile Max percentile Number of epochs per TCP connection - genre, mobile, video,content 90 percentile bytes per TCP Connection client - genre, mobile, content Max percentile push per TCP Connection client - genre, mobile, content, video Max percentile Number of Bytes sent by client per epoch -video,mobile,content,genre Max percentile push per TCP Connection server - genre, content, mobile 90 percentile bytes per IPPair client - genre,mobile,content,video Max percentile bytes per TCP Connection client - mobile,content,video Max percentile bytes per IPPair client - genre,mobile,content,video Max percentile bytes per TCP Connection server - mobile,content,video Max percentile bytes per IPPair server - genre,mobile,content,video 90 percentile bytes per TCP Connection server - genre,mobile,content,video 90 percentile bytes per IPPair server - genre,mobile,content,video 50 percentile Number of Bytes sent by server per epoch - video, mobile

Group 3: **Min bytes per TCP connection client - mobile** Min bytes per TCP connection server - mobile Min bytes per IPPair client - mobile Min bytes per IPPair client - mobile

Group 4: **Average Number of push packets per TCP connection Client Sends - genre, content** 50 percentile bytes per TCP Connection client - genre, content Average Number of push packets per TCP connection Server Sends - genre Average Number of bytes per TCP connection Client Sends - content, genre 50 percentile push per TCP Connection server - content, genre 50 percentile Number of epochs per TCP connection - content, genre 50 percentile push per TCP Connection client - content, genre Average Number of epochs per TCP connection - content, genre Average Number of bytes per TCP connection Server Sends - genre, mobile, content 50 percentile bytes per TCP Connection server - genre, content

Group 5 **Bidirectional Packets - video, mobile, genre, content** Server Sends Packets - video, mobile, genre, content Bidirectional Bytes - video, mobile, genre, content Server Sends Bytes - video, mobile, genre, content

Group 6 **Maximum percentile bytes per epoch by server- video, mobile** 90 percentile bytes per epoch by server- video, mobile 75 percentile bytes per epoch by server - video Average bytes per epoch by server - video

Group 7 **90 percentile Number of Bytes sent by client per epoch - genre, content, mobile** 75 percentile Number of Bytes sent by client per epoch - genre, content, mobile Max percentile Number of Bytes sent by client per epoch - genre, content, mobile

APPENDIX 10: CLASSIFICATION PERFORMANCE FOR KNN FOR DIFFERENT VALUES OF K

This appendix includes the classification performance for the KNN methods tested in Chapter 4 for values of K ranging from 3 to 10.

- Targeted Device Based Labeling
  - KNN: Euclidean (K = 3), Micro F Score = 7.801724e-01, Macro F Score = 7.151038e-01
  - KNN: City Block (K = 3), Micro F Score = 8.006897e-01, Macro F Score = 7.432232e-01
  - KNN: Cosine (K = 3), Micro F Score = 7.743103e-01, Macro F Score = 7.037417e-01
  - KNN: Correlation (K = 3), Micro F Score = 7.760345e-01, Macro F Score = 7.079799e-01
  - KNN: Euclidean (K = 4), Micro F Score = 8.353448e-01, Macro F Score = 7.834862e-01
  - KNN: City Block (K = 4), Micro F Score = 8.444828e-01, Macro F Score = 7.975839e-01
  - KNN: Cosine (K = 4), Micro F Score = 8.320690e-01, Macro F Score = 7.777322e-01
  - KNN: Correlation (K = 4), Micro F Score = 8.310345e-01, Macro F Score = 7.768153e-01
  - KNN: Euclidean (K = 5), Micro F Score = 8.293103e-01, Macro F Score = 7.734458e-01
  - KNN: City Block (K = 5), Micro F Score = 8.348276e-01, Macro F Score = 7.827649e-01
  - KNN: Cosine (K = 5), Micro F Score = 8.229310e-01, Macro F Score = 7.632376e-01
  - KNN: Correlation (K = 5), Micro F Score = 8.205172e-01, Macro F Score = 7.602077e-01
  - KNN: Euclidean (K = 6), Micro F Score = 8.327586e-01, Macro F Score = 7.776919e-01
  - KNN: City Block (K = 6), Micro F Score = 8.406897e-01, Macro F Score = 7.899215e-01
  - KNN: Cosine (K = 6), Micro F Score = 8.300000e-01, Macro F Score = 7.729737e-01
  - KNN: Correlation (K = 6), Micro F Score = 8.275862e-01, Macro F Score = 7.700085e-01
  - KNN: Euclidean (K = 7), Micro F Score = 7.717241e-01, Macro F Score = 6.898190e-01
  - KNN: City Block (K = 7), Micro F Score = 7.896552e-01, Macro F Score = 7.187916e-01
  - KNN: Cosine (K = 7), Micro F Score = 7.646552e-01, Macro F Score = 6.796516e-01

- KNN: Correlation (K = 7), Micro F Score = 7.644828e-01, Macro F Score = 6.802773e-01

- KNN: Euclidean (K = 8), Micro F Score = 8.081034e-01, Macro F Score = 7.395466e-01
- KNN: City Block (K = 8), Micro F Score = 8.248276e-01, Macro F Score = 7.649577e-01
- KNN: Cosine (K = 8), Micro F Score = 8.050000e-01, Macro F Score = 7.341270e-01
- KNN: Correlation (K = 8), Micro F Score = 8.051724e-01, Macro F Score = 7.348018e-01

- KNN: Euclidean (K = 9), Micro F Score = 8.013793e-01, Macro F Score = 7.286198e-01
- KNN: City Block (K = 9), Micro F Score = 8.179310e-01, Macro F Score = 7.544366e-01
- KNN: Cosine (K = 9), Micro F Score = 7.932759e-01, Macro F Score = 7.160006e-01
- KNN: Correlation (K = 9), Micro F Score = 7.941379e-01, Macro F Score = 7.174323e-01

- KNN: Euclidean (K = 10), Micro F Score = 8.091379e-01, Macro F Score = 7.393919e-01
- KNN: City Block (K = 10), Micro F Score = 8.270690e-01, Macro F Score = 7.672785e-01
- KNN: Cosine (K = 10), Micro F Score = 8.048276e-01, Macro F Score = 7.323916e-01
- KNN: Correlation (K = 10), Micro F Score = 8.034483e-01, Macro F Score = 7.309758e-01

- Video Streaming Based Labeling

  - KNN: Euclidean (K = 3), Micro F Score = 9.941379e-01, Macro F Score = 9.826433e-01
  - KNN: City Block (K = 3), Micro F Score = 9.955172e-01, Macro F Score = 9.866700e-01
  - KNN: Cosine (K = 3), Micro F Score = 9.936207e-01, Macro F Score = 9.810669e-01
  - KNN: Correlation (K = 3), Micro F Score = 9.932759e-01, Macro F Score = 9.800142e-01

  - KNN: Euclidean (K = 4), Micro F Score = 9.948276e-01, Macro F Score = 9.847117e-01
  - KNN: City Block (K = 4), Micro F Score = 9.958621e-01, Macro F Score = 9.877115e-01
  - KNN: Cosine (K = 4), Micro F Score = 9.946552e-01, Macro F Score = 9.841619e-01
  - KNN: Correlation (K = 4), Micro F Score = 9.943103e-01, Macro F Score = 9.831400e-01

  - KNN: Euclidean (K = 5), Micro F Score = 9.941379e-01, Macro F Score = 9.826154e-01

  - KNN: City Block (K = 5), Micro F Score = 9.943103e-01, Macro F Score = 9.830298e-01
  - KNN: Cosine (K = 5), Micro F Score = 9.936207e-01, Macro F Score = 9.810394e-01
  - KNN: Correlation (K = 5), Micro F Score = 9.937931e-01, Macro F Score = 9.815395e-01
  - KNN: Euclidean (K = 6), Micro F Score = 9.950000e-01, Macro F Score = 9.851404e-01
  - KNN: City Block (K = 6), Micro F Score = 9.950000e-01, Macro F Score = 9.851052e-01
  - KNN: Cosine (K = 6), Micro F Score = 9.944828e-01, Macro F Score = 9.836139e-01
  - KNN: Correlation (K = 6), Micro F Score = 9.944828e-01, Macro F Score = 9.835917e-01
  - KNN: Euclidean (K = 7), Micro F Score = 9.932759e-01, Macro F Score = 9.799172e-01
  - KNN: City Block (K = 7), Micro F Score = 9.934483e-01, Macro F Score = 9.803915e-01
  - KNN: Cosine (K = 7), Micro F Score = 9.925862e-01, Macro F Score = 9.779330e-01
  - KNN: Correlation (K = 7), Micro F Score = 9.927586e-01, Macro F Score = 9.784054e-01
  - KNN: Euclidean (K = 8), Micro F Score = 9.939655e-01, Macro F Score = 9.819816e-01
  - KNN: City Block (K = 8), Micro F Score = 9.941379e-01, Macro F Score = 9.824647e-01
  - KNN: Cosine (K = 8), Micro F Score = 9.929310e-01, Macro F Score = 9.789600e-01
  - KNN: Correlation (K = 8), Micro F Score = 9.927586e-01, Macro F Score = 9.784323e-01
  - KNN: Euclidean (K = 9), Micro F Score = 9.936207e-01, Macro F Score = 9.809494e-01
  - KNN: City Block (K = 9), Micro F Score = 9.934483e-01, Macro F Score = 9.803588e-01
  - KNN: Cosine (K = 9), Micro F Score = 9.932759e-01, Macro F Score = 9.799869e-01
  - KNN: Correlation (K = 9), Micro F Score = 9.927586e-01, Macro F Score = 9.784323e-01
  - KNN: Euclidean (K = 10), Micro F Score = 9.939655e-01, Macro F Score = 9.819816e-01
  - KNN: City Block (K = 10), Micro F Score = 9.934483e-01, Macro F Score = 9.803588e-01
  - KNN: Cosine (K = 10), Micro F Score = 9.934483e-01, Macro F Score = 9.805134e-01
  - KNN: Correlation (K = 10), Micro F Score = 9.927586e-01, Macro F Score = 9.784323e-01

- Alexa Genre Based Labeling

- KNN: Euclidean (K = 3), Micro F Score = 4.803448e-01, Macro F Score = 3.711311e-01
- KNN: City Block (K = 3), Micro F Score = 5.168966e-01, Macro F Score = 4.016612e-01
- KNN: Cosine (K = 3), Micro F Score = 4.706897e-01, Macro F Score = 3.586735e-01
- KNN: Correlation (K = 3), Micro F Score = 4.700000e-01, Macro F Score = 3.612125e-01

- KNN: Euclidean (K = 4), Micro F Score = 5.036207e-01, Macro F Score = 3.821891e-01
- KNN: City Block (K = 4), Micro F Score = 5.589655e-01, Macro F Score = 4.375162e-01
- KNN: Cosine (K = 4), Micro F Score = 5.013793e-01, Macro F Score = 3.718116e-01
- KNN: Correlation (K = 4), Micro F Score = 4.979310e-01, Macro F Score = 3.760411e-01

- KNN: Euclidean (K = 5), Micro F Score = 5.003448e-01, Macro F Score = 3.793015e-01
- KNN: City Block (K = 5), Micro F Score = 5.500000e-01, Macro F Score = 4.297137e-01
- KNN: Cosine (K = 5), Micro F Score = 4.948276e-01, Macro F Score = 3.656709e-01
- KNN: Correlation (K = 5), Micro F Score = 4.944828e-01, Macro F Score = 3.623283e-01

- KNN: Euclidean (K = 6), Micro F Score = 5.136207e-01, Macro F Score = 3.813165e-01
- KNN: City Block (K = 6), Micro F Score = 5.662069e-01, Macro F Score = 4.506725e-01
- KNN: Cosine (K = 6), Micro F Score = 5.108621e-01, Macro F Score = 3.886743e-01
- KNN: Correlation (K = 6), Micro F Score = 5.148276e-01, Macro F Score = 3.75584e-01

- KNN: Euclidean (K = 7), Micro F Score = 4.784483e-01, Macro F Score = 3.821933e-01
- KNN: City Block (K = 7), Micro F Score = 5.227586e-01, Macro F Score = 4.130418e-01
- KNN: Cosine (K = 7), Micro F Score = 4.710345e-01, Macro F Score = 3.891266e-01
- KNN: Correlation (K = 7), Micro F Score = 4.691379e-01, Macro F Score = 3.865641e-01

- KNN: Euclidean (K = 8), Micro F Score = 4.913793e-01, Macro F Score = 3.7823229e-01
- KNN: City Block (K = 8), Micro F Score = 5.479310e-01, Macro F Score = 4.436172e-01
- KNN: Cosine (K = 8), Micro F Score = 4.863793e-01, Macro F Score = 3.877852e-01
- KNN: Correlation (K = 8), Micro F Score = 4.853448e-01, Macro F Score = 3.864499e-01

- KNN: Euclidean (K = 9), Micro F Score = 4.586207e-01, Macro F Score = 3.801921e-01
- KNN: City Block (K = 9), Micro F Score = 5.077586e-01, Macro F Score = 3.800443e-01
- KNN: Cosine (K = 9), Micro F Score = 4.532759e-01, Macro F Score = 3.880305e-01
- KNN: Correlation (K = 9), Micro F Score = 4.534483e-01, Macro F Score = 3.865609e-01

- KNN: Euclidean (K = 10), Micro F Score = 4.706897e-01, Macro F Score = 3.819095e-01
- KNN: City Block (K = 10), Micro F Score = 5.222414e-01, Macro F Score = 3.801228e-01
- KNN: Cosine (K = 10), Micro F Score = 4.648276e-01, Macro F Score = 3.876880e-01
- KNN: Correlation (K = 10), Micro F Score = 4.639655e-01, Macro F Score = 3.859975e-01

- Navigation Based Labeling

  - KNN: Euclidean (K = 3), Micro F Score = 6.234483e-01, Macro F Score = 6.297429e-01
  - KNN: City Block (K = 3), Micro F Score = 6.508621e-01, Macro F Score = 6.569172e-01
  - KNN: Cosine (K = 3), Micro F Score = 6.122414e-01, Macro F Score = 6.174243e-01
  - KNN: Correlation (K = 3), Micro F Score = 6.141379e-01, Macro F Score = 6.192414e-01

  - KNN: Euclidean (K = 4), Micro F Score = 6.706897e-01, Macro F Score = 6.671849e-01
  - KNN: City Block (K = 4), Micro F Score = 6.948276e-01, Macro F Score = 6.941832e-01
  - KNN: Cosine (K = 4), Micro F Score = 6.648276e-01, Macro F Score = 6.615954e-01
  - KNN: Correlation (K = 4), Micro F Score = 6.629310e-01, Macro F Score = 6.598777e-01

  - KNN: Euclidean (K = 5), Micro F Score = 6.631034e-01, Macro F Score = 6.622070e-01
  - KNN: City Block (K = 5), Micro F Score = 6.887931e-01, Macro F Score = 6.897249e-01
  - KNN: Cosine (K = 5), Micro F Score = 6.615517e-01, Macro F Score = 6.600786e-01
  - KNN: Correlation (K = 5), Micro F Score = 6.637931e-01, Macro F Score = 6.623127e-01

  - KNN: Euclidean (K = 6), Micro F Score = 6.789655e-01, Macro F Score = 6.772107e-01
  - KNN: City Block (K = 6), Micro F Score = 6.929310e-01, Macro F Score = 6.917707e-01
  - KNN: Cosine (K = 6), Micro F Score = 6.736207e-01, Macro F Score = 6.725245e-01

- KNN: Correlation (K = 6), Micro F Score = 6.682759e-01, Macro F Score = 6.668703e-01

- KNN: Euclidean (K = 7), Micro F Score = 5.924138e-01, Macro F Score = 5.895979e-01
- KNN: City Block (K = 7), Micro F Score = 6.120690e-01, Macro F Score = 6.081146e-01
- KNN: Cosine (K = 7), Micro F Score = 5.872414e-01, Macro F Score = 5.831213e-01
- KNN: Correlation (K = 7), Micro F Score = 5.812069e-01, Macro F Score = 5.761234e-01

- KNN: Euclidean (K = 8), Micro F Score = 6.301724e-01, Macro F Score = 6.246410e-01
- KNN: City Block (K = 8), Micro F Score = 6.506897e-01, Macro F Score = 6.457086e-01
- KNN: Cosine (K = 8), Micro F Score = 6.265517e-01, Macro F Score = 6.206297e-01
- KNN: Correlation (K = 8), Micro F Score = 6.232759e-01, Macro F Score = 6.167180e-01

- KNN: Euclidean (K = 9), Micro F Score = 5.910345e-01, Macro F Score = 5.801763e-01
- KNN: City Block (K = 9), Micro F Score = 6.246552e-01, Macro F Score = 6.173230e-01
- KNN: Cosine (K = 9), Micro F Score = 5.855172e-01, Macro F Score = 5.740050e-01
- KNN: Correlation (K = 9), Micro F Score = 5.832759e-01, Macro F Score = 5.714570e-01

- KNN: Euclidean (K = 10), Micro F Score = 6.134483e-01, Macro F Score = 6.047412e-01
- KNN: City Block (K = 10), Micro F Score = 6.406897e-01, Macro F Score = 6.335305e-01
- KNN: Cosine (K = 10), Micro F Score = 6.058621e-01, Macro F Score = 5.950387e-01
- KNN: Correlation (K = 10), Micro F Score = 6.022414e-01, Macro F Score = 5.913638e-01

APPENDIX 11: LIST OF P-VALUES FOR DISTRIBUTION COMPARISON BETWEEN CLASSIFICATION METHODS

This appendix includes a list of the resulting p-values when performing a Kolmogorov-Smirnov test for features that are useful for simulation modeling. The p-values presented below are the result of conducting statistical tests between the ground truth distribution of these features with the classification procedures.

The results shown below are organized according to: Traffic Feature, Classification Method, Classification Label, P-value

- Number of TCP Connections
  - KNN City Block Distance (K = 1)
    - Mobile Optimized, 7.969499e-01
    - Traditional, 1.000000e+00
  - Classification Trees - Deviance
    - Mobile Optimized, 9.952932e-01
    - Traditional, 1.000000e+00
  - Naive Bayes - Multinomial
    - Mobile Optimized, 8.595523e-48
    - Traditional, 3.727902e-12
  - LDA - Linear
    - Mobile Optimized, 1.728203e-09
    - Traditional, 1.846144e-49
  - Random Guessing
    - Mobile Optimized, 6.127685e-51
    - Traditional, 1.335308e-11
  - KNN City Block Distance (K = 1)
    - Video, 1.000

  - Nonvideo, 1.000000e+00

- Classification Trees - Deviance
  - Video, 1.000
  - Nonvideo, 1.000000e+00

- Naive Bayes - Multinomial
  - Video, 1.839656e-01
  - Nonvideo, 6.564360e-02

- LDA - Linear
  - Video, 8.146736e-12
  - Nonvideo, 2.954539e-01

- Random Guessing
  - Video, 5.642288e-02
  - Nonvideo, 5.176586e-36

- KNN City Block Distance (K = 1)
  - Business, 9.828392e-01
  - Adult, 9.640330e-01
  - Arts, 9.994220e-01
  - Computers, 8.773017e-01
  - Games, 1.000000e+00
  - Health, 9.776655e-01
  - Home, 8.737124e-01
  - Kids and Teens, 3.087469e-01
  - News, 6.254663e-01
  - Recreation, 9.487841e-01
  - Reference, 9.901704e-01

  - Regional, 9.799545e-01
  - Science, 3.827947e-01
  - Shopping, 9.997474e-01
  - Society, 9.841684e-01
  - Sports, 9.903195e-01
  - World, 9.437779e-01

- Classification Trees - Deviance
  - Business, 7.205575e-01
  - Adult, 8.350765e-01
  - Arts, 6.039600e-01
  - Computers, 8.525914e-01
  - Games, 9.311919e-01
  - Health, 4.063501e-01
  - Home, 9.871593e-01
  - Kids and Teens, 2.663928e-01
  - News, 7.400380e-01
  - Recreation, 3.456606e-01
  - Reference, 8.384995e-01
  - Regional, 3.694098e-01
  - Science, 2.094731e-01
  - Shopping, 7.081727e-01
  - Society, 8.574617e-01
  - Sports, 7.475772e-01
  - World, 6.671081e-01

- Naive Bayes - Multinomial
  - Business, 1.798867e-02
  - Adult, 6.195595e-02

    - Arts, 1.074055e-09
    - Computers, 2.203120e-06
    - Games, 4.551635e-01
    - Health, 7.009434e-04
    - Home, 3.061002e-01
    - Kids and Teens, 1.541953e-02
    - News, 3.210848e-06
    - Recreation, 2.117827e-01
    - Reference, 7.943265e-07
    - Regional, 5.751698e-01
    - Science, 8.384923e-05
    - Shopping, 4.161680e-04
    - Society, 2.626440e-02
    - Sports, 5.763426e-01
    - World, 7.399185e-01
  - LDA - Linear
    - Business, 4.560472e-01
    - Adult, 1.949473e-06
    - Arts, 1.676528e-10
    - Computers, 3.210711e-12
    - Games, 1.082525e-01
    - Health, 3.345310e-02
    - Home, 8.221657e-01
    - Kids and Teens, 3.257351e-02
    - News, 3.448062e-06
    - Recreation, 1.610210e-01
    - Reference, 8.303009e-02

    - Regional, 1.562155e-02
    - Science, 6.817367e-01
    - Shopping, 1.868109e-10
    - Society, 1.936138e-01
    - Sports, 6.635939e-03
    - World, 8.308815e-02

- Random Guessing
    - Business, 3.923436e-07
    - Adult, 1.426969e-05
    - Arts, 3.590498e-06
    - Computers, 6.640714e-02
    - Games, 4.185450e-03
    - Health, 2.696946e-01
    - Home, 2.694666e-05
    - Kids and Teens, 9.499657e-08
    - News, 3.252373e-04
    - Recreation, 1.887119e-01
    - Reference, 7.046161e-01
    - Regional, 8.376703e-01
    - Science, 6.782937e-01
    - Shopping, 8.877651e-02
    - Society, 4.834888e-01
    - Sports, 1.232587e-04
    - World, 2.593929e-03

- KNN City Block Distance (K = 1)
    - Homepage, 9.827662e-01
    - Clickable Content, 4.800181e-01

        - Search Results, 6.814027e-01

    - Classification Trees - Deviance
        - Homepage, 9.994633e-01
        - Clickable Content, 9.553413e-01
        - Search Results, 9.187428e-01
    - Naive Bayes - Multinomial
        - Homepage, 1.427481e-01
        - Clickable Content, 7.776364e-11
        - Search Results, 1.448096e-05
    - LDA - Linear
        - Homepage, 8.312447e-09
        - Clickable Content, 1.066514e-28
        - Search Results, 4.557833e-18
    - Random Guessing
        - Homepage, 1.911268e-03
        - Clickable Content, 3.991731e-06
        - Search Results, 1.106638e-02

- Number of Bidirectional Bytes
    - KNN City Block Distance (K = 1)
        - Mobile Optimized, 4.774819e-01
        - Traditional, 9.999932e-01
    - Classification Trees - Deviance
        - Mobile Optimized, 8.552577e-01
        - Traditional, 9.996620e-01

- Naive Bayes - Multinomial
    - Mobile Optimized, 3.199858e-75
    - Traditional, 1.585956e-166
- LDA - Linear
    - Mobile Optimized, 1.506361e-19
    - Traditional, 4.646497e-89
- Random Guessing
    - Mobile Optimized, 7.094954e-92
    - Traditional, 9.634609e-24
- KNN City Block Distance (K = 1)
    - Video, 1
    - Nonvideo, 1.000000e+00
- Classification Trees - Deviance
    - Video, 1
    - Nonvideo, 1.000000e+00
- Naive Bayes - Multinomial
    - Video, 1.247966e-03
    - Nonvideo, 2.645609e-22
- LDA - Linear
    - Video, 4.207618e-04
    - Nonvideo, 7.021330e-01
- Random Guessing
    - Video, 1.281741e-09
    - Nonvideo, 1.605416e-207

- KNN City Block Distance (K = 1)
    - Business, 9.940872e-01
    - Adult, 9.772871e-01
    - Arts, 9.999134e-01
    - Computers, 9.999562e-01
    - Games, 9.982084e-01
    - Health, 9.999992e-01
    - Home, 6.048471e-01
    - Kids and Teens, 7.808956e-01
    - News, 5.558294e-01
    - Recreation, 9.921193e-01
    - Reference, 9.980502e-01
    - Regional, 1.000000e+00
    - Science, 8.372938e-01
    - Shopping, 9.821134e-01
    - Society, 1.000000e+00
    - Sports, 9.998535e-01
    - World, 9.999998e-01
- Classification Trees - Deviance
    - Business, 2.916434e-01
    - Adult, 9.965339e-01
    - Arts, 9.941741e-01
    - Computers, 9.963833e-01
    - Games, 1.915453e-01
    - Health, 9.749530e-01
    - Home, 9.945730e-01
    - Kids and Teens, 8.040408e-01

    - News, 9.587885e-01
    - Recreation, 9.949878e-01
    - Reference, 4.182255e-01
    - Regional, 8.047613e-01
    - Science, 1.284729e-02
    - Shopping, 5.542601e-01
    - Society, 6.758099e-01
    - Sports, 6.778845e-01
    - World, 1.067033e-01

  - Naive Bayes - Multinomial
    - Business, 1.672599e-01
    - Adult, 2.541755e-04
    - Arts, 5.219294e-40
    - Computers, 2.843006e-03
    - Games, 2.616177e-01
    - Health, 0
    - Home, 6.231220e-01
    - Kids and Teens, 1.541953e-02
    - News, 4.391006e-03
    - Recreation, 8.943972e-01
    - Reference, 5.290266e-10
    - Regional, 4.547292e-02
    - Science, 3.274436e-09
    - Shopping, 4.123988e-19
    - Society, 5.904710e-03
    - Sports, 1.392094e-02
    - World, 1.501536e-01

- LDA - Linear
    - Business, 2.573437e-01
    - Adult, 4.490945e-04
    - Arts, 7.667922e-15
    - Computers, 2.122138e-09
    - Games, 9.325108e-07
    - Health, 3.978250e-04
    - Home, 1.843479e-02
    - Kids and Teens, 7.919902e-03
    - News, 9.560907e-10
    - Recreation, 6.869806e-02
    - Reference, 1.501172e-02
    - Regional, 1.169344e-02
    - Science, 2.183749e-08
    - Shopping, 4.933331e-13
    - Society, 3.742857e-02
    - Sports, 8.289128e-04
    - World, 1.167155e-02
- Random Guessing
    - Business, 6.207791e-10
    - Adult, 1.755783e-25
    - Arts, 5.265419e-08
    - Computers, 9.662568e-02
    - Games, 8.530924e-08
    - Health, 3.982158e-03
    - Home, 1.627928e-03
    - Kids and Teens, 1.029429e-11

  - News, 3.367470e-01
  - Recreation, 9.287807e-02
  - Reference, 4.431061e-05
  - Regional, 1.070639e-01
  - Science, 1.948593e-16
  - Shopping, 2.443829e-02
  - Society, 1.020441e-02
  - Sports, 7.846725e-04
  - World, 8.004550e-02

- KNN City Block Distance (K = 1)
  - Homepage, 8.980712e-01
  - Clickable Content, 9.863119e-01
  - Search Results, 4.977701e-01

- Classification Trees - Deviance
  - Homepage, 7.249996e-01
  - Clickable Content, 9.998357e-01
  - Search Results, 9.788912e-01

- Naive Bayes - Multinomial
  - Homepage, 1.417794e-34
  - Clickable Content, 2.509826e-130
  - Search Results, 6.143993e-04

- LDA - Linear
  - Homepage, 3.359221e-17
  - Clickable Content, 1.411337e-99
  - Search Results, 1.404193e-05

  - Random Guessing
    - Homepage, 6.826242e-14
    - Clickable Content, 8.066855e-11
    - Search Results, 1.001127e-31

- Number of Servers
  - KNN City Block Distance (K = 1)
    - Mobile Optimized, 8.950941e-01
    - Traditional, 1.000000e+00
  - Classification Trees - Deviance
    - Mobile Optimized, 9.901304e-01
    - Traditional, 9.999999e-01
  - Naive Bayes - Multinomial
    - Mobile Optimized, 6.696808e-48
    - Traditional, 1.672604e-06
  - LDA - Linear
    - Mobile Optimized, 2.009603e-10
    - Traditional, 7.607447e-40
  - Random Guessing
    - Mobile Optimized, 3.745382e-46
    - Traditional, 2.270249e-11
  - KNN City Block Distance (K = 1)
    - Video, 1
    - Nonvideo, 1
  - Classification Trees - Deviance

  - Video, 1
  - Nonvideo, 1

- Naive Bayes - Multinomial
  - Video, 1.893484e-01
  - Nonvideo, 2.123292e-02

- LDA - Linear
  - Video, 1.000000e+00
  - Nonvideo, 1.000000e+00

- Random Guessing
  - Video, 5.651089e-01
  - Nonvideo, 2.565372e-08

- KNN City Block Distance (K = 1)
  - Business, 9.998722e-01
  - Adult, 9.514054e-01
  - Arts, 9.992799e-01
  - Computers, 8.738320e-01
  - Games, 1.000000e+00
  - Health, 9.768581e-01
  - Home, 9.247056e-01
  - Kids and Teens, 2.932152e-01
  - News, 9.819199e-01
  - Recreation, 9.854286e-01
  - Reference, 9.346716e-01
  - Regional, 9.973532e-01
  - Science, 8.385739e-01
  - Shopping, 9.999961e-01

  - Society, 9.765766e-01
  - Sports, 7.860754e-01
  - World, 6.851466e-01

- Classification Trees - Deviance
  - Business, 6.624027e-01
  - Adult, 8.547788e-01
  - Arts, 7.528675e-01
  - Computers, 6.460346e-01
  - Games, 5.245912e-01
  - Health, 4.164604e-01
  - Home, 9.967455e-01
  - Kids and Teens, 2.262396e-01
  - News, 5.416420e-01
  - Recreation, 7.190816e-01
  - Reference, 8.722464e-01
  - Regional, 7.919526e-01
  - Science, 1.203945e-02
  - Shopping, 9.006634e-01
  - Society, 9.788866e-01
  - Sports, 7.934647e-01
  - World, 8.883390e-01
- Naive Bayes - Multinomial
  - Business, 3.757601e-01
  - Adult, 2.728743e-01
  - Arts, 1.374716e-05
  - Computers, 1.119869e-13
  - Games, 1.072337e-01

    - Health, 0
    - Home, 5.822085e-01
    - Kids and Teens, 2.297604e-01
    - News, 5.045282e-07
    - Recreation, 2.632904e-01
    - Reference, 3.137847e-09
    - Regional, 4.826011e-01
    - Science, 1.998443e-02
    - Shopping, 2.434695e-11
    - Society, 1.436340e-01
    - Sports, 1.913320e-02
    - World, 7.399185e-01
  - LDA - Linear
    - Business, 8.939208e-02
    - Adult, 7.348541e-02
    - Arts, 3.181326e-08
    - Computers, 4.536154e-12
    - Games, 1.764841e-02
    - Health, 5.681021e-03
    - Home, 4.353342e-02
    - Kids and Teens, 6.334784e-03
    - News, 7.072774e-11
    - Recreation, 4.037785e-01
    - Reference, 1.001810e-03
    - Regional, 6.118861e-06
    - Science, 1.992831e-13
    - Shopping, 4.344587e-13

    - Society, 6.119660e-02
    - Sports, 1.164970e-01
    - World, 1.160040e-03

  - Random Guessing
    - Business, 9.771616e-07
    - Adult, 1.651121e-03
    - Arts, 2.263240e-05
    - Computers, 1.929651e-01
    - Games, 3.688511e-05
    - Health, 1.034780e-01
    - Home, 5.212033e-06
    - Kids and Teens, 7.991845e-10
    - News, 6.583784e-07
    - Recreation, 8.179732e-01
    - Reference, 3.686094e-02
    - Regional, 7.088230e-02
    - Science, 4.556930e-12
    - Shopping, 6.056420e-01
    - Society, 3.557411e-01
    - Sports, 3.054775e-08
    - World, 2.194456e-04

  - KNN City Block Distance (K = 1)
    - Homepage, 9.999632e-01
    - Clickable Content, 6.867079e-01
    - Search Results, 5.899846e-01

  - Classification Trees - Deviance

    - Homepage, 9.943930e-01
    - Clickable Content, 9.658874e-01
    - Search Results, 7.485196e-01
  - Naive Bayes - Multinomial
    - Homepage, 6.734300e-01
    - Clickable Content, 1.966331e-02
    - Search Results, 5.398247e-03
  - LDA - Linear
    - Homepage, 8.215819e-06
    - Clickable Content, 2.740834e-14
    - Search Results, 4.688613e-10
  - Random Guessing
    - Homepage, 1.379406e-01
    - Clickable Content, 4.198363e-05
    - Search Results, 5.189463e-04
- Number of Bidirectional Segments
  - KNN City Block Distance (K = 1)
    - Mobile Optimized, 8.681408e-01
    - Traditional, 1.000000e+00
  - Classification Trees - Deviance
    - Mobile Optimized, 9.365595e-01
    - Traditional, 9.999792e-01
  - Naive Bayes - Multinomial
    - Mobile Optimized, 7.029645e-75

  - Traditional, 7.932697e-171

- LDA - Linear
  - Mobile Optimized, 2.750753e-20
  - Traditional, 1.365849e-105

- Random Guessing
  - Mobile Optimized, 3.762549e-91
  - Traditional, 1.622602e-23

- KNN City Block Distance (K = 1)
  - Video, 1
  - Nonvideo, 1

- Classification Trees - Deviance
  - Video, 1
  - Nonvideo, 1.000000e+00

- Naive Bayes - Multinomial
  - Video, 1.656106e-03
  - Nonvideo, 2.133030e-20

- LDA - Linear
  - Video, 1.000000e+00
  - Nonvideo, 6.640748e-01

- Random Guessing
  - Video, 8.414154e-10
  - Nonvideo, 4.776861e-207

- KNN City Block Distance (K = 1)
  - Business, 9.988493e-01

  - Adult, 9.532255e-01
  - Arts, 9.993482e-01
  - Computers, 1.000000e+00
  - Games, 9.982084e-01
  - Health, 9.991395e-01
  - Home, 5.510404e-01
  - Kids and Teens, 7.377679e-01
  - News, 7.559213e-01
  - Recreation, 9.965709e-01
  - Reference, 9.869087e-01
  - Regional, 9.999974e-01
  - Science, 9.384791e-01
  - Shopping, 9.981183e-01
  - Society, 9.999999e-01
  - Sports, 9.987011e-01
  - World, 9.954277e-01

- Classification Trees - Deviance
  - Business, 6.398089e-01
  - Adult, 9.579182e-01
  - Arts, 8.554191e-01
  - Computers, 9.697956e-01
  - Games, 1.915453e-01
  - Health, 5.772498e-01
  - Home, 9.209584e-01
  - Kids and Teens, 4.762603e-01
  - News, 8.167495e-01
  - Recreation, 9.926633e-01

    - Reference, 6.691148e-01
    - Regional, 7.421163e-01
    - Science, 5.110980e-02
    - Shopping, 5.685722e-01
    - Society, 6.338860e-01
    - Sports, 8.533837e-01
    - World, 1.067033e-01

- Naive Bayes - Multinomial
    - Business, 9.195638e-02
    - Adult, 2.541755e-04
    - Arts, 4.613082e-41
    - Computers, 4.955141e-02
    - Games, 2.616177e-01
    - Health, 0
    - Home, 9.622295e-01
    - Kids and Teens, 1.541953e-02
    - News, 2.458541e-03
    - Recreation, 3.653853e-01
    - Reference, 2.471514e-11
    - Regional, 5.541286e-03
    - Science, 6.783779e-01
    - Shopping, 2.780317e-18
    - Society, 5.613111e-03
    - Sports, 1.367964e-01
    - World, 1.501536e-01

- LDA - Linear
    - Business, 1.008631e-01

    - Adult, 4.490945e-04
    - Arts, 5.128302e-16
    - Computers, 1.862155e-12
    - Games, 1.932733e-06
    - Health, 1.445641e-02
    - Home, 1.299254e-01
    - Kids and Teens, 2.268978e-02
    - News, 2.598492e-11
    - Recreation, 3.394833e-01
    - Reference, 1.881004e-01
    - Regional, 1.008519e-02
    - Science, 1.283988e-01
    - Shopping, 8.374265e-13
    - Society, 1.762935e-04
    - Sports, 1.196946e-05
    - World, 2.971282e-02

  - Random Guessing
    - Business, 8.161453e-10
    - Adult, 1.039321e-18
    - Arts, 1.194699e-08
    - Computers, 7.911335e-02
    - Games, 8.530924e-08
    - Health, 1.310853e-02
    - Home, 2.433564e-03
    - Kids and Teens, 1.881920e-11
    - News, 3.721474e-01
    - Recreation, 5.020308e-02

    - Reference, 1.797532e-03
    - Regional, 5.742149e-02
    - Science, 9.283746e-01
    - Shopping, 2.443829e-02
    - Society, 9.896470e-03
    - Sports, 7.671102e-04
    - World, 5.080001e-02

  - KNN City Block Distance (K = 1)
    - Homepage, 9.934163e-01
    - Clickable Content, 9.789054e-01
    - Search Results, 4.506664e-01

  - Classification Trees - Deviance
    - Homepage, 8.363577e-01
    - Clickable Content, 9.947511e-01
    - Search Results, 9.922539e-01

  - Naive Bayes - Multinomial
    - Homepage, 6.506030e-21
    - Clickable Content, 2.922743e-126
    - Search Results, 9.716675e-03

  - LDA - Linear
    - Homepage, 1.453289e-09
    - Clickable Content, 5.911520e-104
    - Search Results, 1.163678e-12

  - Random Guessing
    - Homepage, 3.010576e-06

        - Clickable Content, 1.132606e-12
        - Search Results, 3.511810e-28

- Number of Objects/Epochs
    - KNN City Block Distance (K = 1)
        - Mobile Optimized, 7.944660e-01
        - Traditional, 1.000000e+00
    - Classification Trees - Deviance
        - Mobile Optimized, 8.007933e-01
        - Traditional, 9.990047e-01
    - Naive Bayes - Multinomial
        - Mobile Optimized, 6.537436e-74
        - Traditional, 5.043031e-38
    - LDA - Linear
        - Mobile Optimized, 2.265689e-17
        - Traditional, 1.494170e-64
    - Random Guessing
        - Mobile Optimized, 4.122601e-79
        - Traditional, 2.928726e-17
    - KNN City Block Distance (K = 1)
        - Video, 1
        - Nonvideo, 1.000000e+00
    - Classification Trees - Deviance
        - Video, 1
        - Nonvideo, 1.000000e+00

- Naive Bayes - Multinomial
  - Video, 1.692669e-02
  - Nonvideo, 2.327775e-03
- LDA - Linear
  - Video, 1
  - Nonvideo, 9.502916e-01
- Random Guessing
  - Video, 1.903623e-03
  - Nonvideo, 2.595300e-59
- KNN City Block Distance (K = 1)
  - Business, 9.835103e-01
  - Adult, 9.977981e-01
  - Arts, 9.992629e-01
  - Computers, 9.998671e-01
  - Games, 1.000000e+00
  - Health, 6.871176e-01
  - Home, 5.510404e-01
  - Kids and Teens, 9.472889e-01
  - News, 9.615116e-01
  - Recreation, 8.781801e-01
  - Reference, 9.973954e-01
  - Regional, 9.999993e-01
  - Science, 2.338744e-01
  - Shopping, 9.218577e-01
  - Society, 9.999919e-01
  - Sports, 1.000000e+00

  - World, 9.998394e-01

- Classification Trees - Deviance
  - Business, 5.072299e-01
  - Adult, 9.997564e-01
  - Arts, 9.774073e-01
  - Computers, 9.874899e-01
  - Games, 1.028227e-01
  - Health, 8.473341e-01
  - Home, 2.137905e-01
  - Kids and Teens, 2.663928e-01
  - News, 9.052003e-01
  - Recreation, 4.139153e-01
  - Reference, 3.797396e-01
  - Regional, 9.420360e-01
  - Science, 7.774923e-02
  - Shopping, 6.093171e-01
  - Society, 9.904457e-01
  - Sports, 9.569115e-01
  - World, 3.054084e-01

- Naive Bayes - Multinomial
  - Business, 3.104792e-01
  - Adult, 4.647533e-02
  - Arts, 2.196994e-07
  - Computers, 2.830330e-02
  - Games, 3.829776e-01
  - Health, 0
  - Home, 1.211001e-02

  - Kids and Teens, 1.541953e-02
  - News, 2.086029e-01
  - Recreation, 6.660552e-02
  - Reference, 1.905006e-11
  - Regional, 3.324938e-04
  - Science, 8.119221e-05
  - Shopping, 1.700183e-06
  - Society, 8.020590e-06
  - Sports, 2.343055e-02
  - World, 3.788332e-01

- LDA - Linear
  - Business, 1.866651e-01
  - Adult, 4.786060e-02
  - Arts, 1.520112e-21
  - Computers, 5.198083e-13
  - Games, 2.910778e-04
  - Health, 2.418753e-06
  - Home, 4.325192e-05
  - Kids and Teens, 6.475561e-05
  - News, 1.883162e-15
  - Recreation, 4.681125e-01
  - Reference, 6.709096e-03
  - Regional, 3.365982e-04
  - Science, 4.886120e-08
  - Shopping, 8.388555e-14
  - Society, 4.802667e-01
  - Sports, 3.365486e-03

    - World, 4.449033e-01

  - Random Guessing
    - Business, 4.139337e-12
    - Adult, 3.512530e-08
    - Arts, 7.852916e-08
    - Computers, 3.450601e-02
    - Games, 3.521883e-08
    - Health, 5.698458e-03
    - Home, 6.842047e-07
    - Kids and Teens, 1.259717e-11
    - News, 3.618618e-04
    - Recreation, 1.676714e-01
    - Reference, 3.709225e-02
    - Regional, 4.657560e-02
    - Science, 2.212655e-04
    - Shopping, 1.497245e-04
    - Society, 1.837060e-01
    - Sports, 2.657635e-09
    - World, 3.131206e-03

  - KNN City Block Distance (K = 1)
    - Homepage, 9.999554e-01
    - Clickable Content, 9.954072e-01
    - Search Results, 9.802824e-01

  - Classification Trees - Deviance
    - Homepage, 9.914308e-01
    - Clickable Content, 9.793335e-01

    - Search Results, 9.999998e-01

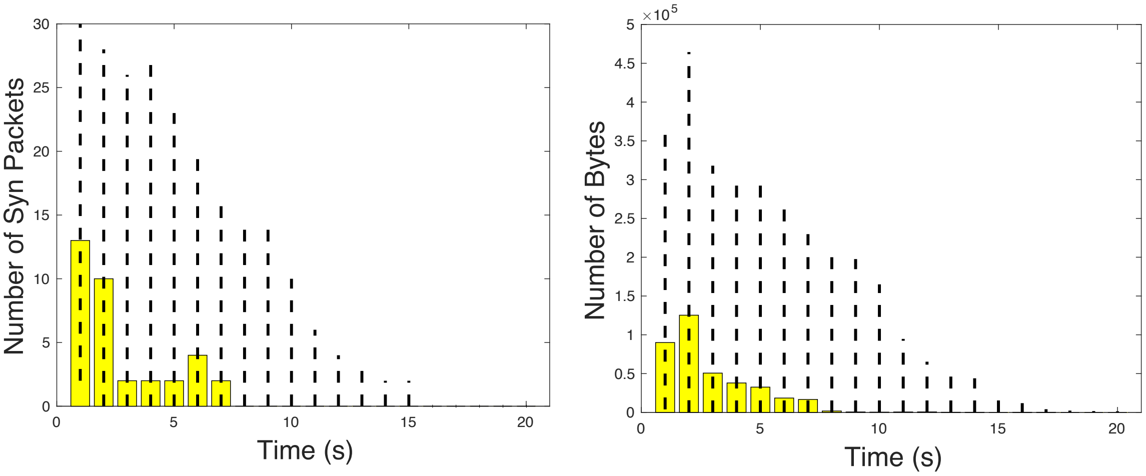
- Naive Bayes - Multinomial
    - Homepage, 6.402773e-07
    - Clickable Content, 3.113979e-29
    - Search Results, 1.778612e-02

- LDA - Linear
    - Homepage, 1.097070e-03
    - Clickable Content, 1.515659e-59
    - Search Results, 9.206878e-07

- Random Guessing
    - Homepage, 2.090200e-04
    - Clickable Content, 6.080219e-05
    - Search Results, 1.055630e-08

APPENDIX 12: TIME SERIES CHARACTERISTICS OF WEB PAGE TRAFFIC

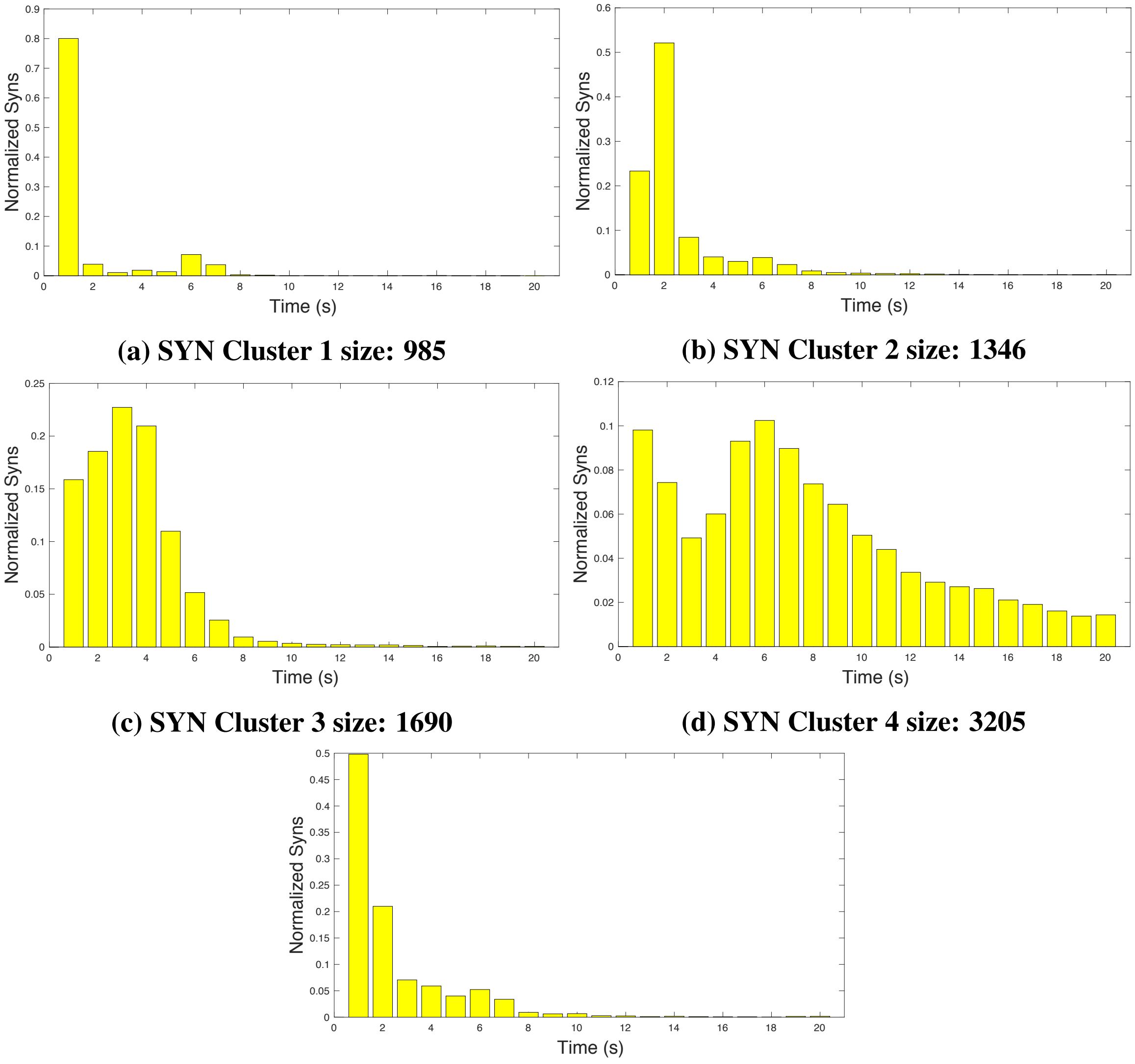
This appendix includes plots which characterize the time series generated by downloading web pages using the Google Chrome v 37.0.2062.124, Firefox v 32.0.2, and Opera v 25.0 browsers on a box running Windows 7. These plots supplement the plots provided in Section 5.1 of Chapter 5.



(a)

(b)

Figure 6.7: Overview of the variability in the magnitude of the traffic features across web pages - Windows 7 operating system.



(a) SYN Cluster 1 size: 985

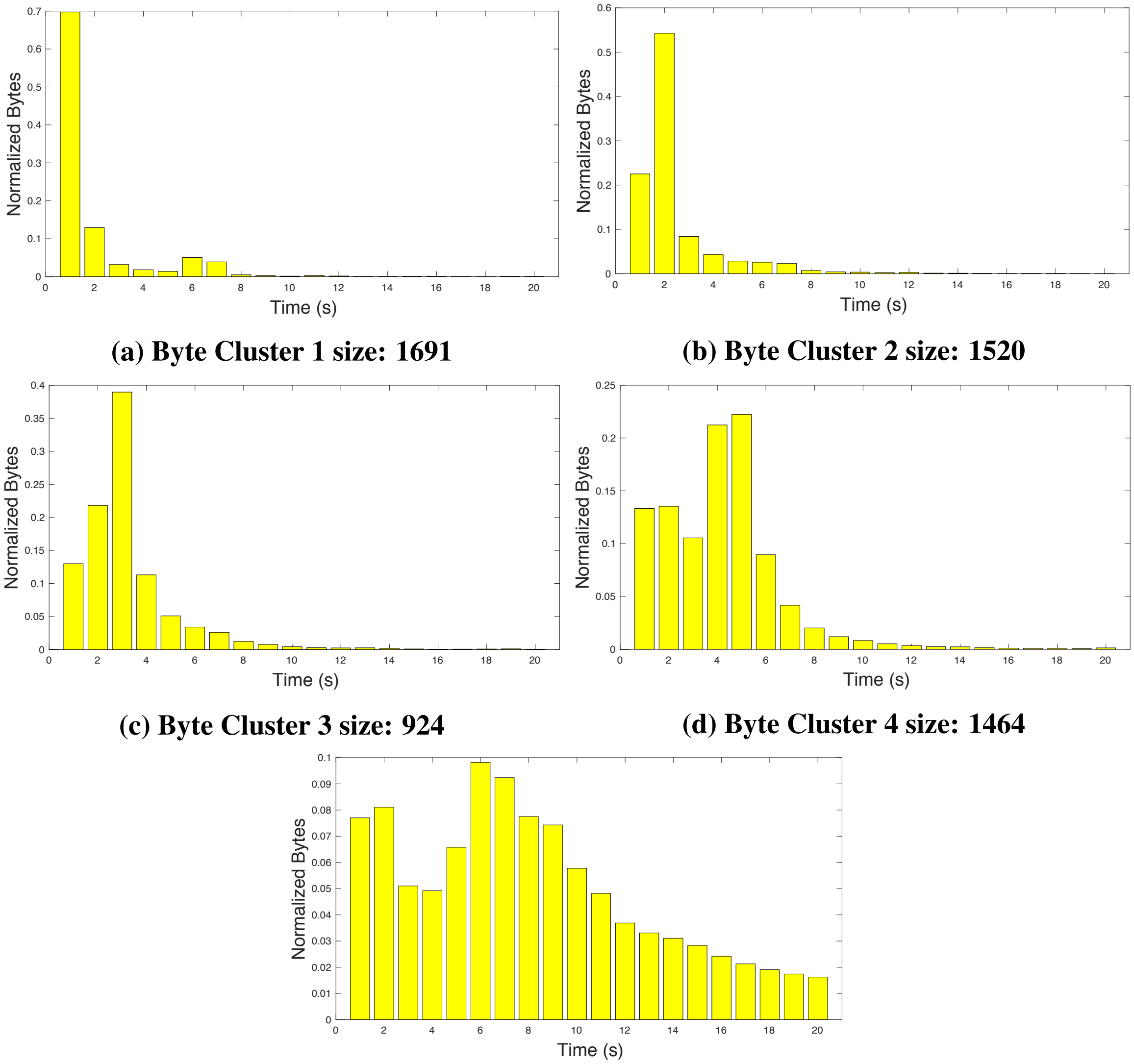
(b) SYN Cluster 2 size: 1346

(c) SYN Cluster 3 size: 1690

(d) SYN Cluster 4 size: 3205

(e) SYN Cluster 5 size: 1246

Figure 6.8: K-mean cluster centroids for web page download time series (SYNs) - Windows 7 operating system.



(a) Byte Cluster 1 size: 1691

(b) Byte Cluster 2 size: 1520

(c) Byte Cluster 3 size: 924

(d) Byte Cluster 4 size: 1464

(e) Byte Cluster 5 size: 2958

Figure 6.9: K-mean cluster centroids for web page download time series (Bytes) - Windows 7 operating system.

APPENDIX 13: WEB PAGE SEGMENTATION PERFORMANCE

This appendix includes performance results for the web page segmentation methods when the resolution parameter is adjusted. Web page segmentation approach performance is described in Chapter 5.

TABLE 6.1: Optimal Parameter Settings Used for Web Page Segmentation Approaches for r = 100ms

Segmentation Approach	Optimal Parameters
Idle-time method (Number of SYNs) Idle-time method (Number of Bytes)	I = 9s I = 10s
Idle-time method with threshold (Number of SYNs) Idle-time method with threshold (Number of Bytes)	I = 9s; T4 I = 3s; T=3000
Heuristic Change Detection Method (Number of SYNs) Heuristic Change Detection Method (Number of Bytes)	I=0s; T=4; L=8s I=0s; T=6000; L=9s
Fused Lasso (Number of SYNs) Fused Lasso (Number of Bytes)	λ = .1; μ=1.25; k = 5s λ = 1000; μ=10000; k = 5s
HMM (Number of SYNs)	k = 5s; $A_{1,1}$=.9; $A_{1,2}$=.1; $A_{2,1}$ =.05; $A_{2,2}$ =.95 $B_{1,1}$=.99; $B_{1,2}$=.01; $B_{2,1}$ =.8; $B_{2,2}$ =.2
HMM (Number of Bytes)	k = 5s; $A_{1,1}$=.87; $A_{1,2}$=.13; $A_{2,1}$ =.05; $A_{2,2}$ =.95 $B_{1,1}$=.97; $B_{1,2}$=.03; $B_{2,1}$ =.79; $B_{2,2}$ =.21

TABLE 6.2: Optimal Parameter Settings Used for Web Page Segmentation Approaches for r = 500ms

Segmentation Approach	Optimal Parameters
Idle-time method (Number of SYNs) Idle-time method (Number of Bytes)	I = 9s I = 10s
Idle-time method with threshold (Number of SYNs) Idle-time method with threshold (Number of Bytes)	I = 9s; T=6 I = 3.5s; T=5000
Heuristic Change Detection Method (Number of SYNs) Heuristic Change Detection Method (Number of Bytes)	I=0s; T=6; L=8s I=0s; T=9000; L=9s
Fused Lasso (Number of SYNs) Fused Lasso (Number of Bytes)	λ = .3; μ=3.25; k = 5s λ = 3000; μ=18000; k = 5s
HMM (Number of SYNs)	k = 5s; $A_{1,1}$=.9; $A_{1,2}$=.1; $A_{2,1}$ =.05; $A_{2,2}$ =.95 $B_{1,1}$=.99; $B_{1,2}$=.01; $B_{2,1}$ =.8; $B_{2,2}$ =.2
HMM (Number of Bytes)	k = 5s; $A_{1,1}$=.87; $A_{1,2}$=.13; $A_{2,1}$ =.03; $A_{2,2}$ =.97 $B_{1,1}$=.98; $B_{1,2}$=.02; $B_{2,1}$ =.77; $B_{2,2}$ =.23

TABLE 6.3: Optimal Parameter Settings Used for Web Page Segmentation Approaches for r = 1000ms

Segmentation Approach	Optimal Parameters
Idle-time method (Number of SYNs) Idle-time method (Number of Bytes)	I = 9s I = 10s
Idle-time method with threshold (Number of SYNs) Idle-time method with threshold (Number of Bytes)	I = 9s; T=6 I = 3.5s; T=5000
Heuristic Change Detection Method (Number of SYNs) Heuristic Change Detection Method (Number of Bytes)	I=0s; T=6; L=8s I=0s; T=12000; L=9s
Fused Lasso (Number of SYNs) Fused Lasso (Number of Bytes)	λ = .35; μ=3.4; k = 5s λ = 5000; μ=20000; k = 5s
HMM (Number of SYNs)	k = 5s; $A_{1,1}$=.9; $A_{1,2}$=.1; $A_{2,1}$ =.04; $A_{2,2}$ =.96 $B_{1,1}$=.99; $B_{1,2}$=.01; $B_{2,1}$ =.82; $B_{2,2}$ =.18
HMM (Number of Bytes)	k = 5s; $A_{1,1}$=.87; $A_{1,2}$=.13; $A_{2,1}$ =.02; $A_{2,2}$ =.98 $B_{1,1}$=.99; $B_{1,2}$=.01; $B_{2,1}$ =.75; $B_{2,2}$ =.25

TABLE 6.4: Impact of Resolution Parameter (r) on Web Page Segmentation Performance - Number of SYNs

Segmentation Approach	r (s)	TPR/Recall (SYNs)	FPR (SYNs)	Precision	F-score
Idle-time method	1.000	.4148	.0869	.8268	.5524
Idle-time method with threshold	1.000	.4016	.0328	.9245	.5600
Heuristic Change Detection	1.000	.7115	.1970	.7832	.7456
Fused Lasso	1.000	.8129	.0776	.9129	.8600
HMM	1.000	.5113	.0594	.8959	.6510
Idle-time method	.500	.4148	.0869	.8268	.5524
Idle-time method with threshold	.500	.4112	.0471	.8972	.5639
Heuristic Change Detection	.500	.7218	.2011	.4148	.0869
Fused Lasso	.500	.8155	.0610	.9304	.8692
HMM	.500	.5028	.0541	.9029	.6459
Idle-time method	.250	.4148	.0869	.8268	.5524
Idle-time method with threshold	.250	.4016	.0328	.9245	.5600
Heuristic Change Detection	.250	.7320	.2038	.7822	.7563
Fused Lasso	.250	.8217	.0629	.9289	.8720
HMM	.250	.5079	.0584	.8969	.6485
Idle-time method	.100	.4148	.0869	.8268	.5524
Idle-time method with threshold	.100	.4328	.0514	.8939	.5832
Heuristic Change Detection	.100	.7412	.2263	.7822	.7563
Fused Lasso	.100	.8376	.0720	.9208	.8773
HMM	.100	.5215	.0619	.8939	.6587

TABLE 6.5: Impact of Resolution Parameter (r) on Web Page Segmentation Performance - Number of Bytes

Segmentation Approach	r (s)	TPR/Recall (Bytes)	FPR (Bytes)	Precision	F-score
Idle-time method	1.000	.1092	.0723	.6017	.1847
Idle-time method with threshold	1.000	.9127	.7462	.5502	.6865
Heuristic Change Detection Method	1.000	.8417	.5518	.6040	.7033
Fused Lasso	1.000	.8012	.1655	.8288	.8147
HMM	1.000	.0234	.0098	.7048	.0453
Idle-time method	.500	.0911	.0752	.5478	.1562
Idle-time method with threshold	.500	.8987	.7688	.5390	.6738
Heuristic Change Detection Method	.500	.8289	.5374	.6067	.7006
Fused Lasso	.500	.8099	.1634	.8321	.8209
HMM	.500	.0241	.0105	.6965	.0460
Idle-time method	.250	.0889	.0770	.5359	.1525
Idle-time method with threshold	.250	.8951	.7623	.5401	.6737
Heuristic Change Detection Method	.250	.8393	.5476	.6052	.7033
Fused Lasso	.250	.8115	.1863	.8133	.8124
HMM	.250	.0257	.0107	.7060	.0496
Idle-time method	.100	.0986	.0836	.5412	.1668
Idle-time method with threshold	.100	.9187	.7916	.5372	.6779
Heuristic Change Detection Method	.100	.8551	.5520	.6077	.7105
Fused Lasso	.100	.8389	.2092	.8004	.8192
HMM	.100	.0266	.0113	.7019	.0513

TABLE 6.6: Impact of *s* Parameter on Web Page Segmentation for Precision and F-score Metrics (*r* = 250ms)

Segmentation Approach	*s* (seconds)	SYNS		Bytes	
	s (seconds)	Precision	F-score	Precision	F-score
Idle-time method	.250	.8268	.5524	.5359	.1525
Idle-time method (threshold)	.250	.9245	.5600	.5401	.6737
Heuristic Change Detection	.250	.7822	.7563	.6052	.7033
Fused Lasso	.250	.9289	.8720	.8133	.8124
HMM	.250	.8969	.6485	.7060	.0496
Idle-time method	.500	.8397	.5611	.5401	.1537
Idle-time method (threshold)	.500	.9468	.5734	.5404	.6740
Heuristic Change Detection	.500	.7919	.7657	.6064	.7047
Fused Lasso	.500	.9326	.8755	.8165	.8156
HMM	.500	.9207	.6658	.7143	.0502
Idle-time method	1	.8646	.5777	.6016	.1712
Idle-time method (threshold)	1	.9553	.5786	.5429	.6772
Heuristic Change Detection	1	.8135	.7865	.6182	.7184
Fused Lasso	1	.9513	.8930	.8248	.8239
HMM	1	.9454	.6837	.7692	.0540
Idle-time method	5	.9121	.6094	..6655	.1894
Idle-time method (threshold)	5	.9609	.5819	.5758	.7182
Heuristic Change Detection	5	.8700	.8411	.6354	.7384
Fused Lasso	5	.9634	.9044	.8773	.8734
HMM	5	.9767	.7063	.8077	.0567

REFERENCES

Why can't we all use chromium instead of google chrome? http://www.techdrivein.com/2010/05/why-cant-we-all-use-chromium-instead-of.html. Accessed: 2016-02-15.

Chromium code reviews. https://codereview.chromium.org/6577021. Accessed: 2017-01-02.

Websites blocked in mainland china. en.wikipedia.orgwikiWebsites_blocked_in_mainland_China. Accessed: 2015-07-30.

Developer's guide. https://developer.chrome.com/extensions/devguide. Accessed: 2017-01-11.

Turning off auto updates in google chrome. https://www.chromium.org/administrators/turning-off-auto-updates. Accessed: 2017-04-20.

Html5 reference: The syntax, vocabulary and apis of html5. http://dev.w3.org/html5/html-author/. Accessed: 2015-04-30.

Matlab. http://www.mathworks.com/help/stats/classify.html. Accessed: 2013-02-19.

Mashable. http://mashable.com/2013/08/20/mobile-web-traffic/, a. Accessed: 2013-11-27.

Make sure your site's ready for mobile-friendly google search results. https://support.google.com/adsense/answer/6196932?hl=en, b. Accessed: 2015-05.

2013 mobile device survey. http://www.scmagazine.com/2013-mobile-device-survey/slideshow/1222/, c. Accessed: 2014-05-04.

Nmap network scanning: Os detection. https://nmap.org/book/man-os-detection.html. Accessed: 2017-01-13.

Opera version history. http://www.opera.com/docs/history/. Accessed: 2016-02-15.

Privacy policy legislation and requirements by country. http://privacypolicies.com/blog/privacy-law-by-country/. Accessed: 2016-05-31.

Web development tools. https://developer.apple.com/safari/tools/. Accessed: 2017-01-11.

10 alternative web browsers for ubuntu linux. https://www.starryhope.com/10-alternative-browsers-for-ubuntu-linux/, a. Accessed: 2014-05-19.

Statcounter. http://gs.statcounter.com/, b. Accessed: 2013-06-30.

Wikimedia. http://stats.wikimedia.org/archive/squid_reports/2013-06/SquidReportClients.htm. Accessed: 2013-06-30.

Zend server. http://www.zend.com/en/products/server/downloads#. Accessed: 2016-03-25.

Privacy policy legislation & requirements by country. URL `http://privacypolicies.com/blog/privacy\-law\-by\-country/`. Accessed: 2017-01-13.

Giuseppe Aceto, Alberto Dainotti, Walter De Donato, and Antonio Pescap. Portload: taking the best of two worlds in traffic classification. In *INFOCOM IEEE Conference on Computer Communications Workshops, 2010*, pages 1–5. IEEE, 2010.

Sohaib Ahmad, Abdul Lateef Haamid, Zafar Ayyub Qazi, Zhenyu Zhou, Theophilus Benson, and Ihsan Ayyub Qazi. A view from the other side: Understanding mobile phone characteristics in the developing world. In *Proceedings of the 2016 ACM on Internet Measurement Conference*, pages 319–325. ACM, 2016.

Shane Alcock and Richard Nelson. Libprotoident: Traffic classification using lightweight packet inspection. *WAND Network Research Group, Tech. Rep*, 2012.

Manos Antonakakis, Roberto Perdisci, Wenke Lee, Nikolaos Vasiloglou II, and David Dagon. Detecting malware domains at the upper dns hierarchy. In *USENIX Security Symposium*, page 16, 2011.

Grenville Armitage and Jason But. Netsniff, 2007. URL `http://caia.swin.edu.au/ice/tools/netsniff/index.html`.

Hadi Asghari, Michel Van Eeten, Johannes M Bauer, and Milton Mueller. Deep packet inspection: Effects of regulation on its deployment by internet providers, 2013.

Paul Barford and Mark Crovella. Generating representative web workloads for network and server performance evaluation. *ACM SIGMETRICS Performance Evaluation Review*, 26(1):151–160, 1998.

Paul Barford, Jeffery Kline, David Plonka, and Amos Ron. A signal analysis of network traffic anomalies. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 71–82. ACM, 2002.

Naimul Basher, Aniket Mahanti, Anirban Mahanti, Carey Williamson, and Martin Arlitt. A comparative analysis of web and peer-to-peer traffic. In *Proceedings of the 17th international conference on World Wide Web*, pages 287–296. ACM, 2008.

Mike Belshe, Martin Thomson, and Roberto Peon. Hypertext transfer protocol version 2, 2015.

Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 49–62. ACM, 2009.

Zachary S Bischof, John P Rula, and Fabián E Bustamante. In and out of cuba: Characterizing cuba's connectivity. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, pages 487–493. ACM, 2015.

Christopher M Bishop. Pattern recognition and machine learning (information science and statistics). 2007.

Kevin Bleakley and Jean-Philippe Vert. The group fused lasso for multiple change-point detection. *arXiv preprint arXiv:1106.4199*, 2011.

Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. Ubicrawler: A scalable fully distributed web crawler. *Software: Practice and Experience*, 34(8):711–726, 2004.

José Borges and Mark Levene. A dynamic clustering-based markov model for web usage mining. *arXiv preprint cs/0406032*, 2004.

Julie Bort. By 2017, we'll each have 5 internet devices(and more predictions from cisco). http://www.businessinsider.com/cisco-predicts-mobile-2013-5?op=1. Accessed: 2014-05-04.

Ed Bott. Despite automatic updates, old browsers are still a problem. http://www.zdnet.com/article/despite-automatic-updates-old-browsers-are-still-a-problem/. 2014-01-06.

Anna Bouch, Allan Kuchinsky, and Nina Bhatti. Quality is in the eye of the beholder: meeting users' requirements for internet quality of service. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 297–304. ACM, 2000.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

Daniela Brauckhoff, Bernhard Tellenbach, Arno Wagner, Martin May, and Anukool Lakhina. Impact of packet sampling on anomaly detection metrics. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 159–164. ACM, 2006.

Jon Brodkin. Netflix performance on verizon and comcast has been dropping for months. http://arstechnica.com/information-technology/2014/02/netflix-performance-on-verizon-and-comcast-has-been-dropping-for-months/. Accessed: 2014-05-04.

Craig Buckler. Browser trends march 2016: Operating system surprises. http://www.sitepoint.com/browser-trends-march-2016-operating-system-surprises/. Accessed: 2016-04-16.

Phillip Bump. Half of internet traffic in north america is just to watch netflix and youtube. http://www.thewire.com/technology/2013/05/netflix-youtube-traffic/65210/. Accessed: 2014-05-04.

Michael Butkiewicz, Harsha V Madhyastha, and Vyas Sekar. Understanding website complexity: measurements, metrics, and implications. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 313–328. ACM, 2011.

Igor Cadez, David Heckerman, Christopher Meek, Padhraic Smyth, and Steven White. Model-based clustering and visualization of navigation patterns on a web site. *Data Mining and Knowledge Discovery*, 7 (4):399–424, 2003.

Xiang Cai, Xin Cheng Zhang, Brijesh Joshi, and Rob Johnson. Touching from a distance: Website fingerprinting attacks and defenses. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 605–616. ACM, 2012.

Thomas Callahan, Mark Allman, and Michael Rabinovich. On modern dns behavior and properties. *ACM SIGCOMM Computer Communication Review*, 43(3):7–15, 2013.

Tom Callahan, Mark Allman, and Vern Paxson. A longitudinal view of http traffic. In *Passive and Active Measurement*, pages 222–231. Springer, 2010.

Davide Canali, Marco Cova, Giovanni Vigna, and Christopher Kruegel. Prophiler: a fast filter for the large-scale detection of malicious web pages. In *Proceedings of the 20th international conference on World wide web*, pages 197–206. ACM, 2011.

Valentín Carela-Español, Tomasz Bujlow, and Pere Barlet-Ros. Is our ground-truth for traffic classification reliable? In *Passive and Active Measurement*, pages 98–108. Springer, 2014.

Abdelberi Chaabane, Yuan Ding, Ratan Dey, Mohamed Ali Kaafar, and Keith W Ross. A closer look at third-party osn applications: Are they leaking your personal information? In *Passive and Active Measurement*, pages 235–246. Springer, 2014.

Zachary Chase Lipton, Charles Elkan, and Balakrishnan Narayanaswamy. Thresholding classifiers to maximize f1 score. *arXiv preprint arXiv:1402.1892*, 2014.

HC Chen and SW Chen. A moving average based filtering system with its application to real-time qrs detection. In *Computers in Cardiology, 2003*, pages 585–588. IEEE, 2003.

Jianqing Chen and Jan Stallaert. An economic analysis of online advertising using behavioral targeting. *Mis Quarterly*, 38(2):429–449, 2014.

Jiu Jun Chen, Ji Gao, Jun Hu, and Bei Shui Liao. Dynamic mining for web navigation patterns based on markov model. In *Computational and Information Science*, pages 806–811. Springer, 2005.

Ruichuan Chen, Istemi Ekin Akkus, and Paul Francis. Splitx: High-performance private analytics. In *ACM SIGCOMM Computer Communication Review*, pages 315–326. ACM, 2013.

Zhicong Cheng, Bin Gao, and Tie-Yan Liu. Actively predicting diverse search intent from user browsing behaviors. In *Proceedings of the 19th international conference on World wide web*, pages 221–230. ACM, 2010.

Flavio Chierichetti, Ravi Kumar, Prabhakar Raghavan, and Tamás Sarlós. Are web users really markovian? In *WWW*, pages 609–618, 2012.

Junghoo Cho and Hector Garcia-Molina. The evolution of the web and implications for an incremental crawler, 1999.

Junghoo Cho, Hector Garcia-Molina, and Lawrence Page. Efficient crawling through url ordering, 1998.

David R Choffnes and Fabián E Bustamante. Taming the torrent: a practical approach to reducing cross-isp traffic in peer-to-peer systems. In *ACM SIGCOMM Computer Communication Review*, pages 363–374. ACM, 2008.

Ben Choi and Zhongmei Yao. Web page classification*. In *Foundations and Advances in Data Mining*, pages 221–274. Springer, 2005.

Hyoung-Kee Choi and John O Limb. A behavioral model of web traffic. In *Network Protocols, 1999.(ICNP'99) Proceedings. Seventh International Conference on*, pages 327–334. IEEE, 1999.

Mikkel Christiansen, Kevin Jeffay, David Ott, and F Donelson Smith. Tuning red for web traffic. In *ACM SIGCOMM Computer Communication Review*, pages 139–150. ACM, 2000.

Brent Chun, David Culler, Timothy Roscoe, Andy Bavier, Larry Peterson, Mike Wawrzoniak, and Mic Bowman. Planetlab: an overlay testbed for broad-coverage services. *ACM SIGCOMM Computer Communication Review*, 33(3):3–12, 2003.

Benoit Claise. Cisco systems netflow services export version 9, 2004.

Gerald Combs et al. Wireshark. *Web page: http://www. wireshark. org/last modified*, pages 12–02, 2007.

Phorm Corporation. Phorm. http://www.phorm.com. URL `http://www.phorm.com`.

Miguel Costa and Mário J Silva. Evaluating web archive search systems. In *Web Information Systems Engineering-WISE 2012*, pages 440–454. Springer, 2012.

Scott E Coull, Michael P Collins, Charles V Wright, Fabian Monrose, and Michael K Reiter. On web browsing privacy in anonymized netflows. In *USENIX Security*, 2007.

Manuel Crotti, Maurizio Dusi, Francesco Gringoli, and Luca Salgarelli. Traffic classification through simple statistical fingerprinting. *ACM SIGCOMM Computer Communication Review*, 37(1):5–16, 2007.

Mark E Crovella and Azer Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *Networking, IEEE/ACM Transactions on*, 5(6):835–846, 1997.

Joan Daemen and Vincent Rijmen. Aes proposal: Rijndael. 1999.

George Danezis. Traffic analysis of the http protocol over tls. http://www0.cs.ucl.ac.uk/staff/G.Danezis/papers/TLSanon.pdf. Accessed: 2016-02-28.

Nienke de Boer, Marijtje van Leeuwen, Ruud van Luijk, Kim Schouten, Flavius Frasincar, and Damir Vandic. Identifying explicit features for sentiment analysis in consumer reviews. In *Web Information Systems Engineering–WISE 2014*, pages 357–371. Springer, 2014.

J-Y Delort, Bernadette Bouchon-Meunier, and Maria Rifqi. Enhanced web document summarization using hyperlinks. In *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pages 208–215. ACM, 2003.

L Deogioanni et al. Windump: tcpdump for windows. *Politecnico di Torino, Italia, March*, 2000.

Luca Deri, Mario Martinelli, Tomasz Bujlow, and Alfredo Cardigliano. ndpi: Open-source high-speed deep packet inspection. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International*, pages 617–622. IEEE, 2014.

Christian Dewes, Arne Wichmann, and Anja Feldmann. An analysis of internet chat systems. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 51–64. ACM, 2003.

Mohan Dhawan, Justin Samuel, Renata Teixeira, Christian Kreibich, Mark Allman, Nicholas Weaver, and Vern Paxson. Fathom: a browser-based network measurement platform. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 73–86. ACM, 2012.

Python Documentation. urllib2 extensible library for opening urls. https://docs.python.org/2/library/urllib2.html. Accessed: 2016-04-16.

Fred Douglis, Anja Feldmann, Balachander Krishnamurthy, and Jeffrey C Mogul. Rate of change and other metrics: a live study of the world wide web. In *USENIX Symposium on Internet Technologies and Systems*, volume 119, 1997.

Idilio Drago, Marco Mellia, Maurizio M Munafo, Anna Sperotto, Ramin Sadre, and Aiko Pras. Inside dropbox: understanding personal cloud storage services. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 481–494. ACM, 2012.

Kevin P Dyer, Scott E Coull, Thomas Ristenpart, and Thomas Shrimpton. Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 332–346. IEEE, 2012.

Jeffrey Erman and Kadangode K Ramakrishnan. Understanding the super-sized traffic of the super bowl. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 353–360. ACM, 2013.

Jeffrey Erman, Martin Arlitt, and Anirban Mahanti. Traffic classification using clustering algorithms. In *Proceedings of the 2006 SIGCOMM workshop on Mining network data*, pages 281–286. ACM, 2006.

Jeffrey Erman, Anirban Mahanti, Martin Arlitt, Ira Cohen, and Carey Williamson. Semi-supervised network traffic classification. In *ACM SIGMETRICS Performance Evaluation Review*, volume 35, pages 369–370. ACM, 2007a.

Jeffrey Erman, Anirban Mahanti, Martin Arlitt, and Carey Williamson. Identifying and discriminating between web and peer-to-peer traffic in the network core. In *Proceedings of the 16th international conference on World Wide Web*, pages 883–892. ACM, 2007b.

Jeffrey Erman, Vijay Gopalakrishnan, Rittwik Jana, and KK Ramakrishnan. Towards a spdy'ier mobile web? In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pages 303–314. ACM, 2013.

Jinliang Fan, Jun Xu, and Mostafa H. Ammar. Crypto-pan: Cryptography-based prefix-preserving anonymization, 2004a.

Jinliang Fan, Jun Xu, Mostafa H Ammar, and Sue B Moon. Prefix-preserving ip address anonymization: measurement-based security evaluation and a new cryptography-based scheme. *Computer Networks*, 46 (2):253–272, 2004b.

Dennis Fetterly, Mark Manasse, Marc Najork, and Janet Wiener. A large-scale study of the evolution of web pages. In *Proceedings of the 12th international conference on World Wide Web*, pages 669–678. ACM, 2003.

Roy Fielding, Jim Gettys, Jeffrey Mogul, Henrik Frystyk, Larry Masinter, Paul Leach, and Tim Berners-Lee. Hypertext transfer protocol–http/1.1, 1999.

G David Forney Jr. The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, 1973.

Ned Freed, Murray Kucherawy, Mark Baker, and Bjoern Hoehrmann. Media types. http://www.iana.org/assignments/media-types/media-types.xhtml, 2015.

Jane Fridlyand, Antoine M Snijders, Dan Pinkel, Donna G Albertson, and Ajay N Jain. Hidden markov models approach to the analysis of array cgh data. *Journal of multivariate analysis*, 90(1):132–153, 2004.

Dennis F Galletta, Raymond Henry, Scott McCoy, and Peter Polak. Web site delays: How tolerant are users? *Journal of the Association for Information Systems*, 5(1):1, 2004.

E Gavaletz, D Hamon, and J Kaur. Comparing in-browser methods of measuring resource load times. In *W3C Workshop on Web Performance 8*, 2012.

Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. Youtube traffic characterization: A view from the edge. http://www.hpl.hp.com/techreports/2007/HPL-2007-119.pdf. Accessed: 2014-03-13.

Google. Pagespeed tools. https://developers.google.com/speed/pagespeed/. Accessed: 2014-03-13.

Aberdeen Group. Why web performance matters: Is your site driving customers away? http://www.mcrinc.com/Documents/Newsletters/ 201110_why_web_performance_matters.pdf, 2011. Accessed: 2013-06-18.

Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection. In *USENIX Security Symposium*, pages 139–154, 2008.

Saikat Guha and Paul Francis. Characterization and measurement of tcp traversal through nats and firewalls. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, pages 18–18. USENIX Association, 2005.

Patrick Haffner, Subhabrata Sen, Oliver Spatscheck, and Dongmei Wang. Acas: automated construction of application signatures. In *Proceedings of the 2005 ACM SIGCOMM workshop on Mining network data*, pages 197–202. ACM, 2005.

Mark A Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

John A Hartigan. *Clustering algorithms*. John Wiley & Sons, Inc., 1975.

Yeye He and Jeffrey F Naughton. Anonymization of set-valued data via top-down, local generalization. *Proceedings of the VLDB Endowment*, 2(1):934–945, 2009.

Jonathan Hedley. jsoup: Java html parser, 2010.

Felix Hernandez-Campos. *Generation and validation of empirically-derived TCP application workloads*. PhD thesis, University of North Carolina at Chapel Hill, 2006.

Félix Hernández-Campos, Kevin Jeffay, and F Donelson Smith. Tracking the evolution of web traffic: 1995-2003. In *Modeling, Analysis and Simulation of Computer Telecommunications Systems, 2003. MASCOTS 2003. 21st IEEE/ACM International Symposium on*, pages 16–25. IEEE, 2003a.

Félix Hernández-Campos, AB Nobel, FD Smith, and K Jeffay. Statistical clustering of internet communication patterns. *computing science and statistics*, 35, 2003b.

Dominik Herrmann, Rolf Wendolsky, and Hannes Federrath. Website fingerprinting: attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, pages 31–42. ACM, 2009.

Allan Heydon and Marc Najork. Mercator: A scalable, extensible web crawler. *World Wide Web*, 2(4): 219–229, 1999.

Junxian Huang, Qiang Xu, Birjodh Tiwana, Z Morley Mao, Ming Zhang, and Paramvir Bahl. Anatomizing application performance differences on smartphones. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 165–178. ACM, 2010.

IDC. Smartphone os market share, 2015 q2. http://www.idc.com/prodserv/smartphone-os-market-share.jsp. Accessed: 2015-05-12.

Sunghwan Ihm and Vivek S Pai. Towards understanding modern web traffic. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 295–312. ACM, 2011.

Alexa Internet Inc. Alexa. http://www.alexa.com. Accessed: 2013-02-19.

Internetlivestats.com. Total number of websites. http://www.internetlivestats.com/total-number-of-websites/, 2015. Accessed: 2015-01-17.

Gregoire Jacob, Engin Kirda, Christopher Kruegel, and Giovanni Vigna. Pubcrawl: Protecting users and businesses from crawlers. In *Presented as part of the 21st USENIX Security Symposium*, pages 507–522, Berkeley, CA, 2012. USENIX. ISBN 978-931971-95-9. URL `https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/jacob`.

V Jacobsen, Craig Leres, and Steven McCanne. Tcpdump/libpcap, 2005.

Van Jacobson, Craig Leres, and S McCanne. The tcpdump manual page. *Lawrence Berkeley Laboratory, Berkeley, CA*, 1989.

Bernard J Jansen and Amanda Spink. How are we searching the world wide web? a comparison of nine search engine transaction logs. *Information Processing & Management*, 42(1):248–263, 2006.

Bernard J Jansen, Danielle L Booth, and Amanda Spink. Determining the informational, navigational, and transactional intent of web queries. *Information Processing & Management*, 44(3):1251–1266, 2008.

Louvel Jerome. Can the rise of spdy threaten http? http://blog.restlet.com/2011/10/06/can-the-rise-of-spdy-threaten-http/. Accessed: 2015-12-6.

Wolfgang John and Sven Tafvelin. Analysis of internet backbone traffic and header anomalies observed. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 111–116. ACM, 2007.

Tyler Johnson and Patrick Seeling. Desktop and mobile web page comparison: Characteristics, trends, and implications. *Communications Magazine, IEEE*, 52(9):144–151, 2014.

Diana Joumblatt, Oana Goga, Renata Teixeira, Jaideep Chandrashekar, and Nina Taft. Characterizing end-host application performance across multiple networking environments. In *INFOCOM, 2012 Proceedings IEEE*, pages 2536–2540. IEEE, 2012.

Jaeyeon Jung, Emil Sit, Hari Balakrishnan, and Robert Morris. Dns performance and the effectiveness of caching. *Networking, IEEE/ACM Transactions on*, 10(5):589–603, 2002.

Thomas Karagiannis, Andre Broido, Michalis Faloutsos, et al. Transport layer identification of p2p traffic. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 121–134. ACM, 2004.

Thomas Karagiannis, Konstantina Papagiannaki, and Michalis Faloutsos. Blinc: multilevel traffic classification in the dark. In *ACM SIGCOMM Computer Communication Review*, pages 229–240. ACM, 2005.

Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. Segmenting time series: A survey and novel approach. *Data mining in time series databases*, 57:1–22, 2004.

Hyunchul Kim, Kimberly C Claffy, Marina Fomenkov, Dhiman Barman, Michalis Faloutsos, and KiYoung Lee. Internet traffic classification demystified: myths, caveats, and the best practices. In *Proceedings of the 2008 ACM CoNEXT conference*, page 11. ACM, 2008.

John Kirstoff. Ip-anonymous, 2005. URL `http://search.cpan.org/dist/IP-Anonymous/`.

Eddie Kohler. Ipsumdump. URL `http://www.read.seas.harvard.edu/~kohler/ipsumdump/`. Accessed: 2017-01-13.

Dimitris Koukis, Spyros Antonatos, Demetres Antoniades, Evangelos P Markatos, and Panagiotis Trimintzios. A generic anonymization framework for network traffic. In *Communications, 2006. ICC'06. IEEE International Conference on*, volume 5, pages 2302–2309. IEEE, 2006.

Srinivas Krishnan and Fabian Monrose. An empirical study of the performance, security and privacy implications of domain name prefetching. In *Dependable Systems & Networks (DSN), 2011 IEEE/IFIP 41st International Conference on*, pages 61–72. IEEE, 2011.

James F Kurose. *Computer Networking: A Top-Down Approach Featuring the Internet, 3/E.* Pearson Education India, 2005.

C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. In *Proc. ACM SIGCOMM*, Aug 2010.

Anukool Lakhina, Mark Crovella, and Christophe Diot. Mining anomalies using traffic feature distributions. In *ACM SIGCOMM Computer Communication Review*, pages 217–228. ACM, 2005.

Nikos Laoutaris, Alex Pentland, and Krishna Gummadi. Data transparency lab. URL `http://www.datatransparencylab.org/`. Accessed: 2017-01-13.

Cornell University Law. Interception and disclosure of wire, oral, or electronic communications prohibited. URL `http://www.law.cornell.edu/uscode/text/18/2511`.

Long Le, Jay Aikat, Kevin Jeffay, and F Donelson Smith. The effects of active queue management and explicit congestion notification on web performance. *IEEE/ACM Transactions on Networking (TON)*, 15 (6):1217–1230, 2007.

Stevens Le Blond, David Choffnes, Wenxuan Zhou, Peter Druschel, Hitesh Ballani, and Paul Francis. Towards efficient traffic-analysis resistant anonymity networks. In *ACM SIGCOMM Computer Communication Review*, pages 303–314. ACM, 2013.

Will E Leland, Murad S Taqqu, Walter Willinger, and Daniel V Wilson. On the self-similar nature of ethernet traffic (extended version). *Networking, IEEE/ACM Transactions on*, 2(1):1–15, 1994.

Nektarios Leontiadis, Tyler Moore, and Nicolas Christin. Measuring and analyzing search-redirection attacks in the illicit online prescription drug trade. In *USENIX Security Symposium*, 2011.

Justin Levandoski, Ethan Sommer, and Matthew Strait. Application layer packet classifier for linux, 2008.

Tai-Ching Li, Huy Hang, Michalis Faloutsos, and Petros Efstathopoulos. Trackadvisor: Taking back browsing privacy from third-party trackers. In *Passive and Active Measurement*, pages 277–289. Springer, 2015a.

Zhenhua Li, Christo Wilson, Tianyin Xu, Yao Liu, Zhen Lu, and Yinlong Wang. Offline downloading in china: A comparative study. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, pages 473–486. ACM, 2015b.

Marc Liberatore and Brian Neil Levine. Inferring the source of encrypted http connections. In *Proceedings of the 13th ACM conference on Computer and communications security*, pages 255–263. ACM, 2006.

Peter Likarish, Eunjin EJ Jung, and Insoon Jo. Obfuscated malicious javascript detection using classification techniques. In *Malicious and Unwanted Software (MALWARE), 2009 4th International Conference on*, pages 47–54. IEEE, 2009.

Yeon-sup Lim, Hyun-chul Kim, Jiwoong Jeong, Chong-kwon Kim, Ted Taekyoung Kwon, and Yanghee Choi. Internet traffic classification demystified: on the sources of the discriminative power. In *Proceedings of the 6th International COnference*, page 9. ACM, 2010.

Tianyi Lin, Wentao Tian, Qiaozhu Mei, and Hong Cheng. The dual-sparse topic model: mining focused topics and focused terms in short text. In *Proceedings of the 23rd international conference on World wide web*, pages 539–550. ACM, 2014.

Chao Liu, Ryen W White, and Susan Dumais. Understanding web browsing behaviors through weibull analysis of dwell time. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 379–386. ACM, 2010.

Gabriel Maciá-Fernández, Yong Wang, Rafael Rodríguez-Gómez, and Aleksandar Kuzmanovic. Isp-enabled behavioral ad targeting without deep packet inspection. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

Bruce A Mah. An empirical model of http network traffic. In *INFOCOM'97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution., Proceedings IEEE*, volume 2, pages 592–600. IEEE, 1997.

Anthony McGregor, Mark Hall, Perry Lorier, and James Brunskill. Flow clustering using machine learning techniques. In *Passive and Active Network Measurement*, pages 205–214. Springer, 2004.

Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. Crowd-assisted search for price discrimination in e-commerce: First results. In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pages 1–6. ACM, 2013.

Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10):1615–1630, 2005.

Brad Miller, Ling Huang, Anthony D Joseph, and J Doug Tygar. I know why you went to the clinic: Risks and realization of https traffic analysis. In *Privacy Enhancing Technologies*, pages 143–163. Springer, 2014.

Chris Milling, Constantine Caramanis, Shie Mannor, and Sanjay Shakkottai. Network forensics: random infection vs spreading epidemic. *ACM SIGMETRICS Performance Evaluation Review*, 40(1):223–234, 2012.

Greg Minshall. Tcpdpriv command manual, 1996.

Andrew W Moore and Konstantina Papagiannaki. Toward the accurate identification of network applications. In *Passive and Active Network Measurement*, pages 41–54. Springer, 2005.

Andrew W Moore and Denis Zuev. Internet traffic classification using bayesian analysis techniques. In *ACM SIGMETRICS Performance Evaluation Review*, pages 50–60. ACM, 2005.

David Moore, Ken Keys, Ryan Koga, Edouard Lagache, and Kimberly C Claffy. The coralreef software suite as a tool for system and network administrators. In *Proceedings of the 15th USENIX conference on System administration*, pages 133–144. USENIX Association, 2001.

Mozilla. Firebug: Web development evolved. http://getfirebug.com. Accessed: 2014-03-13.

Marc Najork and Janet L Wiener. Breadth-first crawling yields high-quality pages. In *Proceedings of the 10th international conference on World Wide Web*, pages 114–118. ACM, 2001.

Christopher Neasbitt, Roberto Perdisci, Kang Li, and Terry Nelms. Clickminer: Towards forensic reconstruction of user-browser interactions from network traces. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 1244–1255. ACM, 2014.

Christopher Neasbitt, Bo Li, Roberto Perdisci, Long Lu, Kapil Singh, and Kang Li. Webcapsule: Towards a lightweight forensic engine for web browsers. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 133–145. ACM, 2015.

Terry Nelms, Roberto Perdisci, Manos Antonakakis, and Mustaque Ahamad. Webwitness: investigating, categorizing, and mitigating malware download paths. In *24th USENIX Security Symposium (USENIX Security 15)*, pages 1025–1040, 2015.

B Newton, K Jeffay, and J Aikat. The continued evolution of the web. In *Modeling, Analysis and Simulation of Computer Telecommunications Systems, 2013. MASCOTS 2013. 11th IEEE/ACM International Symposium on*. IEEE, 2013.

Henrik Frystyk Nielsen, James Gettys, Anselm Baird-Smith, Eric Prud'hommeaux, Håkon Wium Lie, and Chris Lilley. Network performance effects of http/1.1, css1, and png. In *ACM SIGCOMM Computer Communication Review*, pages 155–166. ACM, 1997.

NoLock. Android apps on google play - nolock. https://play.google.com/store/apps/. Accessed: 2015-05-06.

K Nose-Filho, ADP Lotufo, and CR Minussi. Preprocessing data for short-term load forecasting with a general regression neural network and a moving average filter. In *PowerTech, 2011 IEEE Trondheim*, pages 1–7. IEEE, 2011.

Greg R Notess. The wayback machine: The web's archive. *ONLINE-WESTON THEN WILTON-*, 26(2): 59–61, 2002.

George Nychis, Vyas Sekar, David G Andersen, Hyong Kim, and Hui Zhang. An empirical evaluation of entropy-based traffic anomaly detection. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 151–156. ACM, 2008.

P. Ohm, D. Sicker, and D. Grunwald. Legal issues surrounding monitoring during network research (invited paper). In *Proc. ACM IMC*, 2007.

Lawrence Page and Sergey Brin. Pagerank, an eigenvector based ranking approach for hypertext. In *21st Annual ACM/SIGIR International Conference on Research and Development in Information Retrieval*, 1998.

Andriy Panchenko, Lukas Niessen, Andreas Zinnen, and Thomas Engel. Website fingerprinting in onion routing based anonymization networks. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 103–114. ACM, 2011.

Ruoming Pang and Vern Paxson. A high-level programming environment for packet trace anonymization and transformation. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 339–351. ACM, 2003.

Vern Paxson. Bro: a system for detecting network intruders in real-time. *Computer networks*, 31(23): 2435–2463, 1999.

Vern Paxson, Mihai Christodorescu, Mobin Javed, Josyula Rao, Reiner Sailer, Douglas Lee Schales, Mark Stoecklin, Kurt Thomas, Wietse Venema, and Nicholas Weaver. Practical comprehensive bounds on surreptitious communication over dns. In *Presented as part of the 22nd USENIX Security Symposium*, pages 17–32, Berkeley, CA, 2013. USENIX. ISBN 978-1-931971-03-4. URL `https://www.usenix.org/conference/usenixsecurity13/technical-sessions/papers/paxson`.

pcap2har. Pcap web performance analyzer. http://pcapperf.appspot.com. Accessed: 2014-03-13.

Maria Soledad Pera, Rani Qumsiyeh, and Yiu-Kai Ng. An unsupervised sentiment classifier on summarized or full reviews. In *Web Information Systems Engineering–WISE 2010*, pages 142–156. Springer, 2010.

L. Popa, A. Ghodsi, and I. Stoica. Http as the narrow waist of the future internet. In *Proc. 9th ACM Workshop on Hot Topics in Networks (Hotnets-IX)*, Oct 2010.

Xiaoguang Qi and Brian D Davison. Web page classification: Features and algorithms. *ACM Computing Surveys (CSUR)*, 41(2):12, 2009.

Lawrence Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Ashwin Rao, Arnaud Legout, Yeon-sup Lim, Don Towsley, Chadi Barakat, and Walid Dabbous. Network characteristics of video streaming traffic. In *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*, page 25. ACM, 2011.

Andrew Reed and Jay Aikat. Modeling, identifying, and simulating dynamic adaptive streaming over http. In *Network Protocols (ICNP), 2013 21st IEEE International Conference on*, pages 1–2. IEEE, 2013.

The Register. Mandatory http 2.0 encryption proposal sparks hot debate. http://www.theregister.co.uk/. Accessed: 2014-05-04.

Scorecard Research. Scorecard research. http://www.scorecardresearch.com/Preferences.aspx. Accessed: 2016-01-15.

Leonard Richardson. Beautiful soup documentation, 2015.

Haakon Ringberg, Augustin Soule, Jennifer Rexford, and Christophe Diot. Sensitivity of pca for traffic anomaly detection. In *ACM SIGMETRICS Performance Evaluation Review*, pages 109–120. ACM, 2007.

Martin Roesch et al. Snort: Lightweight intrusion detection for networks. In *LISA*, pages 229–238, 1999.

Lior Rokach and Oded Maimon. *Data mining with decision trees: theory and applications*. World scientific, 2014.

Matthew Roughan, Subhabrata Sen, Oliver Spatscheck, and Nick Duffield. Class-of-service mapping for qos: a statistical signature-based approach to ip traffic classification. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 135–148. ACM, 2004.

Shruti Sanadhya, Raghupathy Sivakumar, Kyu-Han Kim, Paul Congdon, Sriram Lakshmanan, and Jatinder Pal Singh. Asymmetric caching: improved network deduplication for mobile devices. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 161–172. ACM, 2012.

Sean Sanders and Jasleen Kaur. On the variation in webpage download traffic across different client types, May 2014a. URL `https://www.dropbox.com/s/hll5x9lx0vqf1ky/Browsers.pdf`.

Sean Sanders and Jasleen Kaur. On the variation in web page download traffic across different client types. In *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, pages 495–497. IEEE, 2014b.

Sean Sanders and Jasleen Kaur. The influence of client platform on web page content: Measurements, analysis, and implications. In *Proceedings of 16th International Conference on Web Information System Engineering*. IEEE, 2015a.

Sean Sanders and Jasleen Kaur. Can web pages be classified using anonymized tcp/ip headers? In *(to appear in) Proceedings of IEEE INFOCOM*. IEEE, 2015b.

Sean Sanders and Jasleen Kaur. Webpage boundary detection and classification using anonymized tcp/ip headers, 2015c. IRB Study #13-4037: University of North Carolina at Chapel Hill.

Sandvine. Global internet phenomena report. https://www.sandvine.com/downloads/general/global-internet-phenomena/2013/2h-2013-global-internet-phenomena-report.pdf. Accessed: 2014-05-04.

Dominik Schatzmann, Wolfgang Mühlbauer, Thrasyvoulos Spyropoulos, and Xenofontas Dimitropoulos. Digging into https: flow-based classification of webmail traffic. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 322–327. ACM, 2010.

Fabian Schneider, Sachin Agarwal, Tansu Alpcan, and Anja Feldmann. The new web: Characterizing ajax traffic. In *Passive and Active Network Measurement*, pages 31–40. Springer, 2008.

Bruce Schneier. Attacking tor: how the nsa targets users online anonymity. *The Guardian*, 4, 2013.

Jakob Schroter. Client-side performance optimizations. http://www.slideshare.net/jakob.schroeter/clientside-web-performance-optimization, 2011. Accessed: 2014-03-13.

Christian Seifert, Ian Welch, and Peter Komisarczuk. Identification of malicious web pages with static heuristics. In *Telecommunication Networks and Applications Conference, 2008. ATNAC 2008. Australasian*, pages 91–96. IEEE, 2008.

SeleniumHQ. Seleniumhq browser automation. http://www.seleniumhq.org/. Accessed: 2016-04-02.

Subhabrata Sen and Jia Wang. Analyzing peer-to-peer traffic across large networks. *IEEE/ACM Transactions on Networking (ToN)*, 12(2):219–232, 2004.

Subhabrata Sen, Oliver Spatscheck, and Dongmei Wang. Accurate, scalable in-network identification of p2p traffic using application signatures. In *Proceedings of the 13th international conference on World Wide Web*, pages 512–521. ACM, 2004.

Muhammad Zubair Shafiq, Lusheng Ji, Alex X Liu, Jeffrey Pang, and Jia Wang. A first look at cellular machine-to-machine traffic: large scale measurement and characterization. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 65–76. ACM, 2012.

Chaofan Shen and Leijun Huang. On detection accuracy of l7-filter and opendpi. In *Networking and Distributed Computing (ICNDC), 2012 Third International Conference on*, pages 119–123. IEEE, 2012.

Dou Shen, Zheng Chen, Qiang Yang, Hua-Jun Zeng, Benyu Zhang, Yuchang Lu, and Wei-Ying Ma. Webpage classification through summarization. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 242–249. ACM, 2004.

Douglas C Sicker, Paul Ohm, and Dirk Grunwald. Legal issues surrounding monitoring during network research. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 141–148. ACM, 2007.

Fernando Silveira, Christophe Diot, Nina Taft, and Ramesh Govindan. Astute: Detecting a different class of traffic anomalies. In *ACM SIGCOMM Computer Communication Review*, pages 267–278. ACM, 2010.

F Donelson Smith, Félix Hernández Campos, Kevin Jeffay, and David Ott. What tcp/ip protocol headers can tell us about the web. In *ACM SIGMETRICS Performance Evaluation Review*, pages 245–256. ACM, 2001.

John R Smith and Shih-Fu Chang. Visually searching the web for content. *IEEE multimedia*, pages 12–20, 1997.

Steve Souders. I'm now at speedcurve. http://www.stevesouders.com/blog/2013/11/07/prebrowsing/. Accessed: 2016-02-08.

Augustin Soule, Kavé Salamatian, and Nina Taft. Combining filtering and statistical methods for anomaly detection. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, pages 31–31. USENIX Association, 2005.

Ellen Spertus. Parasite: Mining structural information on the web. *Computer Networks and ISDN Systems*, 29(8):1205–1215, 1997.

Myra Spiliopoulou, Bamshad Mobasher, Bettina Berendt, and Miki Nakagawa. A framework for the evaluation of session reconstruction heuristics in web-usage analysis. *Informs journal on computing*, 15(2): 171–190, 2003.

William Stallings. *Cryptography and network security: principles and practices*. Pearson Education India, 2006.

Statista. Digital advertising spending worldwide from 2012 to 2018 (in billion u.s. dollars). http://www.statista.com/statistics/237974/online-advertising-spending-worldwide/. Accessed: 2016-07-10.

StayAwake. Android apps on google play - stayawake. https://play.google.com/store/apps/. Accessed: 2015-05-06.

Qixiang Sun, Daniel R Simon, Yi-Min Wang, Wilf Russell, Venkata N Padmanabhan, and Lili Qiu. Statistical identification of encrypted web browsing traffic. In *Security and Privacy, 2002. Proceedings. 2002 IEEE Symposium on*, pages 19–30. IEEE, 2002.

Florian Tegeler, Xiaoming Fu, Giovanni Vigna, and Christopher Kruegel. Botfinder: Finding bots in network traffic without deep packet inspection. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pages 349–360. ACM, 2012.

Robert Tibshirani and Pei Wang. Spatial smoothing and hot spot detection for cgh data using the fused lasso. *Biostatistics*, 9(1):18–29, 2008.

Joe Touch, M Kojo, E Lear, A Mankin, K Ono, M Stiemerling, and L Eggert. Service name and transport protocol port number registry. *The Internet Assigned Numbers Authority (IANA)*, 2013.

Paul F Tsuchiya and Tony Eng. Extending the ip internet through address reuse. *ACM SIGCOMM Computer Communication Review*, 23(1):16–33, 1993.

Víctor Uceda, Miguel Rodríguez, Javier Ramos, José Luis García-Dorado, and Javier Aracil. Selective capping of packet payloads for network analysis and management. In *Traffic Monitoring and Analysis*, pages 3–16. Springer, 2015.

Luis Vieira. Html5 prefetch: Predict users actions and optimistically load resources ahead of time for better performance. https://medium.com/luisvieira_gmr/ html5-prefetch-1e54f6dda15d#ẋ751ed3bd. Accessed: 2016-02-08.

W3Techs. Usage of http/2 for websites. http://w3techs.com/technologies/details/ce-http2/all/all. Accessed: 2015-12-6.

Gang Wang, Tristan Konolige, Christo Wilson, Xiao Wang, Haitao Zheng, and Ben Y. Zhao. You are how you click: Clickstream analysis for sybil detection. In *Presented as part of the 22nd USENIX Security Symposium*, pages 241–256, Berkeley, CA, 2013. USENIX. ISBN 978-1-931971-03-4. URL `https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/wang`.

Xiao Sophia Wang, Aruna Balasubramanian, Arvind Krishnamurthy, and David Wetherall. How speedy is spdy. In *Proc. of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 387–399, 2014.

Yi-Min Wang, Doug Beck, Xuxian Jiang, Roussi Roussev, Chad Verbowski, Shuo Chen, and Sam King. Automated web patrol with strider honeymonkeys. In *Proceedings of the 2006 Network and Distributed System Security Symposium*, pages 35–49, 2006.

HTTP Watch. Httpwatch. http://www.httpwatch.com. Accessed: 2014-03-13.

Michele C Weigle, Prashanth Adurthi, Félix Hernández-Campos, Kevin Jeffay, and F Donelson Smith. Tmix: a tool for generating realistic tcp application workloads in ns-2. *ACM SIGCOMM Computer Communication Review*, 36(3):65–76, 2006.

Kilian Weinberger, Anirban Dasgupta, John Langford, Alex Smola, and Josh Attenberg. Feature hashing for large scale multitask learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1113–1120. ACM, 2009.

Andrew M White, Austin R Matthews, Kevin Z Snow, and Fabian Monrose. Phonotactic reconstruction of encrypted voip conversations: Hookt on fon-iks. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 3–18. IEEE, 2011.

Andrew M White, Srinivas Krishnan, Michael Bailey, Fabian Monrose, and Phillip Porras. Clear and present data: Opaque traffic and its security implications for the future. *NDSS. The Internet Society*, pages 24096–1, 2013.

Walter Willinger, Murad S Taqqu, Robert Sherman, and Daniel V Wilson. Self-similarity through high-variability: statistical analysis of ethernet lan traffic at the source level. *Networking, IEEE/ACM Transactions on*, 5(1):71–86, 1997.

Bin Wu and Ajay D Kshemkalyani. Objective-greedy algorithms for long-term web prefetching. In *Network Computing and Applications, 2004.(NCA 2004). Proceedings. Third IEEE International Symposium on*, pages 61–68. IEEE, 2004.

Guowu Xie, Marios Iliofotou, Ram Keralapura, Michalis Faloutsos, and Antonio Nucci. Subflow: Towards practical flow-level traffic classification. In *INFOCOM, 2012 Proceedings IEEE*, pages 2541–2545. IEEE, 2012.

Guowu Xie, Marios Iliofotou, Thomas Karagiannis, Michalis Faloutsos, and Yaohui Jin. Resurf: Reconstructing web-surfing activity from network traffic. In *IFIP Networking Conference, 2013*, pages 1–9. IEEE, 2013.

Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. Identifying diverse usage behaviors of smartphone apps. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 329–344. ACM, 2011.

Yahoo. Yslow. https://developer.yahoo.com/yslow/. Accessed: 2014-03-13.

He Yan, Ashley Flavel, Zihui Ge, Alexandre Gerber, Dan Massey, Christos Papadopoulos, Hiren Shah, and Jennifer Yates. Argus: End-to-end service anomaly detection and localization from an isp's point of view. In *INFOCOM, 2012 Proceedings IEEE*, pages 2756–2760. IEEE, 2012.

Jun Yan, Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen. How much can behavioral targeting help online advertising? In *Proceedings of the 18th international conference on World wide web*, pages 261–270. ACM, 2009.

Ting-Fang Yen, Xin Huang, Fabian Monrose, and Michael K Reiter. Browser fingerprinting from coarse traffic summaries: Techniques and implications. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 157–175. Springer, 2009.

Yasir Zaki, Jay Chen, Thomas Pötsch, Talal Ahmad, and Lakshminarayanan Subramanian. Dissecting web latency in ghana. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 241–248. ACM, 2014.

Aonan Zhang, Jun Zhu, and Bo Zhang. Sparse online topic models. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1489–1500. International World Wide Web Conferences Steering Committee, 2013.

Renjie Zhou, Samamon Khemmarat, and Lixin Gao. The impact of youtube recommendation system on video views. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 404–410. ACM, 2010.