Resource management strategies for Mobile Web-based services

Claudia Canali, Michele Colajanni, Riccardo Lancellotti
Department of Information Engineering
University of Modena and Reggio Emilia
Modena 41100, Italy
Email:{claudia.canali, michele.colajanni, riccardo.lancellotti}@unimore.it

Abstract—The great diffusion of Mobile Web-enabled devices allows the implementation of novel personalization, location and adaptation services that will place unprecedented strains on the server infrastructure of the content provider. This paper has a twofold contribution. First, we analyze the five-years trend of Mobile Web-based applications in terms of workload characteristics of the most popular services and their impact on the server infrastructures. As the technological improvements at the server level in the same period of time are insufficient to face the computational requirements of the future Mobile Web-based services, we propose and evaluate adequate resource management strategies. We demonstrate that pre-adaptating a small fraction of the most popular resources can reduce the response time up to one third thus facing the increased computational impact of the future Mobile Web services.

Index Terms—Mobile Web, Multimedia resources, Content Adaptation, Performance evaluation.

I. Introduction

The advent of the Mobile Web has significantly changed the Web scenario, because a growing number of users now accesses vast amounts of content at any time, from any location and through any class of devices. According to recent research reports, the mobile browsing will grow over 900% by 2013 [1]. Other major changes are expected in terms of novel forms of traffic and services. For example, photo and video sharing services (e.g., YouTube, Flickr) are causing an explosion of demand for multimedia content. These trends will determine a future mobile Web scenario characterized by a large amount of heterogeneous contents, mainly consisting of multimedia resources (e.g., [1], [2], [3]), that will have to be tailored to user preferences and to device capabilities on-the-fly at the moment of the client request [4], [5], [6]. Designing server architectures that will support future Mobile Web-based services requires an initial evaluation of the workload and computational impact. Moreover, we should investigate which resource management strategies are adequate to the novel mobile Web scenario.

For the first goal, this paper provides a quantitative estimation of the evolution of mobile Web workload characteristics and server infrastructures in the next five years (2008-2013). As the large variety of the Mobile Web-based services prevents the identification of the characteristics of "the typical Web site" of the future, we consider the 100 worldwide most popular Web sites [7] and from this list we focus on the categories

of *online-news* and *social-multimedia* Web sites, because they cover almost 80% of the sites and are likely to remain very popular also in the Mobile Web [2]. For both categories, our analysis indicates that the technological evolution at the server level will not counterbalance the increase of computational demand of the Mobile Web-based services, to the extent that supporting these services would require large replications at the server and resource level.

For the second goal, we propose a resource management strategy based on the pre-generation of adapted versions of the multimedia resources. Applying pre-generation to all multimedia resources is clearly unfeasible because it would increase the size of the working set, the volatility of content would result in a waste of computational and storage resources, and it would be impossible to guarantee consistency of all replicated resources. Hence, we suggest to limit the pre-generation to a small fraction of the most popular multimedia resources. This approach is mainly motivated by the observation that the popularity of multimedia resources in Mobile Web-based services follows a Zipf-like distribution [8], [9]. Furthermore, the expected evolution of the mobile clients is going towards powerful devices with medium-large connections, and this trend helps reducing the device heterogeneity. Through extensive simulation we show that pre-generating adapted versions to a limited fraction (less than 15%) of popular resources is sufficient to reduce to almost one third the response time of Mobile Web-based services for different evolutionary trends. Negligible performance gains are achievable by applying pre-generation to a larger fraction of resources.

The remainder of this paper is organized as follows. Section II describes the Mobile Web-based services. Section III provides an insight on the evolution of these services and evaluates their impact on the server infrastructures. Section IV describes and evaluates the benefits of the proposed resource management technique. Section V discusses the related work and Section VI concludes with some final remarks.

II. Mobile Web-based services

We have already observed that the focus of this paper is on online-news and social-multimedia sites that will be accessed even through mobile Web-enabled devices.

The *online-news* category includes information portals, such as online newspapers and news broadcasting sites, that offer

online information such as events, stock quotes, and sports results. These Web sites typically deliver news in the form of text-based resources and images, that account for 60% and 35% of the requests respectively, although there is a growing tendency to deliver also audio and video content, today limited to 5% of the resources, as confirmed by a recent research report (www.stateofthenewsmedia.org/2007).

The *social-multimedia* category includes Web sites that represent a new form of user communication and interactivity, that is a qualifying characteristics of the so-called Web 2.0. Typical examples are forums, blogs and content sharing sites where users exchange opinions, stories and files (e.g., MySpace, Flickr, Youtube). To understand the workload composition of these Web sites, we download and analyze 10000 pages from MySpace, that currently represents the most popular social-multimedia site [10]. We find that almost the 54% of the resources are in the form of text-based content, while multimedia content accounts up to the 46% of the resources, as many present communications among the users involve exchanges of images, audio or video resources [9].

Each user request for online-news or social-multimedia Web sites may involve different operations of dynamic generation and adaptation with the goal of tailoring the required content to user preferences and device capabilities. We find convenient to distinguish the offered services based on the content type, that may be text-based (e.g., HTML) or multimedia content (e.g., images, audio, video).

Text-based content is dynamically generated from information stored in database(s). As a typical example, we refer to the request of an article in an online news Web site. The user request determines the retrieval of the article from a database containing news and the generation of an HTML page that contains also information/banner derived by the user profile. Specifically, the page layout may be set according to the user preferences. A recommendation list may be generated to suggest other related articles that match the user interests. The typical service time for the generation of a text-based resource is in the order of hundreds of milliseconds for content generation, while the most expensive operations, such as data mining techniques to associate recommendations to specific user preferences, are carried out offline and do not contribute to service time [11], [3].

Multimedia content must be often adapted to match the characteristics of the mobile device in terms of computational power, rendering capabilities and network connection. However, content adaptation may be driven by specific settings of user preferences: for example, a color-blind user may require specific enhancements on images and videos to match its impaired vision [12]. Adaptation of multimedia resources may involve different transformations. Images are adapted by scaling the size modifying colors, cropping irrelevant part or increasing the compression factor. Audio can be recoded with lower bit rate, while common video transformations concern frame size or color modification. The typical service time for adapting a multimedia resource depends on the resource size and may reach up to one or more seconds when complex operations are carried out on audio and video clips of several megabytes.

III. Evolution of Mobile Web-based services

For the design of the future server infrastructures, it is important to analyze how the workload characteristics of the Mobile Web-based services will evolve in the next five years, and to evaluate the computational impact of the future applications.

A. Evolutionary trends

There are three workload characteristics that may affect the computational demand of the Mobile Web-based services: *workload composition*, *size of workload resources* and *workload intensity*. The evolution of the workload composition and resource size is based on a performance study on the Mobile Web-based service by the same authors [3]. We report here the results that are most significant for the evaluation of the computational impact on server infrastructures.

The workload composition in terms of text-based and multimedia resources is characterized by an increasing amount of multimedia content for both online-news and social-multimedia Web sites [3], especially video and audio resources, as shown in Table I. However, while in the case of online-news almost half of the offered content will remain in the form of text-based resources, the social-multimedia workload will consist for the 60% of multimedia content, with a 20% of audio and video resources. The large presence of multimedia resources is likely to cause a computational increase more significant for social-multimedia than for online-news Web sites.

TABLE I
EXPECTED EVOLUTION OF WORKLOAD COMPOSITION

Resource type	Online-news class		Social-multimedia class	
	Nowadays	2013	Nowadays	2013
Text-based	60%	49%	54%	41%
Image	35%	38%	38%	40%
Audio/video	5%	13%	8%	19%

For the analysis of workload resource size, we consider multimedia resources because their size affects the computational cost of the adaptation [13], [3], while the size of text-based resources is not correlated to the costs of content generation. The median resource size is expected to increase per year by 12% for images and by 16% for audio and video resources [3].

Preliminary experiments determined that the workload intensity has the most relevant impact on system performance. For this reason, we choose to consider two evolutionary scenarios for each Web site category, namely *low-growth* and *high-growth* scenarios, that assume different increases of the workload intensity.

For the online-news Web sites, the low-growth scenario assumes an increase of 20% per year of the workload intensity, that is mainly related to the presence of mobile users [1]. For traditional Web accesses, we observe no contribution to the service growth. Indeed, according to recent studies

(www.stateofthenewsmedia.org/2007) the percentage of Web users that get news online seem to be stabilized since 2000. For the high-growth scenario, we consider that the workload intensity increases by 35% per year as for the most popular sites. This value is motivated by recent research reports (www.stateofthenewsmedia.org/2007) on news sites such as some TVs (e.g., CNN, MSNBC) and portals (e.g, Google News, Yahoo!), that aggregate material from many different content producers.

Social-multimedia sites achieved a more recent popularity and the number of mobile users accessing these sites is expected to grow extremely fast in the next five years. We assume that for the high-growth scenario the requests will augment at a rate of about 55% per year [14], [15]. We also consider a low-growth scenario where the workload intensity is expected to increase by 40% per year.

B. Performance evaluation

To evaluate the performance of the server infrastructures that support present and future Mobile Web-based services we use a discrete event simulator based on the Omnet++ framework [16]. We model a system consisting of multiple clients that issue requests to a server infrastructure. A service time is determined for each request depending on the type of resource: text-based, image or audio/video. For the current scenario, the service times to dynamically generate a text-based resource follow an empirical distribution obtained by preliminary experiments, with a median of 220 ms. For multimedia resources, the service times are based on [3]. Since the time to adapt a multimedia resource is proportional to the resource size [6], [13], we consider a service time for MB, that is 730 ms for images and 1054 ms for audio/video. To model future server infrastructures with more powerful CPU, we assume that the server computational power will continue to increase according to the Moore Law for the next five years. If the computational power doubles every 18 months, we can expect that in the same period the service times can be reduced by a factor of 8. The multimedia resource size in the current scenario is assumed to have a median size of 900 KB, 3MB, 8 MB for images, audio and video, respectively [3]. According to the annual growth rate, in the future scenarios the median size will become equal to 1.6 MB, 6 MB and 17 MB for images, audio and video, respectively.

In the current scenario, the server receives requests at the rate of 12 per second for both online-news and social-multimedia Web sites. For the future scenarios, we applied the annual growth rates described in Section III-A arriving to rates of 30 and 54 req/s for online-news sites in low and high-growth scenarios respectively, and to 64 and 106 for social-multimedia sites in low- and high-growth scenarios, respectively. The resource popularity is based on a Zipf distribution with the α parameter set to 1 [8] or to 0.84 [9], according to the state of the art in worklod characterization for novel Web-based services. Each multimedia resource may be requested in 20 different versions in the current scenario, while in the future scenarios there will exist 4 versions for each resource, due to the expected reduction of the device heterogeneity. The

probability of being requested is uniformly distributed among the existing versions of a multimedia resource.

The performance evaluation considers two main metrics: response time and CPU utilization. The response time measures the time elapsed between the instant in which the client issues the request and the moment when it receives all the files in response. Actually, this response time intrerpretation holds for text-based and image content. On the other hand, for audio and video content, the response time ends when the client receives the first 1.5MB of the required resource, that represents a common buffer size for media players supporting HTTP streaming. We should consider that these multimedia resources are not entirely adapted before being delivered to the client, but adaptation and delivery is done on a chunk-by-chunk basis. We model in this way the play-while-downloading behaviour typical of the widely used HTTP pseudo-streaming [17]. The CPU utilization measures the fraction of time the CPU is busy with respect to the observation period.

Figure 1 shows the cumulative distribution of the response time for the online-news Web sites in the three considered scenarios: current, low-growth and high-growth. Opposite behaviors can be observed for the low- and high-growth scenarios, that achieve respectively better and worse performance with respect to the current scenario. In the low-growth case, the more powerful CPU of future server infrastructures can satisfy the higher number of requests with lower response times than in the current scenario. On the other hand, in the high-growth scenario the technological improvements cannot counterbalance the increase of computational requirements of Mobile Web-based services and the performance achieved are significantly worse than in the current scenario.
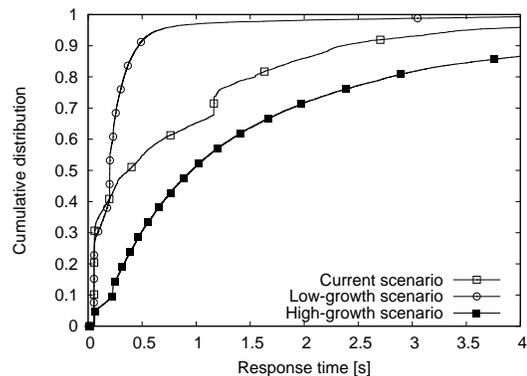

Fig. 1.   Expected evolution of response time for online-news Web sites

Figure 2 shows the cumulative distribution of the response time for the social-multimedia Web sites for the current and future scenarios. The performance degradation is much evident for both the future scenarios, with a 90-percentile of the response time in the high-growth scenario that is almost 6 times higher with respect to the current scenario. Even if the experienced growth will be lower than that expected, an efficient support of these Mobile Web-based services would require a reduction of the amount of on-the-fly multimedia adaptations.
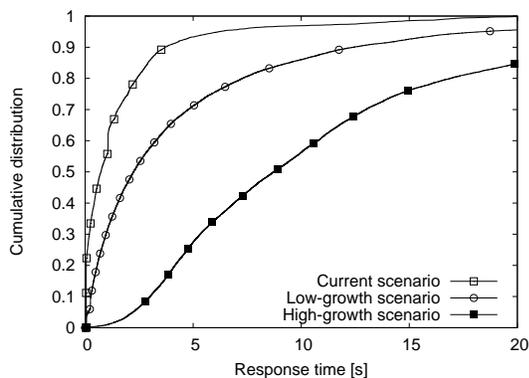
Fig. 2. Expected evolution of response time for social-multimedia Web sites

The values of CPU utilization, shown in Table II, confirm the impact of the evolutionary trends of workload characteristics and technologies. While the performance trend of online-news Web sites will depend on the actual increase of the workload intensity, the computational requirements for social-multimedia Web sites are expected to significantly grow for both evolutionary scenarios. For the social-multimedia context, the computational demand is hardly affordable for the server infrastructure even in the most optimistic hypothesis that the workload intensity will grow less than that predicted by specialized research reports.

TABLE II
AVERAGE CPU UTILIZATION

Web	Scenario		
sites	Current	Low-growth	High-growth
News	0.45	0.43	0.88
Social	0.57	0.89	0.99

IV. RESOURCE MANAGEMENT STRATEGIES

The analysis of the evolutionary trends of Mobile Web-based services evidences a performance issue for future server infrastructures. To support future Mobile Web-based services with adequate performance, there is the need for resource management strategies that reduce the computational cost of each request on the server infrastructures. To this aim the most common approach is to replicate servers and/or resources, but in this paper we focus on resource management.

A. Content pre-generation

Our analysis shows that the performance issues of future Mobile Web-based services are mainly related to the adaptation of multimedia content. Text-based content generation is not critical for performance, because the most expensive tasks such as data mining are already carried out offline. Hence, resource management strategies aiming to reduce the cost of future Mobile Web-based services should avoid whenever possible expensive on-the-fly adaptation of multimedia resources.

The most straightforward solution is the offline pre-adaption of all multimedia resources, that can be delivered at the moment of the client request without costly on-the-fly generation.

However, this approach is clearly unfeasible because the costs in terms of computational and storage requirements would be extremely high, especially if we consider the huge increase of the working set size that is expected in the next future [9], [2]. The problem is even more critical if we consider that content in online-news and social-multimedia Web sites is highly volatile. Indeed, resources are typically characterized by a short lifespan because they concern real-world events or popular hot topics for which users interest rapidly subsides. Content volatility decreases the effectiveness of pre-generating adapted resources and increases the costs to maintain the consistency of the pre-generated versions.

We propose an alternative approach where pre-generation is applied only to a limited fraction of popular multimedia resources. Our proposal is motivated by the expected evolution of the mobile clients and by the popularity distribution of multimedia resources in Mobile Web-based services. The evolution of client device is leading towards powerful devices with medium-large connections. This trend has a positive effect on the server side because it helps to reduce the device heterogeneity [3], [18]. We may expect that in the next future the server infrastructures will not need to adapt resources exactly for every type of client device, as it happens now, but they could generate few adapted versions, where each version can be consumed by different device types. As for popularity distribution in Mobile Web-based services, recent studies found that the resource popularity follows a Zipf-like distribution with an α parameter between 0.8 and 1 [8], [9], hence we may expect that many requests refer to a limited number of popular resources. Pre-generating adapted versions for a small fraction of popular resources should be sufficient to significantly reduce the need for on-the-fly adaptation.

It is important to observe that the fraction of popular resources which have to be considered for pre-generation depends on the workload characteristics of each Mobile Web-based service. The proposed resource management strategy aims to identify the minimum fraction of resources to pre-generate to avoid adaptations that would uselessly increase computational and storage waste because of the Zipf-like popularity distribution and the high volatility of resources. We consider that the server infrastructure knows which resources are the most popular in the working set. This assumption is motivated by the goal of understanding to which extent the pre-generation of adapted versions for a limited number of resources may actually improve the user-perceived performance and reduce the computational requirements of Mobile Web-based services. The issue of identifying the resources belonging to the popular "hot set" would deserve a separate study.

B. Performance evaluation

We evaluate the proposed resource management strategy through our simulator. Consistently with recent literature, we consider two possible values for the Zipf α parameter of the resource popularity distribution, that are 1 [8] and 0.84 [9]. We consider the 90-percentile of response time and the average CPU utilization of the server as the main performance metrics.

Figures 3, 4 and 5 show the 90-percentile of the response time as a function of the fraction of pre-generated resources for the online-news and social-multimedia Web sites in the future scenarios. We do not show the results for the online-news sites in the low-growth scenario because, due to the low computational requirements, client requests can be serviced through on-the-fly content adaptation and pre-generation is useless.
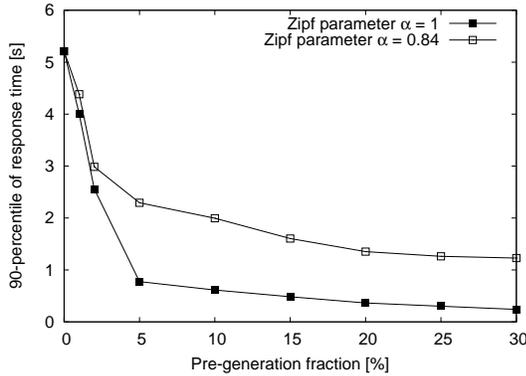

Fig. 3.   High-growth scenario for online-news Web sites
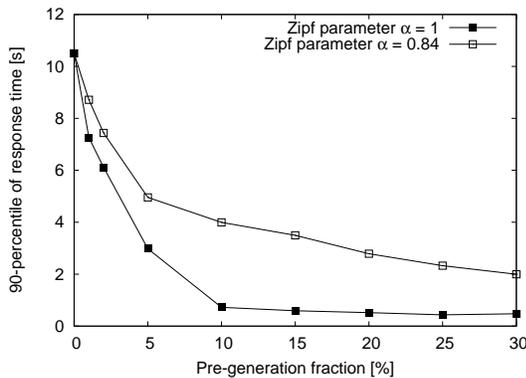

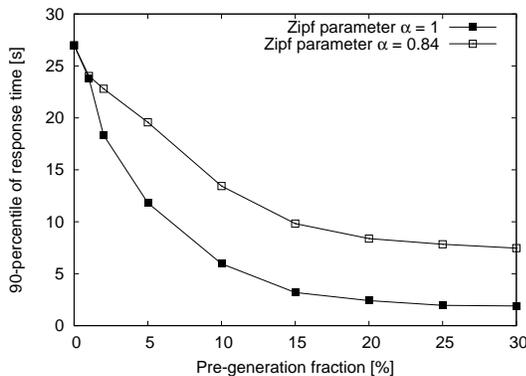Fig. 4.   Low-growth scenario for social-multimedia Web sites

Fig. 5.   High-growth scenario for social-multimedia Web sites

For every considered scenario and for every considered value of Zipf α parameter, we can observe that pre-generation of the most popular resources provides significant performance improvements. For each scenario, the response time drops fast in the left part of the graph, while, as the amount of pre-generated resource grows, the performance gain is reduced, becoming almost negligible as the amount of pre-generation grows beyond 20%.

Another result common to all experiment is the impact of the Zipf α parameter on the user-perceived performance. In the case of high popularity skewness (i.e., $\alpha = 1$), pre-generating a small amount of of very popular resource is sufficient to satisfy a large number of request, thus providing a significant performance gain. On the other hand, when the popularity is less skewed (i.e., $\alpha = 0.84$), the heavy tailed effects are more evident as there is a higher amount of requests for non popular resources, resulting in higher response times.

Figure 3 refers to the high-growth scenario for the online-news category of Web sites. We observe that the response time experiences a sudden reduction as the content pre-generation fraction grows up to 5% for both considered values of the Zipf parameter. This result can be explained by considering the CPU utilization, not reported for space reasons, that passes from a value over 90% to 70% as the amount of pre-generated resources grows up to 5%.

Figures 4 and 5 refer to the social-multimedia Web sites for the low- and high-growth scenarios, respectively. Due to the growth in the workload intensity of these sites, the 90-percentile of response time is higher than that of the online-news sites. The computational demand of future social-multimedia Web sites determines a situation of overload that may require to pre-generate up to the 15% of the working set, as confirmed by the CPU utilization in Table III.

We may conclude that pre-generation may effectively reduce the response times of the Mobile Web-based services by pre-generating a small fraction of popular resources (at most 15%) even in the most demanding scenarios.

TABLE III
AVERAGE CPU UTILIZATION - HIGH-GROWTH SCENARIO ($\alpha = 1$)

Web sites	Content pre-generation fraction						
	0%	5%	10%	15%	20%	25%	30%
News	0.88	0.71	0.58	0.51	0.45	0.40	0.34
Social	0.99	0.91	0.86	0.75	0.68	0.62	0.55

V. RELATED WORK

Recent research about the Mobile Web actually spans different areas, ranging from the characterization of novel Mobile Web-based services to the research on content adaptation, caching and delivery for the Mobile Web. Our analysis of the service evolutionary trends takes advantage of recent results on the characterization of Mobile Web-based services based on social user interactions [19], [9]. Specifically, we exploit the analysis on multimedia resource popularity in social-oriented services, that have been proved to follow a Zipf-like distribution [9]. This knowledge motivates our proposal of resource management strategies, that focuses on a limited

set of popular resources to improve the overall performance of novel Mobile Web-based services.

A preliminary analysis of how the performance requirements of Mobile Web-based services will evolve in the next future was carried out in [3], that represents a starting point for the evaluation presented in this paper. In [3] the impact of future services on server infrastructures was analyzed from an analytical point of view for one evolutionary scenario. On the other hand, in this paper we consider a more complete range of evolutionary scenarios for Mobile Web-based services and we test the performance impact by using a simulation framework, that allows us to evaluate the performance of server infrastructures with the more powerful CPUs of the future. Another major difference is the proposal and the evaluation of resource management strategies based on offline pre-generation of adapted content to efficiently support future Mobile Web-based services.

Although not so popular in the current Web scenario, offline pre-generation is a strategy used by some Web portals, such as AvantGo [20], to deliver content to mobile users: the portal pre-generates a set of defined resources to reduce the response time perceived by the users. This solution limits adaptation just to few defined Web sites, whose multimedia resources are all pre-generated. On the other hand, we suggest to pre-generate adapted versions just for the most popular resources. The proposed pre-generation technique represents a more general solution, that is affordable for servicing resources belonging to several Web sites thanks to the limited number of resources composing the popular hot set.

In order to tailor contents to mobile devices, several studies propose solutions that are based on on-the-fly adaptation [21], [5]. These approaches may be integrated with caching strategies [22], [6], [13] that typically exploit a sort of utility function to determine whether or not it is convenient to cache an adapted version of a certain resource. However, all these proposals consider a traditional Web scenario, with a limited amount of multimedia resources and a small fraction of requests that come from mobile devices and, consequently, require adaptation. In a future Mobile Web scenario, solutions that are based just on on-the-fly content adaptation may be not viable approaches.

VI. CONCLUSIONS

Mobile Web-based applications are placing increasing computational requirements on the server infrastructure due to the need of tailoring contents to user preferences and device capabilities. In this paper, we analyze the performance costs of present and future Mobile Web-based services taking into account the CPU server evolution and the main evolution trends of Web workload characteristics for the period 2008-2013. We show that the computational demand of Mobile Web-based services is likely to grow to the point where supporting on-the-fly adaptation for every resource becomes not affordable for server infrastructures. We propose resource management strategies based on pre-generation of popular contents to reduce the computational demand of future Mobile

Web-based services. The proposed resource management exploits the popularity distribution of Mobile Web workloads and represents a valuable solution for future server infrastructures to support Mobile Web-based services, even because the future mobile client devices will be able to consume a wider range of contents. We demonstrate that pre-generating just small fraction of popular resources (up to 15% of the working set) is sufficient to support the future workload even for the most demanding scenarios.

REFERENCES

[1] Gartner, "Mobile Web Trends 2007 to 2011." Research Report, Jun 2007.
[2] Berg Insight AB, "Mobile Internet 2.0." Research Report, May 2007.
[3] C. Canali, M. Colajanni, and R. Lancellotti, "Performance impact of future Mobile-Web based services on the server infrastructure," *IEEE Internet Computing (to appear)*, 2008.
[4] S. Flesca, S. Greco, A. Tagarelli, and E. Zumpano, "Mining user preferences, page content and usage to personalize Website navigation," *World Wide Web*, vol. 8, no. 3, pp. 317–345, Aug. 2005.
[5] G. Singh, "Guest editor's introduction: Content repurposing," *IEEE Multimedia*, vol. 11, no. 1, pp. 20–21, Mar. 2004.
[6] C. Canali, V. Cardellini, and R. Lancellotti, "Content adaptation architectures based on squid proxy server," *World Wide Web Journal*, vol. 9, no. 1, pp. 63–92, 2006.
[7] "Alexa - the Web Information Company," 2008, http://www.alexa.com/.
[8] T. Yamakami, "A zipf-like distribution of popularity and hits in the mobile Web pages with short life time," in *Proc. of the International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'06)*, Taipei, Taiwan, Dec. 2006.
[9] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proc. of the 7th ACM SIGCOMM conference on Internet measurement (IMC'07)*, San Diego, CA, Oct 2007.
[10] e Marketer, "Just How Dominant Is MySpace?. Jan. 2008."
[11] H. H. Sung, "Helping online customers decide through Web personalization," *Intelligent Systems*, vol. 17, no. 6, pp. 34– 43, 2002.
[12] G. Iaccarino, D. Malandrino, M. D. Percio, and V. Scarano, "Efficient edge-services for colorblind users," in *Poster Proc. of WWW 2006*, 2008.
[13] C.-Y. Chang and M.-S. Chen, "On exploring aggregate effect for efficient cache replacement in transcoding proxies," *IEEE Transactions on Parallel and Distributed Systems*, vol. 14, pp. 611–624, Jun. 2003.
[14] Juniper Research, "Mobile User Generated Content - Dating, Social Networking and Personal Content Delivery 2007-2012," Research Report, Aug. 2007.
[15] e Marketer, "Everyone Is Talking About Mobile Social Networking," Research Report, May 2008.
[16] "OMNeT++ Discrete Event Simulation System," 2008, – http://www.omnetpp.org.
[17] L. Guo, S. Chen, Z. Xiao, and X. Zhang, "Analysis of multimedia workloads with implications for Internet streaming," in *WWW '05: Proc. of the 14th International Conference on World Wide Web*, May 2005.
[18] M. Walker, R. Turnbull, and N. Sim, "Future mobile devices ? An overview of emerging device trends, and the impact on future converged services," *BT Technology Journal*, vol. 25, no. 2, pp. 120 – 125, Apr. 2007.
[19] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube traffic characterization: A view from edge," in *Proc. of Internet Measurement Conference (IMC'07)*, Oct. 2007.
[20] "iAnywhere Inc.", "AvantGo," 2005, http://www.avantgo.com/.
[21] J. Rodriguez, B. Dakar, F. L. Marras, C. Schreder, S. Suzuki, and E. Tapera, *New Capabilities in IBM WebSphere Transcoding Publisher Version 3.5 Extending Web Applications to the Pervasive World*. IBM RedBooks, May 2001.
[22] S. Buchholz and T. Buchholz, "Replica placement in adaptive Content Distribution Networks," in *SAC '04: Proc. of the 2004 ACM Symposium on Applied computing*, Mar. 2004.