Web Search and Browsing Behavior under Poor Connectivity

Jay Chen
New York University
715 Broadway 7th Floor
New York, NY 10003 USA
jchen@cs.nyu.edu

Lakshminarayanan Subramanian
New York University
715 Broadway 7th Floor
New York, NY 10003 USA
lakshmi@cs.nyu.edu

Kentaro Toyama
Microsoft Research
196/36 2nd Main
Sadashivnagar, Bangalore
Bangalore, India
kentoy@microsoft.com

CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-246-7/09/04.

Abstract
Web search and browsing have been streamlined for a comfortable experience when the network connection is fast. Existing tools, however, are not optimized for scenarios where connectivity is poor, as is the case for many users in developing regions where fast connections are expensive, rare, or unavailable.

This study examined how users' web search and browsing behavior differs when the connection is slow, and whether users employ techniques to alleviate the problem. In a preliminary study involving 15 subjects on a university campus in Kerala, India, we identify unique mitigating behaviors of users who routinely suffer low-bandwidth or intermittent connections. We examine the challenges faced by these users and find that existing web search and browsing infrastructure is simply incapable of providing a good experience. Finally we outline potential design improvements.

Keywords
Web search, web browsing, world wide web, low bandwidth, intermittent network

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Slow, expensive, or non-existent Internet access is a fact of life for many computer users in the developing world where the physical infrastructure has yet to catch up to the increasing demand. In these regions connectivity is often intermittent due to power cuts [4], and bandwidth is generally an expensive and rare commodity because none of the traditional wire-line connectivity solutions (fiber, broadband and dial-up) are economically viable for rural regions with low user densities [1].

Existing work on bringing the Internet to the developing world generally focuses on either connectivity issues [3], or user interfaces for the illiterate [7]. The recent emergence of new low-cost connectivity solutions using long-range wireless technologies (cellular, WiMax [5], long-distance WiFi [3]), and delay tolerant mechanical backhaul networks (connectivity via busses, motorbikes, etc. driving in a loop) [2, 6] provide hope for rural connectivity. However, even after Internet connectivity is established via any of these physical channels, the link is slow due to either low bandwidth or intermittency as a result of long backhaul delays. Compression and filtering are good options if the network is slightly bandwidth constrained [6, 9], but when the bandwidth is extremely low or the latency is high they fail to provide much benefit. In these situations interactive applications such as web search and web browsing become completely impractical, and it has thus far been unclear how to improve web browsing when the process is completely asynchronous (with latencies on the order of hours or days in the case of mechanical backhauls). Previous work on low bandwidth web access is outdated [12], and focus on performance optimizations, not user behaviors. Conversely, existing literature on web browsing behavior does not examine user behavior in the context of low bandwidth [14].

We recognize that the scale of the preliminary study is limited in scope. The contribution of this work is to identify the problem as pervasive and unsolved in developing countries, and map out the space for further study. While some of our findings are also applicable to mobile phones, most of the research in the space addresses the small form factor rather than the bandwidth [13]. In contrast, we focus our work on an extremely low bandwidth scenario to avoid conflating the two issues.

To understand this problem we investigated a segment of the population who are literate, well-educated, and connected to the Internet, but did not have practical access to high-speed connections. While this subset of users may not be representative of the entire population we believe they provide a lower bound on browsing problems under poor connectivity. The questions we asked were: How do users interact differently with the web behind a slow network link? Are there specific techniques employed to mitigate the effects of the long delay for retrieving web pages? How well do they work? How effective are traditional caching and compression at reducing the delays? We found that a few users attempted to avoid the problem by multitasking, but most users grew frustrated at wasting large proportions of their time. We also found that existing solutions did little to address these long delays.

Study

We studied university students and staff using the Internet from behind a heavily shared bottleneck connection. We observed computer users performing web search on the campus using their computer accounts. For the remainder of this paper we use the term *search* to refer to interactions with the search engine, and *browsing* to refer to other actions associated with the search task. Participants were observed and *videotaped* during the process. Also, *screen-capture video* of their mouse movements and keyboard entries were recorded. We also interviewed participants about boredom levels and general reactions to their experience.

Study Environment

The network available on the campus was an 8Mbps connection shared across 400 machines and over 3000 students, staff, and faculty. The accounts that the participants used were their own, and the university gateway router allowed higher per-packet priority to packets belonging to accounts of faculty and higher bandwidth caps, but none of the participants had access to these accounts. The maximum upload and download bandwidth allocated to students and staff was only 750Kbps in *total*, and during peak hours nearly every machine was being used. Therefore, the worst-case average bandwidth available per machine was approximately 1.9Kbps. This speed is abysmally slow even compared to dial-up (56.6Kbps).

The experiments used an existing machine in the university's computer lab running Windows XP and Internet Explorer 7. The machine and the network were powered by backup generator to avoid power outages. The experiments were performed between 12:00pm and 10:30pm. We did not artificially constrain the time of day as the fluctuations in bandwidth were themselves part of the phenomenon of interest.

Participants

Fifteen participants (11 male) between 19 and 25 years of age were observed in this study. All participants were enrolled in college, college graduates, or completed their masters. Five participants were students, and ten participants were staff. The participants' fields of study included computer science, electrical engineering, commerce, and business management. Every participant was self-reported to be at least moderately experienced with web search, and able to converse and browse the web in English. The Internet connection that the participants were familiar with varied between dialup and broadband.

Procedure

Each participant was first given a simple demographic and search experience questionnaire at the beginning of the study. Questions included level of familiarity with the Internet, average Internet usage per day, and comfort while searching or browsing the web. Then the participant was asked to use the Internet for web search for 7-15 minutes using the search engine of their choice, and pursue any search topics of their choice. Participants were not informed beforehand as to the exact duration of time available so they would not feel rushed. Finally, participants were given a brief semi-structured interview (around 10 minutes) about their search experience. We recognize that our study is limited in size and scope, and plan to use these preliminary results to motivate a more comprehensive study.

Results

In this paper we focused on the participants' web search and browsing behavior in low bandwidth settings. In our study the actual bandwidth varied significantly from 20Kbps in the early to mid afternoon to 200Kbps in the evening. The reason the actual bandwidth was higher than the worst-case is likely because not all machines were accessing the network simultaneously. We briefly summarize our results due to lack of space.

General Behaviors

Search engine result pages tended to load quickly (under 5 seconds) due to the mostly-text content and low latency to the search engine servers. None of the users had complaints about the search result page load times during the experiment. The requests for general pages (pages not provided by the search engine) took varying amounts of time to load depending on server latency, network congestion, and page contents. They ranged from 1 second up to 240 seconds. These values are inclusive of hits in the local web proxy's cache and any compression implemented by the accessed web servers.

Users had no complaints with pages that took less than 5 seconds to load, and were generally only affected once page loading time exceeded 10 seconds. These users were observed sighing, staring at the blank screen, leaning back, and trying to make conversation with the observer while waiting. This behavior along with self-reported boredom or frustration (e.g. "I feel angry") increased as the page loading time increased. When page load times were over a minute several users reported that under normal circumstances they would do something else, give up, or wait to search at a later time or using a faster connection elsewhere. Four users reported that they had access to a home broadband connection, and all users had access to and were aware of a broadband pay-per-use PC café on the campus.

As observed in previous studies [8], we found that users generally preferred to look only at the first page of results, electing to modify their search query rather than go on to a second page. None of the users used any advanced search engine interface, though most were aware of its existence. Two users, when asked, revealed that advanced search was not worth the effort to use, preferring instead to iteratively modify their query. Four users performed image searches during the experiment, and two other users mentioned they normally use the image search when applicable.

How Did Users Seek to Alleviate Wait Times?

Seven (47%) of the users opened multiple windows while browsing (the web browser we used did not have tabbed browsing available), and two more reported that they occasionally would. Six users reported that outside of the experiment they commonly multi-tasked with other activities. These activities included listening to music, using offline applications, reading, or talking to a friend.

How Successful were the Users at Saving Time?

We used our video data to time the total idle and busy durations for each user. A user was defined as busy when he was observed reading the contents of the page or performing any navigational mouse action, and defined as idle for other times. We found that users who opened multiple windows and switched between them wasted less time while browsing. Users who opened multiple windows, but did not switch between them performed about as well as those who used a single window. Only when the page loading rate is faster than the user's reading rate is the page load time disguised. In our discussion we explain why the improved performance exhibited by the opening multiple windows is an artifact of our experimental environment, and not generalizable to all slow network connections.

How Beneficial were Caching and Compression?
For our users the web proxy only had a cache size of 20GB, and a hit rate of 10-25%. On the Internet approximately 27.5% of web servers compress files they serve [11]. To estimate the best-case scenario for compression we compressed all files downloaded by our users during the experiment using a simple compression algorithm (*gzip*). Unsurprisingly, we found that compressing benefits text files the most up to 80% compared to image files which are already compressed. However, text files only represent a small (and diminishing) proportion of total webpage size on the web [10]. From our idle time measurements and interviews the existing caching and compression were unable to mask the network latency for our users. We observe that even with the best possible caching and compression incorporated it is unlikely that the required orders of magnitude improvement would be attained.

Additional Observations
First, we observed that most of the searches performed were for textual information, thus most images were not useful despite taking up the majority of download times. We confirm that, as assumed by previous projects [6, 9], a text-only browsing option could improve satisfaction if images are unwanted. Second, web browsers by default currently only have a progress bar to indicate page loading progress. This feature was not helpful to users because it presented only a vague estimate as to the time until a page finishes loading. One user even claimed to estimate the time to load a page manually based on the rate at which the progress bar filled up indicating that providing the user with an estimated page loading time could allow users do other tasks while waiting. Finally, we noticed that users often had problems entering search queries that were perfectly useful on the first attempt. Assistance for search query construction was not always available, and it when did appear, it was never explicitly used. When asked, users responded that they preferred to iteratively refine their search terms based on previous results.

Discussion
We emphasize that connections in the developing world are typically intermittent and *even slower* than the network in our experiment. These connections have varied bandwidth and latency characteristics due to their emergence from different technologies, and each variation has different implications. We expect that the web browsing experience in these scenarios would be worse than what was experienced by our participants. To that end, we discuss the consequences of different types of slow connections on web search.

In our experiment, even though the bandwidth was severely limited, it was being capped on a per account class basis. This meant that opening many windows effectively increased the amount of bandwidth the user received from the account class' pool. This is not usually the case; more often than not users have individual bandwidth limits (at home or pay-per-byte connections). In these situations opening multiple windows would actually be counterproductive, and pages should be loaded in series to improve the response time.

In higher latency settings the benefits of caching and compression would be drastically reduced. To adapt caching to these settings, the ability to search in the cache and browse offline should be made available. Also, iteratively constructing the most appropriate query would become problematic as the latency increased. Offline automatic assistance with query construction could help the user construct the best query possible the before the network is actually used. Finally, if the latency is high, but the bandwidth is abundant, intelligent prefetching of pages could tradeoff this extra bandwidth for an overall improvement of query response time.

In summary, the design improvements we suggest are:

- Provide explicit estimate of page load time.
- Load pages in series if bandwidth is low.
- Allow a text-only option if bandwidth is low.
- Facilitate local search when latency is high.
- Provide query construction assistance offline.
- Prefetch pages if bandwidth is readily available.

The main contribution of our work is that we formally establish that these problems exist and outline directions for future research to improve web search and browsing under poor connectivity.

Citations

[1] S. Mubaraq, J. Hwang, D. Filippini, R. Moazzami, L. Subramanian, and T. Du. Economic analysis of networking technologies for rural developing regions. *Workshop on Internet Economics*, 2005.

[2] K. Fall. A Delay Tolerant Network Architecture for Challenged Internets, *Computer Communication Review*, 2003.

[3] R. Patra, S. Nedevschi, S. Surana, A. Sheth, L. Subramanian, and E. Brewer. WiLDNet: Design and Implementation of High Performance WiFi Based Long Distance Networks. *NSDI*, 2007.

[4] S. Surana, R. Patra, S. Nedevschi, M. Ramos, L. Subramanian, and E. Brewer. Beyond Pilots: Keeping Rural Wireless Networks Alive. *NSDI*, 2008.

[5] WiMAX forum. http://www.wimaxforum.org.

[6] W. Thies et al. Searching the world wide web in low-connectivity communities. *WWW*, 2002.

[7] A. Franz and B. Milch. Searching the Web by Voice. *International Conference on Computational Linguistics*, 2002.

[8] T. Joachims, L. Granka, Bing Pan, and G. Gay. Accurately Interpreting Clickthrough Data as Implicit Feedback. *In ACM SIGIR*, 2005.

[9] Loband website. http://www.loband.org.

[10] A. B. King. *Website Optimization Secrets.* O'Reilly Media. Sebastopol, CA. 2008.

[11] Port80 Software. http://www.port80software.com.

[12] T. S. Loon, V. Bhargavan. *Alleviating the Latency and Bandwidth Problems in WWW Browsing.* USENIX Symposium on Internet Technologies and Systems, 1997.

[13] P. Ketola, M. Röykkee. *The Three Facets of Usability in Mobile Handsets.* CHI 2001 Workshop: Mobile Communications.

[14] H. Weinreich, H. Obendorf, E. Herder, M. Mayer. Off the Beaten Tracks: Exploring Three Aspects of Web Navigation. WWW 2006.