

An experimental framework to investigate context-aware schemes for content delivery

Pietro Lungaro⁺, Cristobal Viedma⁺, Zary Segall^{+*} and Pavan Kumar^{*}

⁺ Mobile Service Lab, Royal Institute of Technology (KTH), Kista, Stockholm

^{*} Department of Computer Science, University of Maryland, Baltimore County

Email: pietro@kth.se, crvi@kth.se, segall@kth.se, pavan4@umbc.edu

Abstract—A novel experimental approach for investigating the performances of context-aware content delivery schemes is presented in this paper. An innovative testbed, capable of remotely controlling multiple terminals, injecting a wide range of traffic loads in real networks and monitoring different performance measures has been developed and utilized for quantifying both the energy costs and user perceived service quality associated with different context-aware content pre-fetching schemes. In implementation proposed in this paper, the context information required for performing content pre-fetching is extracted and utilized by individual user terminals and does not require any support from mobile operators. The performances of pre-fetching are compared to those of an on-demand content delivery scheme, for both video streaming and file downloading services.

The results show that not only pre-fetching can increase user service appreciation by reducing the time needed to access the information, but it can also significantly lower the amount of energy consumed in user terminals for retrieving the content. Our experiments further indicate that in order to achieve these additional energy gains only limited content prediction capabilities are required, thus making pre-fetching a solid candidate for the provision of a wide range of content types and services in both wide and local area networks.

I. BACKGROUND

In the last couple years we have been witnessing an unprecedented adoption of mobile data services. Key factors behind this success are both the development of powerful user devices such as smartphones and tablets, capable of delivering superior levels of user experience, and flat rate pricing, which has made mobile Internet finally affordable to wider consumer segments. While on one hand the success of mobile data was somehow predictable, this was not the case for what concerns its rate of adoption and the associated traffic growth in mobile networks. Mobile operators expected a much gradual penetration and significantly milder increase in data traffic [1]. iPhone alone took within one and half years about 14% of the smartphones market share, while the Android platform is rapidly catching up. Moreover, each of these new smartphones is considered loading the network with approximately 60 times more data than previous “feature” phones. This combination has caught mobile operators off-guard, with under-dimensioned networks, in most of the cases leading to severe congestion problems [1]. Together with smartphones, a novel way of delivering and consuming mobile content has rapidly emerged: the apps paradigm. The success of the apps can be quantified in terms of both number of downloads from the various “app stores” and in terms of user face-time: recent surveys, e.g. [2], clearly show that apps have rapidly become the main channel for wireless content consumption. Currently the various apps fetch information from their remote content servers via either “push” or “pull” mechanisms and use the mobile network for completing the content delivery to the user terminals. In many cases the transferred content is characterized by some degrees of tolerance to delay and many data transfers are performed as “background” operations. With increasing numbers of apps per phone, and with smartphones gaining considerable shares

of the device market, it is not difficult to understand the extent of the additional traffic volumes injected into the mobile networks. At the same time, also the novel “cloud” paradigm is likely to stress even further the cellular networks. Services already existing, like Dropbox, and the incoming iCloud are essentially novel ways of expanding the storage and processing capabilities of the various devices by performing operations on files and content that are remotely stored. In order to effectively provide cloud-based services it is important that delays between remote storage locations and user devices are limited. As more services are in the cloud and more thin clients enters the market orders of magnitude in data traffic growth are predicted for cellular network.

In order to mitigate the effects of these novel traffic types we proposed and investigated in [3]-[6] a set of context-aware solutions for opportunistic content delivery. The basic idea is to optimize content delivery, in terms of investments in terminal and network resources, by exploiting the times and locations in which there is an excess of resources at the BSs. The content pre-fetching paradigm allows operators to opportunistically wait for users to roam closer to the BSs and/or to less loaded cells and then perform a more efficient content delivery. By de-coupling content delivery from consumption, this approach has the potential to improve the utilization of radio resources and to lower the OPEX and CAPEX required for meeting predetermined levels of user experience.

The importance of opportunistic content delivery has also been addressed in a series of previous publications, mainly targeting pre-fetching solutions in cellular networks, e.g. [7]-[11]. In some cases ([7],[9]), local storage of information at the terminals has been proposed to reduce the consumption of wireless resources for frequently accessed data items. In other investigations, pre-fetching solutions have been suggested for reducing the effects of channel quality fluctuations ([8]) and improving the performances of (streaming) protocols in wireless environments ([10],[11]).

II. PROBLEM

Evaluating the potential benefits of context-aware content delivery schemes, in terms of network and terminal resource utilization, together with user service perception (QoE), is of paramount importance for the success and sustainability of providing future data intensive services in cellular networks. This paper presents a novel experimental approach, developed within the COSEM project at Wireless@KTH. At the core of it is the “COSEM living laboratory” (C2L) testbed, which allows to remotely control a number of mobile devices injecting in real cellular networks “synthetic” workloads and collecting results on all monitored variables, such as bytes sent and received, signal level, current network or battery status among others, from the controlled devices. By supporting different

types of “synthetic” workloads, the C2L testbed is used in this paper to compare the impact on terminal battery consumption and user service perception of both context-aware content pre-fetching and on-demand content delivery. Furthermore, we quantify the magnitude of the potential energy savings that can be achieved by including different types of context information, e.g. signal quality measurements and estimates of attainable datarates, in the content pre-fetching scheduling decisions. These are, in the considered implementation, taken by individual terminals without support from the mobile operators (over-the-top pre-fetching).

Since pre-fetching content that is not accessed by the end users can cause an increased energy consumption in user terminals, in this paper we experimentally quantify the thresholds on content access prediction required to keep the energy costs, introduced by pre-fetching, comparable, or even inferior, to the on-demand content provision.

III. C2L TESTBED

C2L testbed is a novel approach for evaluating service provision in cellular networks in a testbed oriented living laboratory setting. Its goal is to identify, test and evaluate key indicators affecting QoE and QoS for mobile services and networks under real scenario conditions. The mobile workload generator tool allows defining, uploading and controlling tasks in multiple handsets simultaneously. These tasks include downloading a user-controlled sized package, streaming, as well as to download locally and play, video and audio, monitoring different parameters of the devices without performing any activity, etc. The testbed architecture consists of two different components which inter-operate with each other: The *Melange Server*, and the *Melange Mobile Application*. The **Melange Server** process all incoming and outgoing information to the mobile nodes. In particular, it handles the requests sent from and to the terminals such as sending network statistics (push) and getting workload information (pull). It is also in charge of triggering the *Cloud To Device Messaging (C2DM)* server that notifies terminals of new experiments. Additionally, the entire system is exposed to the user through a web-based interface which is served by the Melange Server. This web interface has provisions to manage terminals, control them remotely, generate tasks, workloads and experiments, processing statistics and displaying results. The **Melange Mobile Application** is responsible to execute the required tasks and to record vital information about various parameters such as cell id, location, signal strength, battery capacity, battery voltage, battery drain percentage, network statistics such as Bytes sent and receive, and application usage information. It registers with the C2DM to receive notifications of upcoming experiments and sends the results to the Melange server after the experiment is executed. As we can see in the overall architecture (Figure 1), the researcher can create and deploy an experiment to the testbed via the web interface. Subsequently the Melange server notifies the C2DM server the availability of an experiment. This server will then send an push message to the terminal to notify the availability of an experiment. The terminal pulls the experiment description from the Melange server and starts executing the workload of the experiment. When the execution has finalized and the results are ready, it returns them to the Melange platform. Finally, the researcher can visualize the results on the web interface or download them to process them in a different tool. The testbed works seamlessly and independently of any service provider or device manufacturer

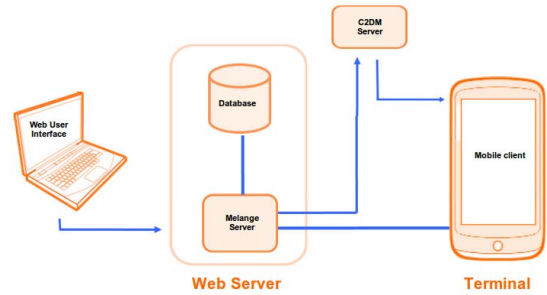


Fig. 1. COSEM Living Laboratory (C2L) Architecture

using the Android platform. The system can manage mobile nodes spanning across different remote locations and allows different combinations of heterogeneous workloads. Additionally, the workload scheduling feature permits the users to deploy workloads at a future time instance.

IV. SERVICES AND CONTENT DELIVERY METHODS

In order to test the performances of context-aware content provision in mobile networks a set of relevant service types have been considered in our experiments. Since significant amount of resources are currently consumed in mobile networks for **videos streaming**, YouTube-like video services have been included in the set of workloads supported by the C2L testbed. **Audio streaming** represents another relevant category of services that are frequently used by many users. Music services like Spotify, Pandora and the incoming iCloud services are constantly increasing in popularity. Even though each individual song does not consume significant amounts of bandwidth and terminal energy, these services are typically left running for several minutes at the time, thus potentially consuming non negligible volumes of data, and associated battery energy, throughout the overall service duration. The third category of services considered in our experiments are **file downloads**. Apart for the well-established email services, with associated download of attachments, there is a clear trend towards widespread adoption of services like Dropbox, effectively extending terminal memory in the cloud.

The standard “on-demand” provision of the aforementioned services is in this paper compared to one in which content is opportunistically pre-fetched in user terminals when and where the network conditions are better suited. Two triggers, representing different degrees of network context-awareness, are considered in this paper: signal quality and average data rate. While information concerning the achievable data rate can only be estimated after starting downloading an initial portion of the intended file (e.g. few KBytes), signal quality information is instead already available in user terminals, e.g. by monitoring the pilot signal-to-noise ratio (CPICH). However, if even if “cheaper”, in terms of terminal energy, signal quality does not fully reflect the levels of congestion and of users’ competition for resources in a given cell.

Apart from potentially reducing energy costs through achieving higher datarates, content pre-fetching can also lower energy by performing content delivery as background operation with terminal screens off. Since screen illumination consumes a significant part of battery capacity, this proposed approach is likely to further reduce energy consumption.

A. Energy Modeling

Since one of the goals of our investigation is assessing the overall energy costs associated with the alternative methods

for service provision, we performed a series of tests for characterizing the individual energy costs of the various phases in which the aforementioned services can be divided.

In particular, in our model a file download service is considered consisting only of a data retrieval phase, which can be performed through the different radio interfaces available in our test terminals (WiFi and HSDPA). Furthermore, when providing this service via pre-fetching the terminal screen is set to off for the whole duration of the download. On the contrary, when the service is provided in an on-demand fashion, the terminal screen is triggered to on, as it is a normal condition in all explicit user interactions with the devices.

From the results obtained in the various tests we have been able to confirm that the amount of energy consumed in a download phase increases linearly with its duration. Different coefficients have been identified, depending on the active wireless interface and on the screen status. With ϵ_w^{on} and ϵ_w^{off} we indicate the amount of energy spent during a second of download activity using WiFi and with screen on and off, respectively. Similarly, with ϵ_h^{on} and ϵ_h^{off} , we refer to the amount of energy spent during a second of download using HSPA and with screen set to on and off, respectively.

For what concerns the video service provision, the results of our tests have shown that the energy cost associated with playing a video on the device screen follows a linear relationship with the duration of the video itself and its angular coefficient is indicated, in this paper, with ϵ_v . Since video pre-fetching consists of two independent phases, first download and then play the video at a different time, its overall energy cost can be computed as the sum of the individual amounts of energy spent in each phase. On the contrary, when streaming on-demand a video, the playing phase is temporally superimposed to the download phase and with only partial overlap. Throughout our tests we have not been able to identify an explicit relationship between the energy cost of a streaming service and its overall duration. We suspect that energy cost varies in function of how the video buffer evolves in time, e.g. the duration of the overlap between video play and data retrieval and if there are some cases of buffer starvation.

B. Pre-fetching and prediction

Exploiting context-aware information can save substantial amounts of energy for completing the delivery of a specific data object to a user terminal. These energy savings can be quantified in terms of number of additional data objects (with the same size of the wanted one) that can be delivered to the terminal while reaching the identical total energy cost to the on-demand content provision. By defining $\bar{E}_s[t, I_i, \bar{R}(t)]$ as the total energy cost for streaming a video I_i of size B_i and duration T_v^i to a user terminal, with average datarate $\bar{R}(t)$ available at time t , and $\bar{E}_{pf}[t, I_i, \bar{R}(t - \Delta_t)]$ as the energy for pre-fetching at time $t - \Delta_t$ and playing the same video, the number of additional items that can be pre-fetched can be computed in the following way:

$$N_{pf} = \frac{\bar{E}_s[t, I_i, \bar{R}(t)] - \bar{E}_{pf}[t - \Delta_t, I_i, \bar{R}(t - \Delta_t)]}{\epsilon_x^{off} \cdot [B_i / \bar{R}(t - \Delta_t)]},$$

where ϵ_x^{off} is equal to ϵ_w^{off} or ϵ_h^{off} depending on the specific radio interface used for pre-fetching. Thus, in order for pre-fetching to deliver improved user experience without increasing energy costs it is required that a user accesses the pool of pre-fetched items with a probability larger or equal

than $1/(N_{pf} + 1)$. This constitutes a threshold on content predictability that can be used to assess, based on past user's access patterns, which content and service types are more suitable to be provided via pre-fetching.

V. INVESTIGATION

A. Performance measures

In order to quantify the energy impact of the providing the considered services via on demand or via context-aware content pre-fetching the **battery share** β_E is considered. This represents the ratio between the overall energy spent for a specific service provision and the total battery capacity.

Since the performances of pre-fetching depend on the capability of predicting which content will be requested by the end users in the future, and this varies with both users and content types, in this investigation we consider the **break-even cache hit ratio** \hat{t}_{pf} as a measure of effectiveness for pre-fetching schemes. Following the notation introduced in Section IV-B, the cache hit ratio is here defined as $\hat{t}_{pf} = 1/(N_{pf} + 1)$.

For what concerns a measure of user experience for streaming services we considered T_{out} the overall **duration of the interrupted playtime**, including both the initial buffering phase and the overall duration of potential interruptions due to buffer starvation. Recent studies have shown that users are particular annoyed by more than one case of video stalling and that about 30 seconds of total video interruption seems to be an hard limit for user perception [12].

B. Experimental settings

A significant number of experiments (over 400) have been performed considering the service types and delivery methods described in Section IV. These included experiments with terminals accessing both HSDPA and WiFi networks, with the screen turned on and off performing different tasks. Various locations within the city of Stockholm, Sweden, have been considered together with different hours of the day, since data rate and traffic variations are highly dependent on both location and time. In particular the locations chosen are one in downtown Stockholm (DT) corresponding to Östermalmstorg, one (NC) very close to the previous location and corresponding to the underground station of Kungsträdgården and finally the location of our lab in a suburban area (SU) of Stockholm.

The HSDPA networks of three of the major mobile operators in Sweden have been used in the study, accessed through Google Nexus One terminals equipped with Android 2.3.4 (Gingerbread) operating system. For the testing terminals the values associated to the coefficients described in Section IV-A have been experimentally determined and are $\epsilon_w^{on}=190.08$ mA, $\epsilon_w^{off}=69.12$ mA, $\epsilon_h^{on}=246.24$ mA and finally $\epsilon_h^{off}=116.64$ mA.

The C2L Melange server is implemented on a Linux Red Hat 4.1.2-48 with an Intel Xeon 64 bit CPU E5620 at 2.40GHz with 29184MB of RAM machine. For performing the streaming tests we considered a specific YouTube video (MP4) characterized by resolution of 640 x 360, with file size of 9.29 MBytes and playtime of 111 s.

C. Results: exploiting geographical traffic variations

A series of video streaming experiments have been targeting the capability of context-aware content delivery schemes to exploit geographical traffic variations, thus leading to improved terminal performances and user experience. Considering the aforementioned locations, all tests have been running between 11:30 and 13:30. This time interval, roughly corresponding to

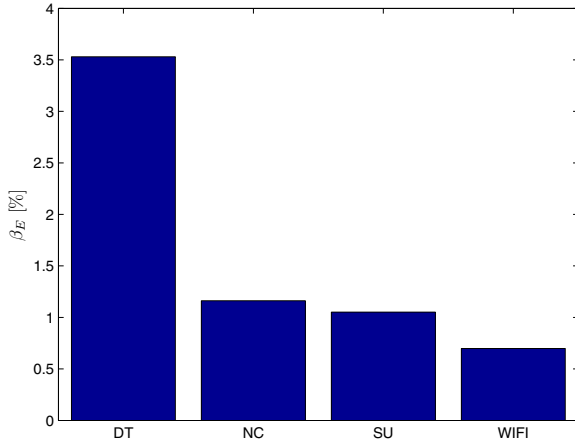


Fig. 2. Average percentage of battery consumption β_E for the video streaming service in different locations. The video is provided via streaming in DT, while for all other locations β_E refers to the energy costs associated with pre-fetching and playing the same video.

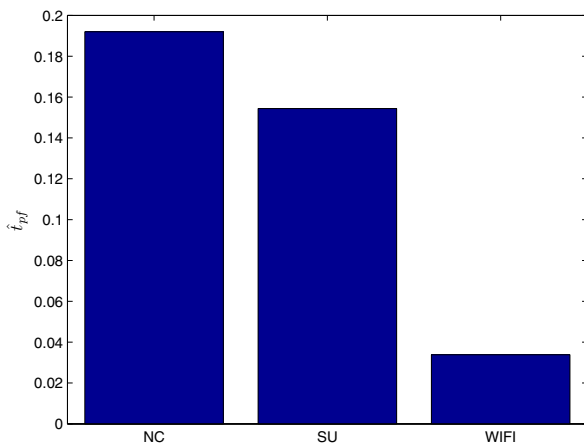


Fig. 3. Break-even cache hit ratios $\hat{t}_{p/f}$ for the three pre-fetching locations when the energy costs of streaming in DT are chosen as reference.

the lunch break, was chosen since it is likely to be one of the periods with higher data traffic in the mobile networks.

In Figure 2 the average percentage of battery consumption β_E is shown for the different locations and different content delivery methods. The video is streamed on demand only for the experiments performed in DT, while in all other location we delivered the same content via pre-fetching: the underlying scenario considers a nomadic user, roaming throughout the town, that requests and consumes the wanted video content only upon reaching DT. The values of β_E for the pre-fetching cases include also the total energy cost associated with playing the video. With WIFI, we refer to the performances obtained utilizing the wireless network deployed in our facilities (SU), however we expect that similar performances can be achieved on other WLANs. The results show that including information on the achievable datarate, in the set of triggers for content pre-fetching, can dramatically reduce overall energy costs.

The break-even cache hit ratios associated with these experiments are illustrated in Figure 3, where the energy costs of streaming in DT are used as reference for evaluating the required content predictability in all considered pre-fetching cases. For the WIFI case, the results show that pre-fetching can lead to both improved user experience (no delivery delay) and energy saving even for objects with a future access probability

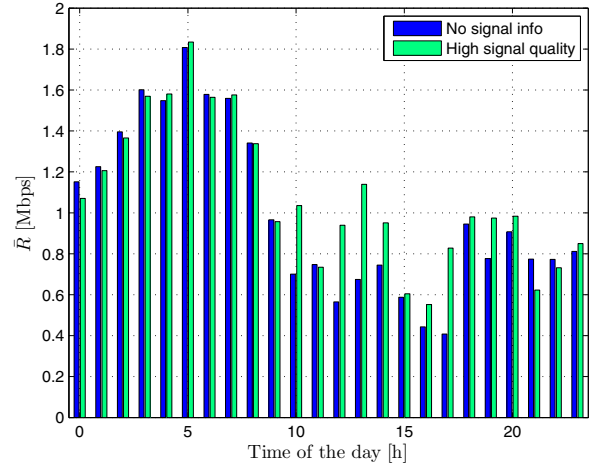


Fig. 4. Expected datarates achieved at location SU during 24 hours, when considering, or not, signal quality information before starting movie download.

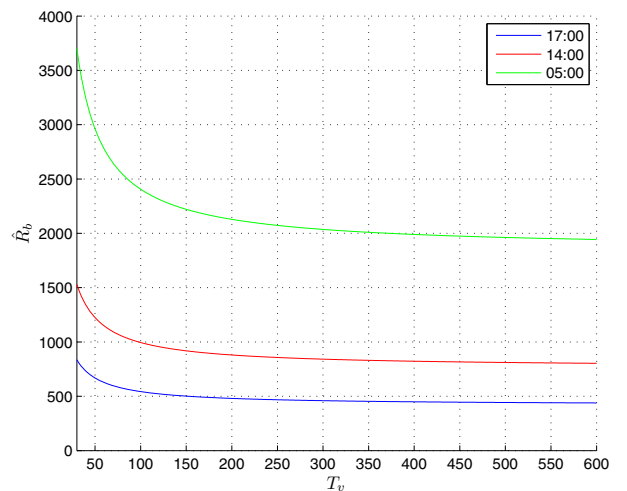


Fig. 5. Combinations of video bitrates \hat{R}_b and playtime durations T_v that achieve $T_{out} = 30$ s, for three selected hours of the day in the SU location.

as low as 3%. Slightly higher, but still below 19%, is the content access predictability required for triggering “energy saving” pre-fetching using HSDPA, and this is obtained in congested hours. Information on videos shared in social networks, video updates from channel subscriptions on YouTube and content pushed through dedicated apps are all likely to support terminal agents in identifying personalized content meeting the required levels of access predictability.

D. Results: exploiting temporal traffic variations

The expected value of the datarate that can be achieved during each hour is shown in Figure 4 for an entire day at location SU. Every 5 minutes, a video download has been triggered in all the controlled terminals, each using an HSDPA connection with a different mobile operator. Every of the displayed values is obtained by averaging all realizations that started after the beginning of a given hour and ended before the following one. Two cases are displayed, one in which signal quality information is considered before starting the video download and one in which this information is not available.

The results clearly show a period with increased traffic activity between 10:00 and 17:00, coinciding with the subset of working hours where the majority of employee are present. In these most congested hours of the day the usage of context information, concerning signal quality, is shown to improve the

TABLE I
OD (HSDPA) VS. PF (WIFI) ENERGY COMPARISON FOR DOWNLOADING

$\bar{E}_s^h/\bar{E}_{pf}^w$	$\bar{R}_{pf}^w=1$	$\bar{R}_{pf}^w=2$	$\bar{R}_{pf}^w=4$	$\bar{R}_{pf}^w=6$
$\bar{R}_{od}^h=0.1$	35.5	70.9	106.4	212.8
$\bar{R}_{od}^h=0.5$	7.1	14.2	21.3	42.6
$\bar{R}_{od}^h=1$	3.5	7.1	10.6	21.3
$\bar{R}_{od}^h=2$	1.8	3.5	5.3	10.6

TABLE II
OD (HSDPA) VS. PF (HSDPA) ENERGY COMPARISON FOR DOWNLOADING

$\bar{E}_s^h/\bar{E}_{pf}^h$	$\bar{R}_{pf}^h=0.1$	$\bar{R}_{pf}^h=0.5$	$\bar{R}_{pf}^h=1$	$\bar{R}_{pf}^h=2$
$\bar{R}_{od}^h=0.1$	2.1	10.4	20.7	41.5
$\bar{R}_{od}^h=0.5$	0.4	2.1	4.1	8.3
$\bar{R}_{od}^h=1$	0.2	1.0	2.1	4.1
$\bar{R}_{od}^h=2$	0.1	0.5	1.0	2.1

average datarate achievable with pre-fetching of about 42%.

The expected datarates shown in Figure 4 have been further used to understand which type of quality can be provided for on-demand video streaming services, while delivering acceptable QoE to the end users. In Figure 5 the combinations of video bitrates \hat{R}_b and playtime durations T_v that can deliver a values of T_{out} smaller than 30 s are shown for three different hours of the day: 17:00 (lowest \bar{R}), 14:00 (highest \bar{R} in the “busy” period) and 05:00 (highest overall \bar{R}). The total durations of the video interruptions are computed considering the average data rates available in the selected hours and comparing their required time to complete the video download ($\hat{R}_b \cdot T_v / \bar{R}$) with the video playtime. Note that the points of the video quality plane below each line are considered feasible, from a QoE perspective, for the expected datarates achievable in its corresponding hour. Considering that the highest video qualities on YouTube, 720x1280 and 1080x1920, are typically associated with bitrates ranging between 2 and 2.9 Mbps and 3.5 and 5 Mbps respectively [13], only medium and low quality video can be effectively provided in the considered location, and not in every hour of the day.

This result confirms content pre-fetching as a very promising solution for circumventing current datarate limitations in wide area networks and for delivering high quality videos with a superior user service perception.

E. Results: Downloading services

Based on the energy coefficients described in Section IV-A, the results concerning the energy costs of pre-fetching versus on demand content provision are presented in Table I and Table II for the file download services. A set of achievable datarates, expressed in Mbps, are considered in both tables, for both content provision methods. In I, the ratio between the energy costs of delivering a file on-demand via HSDPA and pre-fetching on WiFi ($\bar{E}_s^h/\bar{E}_{pf}^w$) is computed for average WiFi rates (\bar{R}_{pf}^w) between 1 and 6 Mbps and average HSDPA rates (\bar{R}_{od}^h) ranging between 0.1 and 2 Mbps. Note that for the case of file download services this measure coincides with the number of bits that can be served by pre-fetching for each bit served on-demand, while obtaining an identical energy cost. The results show that for all evaluated rates pre-fetching delivers significant improvements in terms of energy efficiency. Moreover, being able to provide large amounts of bits at low cost decreases also the requirements on content prediction. In Table II a similar investigation is presented, this time considering pre-fetching on HSDPA network. As expected energy improvements can be achieved only when pre-fetching is performed at rates larger than with on-demand.

In the case of equal rates, the performance variation depends on the different energy costs associated with keeping on, or off, the terminal screens while completing the file download.

VI. CONCLUSION

A novel experimental approach for investigating the performances of context-aware content delivery schemes has been presented in this paper. The developed C2L testbed allows to remotely control a set of mobile terminals, generate a wide range of different service workloads and monitor various performance measures. In the specific studies considered in this paper, the testbed has been used to quantify both the energy costs and the level of QoE that can be achieved, in real networks, by adopting context-aware content pre-fetching schemes at user terminals (OTT). These solutions have been tested for both video streaming and file download services and their performances compared to those obtained with “standard” on-demand service provision methods.

The results showed that not only pre-fetching can increase user service appreciation by reducing the time for accessing content, but it can also dramatically reduce the amount of energy consumed in user terminals. In order to achieve these energy gains only limited content prediction capabilities are required, making pre-fetching a solid candidate for the provision of a wide range of content types in both wide and local area networks. While the majority of performance improvements are obtained by estimating achievable datarates, e.g. with probing packets, signal quality estimation can further improve performances during peak hours.

The limited range of data rates currently provided by wide area networks poses severe limitations to the quality of the user experience for streaming services. Thus, with increased penetration of advanced terminals and more services adopting cloud-based solutions the current on-demand service paradigm is likely not to be sustainable. On the contrary, the alternative context- and content-aware approach considered in this paper is a promising candidate for supporting the success and sustainability of future data intensive services in mobile networks.

REFERENCES

- [1] Roben Farzad, “AT&T’s iPhone Mess”, article on Businessweek, 15th of February 2010.
- [2] Zokem press release on smartphone usage, 14th of February 2011.
- [3] P. Lungaro, Z. Segall and J. Zander, “ContextShift: a Model for Efficient Delivery of Content in Mobile Networks”, in Proceedings of the Wireless Communications & Networking Conference (WCNC10), April 2010.
- [4] P. Lungaro, Z. Segall and J. Zander, “Predictive and Context-based Multimedia Content Delivery in Future Cellular Networks”, in Proceedings of the Vehicular Technology Conference (VTC10), May 2010.
- [5] P. Lungaro, Z. Segall and J. Zander, “Context-aware Resource Schedulers for Opportunistic Content Delivery in Cellular Networks”, in Proceedings of the 3rd international conference on Communication Theory, Reliability, and Quality of Service (CTRQ10), June 2010.
- [6] P. Lungaro, Z. Segall and J. Zander, “ActiveCast - a network and user aware mobile content delivery system”, in Proceedings of the 2nd conference on Ubiquitous and Future Networks (ICUFN10), June 2010.
- [7] D. Barbara, T. Imielinski, “Sleepers and Workaholics: Caching Strategies in Mobile Environments”, in Proceedings of SIGMOD94, 1994.
- [8] S. Gitzenis, N. Bambos, “Power-Controlled Data Prefetching/Caching in Wireless Packet Networks”, in Proceedings of INFOCOM02, 2002.
- [9] Y. Lin et al., “Effects of Cache Mechanism on Wireless Data Access”, IEEE Transactions on Wireless Communications, Vol.2, November 2003.
- [10] F. Fitzek, M. Reisslein, “A Prefetching Protocol for Continuous Media Streaming in Wireless Environments”, IEEE Journal on Selected Areas in Communications, Vol.19, November 2001.
- [11] J. Hu, G. Feng, K. Yeung, “Hierarchical Cache Design for Enhancing TCP Over Heterogeneous Network with Wired and Wireless Links”, IEEE Transactions on Wireless Communications, Vol.2, March 2003.
- [12] Acision Mobile Broadband Research, “Seizing the opportunity in Mobile Broadband - United States Perspective”, January 2011.
- [13] <http://en.wikipedia.org/wiki/YouTube>, last accessed on June 15th 2011.