

A Snapshot of Desktop and Mobile Web Page Characteristics

Troy Johnson

Dept. of Computer Science
Central Michigan University
Mount Pleasant, MI 48859
Email: johns4ta@cmich.edu

Patrick Seeling

Dept. of Computer Science
Central Michigan University
Mount Pleasant, MI 48859
Email: pseeling@ieee.org

Abstract—In this paper, we evaluate the differences of the number of objects comprising and sizes of web pages as delivered to desktop and mobile web clients based on a large publicly available dataset of initial page views. We find that for web pages that are available in both versions, the mobile version exhibits approximately half the number of objects and final size, which interestingly results in a similar size per request. For both we find that a significant amount of objects is not to be cached, which will result in frequent downloads. For

Index Terms—Mobile communication, traffic modeling, data communications

I. INTRODUCTION

Smart mobile devices with ubiquitous connectivity have become popular worldwide. Cisco, Inc. predicts that this will result in steady growth of mobile data for years to come [1]. As mobile users employ their always-present devices to access web-based services, additional predictions include a shift of the main access of the world wide web from fixed desktop to mobile clients. This results in past studies focusing on desktop client web access, such as [2], not reflecting the current state of the web anymore.

To address these changes, more recent studies have found that significant amounts of data traffic can be saved for providers through caching [3]. Investigating the complexity of web pages, the number of loaded objects were found to positively correlate with the web page load times [4]. Furthermore, in their evaluation, the authors found that mobile and non-landing web pages tend to exhibit less complexity, devices.

In this paper, we evaluate the modeling possibilities for web pages as they were accessed with a desktop and mobile (iphone) client with respect to sizes, number of requests, and caching abilities. In the following section, we describe the underlying dataset and data gathering procedures. We then describe the high-level characteristics and approximations through distributions. We conclude with an outlook on future works.

II. WEB PAGE DATASET

We utilize the httparchive.org [5] datasets of captured web performance metrics, which are publicly available. Each of the datasets is generated from initial client views (eliminating

caching) of a broad set of web sites. While the project gathers the statistics of landing pages, in conjunction with the observations in [4], these can be seen as representative upper boundaries of the evaluated characteristics. We utilize the fixed and mobile datasets available for Oct. 1st, 2013. We pre-process the dataset to only contain web sites that are present in both sets and hence represent a comparable subset. Even with this pre-processing, we evaluate the characteristics of 4779 web sites.

III. WEB PAGE CHARACTERISTICS

We evaluate the overall characteristics for sizes and number of object requests, in addition to cache lifetime, before determining statistical distribution characteristics.

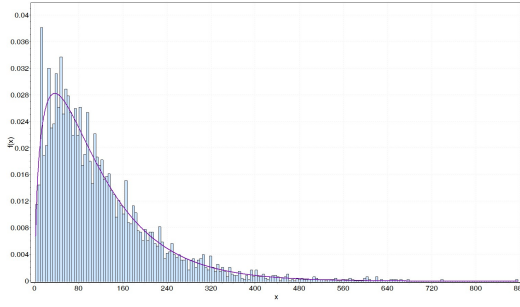
A. Characteristics

We denote the number of object requests for a web page as r and the related average as \bar{r} . Similarly, we evaluate the average size of fixed and mobile web pages as x . We additionally investigate the average request size as $\frac{x}{r}$. Furthermore, we evaluate the number of objects with a zero maximum age in cache (download every time) as c_0 and those with potential for caching as c by presenting their respective fractions. We indicate the variability amongst the 4779 URLs evaluated using the Coefficient of Variation (CoV). The results are provided in Table ?? for the fixed and mobile web versions.

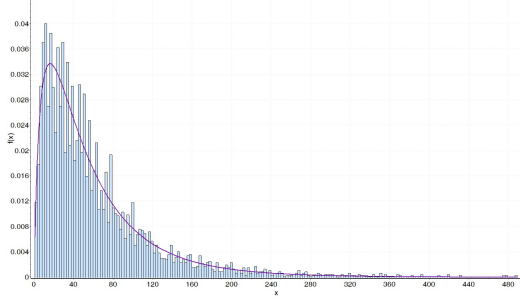
We observe that the average number of requests is approximately twice as high for desktop versions when compared to their mobile counterparts. Similarly, the average desktop web page size is almost twice as large in size as well. The resulting average request sizes are in turn almost the same, independent of the accessing client's type. We furthermore note that the

TABLE I
COMPARISON OF THE AVERAGE NUMBER OF REQUESTS r , PAGE SIZES x , AND MAX. CACHE AGES c FOR FIXED AND MOBILE WEB PAGE VERSIONS.

Metric	Desktop Client		Mobile Client	
	(\cdot)	CoV(\cdot)	(\cdot)	CoV(\cdot)
Numb. req. r	116.4	0.838	61	0.946
Page size x (MB)	1.56	1.263	0.85	1.245
Bytes/req. x/r (kB)	14.76	1.537	14.57	0.869
Zero cache fraction c_0	0.412	0.246	0.450	0.702
Cache fraction c	0.588	0.172	0.550	0.575

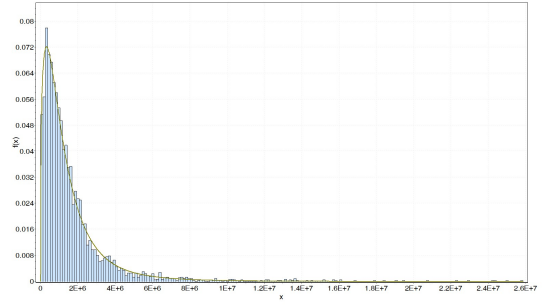


(a) Desktop client

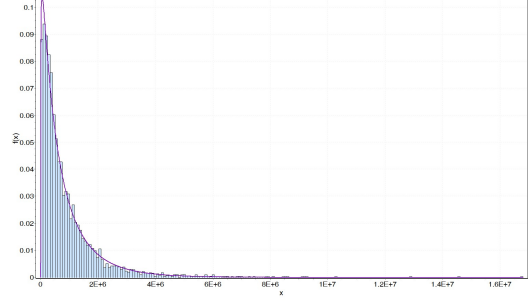


(b) Mobile client

Fig. 1. General Gamma distribution fitting for web page sizes (with 200 illustrated bins).



(a) Desktop client



(b) Mobile client

Fig. 2. General Gamma distribution fitting for web page sizes (with 200 illustrated bins).

variability within the dataset is fairly low and close between the two versions for the number of requests and the average page sizes. We find that the average request sizes are exhibiting a more significant variability for the desktop version. Shifting the view to the possibility of reducing network traffic through caching, we note that the fraction of items that have a max. age of zero as determined through the httparchive.org processing account for a significant portion of all data for both clients alike. These items in turn can be seen as the dynamically changing content requiring frequent refreshes and constitute just above 40% of the items accessed by the web clients.

B. Distribution Fitting

In the following, we evaluate the possibility to fit the individual web page statistics for the number of requests and web page sizes. We utilized the EasyFit distribution fitting software [6] to evaluate the underlying distributions. We find that for both, the General Gamma Distribution provides as reasonable fit of the discovered fundamental characteristics (while noting that other distributions might individually provide a similarly good fit). We illustrate the fitted distributions in Figures 1 and 2 for requests and sizes, respectively.

QQ plots (omitted due to page limits) indicate that the tails exhibit increasing deviations, but the lower ends with the majority of observations are fitted well.

IV. CONCLUSION

The increasing importance of web access using mobile devices triggers additional challenges for networks and investigating the underlying characteristics and differences between

fixed and mobile client web access can yield important insights for optimization and modeling of today's web traffic. We found that there is a strong similarity between the access modes for (i) sizes per request and (ii) the amount of objects that can be cached while the desktop versions exhibit (iii) approximately twice the size and (iv) twice the number of objects. In the future, we will investigate the potential for detailed modeling of the size and request distributions as well as the possibility for correlations between objects that are cacheable and sizes.

REFERENCES

- [1] Cisco, Inc., "Cisco visual networking index: Forecast and methodology, 2012–2017," Tech. Rep., may 2013. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf
- [2] P. Barford and M. Crovella, "Generating representative web workloads for network and server performance evaluation," *SIGMETRICS Perform. Eval. Rev.*, vol. 26, no. 1, pp. 151–160, Jun. 1998. [Online]. Available: <http://doi.acm.org/10.1145/277858.277897>
- [3] S. Ihm and V. S. Pai, "Towards understanding modern web traffic," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, ser. IMC '11. New York, NY, USA: ACM, 2011, pp. 295–312. [Online]. Available: <http://doi.acm.org/10.1145/2068816.2068845>
- [4] M. Butkiewicz, H. Madhyastha, and V. Sekar, "Characterizing web page complexity and its impact," *Networking, IEEE/ACM Transactions on*, vol. PP, no. 99, pp. 1–1, 2013.
- [5] httparchive.org, "Http archive (beta)," July 2013. [Online]. Available: <http://www.httparchive.org>
- [6] MathWave, "Easyfit," 2013.