Performance Evolution of Mobile Web-Based Services

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

Abstract

The widespread diffusion of the Mobile Web opens many interesting design and management issues about the server infrastructures that will have to satisfy the present and future client demand. Future Mobile Web-based services will have growing computational costs because even requests for the same Web resource will require the dynamic generation of different contents that will take into account specific devices, user profiles and contexts. This paper considers the evolution of the Mobile Web workload and of the technology trends of servers and client devices with the goal of evaluating possible bottlenecks and anticipating some management strategies for the server infrastructures that have to support the future services for the Mobile Web.

Keywords: H.3.4 [Information Storage and Retrieval]: Systems and Software - Performance evaluation; H.3.5: Online Information Services - Web-based services; J.9 [Mobile Applications]: Location-dependent and sensitive; J.9: Multimedia applications and multimedia signal processing.

1 Introduction

The great diffusion of Web-enabled mobile client devices, such as laptops, handheld PC, PDA and smart phones has significantly changed the Web scenario, because a growing amount of users can access vast amounts of data in the form of Web content. This evolution is determining the advent of the so called Mobile Web, where future users will access Web content at any time, from any location and through any class of devices. This mobile scenario, coupled with the complexity and diversification of the resources available on the Web, is opening the possibility of Mobile Web-based services that tailor the content to the user preferences, user locations and device capabilities.

For the design of future server infrastructures, it is of utmost importance to consider the performance requirements of Mobile Web-based services and to understand how their impact on the server infrastructure will evolve. Due to the heterogeneity of scenarios the main solutions require that Web content is dynamically generated and adapted on-the-fly at the moment of the client request [2, 5, 13]. However, on-the-fly generation and adaptation of content require computationally expensive operations. These costs are exacerbated by the increasing amount of innovative services offered on the Web. Specifically, the success of video sharing (e.g., YouTube), online games, virtual reality environments (e.g., Second Life) and mobile IPTV is causing a growing demand for graphical and multimedia content. Adapting these novel services to a mobile context will represent one of the major challenges for the future server infrastructures.

In order to evaluate the future performance requirements of Mobile Web-based services, we consider three main factors: the evolution of workload characteristics, the technological improvements at the level of server infrastructures and client devices. The trade-off should be clear: workload evolution and novel services will increase the computational demand of each user request, while the technology improvements of server platforms and client devices will allow the support of more demanding services. Assuming that no disruptive technology, (i.e., technologies with the potential to upset the current computation paradigm) will arise in the next decade, we evidence that the technology evolution may be insufficient to support a widespread diffusion of the Mobile Web-based services. The goal is to anticipate possible future system bottlenecks and to propose some management strategies to cope with them.

2 Mobile Web-based services

A key feature of the Mobile Web is the need of tailoring content to user preferences, user context and client devices. We find convenient to distinguish the three main categories of Mobile Web-based services: *personalization*, *context-awareness* and *content adaptation*.

Personalization aims to generate customized content to tailor the user navigation to his/her particular preferences and needs [5, 15]. Information about the user is typically stored in some database(s) and may be obtained as a consequence of explicit information coming from the user request or may be inferred through run-time or offline analysis of the user behavior (e.g., through data mining on log files of a Web site). This information is the basis to deploy different personalization services, such as subscription to news feeds, customized layout presentation, request filtering, recommendation systems [15], and adaptation to the user navigation style.

Context-awareness represents an innovative class of services that generate Web content on the basis of information about the context of the user. Context should be considered in its broad meaning that may include information about geo-location, ambient location, time of the day or current user activity [13]. This is one of the most innovative information concerning the Mobile Web, where the context and the user position may be continuously traced. Typical examples of context-awareness are location-based applications offering tourist information and commercial advertisements, and services that help the users to discover neighbor resources in their environment, such as devices, friends and members of some social community.

Content adaptation services transform original Web content to match the client device requirements in terms of network connection, computational power, storage and display size. Content adaptation may be applied to *images*, *audio* or *video* resources. Multimedia content represents a growing percentage of Web information, especially after the popularity of YouTube, online games and virtual words, such as Second Life. Hence, the server infrastructure should provide efficient transformations within multimedia types (e.g., changing the color depth of an image, converting from high-fidelity JPEG to low-fidelity GIF format), across multimedia types (e.g., from video item to a set of images) or both of them.

A problem of our study is related to the large variety of Mobile Web-based services that prevent the identification of "the typical Web site". Hence, we find convenient to identify three representative categories deriving from an analysis of the 100 worldwide most popular Web sites, according to the Media Metrix top 500 list (www.mediametrix.com) and the Alexa top 500 sites (www.alexa.com). From these lists we exclude the company sites and identify the *news-oriented* and *social-oriented* Web sites, that cover almost 80% of the sites and are likely to remain very popular also in the Mobile Web [1]. We identify a third category, namely *travel-oriented*, which represents an emerging type of Web site that is expected to have great diffusion among the mobile users [12]. We outline the main services characterizing these three categories of Web sites.

News-oriented. This category includes information portals, such as online newspapers or news broadcasting sites, that offer online information like events, stock quotes, and sports results. These sites typically deliver news in the form of textual 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 (www.stateofthenewsmedia.org). Textual resources are usually customized through personalization and context-aware services to provide users with information of interest based on specific preferences and contexts, while multimedia resources are tailored through some content adaptation services.

Social-oriented. The sites of this category 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). Other significant examples are online gaming and virtual worlds (e.g., Second Life) where users navigate, and socialize in three dimensions. Personalization and context-awareness account for almost 55% of the offered services and aim to suggest specific stories or information to the registered users. These sites rely also on content adaptation services to tailor multimedia contents (now 45% of the resources), as many present communications among the users involve exchanges of images, audio or video resources [3].

Travel-oriented. This category includes Web sites offering location-based information to support users mobility. Typical examples include sites for online travel reservations (e.g., Expedia) or sites offering touristic and traveling information (e.g., www.journeyplanner.org). Almost 65% of the offered services may involve personalization and context-awareness. Other services adapt multimedia resources illustrating touristic localities or events of interest (27% for images, 8% for audio and video).

3 Computational cost of Mobile Web-based services

We evaluate the cost of personalization, context-awareness and content adaptation services with the goal of identifying which service is likely to have the highest computational demand on the server infrastructure.

In our study, we consider as an example of personalization a *recommendation service* that is used by popular sites (e.g., Facebook and YouTube) to rate and categorize content and users. The recommendation service groups users with similar preferences or past click history [15] through offline collaborative filtering and clustering techniques. The resulting associations are stored in a recommendation database. When accessing the site, each user is identified as belonging to a specific group and the corresponding list of appropriate recommendations is retrieved from the database. The typical service time for a personalization service is in the order of one hundred of milliseconds because the mos expensive operations are carried out offline. Online operations in our experiments are implemented through typical open source technologies: Apache Web server, MySQL database servers and PHP applications.

As an example of context-awareness, we implement a *location-based service* that resembles the Food.com service. The service exploits information about the user current location and the time of the day to select a list of restaurants in the neighborhood and their daily menus and show them on the user map. These location-based services have a service time between one and four hundreds of milliseconds, depending on the number of accessed databases.

The computational cost of content adaptation is evaluated by considering services that are offered by popular mobile Web portals, such as AvantGo, that tailor contents of popular sites to the client device requirements (small displays, limited rendering capabilities and low-speed connections). The typical service time for content adaptation is in the order of one or more seconds because complex operations on possibly large files need to be carried out.

Adaptation of *images*, *audio*, and *video* resource requires a separate analysis because service time depends on the resource size. Images are usually adapted through different techniques, such as scaling, cropping irrelevant parts of the image, or increasing the compression factor [2]. In our experiments we evaluate the reduction of the spatial geometry from the original resolution to 480x320, which is a typical display size for a mobile device. In our evaluation we consider images in JPEG format with a median size of 80 KBytes, consistently with the typical working set of Web workloads [4].

Most of the audio resources accessed through the Web are music files in MP3 format with a median size of 3.5 Mbytes [9]. These resources are usually adapted by reducing the bit rate. We recode the audio resources to reduce the original bit rate to 32 Kbps.

If we exclude large videos related to the future mobile television where statistics are not available, the prevalence of video resources on the Web is represented by video clips in MPEG format with a median size of 7 MBytes and a duration of few minutes [9]. Common transformations concern frame size or color depth reduction. In our experiments, we reduce the frame size of short video clips from the original to 480x320.

For content adaptation we rely on open source software: ImageMagick library (www.imagemagick.org) to adapt the images, Lame (lame.sourceforge.net) and Transcode software (www.transcoding.org) for audio and video recoding. Various experiments on different server platforms (including a cluster of virtual servers) evidenced that the orders of magnitude for service times do not differ and that there is no significant difference among the respective results of the service times. Hence, we can state that the main conclusions of our performance evaluation do not depend on the hardware and software platform.

Figure 1 reports the 90 percentile of service time, which denotes that 90% of the requests are served within that value. This metric is more meaningful than average values in a context such as the Web, that is characterized by heavy-tailed distributions at the system and workload level.

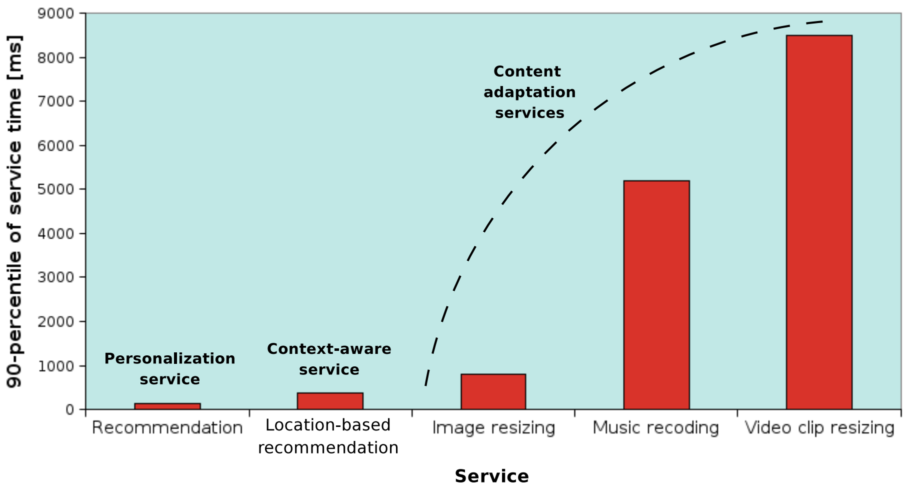


Figure 1: Service times for Mobile Web-based services

Figure 1 summarizes two results that emerge from all our experiments: the service times characterizing Mobile Web may span several orders of magnitude; the lowest service times are related to personalization and context-aware services. The former result was expected, while the latter may surprise because there are complex operations behind these services. The motivation is that the most time consuming tasks for personalization and context-awareness are related to data preparation through data mining, clustering and indexing techniques [15] that can be carried out offline. Hence, at the moment of the client request, the service has only to access one or multiple database(s) that are updated offline.

Another important factor is that the computational cost for personalization and context-aware services is unrelated to the size of the requested resource, because the service time is mainly related to database interactions. On the other hand, the service time for adaptation is directly correlated to the size of the resource being adapted [2]. The low service time for resizing images (with a size in the order of tens of Kbytes) and the high service times related to audio and video files (with a size in the order of multiple Mbytes) confirm this observation. In Table 1 we report the relationship between the resource size and the adaptation time. The first column shows the absolute service times per Mbyte, while the second column shows the normalized service times. Although the service times have a little value in absolute terms because they are obtained on a specific architecture, they demonstrate that the adaptation service times of different resources per unit size are in the same order of magnitude and the differences in Figure 1 are due to the "typical" size of a multimedia resource.

Table 1: Computational cost of content adaptation services

Adaptation service	**Service time per MB [ms]**	**Service time per MB (normalized)**
Image resizing	920	1
Audio recoding	1471	1.59
Video clip resizing	1254	1.36

Evolution of mobile client devices

The first generation of mobile Web-enabled devices was characterized by a wide spectrum of devices with heterogeneous characteristics in terms of network connection, computational power, storage and display size, ranging from laptops to smart phones with very limited capabilities. The technological evolution of the mobile devices is substantially improving the scenario of the mobile client population [1].
The bandwidth available to mobile devices has been greatly increased with the diffusion of 3-G wireless networks, that have determined the evolution from the few Kbps of the GSM technology to bandwidths in the order of Mbps. Further enhancements are expected with the advent of 4-G and WiMax technologies [2, 3].
The significant improvements in computational power have allowed even small devices to consume multimedia contents in several formats [4]. The computational power has also a significant impact on the rendering capabilities of the devices. While the first generation of devices ranged from monochrome to full-color capabilities, modern devices can display at least 16-bit color images, hence previous adaptations from color to B/W videos are now useless.
Processing power and storage resources of mobile devices are increased as well, thus allowing them to consume larger size resources and facilitating some content adaptation at the client side. The display size is likely to experience minor increments due to portability requirements, but the trend is towards devices with at least 3 inches screens and resolutions of 480x320 pixels or higher [1] as the mobile devices are being used to access multimedia resources.
The technological evolution of the mobile clients has a positive consequence on the server side because future server infrastructures will not need to tailor resources *exactly* for every type of client device as it happens now. In the next future, we can expect that different devices will be able to consume the same version of a multimedia resource with no server adaptation and, if necessary, some devices will carry out locally final adjustments. However, even if mobile clients become powerful devices with medium-large connections, limitations on energy and bandwidth will prevent client-side only adaptations.

References

[1] M. Walker et al., “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.

[2] T. B. Zahariadis, “Migration toward 4G wireless communications”, *IEEE Wireless Communications*, vol. 11, no. 3, pp. 6–7, June 2004.

[3] A. Ghosh et al., “Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential”, *IEEE Communications Magazine*, vol. 43, no. 2, pp. 129–136, Feb. 2005.

[4] Y-K. Chen et al., “Trends and challenges with system-on-chip technology for multimedia system design”, *Proc. of Emerging Information Technology Conference*, Santa Clara, CA, Mar. 2005.

4 Evolution of Mobile Web-based services

We study the evolution of workload characteristics and technologies with the goal of analyzing how the performance impact of the Mobile Web-based services is going to change in the next future. We consider the three main characteristics that may affect the service demand (workload intensity, workload composition and size of the workload resources) by taking into account some studies on workload trends [6, 11, 12, 17].

4.1 Workload intensity

The trend of the last years about the workload intensity denotes a stable request rate to online news-oriented sites. The percentage of Web users that get news online has been stabilized since 2000 (www.stateof-thenewsmedia.org/2007), but the requests increase because of mobile users [6]. The workload intensity of the travel-oriented Web sites is expected to grow by 29% per year during the next 6 years [12]. Social-oriented sites have achieved a more recent popularity, but the number of mobile users accessing these sites will grow extremely fast in the next six years [11]. A summary of the expected annual growth of the workload intensity for the considered sites is shown in Figure 2.

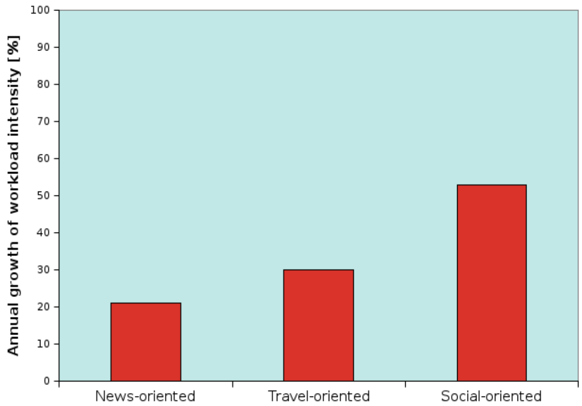


Figure 2: Expected growth of workload intensity for different categories of Web sites

4.2 Workload composition

The mix of content types is an important workload characteristic for evaluating the performance impact of Mobile Web-based services on the server infrastructure. Images, audio and video resources may require content adaptation services; textual resources may involve personalization or context-aware services that tailor, for example, page layout or recommendations based on user preferences and/or contexts.

The workload of the news-oriented scenario is characterized by an increasing amount of multimedia content, especially video and audio resources. An example of news site following this trend is cnn.com, that passed from text/image information in 1997 to the current rich mix of video and audio content. The expected trend of workload composition for the news-oriented scenario is shown in Figure 3(a). The solid lines represent the trend observed from 2004 to 2007, while the dotted lines in Figure 3(a) show the expected evolution until 2013. The data comes from a scientific report on american journalism (www.stateofthenewsmedia.org/2004) for the workload composition in 2004, and from an analysis of the home pages of five top news sites (CNN, Yahoo News, MSNBC, BBC, and AOL News) in June 2007. The trend of the last three years is assumed to continue for the next six years. Even the workload of social-oriented sites appears to evolve rapidly towards multimedia. From Slashdot, that was an early example of blog essentially based on textual content, modern social-oriented Web sites have evolved up to the most recent multimedia sharing sites [7], such as Youtube, where the user interactions occur through video clips. Due to the large variety of sites and the lack of scientific statistics about the percentages of resources composing the workload of social-oriented sites, we find convenient to assume that the workload composition follows multimedia increments similar to these of the news-oriented sites. On the other hand, the workload composition of the travel-oriented sites, that already contain significant percentages of multimedia contents, is not likely to experience abruptly changes.

From these analyses, we can derive that for the next future the news- and social-oriented sites will be characterized by a high increase of audio and video resources with respect to textual resources and images. As a consequence, we can expect that the adaptation of audio and video resources will represent the most expensive service.

4.3 Resource size

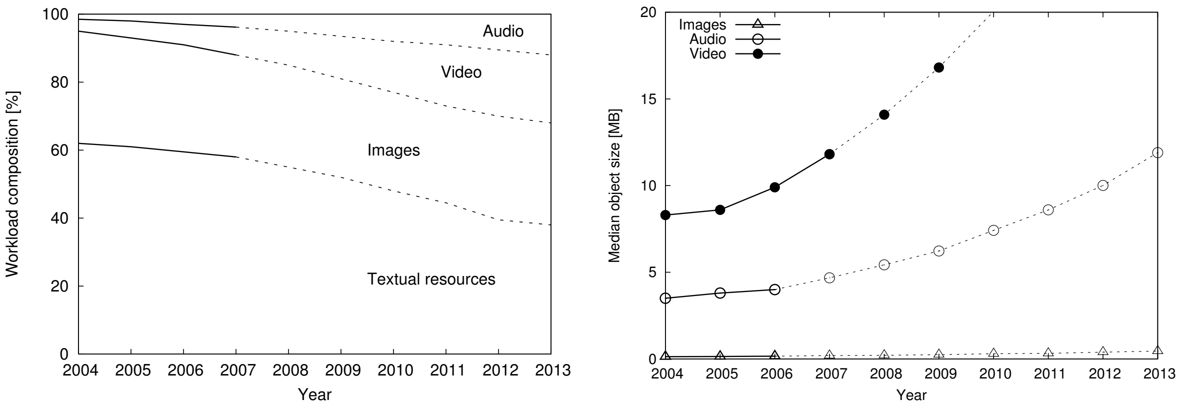
The third workload feature is the resource size. We focus on multimedia resources, because they affect the computational cost of the content adaptation services. To predict the trend of multimedia resource size we rely on multiple studies [2,7,9,10,17]. For images resources we refer to the Web workload characterization of Arlitt *et al.* [17], that identifies the size evolution during the period 1995-2004. We assume that the median image size will follow similar patterns throughout the decade 2004-2013. The expected trend is also confirmed by a more recent study [2]. For audio resources, we refer to measurements carried out in 2004 [9] and in 2006 [10]. For video resources, data come from analysis performed in 2004 [9] and 2007 [7]. For both audio and video resources, we expect that the trend observed from 2004 until now will continue for the next six years. Our measures on resource sizes consider sites that are operated by professional providers that tend to enforce size limits on multimedia resources (especially audio and video resources) to guarantee high quality experiences to the users. For this reason, we introduce a correction in the data obtained from non-professionals sites (such as [17]) by truncating the tail of the resource size distribution.

Figure 3(b) presents the expected trend of the median resource size for each multimedia type. The dotted line denotes the prediction trend for the future years. The median resource size is likely to increase for every multimedia type, but the highest increment concerns audio and video resources.

4.4 Computational demand vs. CPU computational power

From the previous analyses, we can derive that the computational demands of the Mobile Web-based services are going towards a significant growth due to the combination of multiple trends, such as the increments of request rate and resource size, and larger amounts of multimedia contents.

Along with the workload evolution, we should also consider that the computational power of the server platforms will improve as well. To this purpose, we refer to the Moore law, which asserts that the number of transistors on an integrated circuit is expected to double every 18 months. We assume that the same trend described by the Moore law can be applied to the computational power. This assumption is justified by the effect of the increased transistor density coupled with improved capabilities to exploit the internal parallelism due to better technologies on system bus and internal logic. Let us assume that in the next years this trend will remain valid and no disruptive technology will emerge. In this case, the computational costs of Mobile Web-based services are likely to follow a 64-fold decrease in the next decade as a consequence of the increasing computational power of the server platforms. Hence, if we consider the service times for



(a) Workload composition for the news-oriented scenario

(b) Workload resource size: all scenarios

Figure 3: Expected evolution of workload composition and resource size

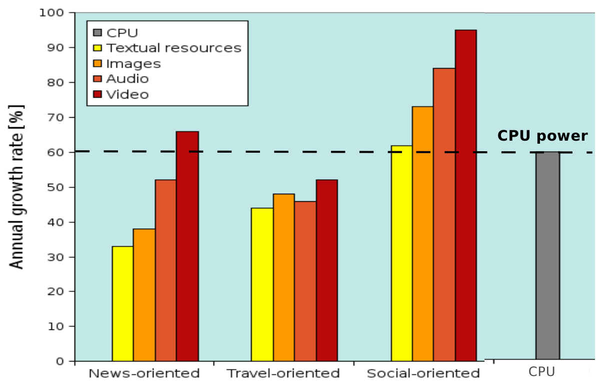
adapting multimedia resources shown in Table 1, in ten years they are expected to decrease between 15 and 20 ms per MB for any considered type of multimedia resource.

To evaluate the performance impact of Mobile Web-based services, we need to consider the combined effects of the trends related to workload intensity, workload composition and median resource size. Figure 4(a) reports the annual growth rate of the computational power that is required to serve each resource type. For example, the annual growth rate related to video content in the social-oriented scenario is 96%. This percentage is obtained by multiplying the growth factor of the median video size (16%) by the increment of the workload intensity for social-oriented sites (51%) and by the change of workload composition for that service, where the fraction of video content grows by 12% every year. In Figure 4(a) we report also the improvements of the CPU computational power per year. The dotted threshold denotes the services that are likely to cause problems at the server side. Figure 4(b) compares the trends of the expected increment of the computational power to serve each resource type and of the CPU power in the decade 2004-2013. We may observe that the computational demand of the services for the travel-oriented scenario is expected to increase slower than the CPU power. On the other hand, for the other scenarios some services are likely to experience a faster growth. In particular, the CPU power improvements do not seem to counterbalance the explosion of the computational demand that is related to the Mobile Web-based services in the social-oriented sites.

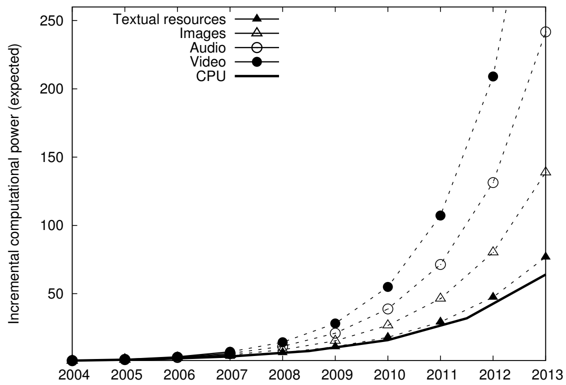
5 Consequences of the performance study

All the performance analyses indicate that the computational demand of Mobile Web-based services may grow faster than the CPU power of the servers to the extent that large investments will be required for supporting a widespread diffusion of mobile accesses. A critical decision for the support of future Mobile Web-based services is about the tasks that have to be carried out on-the-fly at the moment of the client request and tasks that may be performed offline. Personalization and context-aware services already carry out offline the most expensive tasks, and this solution will be enhanced through more efficient data and text mining techniques that will allow the pre-generation of content and server-side caching techniques for specific user profiles.

The adaptation of multimedia resources may have very large computational costs to the point where most present server infrastructures become inadequate. Different solutions are practicable depending on the



(a) All workload scenarios



(b) Social-oriented scenario

Figure 4: Predictive performance study of Mobile Web-based service computational costs vs. CPU computational power

nature of multimedia resources: *pre-recorded* and *live contents*.

5.1 Pre-recorded multimedia content

Pre-recorded contents are maintained on the servers and delivered at request, possibly after some adaptation. The technological evolution of the clients will allow future devices to consume directly or through some client-side adaptation a wide range of resources (see the sidebar “Evolution of mobile client devices”). This scenario opens novel content management and architectural solutions. Now, pre-generating or caching all versions of each pre-recorded multimedia resource is not feasible or convenient because there is a combinatorial number of alternatives deriving from different contents, devices, and network bandwidths. On the other hand, in the next future offline pre-generation and server-side caching of adapted multimedia resources can become a convenient way to limit expensive online adaptations. The effectiveness of caching solutions for multimedia content has already been demonstrated [3]. For example, caching the most popular 10% of the resources satisfies 80% of the requests due to the highly skewed popularity distribution of video-on-demand resources. Moreover, server-side caching is simple to deploy, and avoids the need to re-design the applications to generate offline all the versions of pre-recorded the contents.

Pre-generation and caching save CPU cycles at run-time, but causes a significant augment of storage costs, hence there is an interesting trade-off between CPU and storage investments. Storage requirements will increase because of the growth of the size and number of multimedia resources, but will decrease thanks to the lower number of versions that must be saved. For example, our experiments show that introducing five versions of each resource in a multimedia workload leads to just a twofold growth of the working set size. We confirm that the choice of storing multiple versions of multimedia resources will become a convenient alternative for the next future by comparing the evolution of the storage requirements with the trend of the storage capacity. To this purpose, we consider the social-oriented scenario that is expected to have the highest storage requirements. The storage capacity, in terms of disk density, has increased from 60% to 100% every 12 months during the last 15 years [8] at the same or lower cost per GByte. On the other hand, the storage requirement for the social-oriented scenario is expected to increase by 54% every 12 months for the period of interest. From this simple analysis we can conclude that if the storage capacity continues to experience an increment of at least 60% per year, the pre-generation and caching of adapted resources will became the most convenient alternative to manage pre-recorded multimedia content.

5.2 Live multimedia content

Live contents include event broadcasting, video conferencing, online gaming, for which there is no alternative to expensive on-the-fly adaptations. Server side scalability can be achieved through software and/or system solutions at the local or geographical level. At the software level, transcoding research is proposing novel algorithms to reduce the computational cost of content adaptation (e.g., [16]) and parallelism is often a viable solution. At the system level, modern and future clusters will be able to address most computational issues. However, when a very large number of streams must be delivered on a geographic scale, as in the case of IPTV services, recurring to third party infrastructures may represent a valid alternative. This approach can reduce the computational requirements on the central servers by more than 50% [14]. The two main third party options are not comparable now, but both them may result viable solutions in the next future. The consolidated approach relies on a Content Delivery Network (CDN), such as Akamai, that takes care of content adaptation and distribution through its geographically spread servers. A more innovative solution refers to peer-to-peer systems where the servers of the overlay network can collaborate for multimedia adaptation and delivery. There is yet some resistance to think to peer-to-peer infrastructures in a commercial context because there are various open issues at the level of availability, security and copyright infringements, however the efforts of various research groups and companies may modify a negative halo surrounding these systems.

References

[1] Berg Insight AB. Mobile Internet 2.0. Research Report, May 2007.

[2] C. Canali, V. Cardellini, and R. Lancellotti. Content adaptation architectures based on squid proxy server. *World Wide Web Journal*, 9(1):63–92, 2006.

[3] 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.

[4] S. Chandra, A. Gehani, C. S. Ellis, and A. Vahdat. Transcoding characteristics of Web images. In *Proceedings of Multimedia Computing and Networking (MMCN'01)*, San Jose, CA, Jan. 2001.

[5] S. Flesca, S. Greco, A. Tagarelli, and E. Zumpano. Mining user preferences, page content and usage to personalize Website navigation. *World Wide Web*, 8(3):317–345, Aug. 2005.

[6] Gartner. Mobile Web Trends 2007 to 2011. Research Report, Jun 2007.

[7] 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.

[8] E. Grochowski and R. D. Halem. Technological impact of magnetic hard disk drives on storage systems. *IBM Systems Journal*, 42(2), 2003.

[9] L. Guo, S. Chen, Z. Xiao, and X. Zhang. Analysis of multimedia workloads with implications for Internet streaming. In *WWW '05: Proceedings of the 14th International Conference on World Wide Web*, May 2005.

[10] L. Guo, E. Tan, S. Chen, Z. Xiao, O. Spatscheck, and X. Zhang. Delving into Iinternet streaming media delivery: a quality and resource utilization perspective. In *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, 2006.

[11] Juniper Research. Mobile User Generated Content - Dating, Social Networking and Personal Content Delivery 2007-2012. Aug. 2007.

[12] Juniper Research. Portable Navigation & Wireless Tracking - Western European Markets & Forecasts 2007-2012. Oct. 2007.

[13] E. Kaasinen. User needs for location-aware mobile services. *Personal Ubiquitous Computing*, 7(1):70–79, 2003.

[14] D. Liu, E. Setton, B. Shen, and S. Chen. Pat: Peer-assisted transcoding for overlay streaming to heterogenous devices. In *Proc. of NOSDAV'07*, 2007.

[15] H. H. Sung. Helping online customers decide through Web personalization. *Intelligent Systems*, 17(6):34– 43, 2002.

[16] A. Vetro, C. Christopulos, and H. Sun. Video transcoding architectures and techniques: an overview. *IEEE Signal Processing Magazine*, 20(2), Mar. 2003.

[17] A. Williams, M. Arlitt, C. Williamson, and K. Barker. Web Workload Characterization: Ten Years Later. *Web content delivery*, 2005. (Tang, Xu, Chanson eds.), Springer.