2013 IEEE 14th International Conference on Mobile Data Management

Selective offload and proactive caching of mobile data in LTE-based urban networks

Sabrina Gaito, Dario Maggiorini, Christian Quadri, Gian Paolo Rossi
Computer Science Dept.
Università degli Studi di Milano, Milan, Italy
Email: firstname.lastname@unimi.it

***Abstract*—In this paper we focus on mobile data offloading and propose a solution for the placement of offloading infrastructures that optimizes the trade-off between reducing deployed resources and increasing traffic breakout. We base our analysis on a real dataset of Internet accesses generated, in the city of Milano, by some 50,000 users of an important mobile network operator. The target application we consider is the distribution of digital contents (such as MP3 songs, videos and newspapers) over an urban area. The paper's contributions is showing that offloading of digital contents can be achieved by leveraging people's regular mobility patterns.**

Keywords—Mobile data offload, proactive content caching, LTE Advanced, real data analysis

I. Introduction

Today's core network of mobile operators, as well as the relevant radio access networks, are affected by unprecedented volumes of data traffic that will keep on growing at the estimated mean annual rate of 78% from 2012 to 2016 [1]. Operators are therefore in urgent need of solutions that focus on network capacity scaling, that reduce the rate of expensive infrastructure upgrades and that limit resorting to use pricing as a congestion control tool. The selective offload of data traffic is now considered to be the most promising solution to the problem.

This paper deals with offloading from the perspective of a mobile operator that has planned the deployment of the offload infrastructure in a given metropolitan area based on LTE architecture. Such an operator is in need of a quantitative approach to plan and dimension the placement of the infrastructure, as well as to minimize installation costs while maximizing the volume of data traffic routed outside its core network.

The key premise of the paper is the privileged access to a portion of a real dataset of Internet accesses generated, in the city of Milano, by some 50,000 users of an important mobile network operator. By combining a tightly integrated offload architecture and the analysis of a real dataset, in this paper we show that an offloading infrastructure may be placed by simply leveraging regular human mobility patterns. We provide a preliminary evaluation from the operator's perspective to show that the tradeoff between the offload infrastructure costs and the amount of offloaded data contents can be optimized in a practical scenario. Finally, by assuming to equip the offloading platform with data kiosks, we sketch the design of a proactive content distribution service for the metropolitan area of Milano.

II. Dataset Description

The real dataset we use in the following analysis contains data about the Internet activities of a set of anonymized users of a leading mobile operator in Italy. Data are limited to users moving in the metropolitan area of Milano. The time window of the dataset spans from 14th May to 27th May, for a total of two weeks. The first week has been used to model the service, while the second represents the dataset we adopted to run simulations. Each user's activity is recorded with the following information: date, time, cell ID of the cell where the session started. It is worth observing that the user's location is based on cells where the s/he is active, namely the access cells, or ACs, and we have no information about the user's position during the inactive periods. We preprocessed the dataset in order to identify the users who are relevant to our goal. We consider only those users $active$ each day (a user is active when s/he generates at least one activity per day). Moreover, the selection applies to working days only, because mobility patterns and content requests are very different during the week-end. Finally, as a consequence of the fact that we are considering a service where mobility plays a central role, we limited the users in the dataset to those visiting at least two different cells every day.

After the preprocessing we obtained a dataset of 49,067 users who were active in up to 1,716 cells. This is the set we will consider throughout the paper.

III. The Placement of Offload Infrastructure

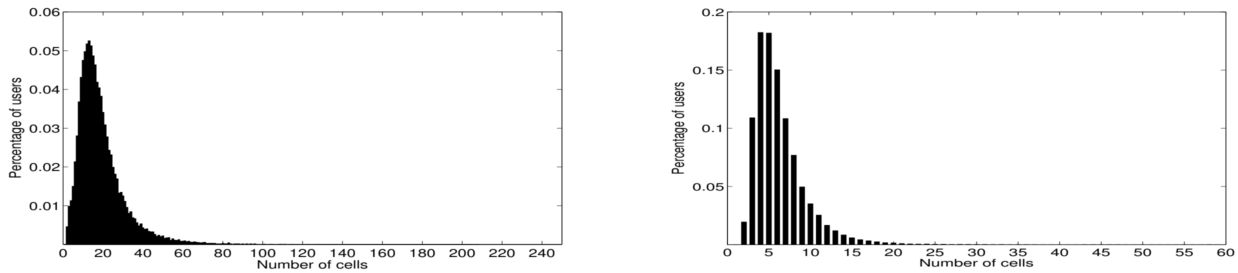
In this section we deal with the problem of deploying the offload infrastructure by equipping the minimum set of operator's cells. The goal is to reduce the installation costs for networking resources while ensuring high volumes of data offloading. Initially, we will assume a $push$ content delivery schema, that is, the service autonomously pushes contents to the users as soon as they happen to be under coverage of cell selected by the operator. To this purpose, a deeper understanding of the user's attitude to generate traffic over the cells would enable us to identify the subset of the 1,716 cells that will be involved in the deployment process.

A. Cell-based Activity Patterns

To approach the described problem, we extract all per-user activity patterns, where an activity pattern is defined as the set of cells where a given user is active.

We initially consider the overall number of cells involved in the activity patterns of users during the first week; for each

978-0-7695-4973-6/13 $26.00 © 2013 IEEE
DOI 10.1109/MDM.2013.39



(a)

(b)

Fig. 1. (a) Histogram of the number of cells accessed by users during the entire week (from 8:00 A.M. to 6:00 P.M.). (b) Histogram of the mean (over a week) number of cells a user accesses during a day (from 8:00 A.M. to 6:00 P.M.).

user we compute the number of different cells he/she accesses during the whole first week and we show the related histogram in Fig. 1(a). The mean value of 19 cells and the median of 16 show that, apart from a tail of users accessing a very large number of cells, most have activity in a few ACs per week. This behavior is even more evident if the same quantity is considered on a daily basis.

In Fig. 1(b) we report the histogram of the mean (over the week) number of ACs accessed by user in a day. In line with similar studies [2], [3], [4], [5], [6], most users are active in very few different locations/cells (mean: 6, median: 5), while a very small percentage of individuals show a higher mobility.

To properly target the deployment phase, the operator must understand whether or not the set of ACs is changing daily. We extract this information by analyzing the set of per-user regularly accessed cells during the considered week. We have found that 60% of users have at least one favorite cell that they access daily.

The above results clearly state that individuals are accustomed to following regular mobility patterns that include a small number of ACs visited daily. This information must be combined with the minimum number of cells required to serve a single user during the entire week. We have found that more than 90% of users require only two cells at most: 60% require one cell, the remaining 30% of users require only two cells.

B. Minimizing the number of cells

In this section we face the issue of finding the best selection of cells where to deploy the offload infrastructure. The set of above per-user preferred cells may be a good candidate for the deployment. In fact, by placing the offload infrastructure and by proactively caching the contents according to a per-user subscription, the operator would certainly be able to daily $push$ the subscribed content to all users, thus ensuring that 100% of the offload data are efficiently routed outside the core network. Nonetheless, the general operator's achievement of the highest traffic breakout at the lowest cost can be pursued by attempting a cell optimization that selects cells shared among different users. For this purpose, we solve an instance of standard formulation of Set Covering applied to the first week of the dataset.

Our minimization problem is constrained to ensure coverage to all users for all days of the chosen week. The problem is defined as follows:

$$\min \sum_{c \in C} x_c \tag{1}$$

$$\sum_{c \in C} a(u,c,d)\ x_c \geq 1 \qquad \forall u \in \textit{Users}, \forall d \in \textit{Days}, \tag{2}$$

$$x_c \in \{0,1\} \tag{3}$$

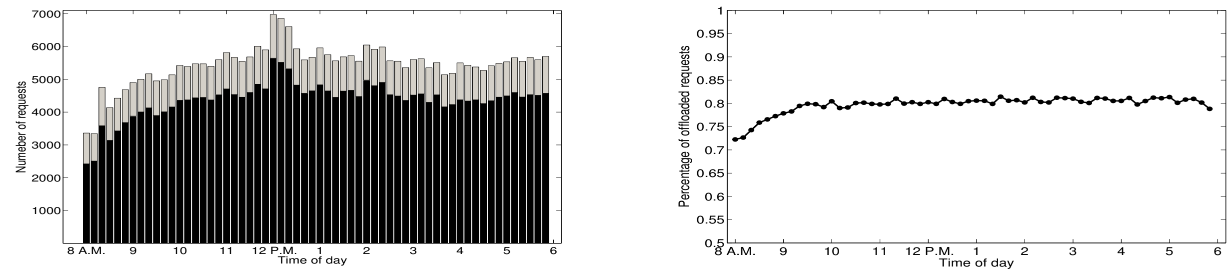
where C is the set of cells accessed by users during the first week, $a(u,c,d)$ is a binary coefficient set to one iff the user u accessed to the cell c the day d, while x_c is a binary variable set to one iff the cell c belongs to the optimal solution. $Users$ is the set of all users, and $Days$ is the set of the five working days of the first week. The constraint of Eq. 2 ensures that each user is covered by at least one cell in each day of the week.

The number of cells resulting from the optimization problem is 1,081, the 63% of the total. In the following we will refer to this set of cells as C_o. Despite the strong constraint, we have obtained an encouraging result which allows the network operator to bound the deployment spending, while covering all users. We evaluated this optimal placement with the dataset of the second week, still considering a content $push$ approach. By leveraging the strong regularity that users show in their activity patterns (as shown in Sec. III-A), the placed infrastructure ensures more than 95% of successful offload of the considered traffic.

IV. A Pull Approach to Content Delivery

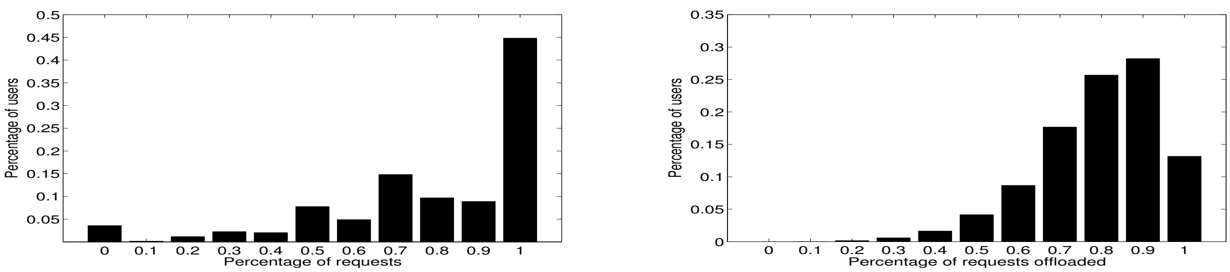
The $push$ approach we considered so far is highly effective but slightly operator centric. In fact, it assumes that users accept a delivery service that autonomously pushes contents to them when they happen to transit under coverage of the cell selected by the operator. This can be effective in practice for a subset of contents but not for their totality. Many users may actually be willing to freely access their contents anytime and anywhere by $pulling$ them from the Internet. In the following, we consider such a $pull$ approach and we evaluate whether the described optimal placement remains effective under changed access conditions. To evaluate this new setting, we randomly select, for each user, 40% of his/her daily cell accesses and we use them to describe the users' requests for digital contents.

As a first performance index, we evaluate the amount of requests we can serve through the offload infrastructure



(a) (b)

Fig. 2. (a) Network traffic load from 8:00 A.M. to 6:00 P.M. of a representative day. Each bar covers an interval of 10 minutes. Black bars show the traffic load of offload network, grey bars show the traffic load of core network. The sum of the two bars show the traffic load of entire network. (b) Percentage of offloaded requests.



(a) (b)

Fig. 3. (a) Histogram of users' percentages of offloaded requests in a single day. (b) Histogram of users' percentages of offloaded requests over the entire week.

and compare them with those served through the operator's core network. In Fig. 2(a) we report the traffic (aggregated in bins of 10 minutes) along a sample day, with all days behaving quite similarly. The black bars report the number of offloaded requests, while the grey bars report the number of requests served by the core network. The sum of both bars is the total number of users' requests. While the overall number of requests grows along the morning, has a peak in the launch pause and then remains quite constant during the afternoon, the percentage of offloaded traffic remains stable, as shown in Fig. 2(b). In fact, the traffic breakout remains quite stable, around 80% of the total. This is a good result from the operator's perspective, for it shows the viability of the placement under different settings of access distribution.

From a user perspective, we assume that a mobile user may experience a somehow improved service efficiency when exploiting the offload platform. This argument makes interesting to estimate how fairly this opportunity is distributed among the set of users. Such a measure highly depends on the considered dataset, however, under our setting, the percentage of requests offloaded in a representative day is reported in Fig. 3(a) showing that about 45% of the users have all requests served by the offload infrastructure, while only few users experience less than 50% of requests offloaded.

We then enlarge the analysis window to the whole second week and report the results in Fig. 3(b). On a weekly basis, the performances are very satisfying; overall, 95% of users can experience improved performances in more than half of their requests when compared to core-routed traffic.

V. The Placement of Data Kiosks

We assume that the offload platform includes data kiosks which we envision to exploit for enhancing mobility support in a news delivery system. In particular, we consider the delivery of newspapers, a very common data content. Contents are characterized by a sort of *durable subscription* [7] where kiosks cache the subscribed contents, independently of whether the user is connected to the associated cell or not, for a specified time of content validity (one day and one week, respectively). For the purposes of this paper, we do not further differentiate subscriptions inside each content category and assume that newspapers are mostly accessed in the time window from 8:00 A.M. to 11:00 A.M.. In this caching system, contents are replaced daily or weekly, according to the class they belong to. We want to define a cell selection process to proactively cache contents based on the knowledge of the regularity of activity patterns of subscribers. The problem to solve in this case is slightly different w.r.t. the aforementioned one: here, we use the first week in the dataset to identify the per-user subset of most visited cells and we place the subscribed content in the associated kiosk. In practice, we leverage the regularity of user mobility to increase the probability of finding the cached content when the user will generate his/her content request during the second week.

The resulting new matching function $f_u(c)$ to each cell and for each user u is as follows:

$$f_u(c) = e^{d_u(c)} \sum_{d \in Days} w_u\left(c, d\right) \tag{4}$$

where c represents a cell, d_u is the number of days in which the user u accesses cell c, $w(c, d)$ is the number of accesses to cell c during the day d and $Days$ is the set of trial days. This function allows us to consider, for every cell and for every user, both the number of accesses in a day, $w_u(c, d)$, and the number of days it has been accessed by the user, $e^{d_u(c)}$. For each user u, cells are ordered by descending matching values and the first n elements are selected as favorite cells. Thus, we denote with C_u the set of the favorite cells of user u. In the following analysis we consider $n = 3$, being other cells not relevant according Section III-A. In the optimization process only the subset of favorite cells belonging to the offload infrastructure ($C_u \cap C_o$) is considered. If the intersection between C_u and C_o is empty, then the user is discarded because it will not benefit from the offloading infrastructure anyway. Thus, the set of cells where minimization will take place is C_f defined as:

$$C_f = C_o \cap \{\cup_{u \in Users} C_u\} \tag{5}$$

As above, we find the optimal set of cells for the whole set of users by solving an instance of the set covering problem defined as follows:

$$\min \sum_{c \in C_f} x_c \tag{6}$$

$$\sum_{c \in C_u \cap C_f} x_c \geq 1 \qquad \forall u \in \text{Users} \tag{7}$$

$$x_c \in \{0, 1\} \tag{8}$$

where x_c is a binary variable set to one iff the cell c belongs to the optimal solution.

Optimization is performed using mobility data of the first week while network access simulation is performed using the second week. User content retrieval is simulated by randomly selecting one cell access per day in the time window planned by the service, since we suppose that each user submits only one request during a day.

After optimization, we are left with 993 cells where caching is required (92% of total infrastructure cells). This scant reduction is due to the limited number of cells accessed during the morning. As a matter of fact, the ratio between the number of cells accessed in the morning and those accessed during the whole day is 0.88, meaning that most cells are accessed later in the day.

By analyzing the performance from a network operator point of view, we obtain the results depicted in Fig. 4. In the figure, we report the traffic (aggregated in bins of 10 minutes) along one representative day, being all days very similar. Black bars show the traffic load of offload network while grey bars show the traffic load of core network; the sum of the two bars is the traffic offered to the entire network. The percentage of traffic offloaded is, in the average, 70% of the total traffic.

From the user perspective, the best download performances are experienced when the content is accessed by means of data kiosk. To evaluate these performances, we compute for each user the percentage of requests satisfied in offload. The related histogram on all users is reported in Fig. 5. As we can see, 35% of users are always satisfied by the offload infrastructure, while more than 70% of users are satisfied more than half of the times.

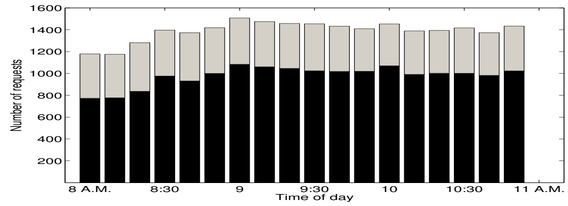


Fig. 4. Network traffic load from 8:00 A.M. to 11:00 A.M. of a representative day.

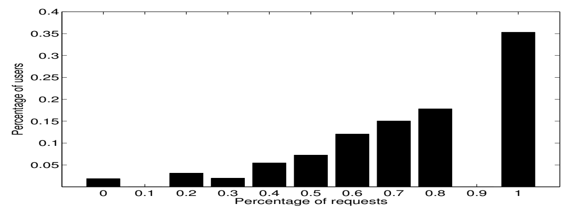


Fig. 5. Histogram of users' requests offloading percentage for newspaper subscription.

VI. Conclusion

This paper presents preliminary analysis of a real dataset of Internet accesses generated, in the city of Milano, by some 50,000 users of an important mobile network operator. On this base we have been able to (*i*) extract a well optimized subset of cells, based on users' activity pattern, where the proactive caching service can be deployed to reduce costs and (*ii*) to evaluate performances for the proposed solution in real urban settings. The preliminary results indicate that offloading can be easily managed and planned by network operators and that the approach needs to be tightly interwoven with the emerging LTE's femtocell architecture.

Acknowledgment

This work was funded by the Italian Ministry for Instruction, University and Research under the PRIN PEOPLENET (2009BZM837) Project.

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016," Tech. Rep., 2012.

[2] G. Colombo, M. Chorley, M. Williams, S. Allen, and R. Whitaker, "You are where you eat: Foursquare checkins as indicators of human mobility and behaviour," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, 2012.

[3] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[4] M. Zignani, S. Gaito, and G. Rossi, "Extracting human mobility and social behavior from location-aware traces," *Wireless Communications and Mobile Computing*, 2012.

[5] M. Zignani and S. Gaito, "Extracting human mobility patterns from gps-based traces," in *Wireless Days (WD), IFIP/IEEE.* , pp. 1–5, 2010.

[6] W. jen Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *INFOCOM*, 2007.

[7] X. Vasilakos, V. A. Siris, G. C. Polyzos, and M. Pomonis, "Proactive selective neighbor caching for enhancing mobility support in information-centric networks," in *ICN workshop on Information-centric networking*, ser. ICN '12. New York, NY, USA: ACM, 2012, pp. 61–66.