Landing on the Mobile Web: From Browsing to Long–Term Modeling

Journal:	*IEEE Communications Magazine*
Manuscript ID:	Draft
Topic or Series:	Open Call Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Johnson, Troy; Central Michigan University, Dept. of Computer Science Seeling, Patrick; Central Michigan University, Dept. of Computer Science
Key Words:	World Wide Web, Web sites, Browsers, Mobile computing, Performance evaluation

Landing on the Mobile Web: From Browsing to Long–Term Modeling

Troy Johnson and Patrick Seeling

Abstract

Browsing the World Wide Web has become a common task performed using personal mobile devices, resulting in significant access network and battery limitation challenges. Efforts to alleviate these challenges are commonly based around approaches incorporating elements of on–device and network optimizations. Energy–efficient mobile web content delivery has, in turn, attracted a significant body of research and practical developments. However, the efforts put forth today might not result in long–term applicable results should the underlying characteristics of the mobile content change drastically over time.

As caching frequently used data locally is a common initial approach employed to limit network traffic and energy expenditures while "on the go," we evaluate the long–term suitability of approximating a basic set of parameters for a cache and request behavior model using a popular large dataset. We present a convenient approach that can employ a general approximation of parameters over time with tolerable impacts. For more than one year's changes in the underlying data set, our model incurs only a maximum 13 % penalty in accuracy when performing long–term prediction of parameter trends. In turn, practitioners and researchers are enabled to readily employ modeling and simulation approaches over a significant time horizon with only slight impacts on their approaches and results.

Index Terms

World Wide Web; Web sites; Browsers; Mobile computing; Performance evaluation

I. Introduction

Mobile users oftentimes need to quickly access information that is provided using the World Wide Web (WWW), rather than in an application–wrapped context. With WWW access predictions indicating that browsing will be predominantly performed by users employing their personal mobile devices, a significant portion of future network traffic will be delivered over the wireless air interfaces of non–stationary personal hand–held devices [1]. Other mobile user habits and a (mobile) web–first strategy for

T. Johnson and P. Seeling are with the Dept. of Computer Science, Central Michigan University, Mount Pleasant, MI 48859, Email: johns4ta@cmich.edu, pseeling@ieee.org

Please direct correspondence to P. Seeling.

new service developments might additionally contribute to the frequency and amounts of data that mobile users request from service landing pages. Additionally, the increased pressure to create applications that can be spread across mobile platforms through the use of web–based technologies, such as HTML5 and JavaScript, increases the likelihood for future applications employing a hybrid strategy of pre–packaged content that is supplemented by requesting specific landing web pages from the WWW. Similarly, if new web–centric services emerge (or new information about an existing service is sought by users, e.g., by sharing information about them in social contexts), mobile users might be reluctant to engage into the process of installing an application to their device and rather try out the web–based version of the service.

Based on this emerging trajectory, we initially presented a comparative overview of the landing web page characteristics differences between desktop and mobile versions in [2], which complements observations made in [3]. As web pages are constituted by a typically large range of individual objects that need to be retrieved, the authors of, e.g., [4], demonstrate the possibility for energy–optimized mobile WWW browsing. Circumventing the need to download individual web objects, caching performed by (mobile) browsers commonly can be considered a first optimization step for improvement as outlined in, e.g., [5]. While some limits exist for individual devices, multi–level caching approaches could yield additional benefits, as recently described in [6]. An interesting approach discussed therein is the notion of decentralized caching, an intuitive example is to exchange content between different device types belonging to the same user to save energy (from user and content provider perspectives) by essentially generating a locally shared cache. The differences as well as similarities between the different versions of web pages inherently offer possibilities for optimization efforts. Information about cache lifetime expirations of web objects in general could be exploited to locally forward content in an opportune fashion; the potential for this approach was evaluated in, e.g., [7] using a popular user browsing behavior model.

The integration of social networking concepts when considering today's mobile shareconomy trajectories [8] increases the importance of providing WWW–based content to mobile users and was outlined in, e.g., [9]. Here, the authors exploit the significant improvements in mobile resources as well as social and other contexts to pro–actively deliver content off–peak, thus reducing the "network crunch."

To facilitate the inclusion of the tendencies observed for the web page objects in research efforts that strive to optimize the energy–efficient mobile web content provisioning, modeling of the underlying content characteristics is typically desirable for a practically suitable evaluation. Employing the large *httparchive.org* data set [10], a model that is able to capture the main facets of the cache behavior a mobile user's browser would encounter over the course of an unconnected time of up to a week was presented in, e.g., [11]. However, long–term trends could have a significant impact on the suitability of efforts made today that employ characteristics of actual content found on the WWW. Given the overarching trends that were presented in [2], here we provide an evaluation of an accessible long–term approach to modeling the mobile landing web–page characteristics with respect to cache implications and resulting network requests.

The remainder of this article is structured as follows. In the subsequent section, we provide an overview of the typical mobile device web cache's content characteristics based on prior works. We continue by describing a convenient methodology for simulating a mobile browser's means of accessing the WWW content represented by web landing pages. In Section III, we evaluate the extrapolation of past observations for the simulation of mobile device web access characteristics with caching, specifically considering long–term trends and implications. We conclude in Section IV.

II. A Mobile Browser's Cache

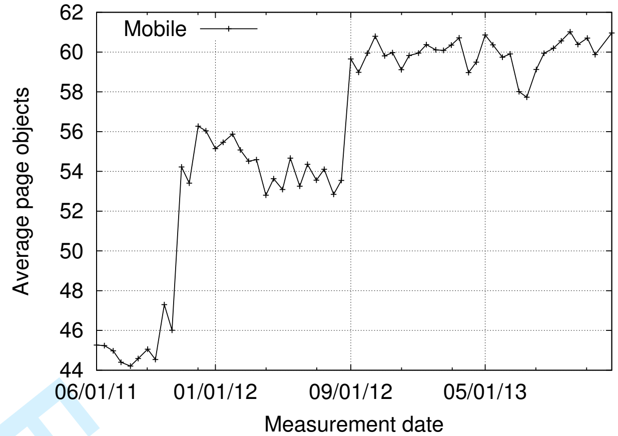
Initially, we consider the contents and the overall simulation of mobile browsing and subsequent cache utilization in this section.

A. Cache Contents

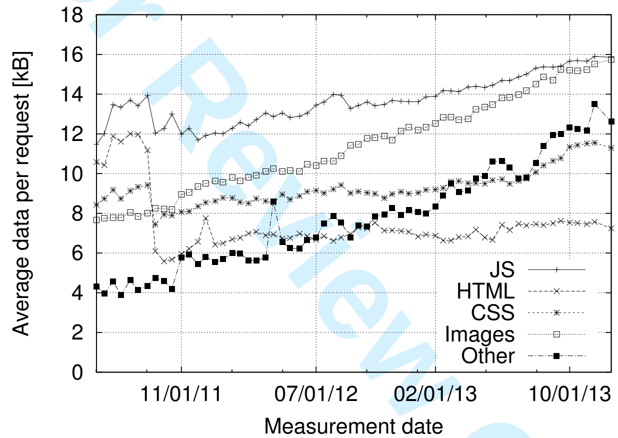
The contents that comprise a mobile browser's assumed cache can be evaluated by considering its composition first. To this extent, the common web landing pages of popular web sites were found to provide an adequate upper limit for complexity considerations. We provide the cache contents as identified in [2] for these landing pages here in Figure 1. We initially note that the number of objects per mobile web page increases at several time instances, commonly coinciding with an update of the underlying *httparchive* data set. Within each updated sequence, however, the number is overall at a stable level, which reaches around 60 objects in October 2013.

For the distribution of the object sizes by type that comprise a web page, we note in Figure 1b that a general upwards trend exists for most object types; they are more independent from the time period than the number of objects in comparison. Overall, HTML content exhibits a steady trend, while all other types of objects commonly embedded feature a steady increase in relative size. Combined with the number of objects per page, this results in a continuous increase of mobile web landing page sizes.

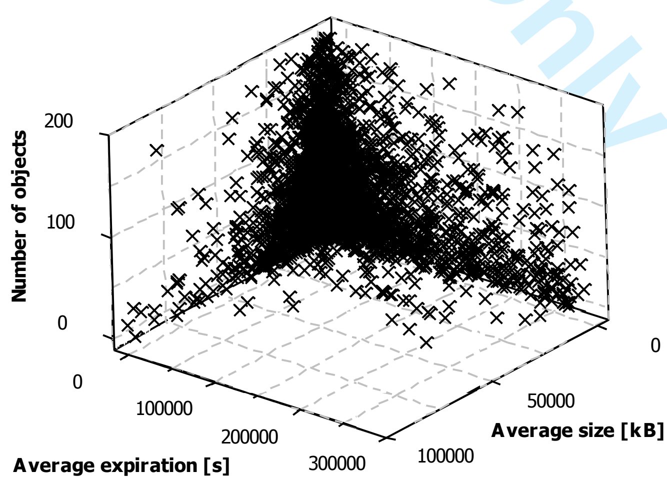
A mobile browser requesting these pages would in turn be required to download a significant amount of this data were general cache expiration times low. Indeed, when considering the categorized object expirations illustrated in Figure 1c for October 15, 2013, we note that a significant portion of objects are immediately expiring or very short lived. Further discussed in [2], these results are fairly stable over time, which makes the distribution of expiration ages and related objects another rather stable component to consider. Thus, a mobile browser's cache would have to expire a significant portion of data, resulting in required re–downloads (with subsequent impacts on air interface use and resulting battery utilization).



(a) Average number of objects per page



(b) Object types and sizes



(c) Object expirations for Oct. 15, 2013

Fig. 1. Web object numbers, sizes by types, and expirations over time encountered for the *httparchive.org* data set, see [2], [11] for more details.

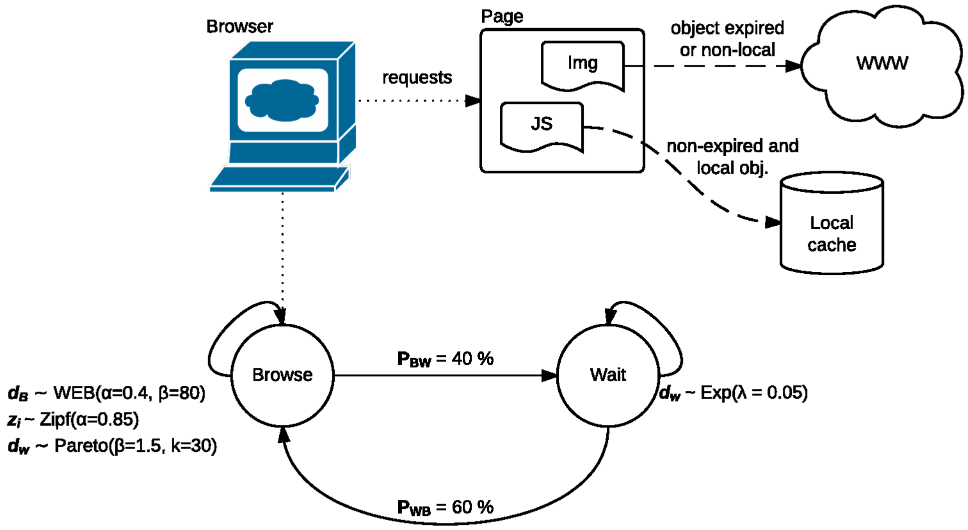


Fig. 2. Overview of the assumed user browsing behavior and content retrieval from local cache or the Internet.

Overall, these trends motivate us to consider a more stable distribution of the page compositions of objects and their expirations, with a focus on size trends over time, as discussed now in greater detail.

B. *Simulating a Mobile Browser's Cache*

We developed a model that captures the main facets of the web object behavior that a mobile user's browser would encounter by separating the sizes and expiration age limitations for web page objects, as little correlation is exhibited amongst them, see [11]. We note that we consider only objects with cache expirations below one week, as those above can be considered static and pre–provisioned in the context of mobile web browsing. Employing this approach, synthetic web pages are generated in a multi–step process. Initially, a page is generated to be of either a current, short–lived, medium–lived, or regular type. Based on this selection, the number of objects for the page drawn randomly, followed by the sizes for the objects. Subsequently, the object expirations are randomly distributed over the objects based on the synthetic page type they fall under. Finally, for a set of generated web pages, a unique popularity index is randomly assigned to each page (as little correlation was found for page compositions and their popularity ranking in the underlying data set). 500 sets of pages are generated to create data sets to be used in a browsing simulation that takes user behavior into account.

We consider the model presented in [12] as a baseline and employ a modified version as described in [13], maintaining the main facets of user behavior modeling. We illustrate the overall model used for simulating a mobile browsing experience in Figure 2.

While the user is in a browsing session, assumed to last a specific amount of time that is Weibull distributed with $d_B \sim WEB(\alpha = 0.4, \beta = 80)$, requests to (different) synthetically generated web pages

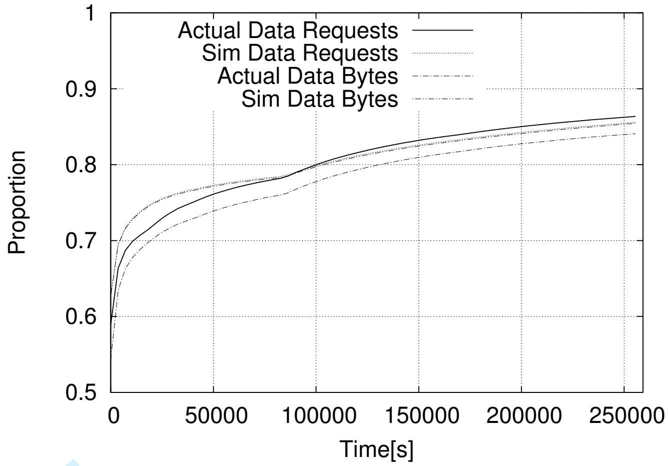


Fig. 3. Overview of the simulated mobile browser required external requests based on expired cache contents employing the underlying original *httparchive* and synthetically generated mobile web landing pages.

follow the popular Zipf distribution for content popularity $z_i \sim Zipf(\alpha = 0.85)$. The mobile user is assumed to continue browsing after a Pareto–distributed waiting time in seconds, see, e.g., [14], for viewing the page (commonly also referred to as user think time, UTT) has passed, $d_w \sim Pareto(\beta = 1.5, k = 30)$. Alternatively, the user could end the currently active browsing session with P_{BW} of 40 %, in which case the overall system enters a waiting state for a period of time. The duration of the waiting state is exponentially distributed with $d_w \sim Exp(\lambda = 0.05)$. We refer the interested reader to [13] for a more thorough description of the simulation of the employed mobile user's web browsing behavior.

C. Initial Simulations of Cache Benefits

We initially employ our simulation model for the approximation of the October 15, 2013 data set's characteristics, starting with an assumed *hot* cache (i.e., the cache is assumed to be fully populated at the beginning of the simulation). Here, we evaluate the overall suitability to determine the relative amounts of objects and associated amounts of data that would require a download instead of being serviceable from the mobile device browser's local cache. We perform 500 synthetic web page data set generations and simulate the mobile browsing over each of these sets 500 times for a duration of three days. We present the resulting graphic comparison in Figure 3, whereby we omit confidence intervals for readability purposes (noting that they typically are within 5 % of the presented averages).

With a significant portion of the objects expiring instantly or almost instantly, we initially note that even for just a very short time offset, a significant amount of data would require downloading. Next, we observe that there is a "kink" in the relative amounts of data requests around a day's border, coinciding with a common web object expiration value on the shorter time scale observable from Figure 1. We additionally note that our model follows this trend, but slightly overestimates the download requirements before the time boundary of a day, while for the remaining time, the model slightly underestimates the number of objects to require a download. While the number of bytes resulting from the actual data also exhibits a “kink,” employing the simulated web pages does result in a slightly increased margin of estimated downloads below a day’s threshold, which narrows as the simulated time increases. Overall, however, we note that our model is within approximately five percent of the actual data simulation

III. Approximating the Cache’s Content over Time

In this section, we consider an approximation of the long–term trends exhibited in prior works and combine those with the simulations of page–level popularities, number of objects, and their characteristics (sizes and expirations).

A. Parameter Estimations for Long Horizons

Following the simulation approach outlined in Section II-C, the entire set of object sizes (i.e., individual web objects not aggregated based on the page they belong to) is approximated using a Weibull distribution before the objects are aggregated to determine the synthetically generated page level characteristics. The number of responses per page, in turn, is approximated by employing a gamma distribution with fixed parameters for each month, effectively plateauing the current level of page complexities for the purpose of our approach; this reflects the general trend observed within the underlying data sets.

We anchor our evaluation around the October 15, 2013 data set as a base to extrapolate the long–term estimation of the object sizes only, as their overarching characteristics have little correlations. Continuing the trend of a bi–weekly basis as provided from the underlying *httparchive.com* data sets, we illustrate the resulting shape and scale parameters for the Weibull distribution employed to generate synthetic object sizes in Figure 4.

We observe a somewhat jittery behavior for shape and scale parameters over time. However, the deviations are all within fairly narrow confines, allowing us to perform a basic approximation of their trends over time. Implementing a basic linear fitting to this set of parameter variability, we note an increase of approximately 1.4 % each half–month for the scale parameter. For the shape parameter, we observe a slower increase of 0.05 % each half–month. In the next steps, we now consider employing these parameter estimates within our model as a long–term trend for the synthetically generated object sizes reaching beyond the time horizon of the parameter estimations.

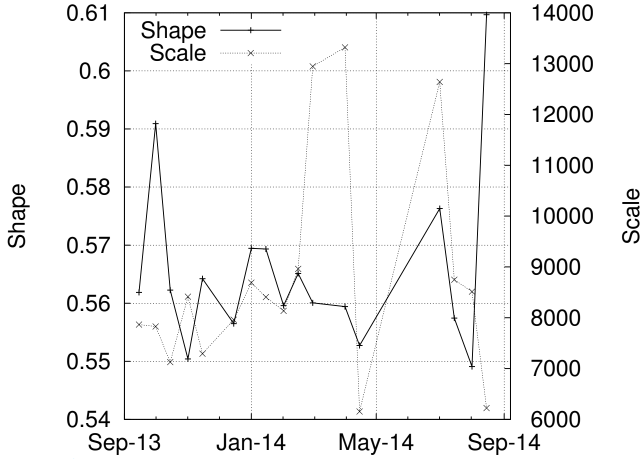


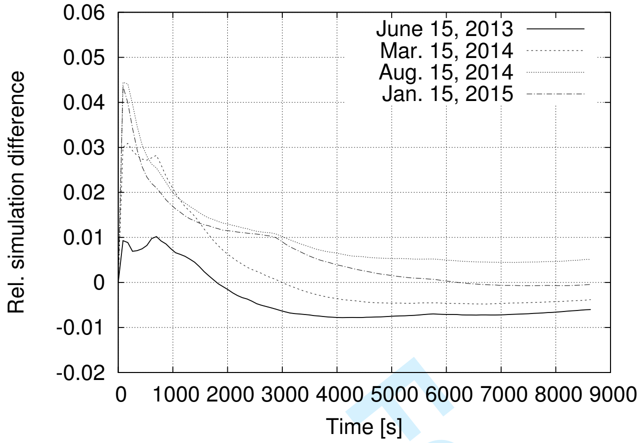
Fig. 4. Estimation of shape and scale parameters for the Weibull distribution used to approximate the overall web object sizes of the data set.

B. Impact on Long Horizons

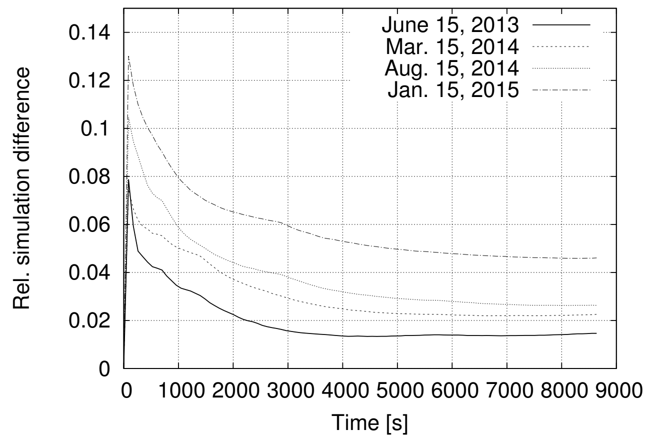
Initially, we commence the simulation with a *hot* cache, i.e., we assume all objects are locally available. We subsequently continue with an evaluation period of up to three days, for which we generate 500 sets of artificially composed web pages. The number of individual web pages in each of the generated sets was chosen to match the numbers in the original source data for the individual months (ranging from 4803 pages for June 15, 2013 to 4713 pages for January 15, 2015).

For each individual web page, the characteristics of page compositions out of objects and object attributes are generated according to our previous modeling, described in greater detail in [11], with parameters of individual distributions replaced by their long–time approximations. Since no strong correlation was found to exist between page rank (popularity) and other characteristics, we randomize the popularity of synthetically generated pages. To account for the impact that the popularity variability can have on the simulated user browsing behavior, each set of artificially generated web pages (500 sets) were employed in 500 simulations (for a combined total of 250,000 simulations at each artificially generated monthly data point). The results were captured in terms of bins with sizes of 30 seconds, 300 seconds, and 3600 seconds for processing. To determine the impact that the approximation of values has on the quality of the simulation, we determine the difference between the simulations employing the approximated parameters and the actual *httparchive* data set, illustrated in Figure 5a. We note that for readability purposes, the results for the 30 second bin sizes were smoothed after generation of the data. We additionally note that for readability purposes, we do not illustrate the upper and lower confidence intervals, as they are very narrow (typically within 5 % of the averages presented).

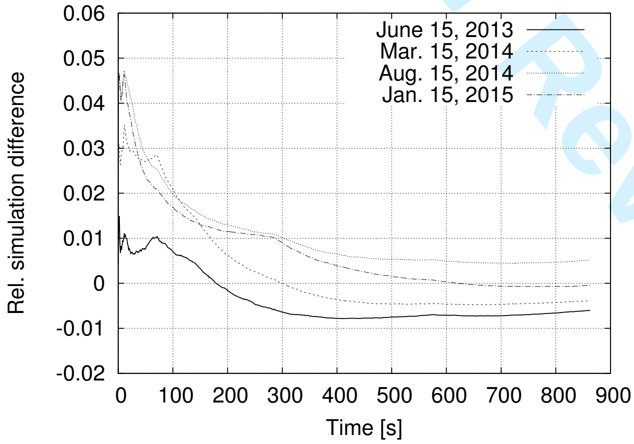
For the impact the approximation has on object requests, we initially evaluate the fine–grained 30



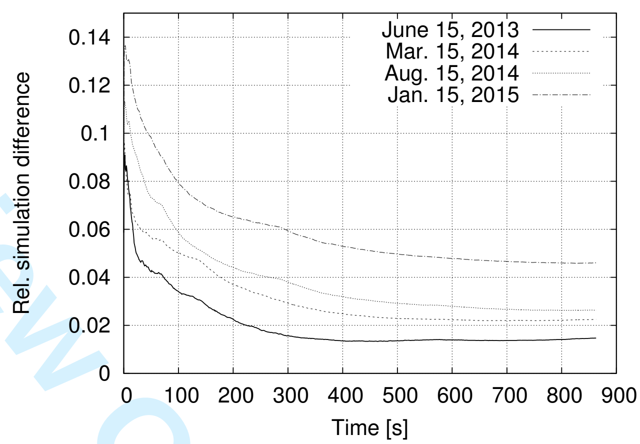
(a) 30 s bins, requests



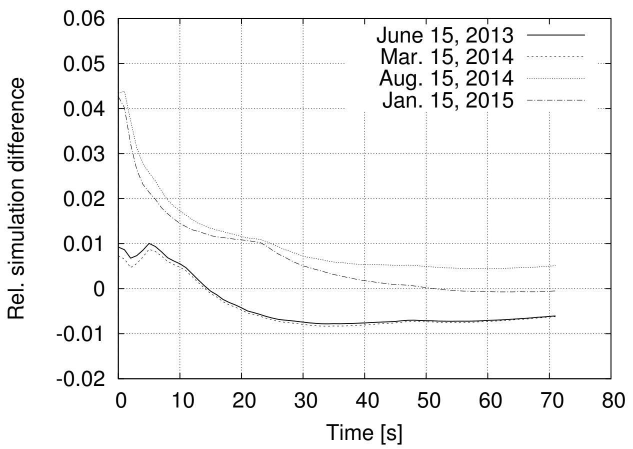
(b) 30 s bins, data



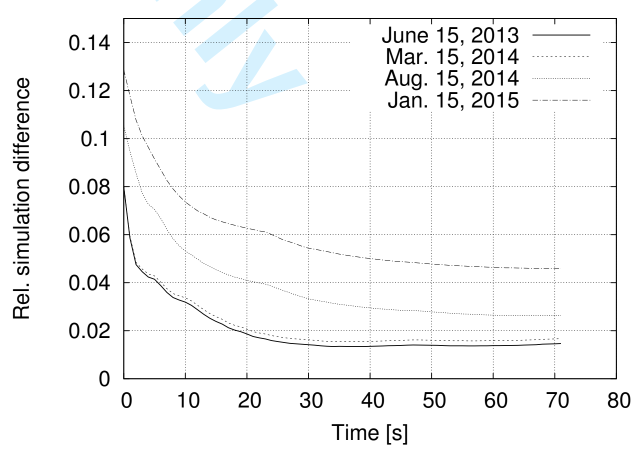
(c) 300 s bins, requests



(d) 300 s bins, data



(e) 3600 s bins, requests



(f) 3600 s bins, data

Fig. 5. Differences between the simulations employing the synthetically generated data sets with linearly approximated parameters for long–term evaluations and the underlying *httparchive* data set.

second bin results. We observe an initial paid increase in the difference, which is followed by a plateau or more pronounced peak, and succeeded by a slower decrease over simulated time observing the cache behavior. Comparing the different snapshots in time for which the comparison was made, we note that the further the simulation using approximations moves away from the anchor point, the higher the impact of approximating the parameters for the web page composition. Overall, however, the level of difference is very narrow at below 5 percent for the highest difference.

Next, we shift our view to the approximation that includes the sizes at fine granularity in Figure 5b. We observe an immediate rise of the differences between the two approaches, followed by a distinct peak, which is trailed by an asymptotic decay. Upon closer inspection of the different time distances from the October 15, 2013 data set used as base, we note that here, a more distinct order can be found than for the number of requests alone. Specifically, the further away the approximation becomes, the higher the differences, as can be expected for such an approach. It is noteworthy, however, that even for a time distance of more than one year from the last point used for estimation of the parameter development over time, our model is still only 13 % off in its peak difference.

We can attribute the additional increase in the difference between approximated and actual parameters used in the two simulation approaches to the additional variability that stems from the simulations of object sizes that constitute individual simulated web pages in the cache. This effect has less impact on the simulation times beyond a day, for which even the furthest evaluated approximation yields less than 8 % difference of the data in the simulated cache.

As different simulation scenarios might require different time horizons, we aggregate the results into larger different bin sizes of 300 s and 3600 s, illustrated in Figure 5 as well. We observe from comparing Figures 5a, 5c, and 5e for the number of cached objects and Figures 5b, 5d, and 5f for the amount of data in the cache, respectively, that few characteristic behavior changes can be observed. We note, however, that the increase in the bin size to one hour results in the next close approximation for March 15, 2014 exhibits significantly lower differences. Thus, when considering smoothed or capacity–level aggregations for simulations, the averaging effects have a significant effect on the result accuracy.

IV. Conclusion

In summary, our approach shows great promise to facilitate the simulation of mobile web landing pages, which can be regarded as reflecting the upper–bound characteristics of actual web pages visited by mobile users. In particular, it enables the evaluation of how a mobile web browser could perform realistic web requests over time in the presence of its local cache. The continuous evolution of the mobile web's characteristics from human interface and interaction design principles are embodied into the web objects to be retrieved by a browser.

Over time, it might not always be possible to predict or derive up–to–date representations of the latest developments to include those in performance evaluations of new mechanisms in the realm of mobile communications. We presented an accessible approach that can employ a general approximation of its parameters over time with only manageable impacts on accuracy. Specifically, for more than one year's changes in the underlying data set, our model incurs only a maximum 13 % penalty in quality. While this showcases that a long–term prediction can be facilitated by our model with reasonable accuracy, it also indicates a sensible result quality of performance evaluations for long time horizons can be achieved by other researchers when approaches operate on the granularity of web page objects.

References

[1] Cisco, Inc., "Cisco visual networking index: Global mobile data traffic forecast update, 2014–2019," Tech. Rep., Feb. 2015.

[2] T. A. Johnson and P. Seeling, "Desktop and mobile web page comparison: Characteristics, trends, and implications," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 144–151, Sep. 2014.

[3] M. Butkiewicz, H. Madhyastha, and V. Sekar, "Characterizing web page complexity and its impact," *IEEE/ACM Transactions on Networking*, vol. 22, no. 3, pp. 943–956, Jun. 2013.

[4] B. Zhao, W. Hu, Q. Zheng, and G. Cao, "Energy-Aware Web Browsing on Smartphones," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 3, pp. 761–774, Mar. 2015.

[5] H. Shen, M. Kumar, S. K. Das, and Z. Wang, "Energy–efficient data caching and prefetching for mobile devices based on utility," *Mobile Networks and Applications*, vol. 10, no. 4, pp. 475–486, Aug. 2005.

[6] M. A. Maddah-Ali and U. Niesen, "Fundamental Limits of Caching," *Information Theory, IEEE Transactions on*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[7] T. A. Johnson and P. Seeling, "Web cache object forwarding from desktop to mobile for energy consumption optimizations," in *Proc. of IEEE Online Conference on Green Communications (OnlineGreenComm)*, Online, Nov. 2014.

[8] M. Katz, D. E. Lucani, F. H. Fitzek, and P. Seeling, "Sharing resources locally and widely: Mobile clouds as the building blocks of the shareconomy," *IEEE Vehicular Technology Magazine*, vol. 9, no. 3, pp. 63–71, Sep. 2014.

[9] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, 2014.

[10] S. Souders, "httparchive.org," Apr. 2015. [Online]. Available: http://www.httparchive.org

[11] T. Johnson and P. Seeling, "Landing page characteristics model for mobile web performance evaluations on object and page levels," in *Proc. of IEEE International Conference on Communications (ICC)*, London, United Kingdom, Jun. 2015.

[12] I. Tsompanidis, A. H. Zahran, and C. J. Sreenan, "Mobile network traffic: a user behavior model," in *Proc. of the 7th IFIP Wireless and Mobile Networking Conference (WMNC)*, Vilamoura, Algarve, Portugal, May 2014, pp. 1–8.

[13] T. Johnson and P. Seeling, "Browsing the mobile web: Device, small cell, and distributed mobile caches," in *Proc. of IEEE ICC Workshop on Cooperative and Cognitive Networks (CoCoNet7)*, London, United Kingdom, Jun. 2015.

[14] G. Anastasi, M. Conti, E. Gregori, and A. Passarella, "Performance comparison of power–saving strategies for mobile web access," *Performance Evaluation*, vol. 53, no. 3–4, pp. 273–294, Aug. 2003.