Last Call for the Buffet: Economics of Cellular Networks

Jeremy Blackburn‡♭, Rade Stanojevic‡, Vijay Erramilli‡, Adriana Iamnitchi♭
Konstantina Papagiannaki‡
Telefonica Research, Barcelona (Spain)‡
University of South Florida♭

ABSTRACT

Voice and data traffic growth over the last several years has become a major challenge for cellular operators with a direct impact on revenues, infrastructure investments, and end-user performance. The economics of these operators depend on various incentives used to attract users in the form of unlimited, buffet-like voice/sms/data packages. However, our understanding of the effects of user behavior under these offerings on operator revenues/costs remains poor. Using two years of detailed usage information of ~ 1 million users across three services, voice, sms and data, combined with payment and cost information, we study how user behavior affects the economics of cellular operators. We discover that around 20% of the users consume more resources than what they pay for and hence are non-profitable. In addition to the individual user behavior, we study how the user interactions in the call graph affect the operator's revenues and cost, drawing on tools from social network analysis. We develop a framework that incorporates both the individual and social user behavior for studying how volume caps influence the revenues and the traffic costs. Using this framework we empirically show that volume caps can increase the difference between the revenues and the traffic costs of the studied operator by a factor of 2, while affecting only 16% of the existing user base.

Categories and Subject Descriptors

K.6.2 [**Management of computing and information systems**]: Pricing and resource allocation

General Terms

Economics

Keywords

User behavior, Network effects, Cross-subsidization

MobiCom'13, September 30-October 4, Miami, FL, USA.

1. INTRODUCTION

The network services industry (wired and wireless) has experienced high growth in every corner of the world in the previous decades. This growth has been fueled by technology advancements, increase in demand for connectivity among end-users as well as pricing mechanisms that can lead to profits and are conducive to attract and engage users. The growth of the subscriber base is seen as one of the most important metrics to reflect success. Towards this end, pricing schemes tend to be simple; a flat rate and contain an abundant amount of communication units (all-you-can-eat buffet plans) to attract a large user base.

The wide-scale adoption and profitability of such pricing schemes is predicated on several premises. First, it has been reported that users prefer flat rate pricing schemes due to their inherent predictability [17]. Second, it is assumed that most users will consume only a small amount of the purchased packages, thus not generating a large aggregate traffic cost [26]. And third, the cost of delivery for many network services on a per service unit basis has been a small fraction of the retail price; common examples include sms service or residential (cable/DSL) broadband where the marginal cost of delivery of a 160-character sms or 1GB of wired data is significantly lower than the retail price the providers charge for the service [16, 24].

However, in recent years many network operators, specially those in mature markets, have seen their revenues and profits saturating or even decaying [30, 28]. Consider mobile broadband that has contributed a relatively small fraction of the total revenues but is projected to be the main cash cow for providers in near future [28]. Fig. 1 depicts an estimation of current and future costs for delivering a gigabyte of mobile data among north American carriers with current retail prices of $10-20, leaving very little margins for profits [1]. There are several reasons for this loss of revenue – intense competition, regulation and user behavior [28].

In this paper, we focus on understanding the effect of user behavior on revenues and traffic costs in cellular networks. In particular, we rely on a large (anonymized) dataset of $\sim 1M$ customers of a cellular operator that consists of their entire usage behavior across voice calls, sms and data usage, as well as their billing details and costs incurred *by* the operator, to answer the following questions: (i) how do users respond to flat/unlimited tariff structures, (ii) how such user behavior affects the revenues and costs of operators and (iii) how operators can optimize their profit under the presence of costly user behavior. Our dataset (Sec. 2) has unique properties to aid us in answering the above questions more accurately.

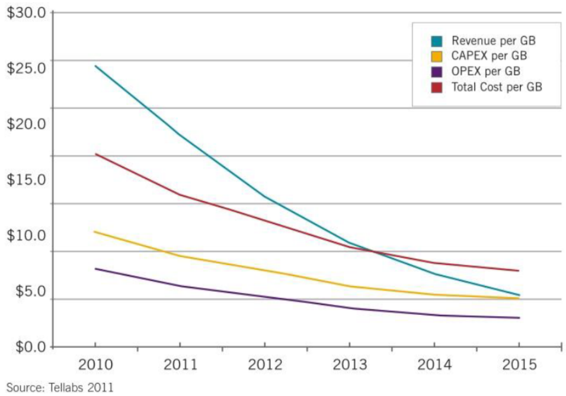

Figure 1: North American mobile data carriers: estimate of costs and revenues per GB of mobile data, 2010-2015, [1].

First, the pricing plans offer unlimited volume units, hence the usage of subscribers is the real communication needs of users and second, the cost information we have per service unit is an accurate proxy for the true cost of delivery.

For the first question, we extract the amount of revenues and traffic costs per customer of the operator from our dataset and we find that 20% of the users lead to more cost to the network than revenue (Sec. 3). This result points to two conclusions – first, unlimited, buffet-type plans attract such unprofitable user behavior and second, this unprofitable user behavior is *cross-subsidized* by other users[1].

Digging deeper into user-behavior, we find that defining the value of a user in the network only in terms of her individual revenue and cost is myopic. We find that by understanding network effects – how users interact with other users in the network and how users attract other users to join the network can give a more accurate value of a user in the network. In order to study network effects, we rely on tools from social network analysis (Sec. 4). Finally, in order to answer the third question, we develop a case for using volume caps as a means to remove unprofitable user behavior and increase profits (Sec. 5).

Our main results can be summarized thus:

- Using the fine-grained usage and billing data from the customers of a mobile virtual network operator with $\sim 1M$ customers, and the traffic cost information, we demonstrate that around 20% of customers generate higher traffic cost than what they pay to the operator and hence are subsidized by other customers.

- We show that the usage of the non-metered services (3G data, sms and on-net calls) is highly (negatively) correlated with the profitability of the customer, and that customer profitability is assortative: connected customers are likely to have similar profitability.

- As with many other services that involve social interaction, the social graph (observed via voice calls and sms) has non-trivial impact on the revenues and the costs of the operator. We quantify the relationship between the social network structure and the operator's revenues and costs.

- Finally, by integrating the individual user behavior and user interaction behavior we develop a framework for studying the impact of volume caps on the operator's revenue and costs. We show that simple capping strategies can increase the difference between the revenues and the traffic costs by almost a factor of 2 while affecting only 16% of the customers.

[1]Unlimited volume in the wired access network have lead to cases where few heavy users were being cross-subsidized by other users and led to practices like throttling/blocking peer-to-peer traffic [5, 11], bringing into focus the notion of net-neutrality [4].

2. BACKGROUND AND DATA

In this section, we first provide background for concepts we use throughout the paper. We then discuss the dataset we study.

2.1 Background

In this paper, we deal with cellular networks. Cellular networks around the world have a very diverse set of pricing plans. In general, pricing plans can be thought of as being on a spectrum, where on one end lies a single flat rate for 'unlimited' usage and on the other end lies pure usage based pricing (UBP) where every unit of service (voice minutes, sms, data) consumed is 'metered' and charged. Most operators, however offer plans that lie somewhere in the middle of the spectrum. The typical offerings are monthly plans – payments are done on a monthly basis and a specific flat rate is charged for a bundle of services – specific volumes of voice/sms/data. 'Unlimited' plans are those bundles where the volumes of different services have no cap on usage every month. If a maximum usage limit is specified, then the plan is referred to as 'flat up to a cap' or just capped. Any volume used more than the volume cap is either metered and the extra volume is charged proportional to the usage, or the service is degraded for the extra volume (eg. 3G to 2G). 'Tiered' plans with different flat-rate prices for different volume caps are used to offer choice to the user. We note here that tiered plans are also a type of usage based pricing (UBP).

We focus on cellular network operators setting the retail prices for the sole purpose of maximizing profits. Pricing can also be dictated by other factors like regulation, but we do not deal with that here. In the process of user acquisition and retention many of the operators choose to offer 'incentives' of various forms including: (1) *handset incentives*, where the operator offers a handset for a discounted price, usually requesting the commitment from the customer to remain within the network for an extended period of time (12-24 months); (2) *traffic incentives*, where the operator offers a number of free service units (eg. voice minutes or sms messages) or an unlimited service quota for a particular service like data. Many operators also offer *social* incentives to attract new customers. These incentives normally come in the form of unlimited volume of voice minutes or sms between users of the same network and are incorporated in the pricing plans/tariffs. Hence, if a user has many of her contacts in a certain network, then it would be rational for her to join this network. Combined with word-of-mouth marketing where customers of an operator explicitly or implicitly advertise the operator, it is clear that a user's activity with other users in the network also plays a role in the revenue/cost structure of an operator.

Tariff	On-net voice	On-net sms	Off-net Voice	SMS	3G data	Price
PAYG	Free	Free	1★/min	0.6★/txt	$\frac{2★}{20MB}$ per day	0
Bundle 1	Free	Free	60min	unlmtd	$\frac{2★}{20MB}$ per day	50★
Bundle 2	Free	Free	250min	unlmtd	unlmtd	100★
Bundle 3	Free	Free	400min	unlmtd	unlmtd	150★
Bundle 4	Free	Free	800min	unlmtd	unlmtd	200★
Bundle 5	Free	Free	1500min	unlmtd	unlmtd	250★

Table 1: Tariffs. Note that SMS is unlimited across all bundles, while 3G data is unlimited across most

2.2 Data

The data used in this paper belongs to a national Mobile Virtual Network Operator(MVNO), with around one million customers. In contrast to conventional Mobile Network Operators (MNO) that own the radio spectrum and cellular network infrastructure, MVNOs utilize the infrastructure of one (or more) MNOs by effectively purchasing the access to network services (voice/sms/data) at wholesale rates. In general, MVNOs provide their own customer and billing service. Similar to most other MVNOs, the one we study targets price-conscious segment of the market.

There are several types of data we use:

- *Usage data.* An anonymized record is stored for *every* call, sms and data session generated (initiated or received) by every customer of the network over a period of 27 months, from late 2009 to early 2012. The usage records contain the calling number, the receiving number (for voice calls and sms), the time-stamp, the duration of the call/data session as well as the volume of the data transmitted in the data session. This enables us to understand usage of different services on a per-user basis.

- *Payment information.* An anonymized record is stored for every service payment by every customer, over the same period of 27 months. We can thus correlate the usage with payment.

- *Traffic cost information.* We have the cost the MVNO has to pay to its cellular provider per minute of a voice call, per sms and per MB of data sent. The real figures are proprietary but one minute of voice calls, one sms and one MB of 3G data wholesale prices are within the same order of magnitude. The cost information helps us figure out the traffic cost on a per customer basis. The cost the MVNO pays to its provider is a linear function of the usage, in that if the MVNO users aggregately generated M minutes of voice calls, S txt messages and D $Mbytes$ of mobile broadband traffic over a billing period, the MVNO must pay

  $$C(M,S,D) = p_M \cdot M + p_S \cdot S + p_D \cdot D,$$

  where p_M, p_S and p_D are the wholesale prices of one minute of voice calls, one sms and one $Mbyte$ of mobile broadband, respectively.

The tariffs offered by the operator are listed in Table 1. The customer can choose between the pay-as-you-go (PAYG) tariff in which each service unit (off-net calls[2], sms and 3G data) are charged separately, or she can purchase one of the bundles that has a fixed number of free off-net voice minutes and unlimited/non-metered[3] data and sms services. Once a bundle is purchased, it expires in one month, and then the user can choose whether to purchase another bundle or continue using the service in a pay-as-you-go manner. At any time, the voice calls and sms between the customers of the network are non-metered.

Why data from a MVNO?

We use data from a MVNO for the following reasons: (i) The data we have has unlimited tariffs. This enables us to study user behavior in the *absence* of caps; the demand put on the network is natural demand and not self-regulated and constrained by the caps/metering [3]. (ii) As mentioned earlier MVNOs purchase the access for network services from a cellular operator that owns the network infrastructure at *wholesale* rates. The wholesale rates closely match the real production cost of the network services and are a good proxy for the *real cost* of the generated traffic, and do not include other overheads like marketing etc. In addition, the MVNO we consider does not offer handset incentives and offers only traffic based incentives – hence any users who joined this MVNO have done so for reasons other than handset incentives. (iii) MVNOs typically have very limited budgets for personnel and customer acquisition/retention. Thus, the overall costs of the MVNOs (including the one we study here) are dominated by the traffic cost [30]. Finally, most of the analysis we perform can easily be extended to data from a cellular operator that owns its infrastructure and radio network. We discuss these issues in more detail in Sec. 7.

3. USER BEHAVIOR: INDIVIDUAL LEVEL

In this section, we analyze user behavior under the different tariffs presented in Sec. 2, in terms of consumption – that gives us cost to the operator and in terms of revenues – what users pay. We focus on individual user behavior and how that relates to revenues/cost.

3.1 How much does each user contribute?

The first question we ask: what is the contribution of each user to the overall revenues and costs of the operator? For this, we look at the payments made by the user and the cost imposed by the user when she uses non-metered services.

We define the balance $B(u)$, for each user u, as the difference between the total revenue she paid for the network services to the operator ($R(u)$) and the total traffic cost she generated for the operator ($C(u)$):

$$B(u) = R(u) - C(u).$$

For each user, we possess the billing records over the 27 month period, from which we derive the revenue the user raised. We use the voice/sms/3G-data usage records to calculate the traffic cost to the provider each user imposes by adding the cost of voice calls, sms and 3G data generated by

[2]We refer to interactions, voice or sms, between two customers of the operator as on-net, and to interactions in which only one end is the customer of the operator as off-net.

[3]The exception is the Bundle 1 that has metered data service.

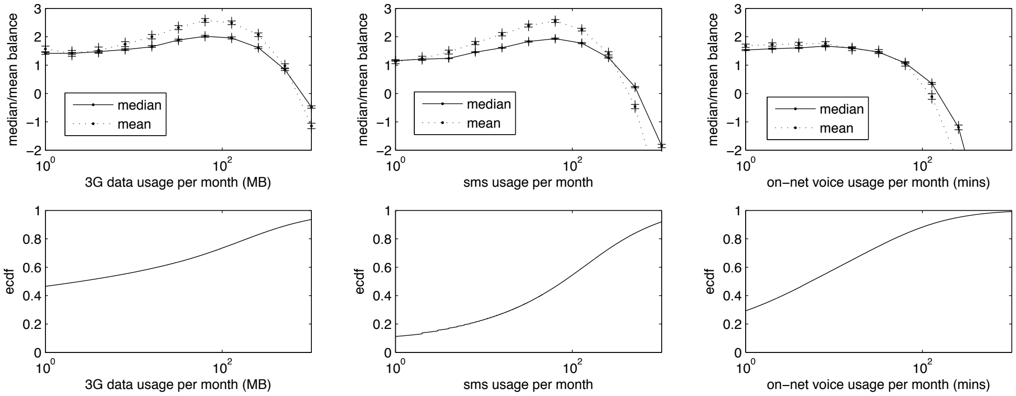

Figure 3: The median/mean per-user balance (normalized) vs. amount of non-metered service (top). Empirical CDF of the amount of non-metered services (bottom) consumed. x-axis is in log-scale.

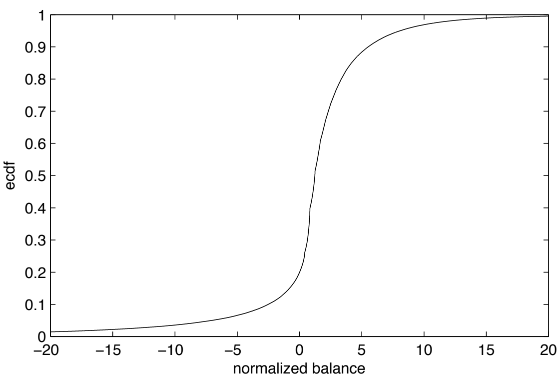

Figure 2: Empirical CDF of per-user balance (normalized to have mean equal to 1). Around 20% of users have negative balance!

the user, charged at the wholesale rates of p_M/min, p_S/sms and p_D/MB, as described in Sec. 2. The balance measures how profitable the user is for the operator. Users with large balance are very profitable for the operator, while users with negative balance generate uncovered costs that must be subsidized by those with positive balance. In Fig. 2 we plot the empirical CDF of the balance normalized[4] to have mean equal to 1, and in the rest of the paper we use use this scalar value $b(\cdot)$ to refer to balance instead of $B(\cdot)$:

$$b(\cdot) = \frac{B(\cdot)}{average(B(\cdot))}.$$

We observe that around 20% of users have **negative** balance, and hence must be subsidized by other users. In addition, while the empirical distribution of $b(\cdot)$ is bell-shaped, the tails are long with the minimum (= -443) and maximum (= 157) values far away from the mean compared to the standard deviation (σ = 6.75). Overall, the total balance of users who are subsidized $|\sum_{u:b(u)<0} b(u)|$ is greater than the whole total balance of all users $\sum_u b(u)$, see Eqn .(1) in Sec. 5.

Balance, as defined above, clearly depends on consumption of non-metered services (3G data, sms and on-net voice calls). One would guess that users who consume more of the non-metered services are less likely to be profitable. In Fig. 3 we quantify this dependence between balance and usage, by evaluating the mean and median balance among users consuming x units of non-metered service (MB of 3G data, sms, mins of on-net calls), for $x \geq 0$. We do not show the instances of very small ($x < 1$) and very large ($x > 1000$) usage, in order to depict the behavior of the majority of the users. The median balance is positive for users that use low amounts of the non-metered services. However, once the usage of the non-metered services crosses a threshold the balance decays below zero. Such thresholds, beyond which the median balance turns negative are: 796MB (for 3G data), 552 sms messages (for texting) and 150 mins (for on-net calls).

[4]We perform normalization in order to protect the proprietary information related to revenues and costs.

	r	p	95% conf.
V_{off} v. Data	0.080	< 0.01	0.077, 0.082
V_{off} v. SMS	0.228	< 0.01	0.225, 0.230
V_{off} v. V_{on}	0.145	< 0.01	0.142, 0.147

Table 2: Correlation coefficients of unmetered service usage.

Main takeaways: We find that 20% of the users have negative balance; imposing a higher cost on the operator than the revenues they generate. We also find, somewhat intuitively, that costs are correlated to usage of non-metered services. We later show in Sec. 5 that significant improvements in the overall profit of the operator are possible by treating these 'subsidized' customers with caps.

3.2 Is off-net voice a good proxy for costs?

Table 1 shows that the charge the customer u pays for the service in most bundles is related to the amount of off-net voice minutes she uses ($V_{off}(u)$), while the cost she generates for the operator also depends on the amount of the non-metered services she uses: on-net calls ($V_{on}(u)$), sms ($S(u)$) and 3G data ($D(u)$). If these service usages were strongly correlated, in other words if one of them could be a good proxy for estimating the other, then metering that one service could align the revenues with the costs users generate. However looking at the consumption of these 4 services we see low levels of correlations between the usage per-service (see Fig. 4 and Table 2), indicating that the total traffic cost per-user (aggregated cost for all services availed) is weakly correlated with the revenues that depend only on the usage of one service ($V_{off}(u)$).

Main takeaways: We observe no correlation between the usage levels per service across the customers of the operator.

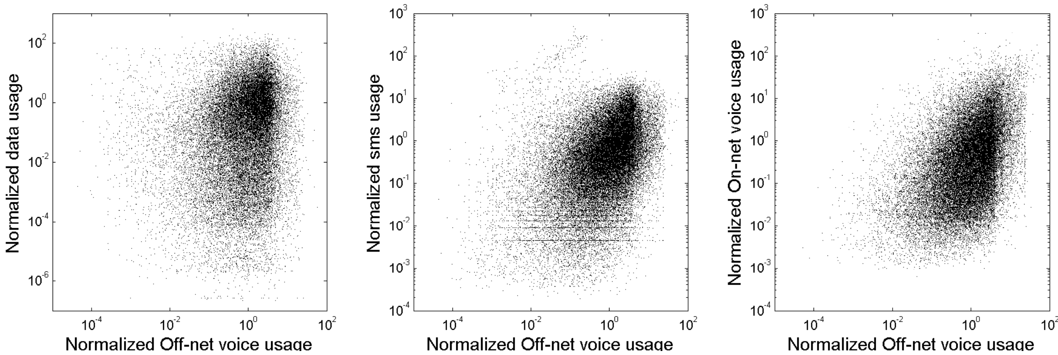

Figure 4: Correlation between the metered Off-net voice usage and non-metered: 3G data, sms and on-net voice usage. Each dot represents a single user.

By metering one service, while the costs depend also on the other three services whose usage is not correlated with the metered one, the revenues can be highly disconnected with the costs, as demonstrated in Fig. 2.

4. USER BEHAVIOR: CALL GRAPH

In the previous section, we looked at user-behavior in isolation and how the behavior in terms of usage impacts revenues and costs of an operator. In this section, we look at user-behavior in the social network defined by the call graph. This is important for two reasons. First, there are social incentives in place – from Table 1 we see that calls and sms made to other users in the network are non-metered and are free – and user behavior with these incentives in place can affect costs and revenues. Second, again because of the social incentives the users in the network can play an active role in the growth (or contraction) of the network, and this facet should be accounted for in a user's balance.

4.1 User balance in the call graph

In Sec. 3, we saw that the balance of a user is negatively correlated with the amount of non-metered services she uses. We expand the analysis to the social network (as represented by all other users contacted via calls/sms) of a user. Intuitively, since the balance is negatively correlated with the amount of on-net minutes (as we see in Fig. 3) we expect to see similar correlation between the balance and the number of on-net contacts (contacts within the same operator) of the user, which we confirm in Fig. 5. We observe that for users with 1 or more on-net contacts, the mean balance is significantly smaller than the median and that for users with 5 or more on-net contacts the mean balance is negative.

The notion of similarity in behavior between users can be captured by the *assortativity coefficient*; which is a measure of correlation between nodes in a graph. Intuitively, it is a measure of users to engage with 'similar' users. In our case, we are interested in assortativity with respect to user balance. Specifically, the assortativity coefficient, defined as the Pearson correlation coefficient of the balance vector between all pairs of connected users, is 0.27, indicating a high assortativeness[5]. In Fig. 6 we plot the dependence between

[5]For example the node-degree in Facebook social network has assortative coefficient of 0.23 that is interpreted as high [8].

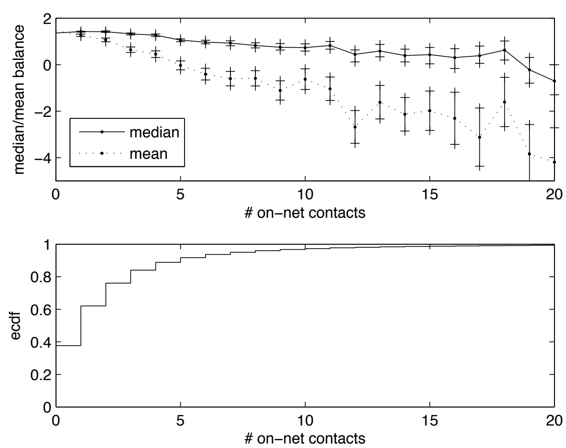

Figure 5: The median/mean balance vs number of contacts. Users with many contacts have low expected balance. Cutoff in the figure at #on-net contacts ≤ 20 is made to increase the readability of the figure. In addition $< 1\%$ of users have #on-net contacts> 20. max#on-net contacts> 100.

the mean and median balance across contacts of users with balance (approximately) equal to b, where we vary b. The monotonic increasing nature of the mean/median contact balance, is another indicator of the high assortativity of the balance in the studied social network.

Main takeaway: We find that heavy users interact more with other heavy users and vice-versa. Hence in our data we see that social incentives sometimes hurt the operator more than help it.

4.2 Number of on-net contacts

In this section we show evidence that social contacts can play a role of attractors in the organic growth of the operator's user base, and quantify such effect using a simple metric. While we note here that establishing pure causality – users joined for social contacts only – is beyond the scope of this work, previous work has looked into the (major) role of network effects on the choice of an operator [2].

The operator we study has nation-wide coverage and serves relatively small fraction ($< 5\%$) of the population in the country. Given our voice calls and sms data we can detect

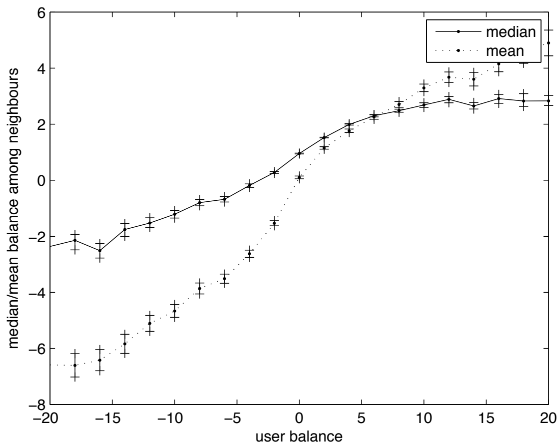

Figure 6: The median/mean balance among the contacts of users with balance $\approx b$, for $b \in [-20, 20]$. User balance is assortative: contacts are likely to have similar balance.

whether a user (on-net or off-net) is a contact with an existing customer of the operator. We define a mobile user v to be the *contact* of a customer u if there is at least one interaction (voice or sms) between u and v in *both* directions[6].

For $k = 0, 1, 2, \ldots$ we use N_k to denote the number of customers of the operator with exactly k contacts in the operator and we use F_k to denote the number of national mobile users (both on-net and off-net), with exactly k contacts in the operator. Hence N_0 denotes the number of existing customers of the operator who have 0 contacts within the operator. Likewise, F_0 denotes the number of all mobile users in the country who have 0 contacts in the operator.

We can compute N_k for any $k \geq 0$ and F_k for $k \geq 1$, from our data since it archives any voice or sms interaction in which one (sending or receiving) party is a customer of the operator. However, computing F_0 is not as straightforward, since we possess no information regarding the communications that happen outside the operator. To estimate F_0 we use the estimate of the total number, T_0 of mobile phone users in the country [32] and use:

$$F_0 = T_0 - \sum_{k \geq 1} F_k.$$

To understand the relationship between the number of contacts of u that are customers of the operator and the likelihood that u is customer of the operator we define the following metric

$$p_k = \frac{N_k}{F_k}, \ k = 0, 1, 2, \ldots$$

Thus, p_k is the fraction of users from the entire mobile user base with k contacts in the operator. In Fig. 7 we report the values of p_k. We can observe that the likelihood of a user with no contacts in the operator ($k = 0$) to be a customer of the operator is small, while such likelihood increases with k, and flattens out around $k = 5$; having more that 5 contacts in the operator does not affect the chances of being the customer of the operator.

[6]Thus the contact relationship is symmetric: if u is contact with v, then v is contact with u.

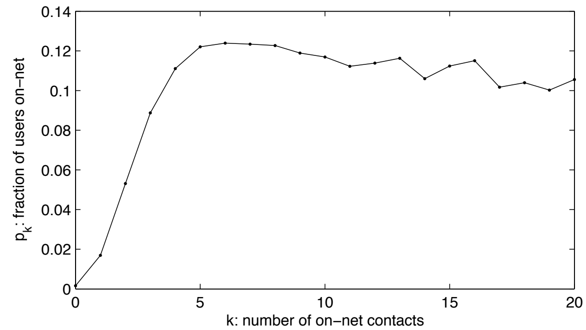

Figure 7: Number of contacts in the operator vs. likelihood of being the customer of operator. Having more contacts in the operator increases chances of being a customer of the operator up to $k = 5$.

Main takeaways: The parameters p_k determine the impact of social ties in the growth of the network from a macroscopic point of view. In Sec. 5 we will use the parameters p_k to evaluate the effects of node removal on its contacts and the impact of such cascade removal process on the revenues and costs of the operator.

4.3 Revisiting user balance: Incorporating social effects

The previous result shows that users within the operator can play a role in attracting other users. In Sec. 3, the *value* of each user is the user's balance. However, it can be the case that a heavy user, with a negative balance can attract other users to the operator who contribute positively. Hence the effective value of the user to the operator should reflect this facet. To examine this, we use the call graph between customers of the operator to model their interactions. The undirected edges of this graph are weighted with the total duration of calls between contacts, a known proxy for social strength [19].

The balance for each node is then distributed between each of its contacts, proportional to the duration of calls (acting as a proxy for social strength). In this way, the aggregate balance across nodes remains the same but each node's balance is now derived entirely from the balance of its contacts and their behavior patterns.

We compare differences between user balance calculated in Sec. 3 and user balance using social relations in Fig. 8, that plots the density distribution of balances in the range [-10, 10]. The network calculation "squeezes" the distribution, resulting in more borderline profitable customers. This points to the fact that there is a fraction of users who are themselves costly users but nonetheless bring in profitable users, underscoring the importance of considering network effects in assigning values to customers.

Main takeaways: Users join the service to take advantage of social incentives, and there are users who at face value are subsidized, but who are profitable with respect to the other service users they interact with.

5. VOLUME CAPS: WHY AND HOW?

In this section we study how volume caps affect the revenues and costs for the operator. We incorporate our findings

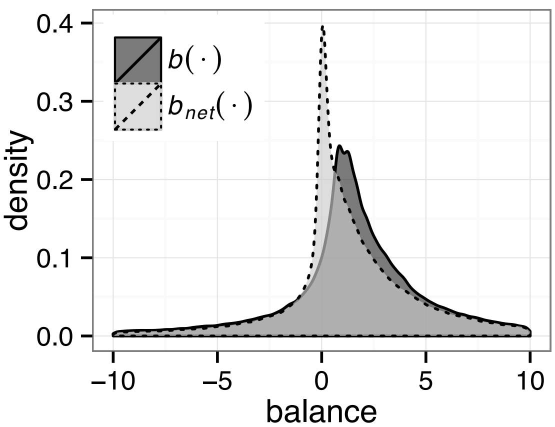

Figure 8: Density of normalized balances with two calculation types: 1) local ($b(\cdot)$), and 2) network ($b_{net}(\cdot)$).

from previous sections in the design of cap-based solutions. To understand the impact of the caps on user behavior, and hence the revenues and costs, we study two models. In the first model the users affected by the cap are only those that cross the cap. The second, more general model also takes into account the social ties between the users to capture the social network effects that can arise due to the introduction of caps.

We calculate optimal volume caps under the two models, with the goal of optimizing profits. We demonstrate that significant improvements with regards to the aggregate balance are possible; by using optimal caps one can expect the relative increase in the total balance of 107% (for the net-oblivious model) and 90% (in the net-aware model) that is close to the maximum of 124% that one can obtain by removing non-profitable users (and them only) from the network (as we show next). We also show that setting caps on only one of the three non-metered services (3G data, sms and on-net calls) can increase the total balance, although the gains in these cases are not as large as they are in the case of capping all three services. Before we describe our solutions, we first discuss why we need caps.

5.1 Why volume caps?

From our findings in Sec. 3, we saw that certain users cost more to the operator than what they contribute to the revenues. A natural question to ask in this context is: what can be the possible increase in the profit of the operator by eliminating these users; users with a negative balance, how much would the total (aggregated over all users) balance change? The following coefficient gives the answer:

$$\gamma = \frac{\sum_{u:b(u)>0} b(u)}{\sum_u b(u)} = 2.24 \qquad (1)$$

The amount spent on subsidizing the users with the negative balance is (1.24×) larger than the total balance, hence eliminating the users has a great potential to increase the aggregate balance without charging more for the service. However, eliminating as a policy is not desirable. First there are contractual obligations. Second, there have been instances of negative press when heavy users have been pushed out of the network [33]. Another solution can be to remove social incentives – however as we saw in Sec. 4, social incentives do help in driving growth and any solution that disrupts the social mechanics could backfire, resulting in the loss of both subsidized users and their subsidizing friends.

Volume caps have been suggested as a *simple* mechanism to alleviate cross-subsidization and increase profits [18]. In addition, it is also posited that volume caps in cellular networks are used for sustainability reasons; curbing runaway growth that can lead to operators being “crushed by capital spending” [18]. Our contribution is to design and study simple caps using real data, while incorporating user behavior.

We stress that more complex solutions with more pricing tiers on segmented user base and price differentiation [12] or has different form of social incentives, may offer even higher gains than those we compute with caps. However we demonstrate that simple cap-based solutions have potential to eliminate the bulk of the cross-subsidizations and hence there is limited room for extracting higher profit without increasing the prices on existing customers. The problem of profit maximization beyond cap-based solutions remains open and is out of scope of the present paper.

5.2 Optimizing caps

To quantify the effects of caps on the users and the operators we use two models, the *net-oblivious* and the *net-aware*, that we describe below.

5.2.1 Net-oblivious model

In the net-oblivious model, we assume that if a user consumes more service units than what the cap offers, she either quits the network or the overage charges compensate for the extra traffic she consumes, thus bringing her balance to zero. More formally, if the user u consumes s_u smses, d_u MB of mobile broadband and f_u minutes of on-net calls, on average and the operator packages have caps of x sms, y MB of mobile broadband and z minutes of free on-net calls, we say that u is *affected* by the caps if she consumes more service than the cap quota in at least one of the three services. In the net-oblivious model: if a user u is affected by the cap we assume that she is not customer of the operator and hence has balance zero, otherwise neither her usage nor charge is affected by the caps and thus her balance remains the same:

$$b_u^{(x,y,z)} = \begin{cases} 0 & \text{if } (s_u > x) \vee (d_u > y) \vee (f_u > z) \\ b(u) & \text{otherwise} \end{cases}$$

The users with large consumption of non-metered services are the ones that are cross-subsidized and putting a cap on how much of the free services they obtain in the package they purchase should reduce the instances of cross-subsidization and increase the total balance $B(x, y, z)$ defined as:

$$B(x, y, z) = \sum_u b_u^{(x,y,z)}. \qquad (2)$$

To find the optimal caps we consider the solution of the optimization problem:

$$(x_{opt}, y_{opt}, z_{opt}) = \underset{x\geq 0, y\geq 0, z\geq 0}{\arg\max}\; B(x, y, z). \qquad (3)$$

One can observe that in this model the user affected by caps either does not change her behavior in terms of their usage and payments in case she remains under the cap, or has zero balance if she crosses the cap.

Data: $G = (V, E)$: Social graph
(x, y, z): caps
s_u, d_u, f_u: average per service usage of customer $u \in V$
p_k: probability that a user with k contacts is customer of the network
Result: Compute set A of affected nodes

```
begin
    A = ∅;
    Ē = E;
    for u ∈ V do
        if s_u > x or d_u > y or f_u > z then
            affected (u);
        end
    end
end
affected (u)
begin
    A = A ∪ u;
    for v : (u, v) ∈ Ē do
        k = #{w : (w, v) ∈ Ē};
        with probability 1 − p_{k−1}/p_k : affected (v) Ē = Ē \ (u, v);
    end
end
```

Algorithm 1: Finding the set A of affected nodes.

Clearly, the relative total balance RTB, relative to the baseline total balance $\sum_u b(u)$, is upper-bounded by

$$\frac{B(x,y,z)}{\sum_u b(u)} \leq \gamma = 2.24$$

from Eq. 1

5.2.2 *Net-aware model*

In Sec. 4.2 we uncovered evidence that social connections impact the growth of the operator's user base. Hence, eliminating the instances of losing customers[7] (those with $balance < 0$), may affect whether their social contacts join the network or not. The net-aware model, is a generalization of the net-oblivious model described above, in which each user affected by the cap can affect her contacts, by pulling them out from the network. The process of pulling out the contacts is recursive, and is driven by the parameters derived in Sec. 4.2 that map the likelihood of user u being the customer of the operator to the number of the contacts u has in the operator's network.

In the net-aware model, we consider the social graph $G = (V, E)$ with the nodes being the customers of the operator and the edges between the nodes if the customers have interacted using the network infrastructure (via voice calls or sms). In this graph, we define two sets of nodes: A contains the set of nodes affected by the cap and A^c those not affected by the cap. The balance of user u is either 0, if she is an affected node, or her original balance, $b(u)$, if she is not affected:

$$\bar{b}_u^{(x,y,z)} = \begin{cases} 0 & \text{if } u \in A \\ b(u) & \text{if } u \in A^c \end{cases}$$

To decide which nodes are affected we perform the following recursive procedure. Every node that uses more than x sms or more than y MB of mobile broadband or more than z minutes of free on-net calls is added to the set A of affected nodes. Each time a node u is added to set A, every contact v of u is added to A with the probability $1 - \frac{p_{k-1}}{p_k}$, where k is the number of contacts of v among yet non-affected nodes (including u) and p_k is the probability that a user (from a pool of *all* mobile users in the country) is the customer of the operator conditioned on the fact that she has k contacts that are customers of the operator. The parameters p_k are estimated in Sec. 4 and provide the basis for the macroscopic behavior model that captures the relationship between the social network of a user and the social pressure that makes her a customer of the operator. Note that the conditional probability that node v with $k > 0$ contacts on-net remains the customer of the operator after one of the contacts leaves the operator is indeed $\frac{p_{k-1}}{p_k}$. Hence the probability that she gets affected is $1 - \frac{p_{k-1}}{p_k}$. The pseudocode of the described procedure is shown in Algorithm 1.

Once the set of affected nodes is computed the total balance is:

$$\bar{B}(x,y,z) = \sum_u \bar{b}_u^{(x,y,z)}. \quad (4)$$

And the optimal caps are those that maximize the total balance:

$$(\bar{x}_{opt}, \bar{y}_{opt}, \bar{z}_{opt}) = \underset{x \geq 0, y \geq 0, z \geq 0}{\arg\max} \; \bar{B}(x,y,z). \quad (5)$$

Empirically, the probabilistic nature of the procedure that determines the set of affected nodes has very small influence on the total balance: in our data, the sample standard deviation of $\bar{B}(x,y,z)$ is 2 orders of magnitude smaller than the sample mean, and hence for the analysis of how caps affect the total balance, we use single instance of the procedure to calculate the total balance $\bar{B}(x,y,z)$.

Remark. Note that in both of the described models, the notion of cap we use refers to capping the average usage, rather than usage at any given month. The inherent variability of per-customer usage volumes, on month-to-month basis, can be resolved simply by allowing customers to migrate non-used parts of the package (mins, sms, MB) towards the next month's cycle.

5.3 Data-driven evaluation

In the previous section, we described two models of implementing caps. Both of our models are rather conservative in that they rely on the assumption that the 'heavy' customers are not likely to generate any profit for the operator, while the non-heavy users remain unaffected by the caps in both the service payment and the service usage. In this section, we will empirically evaluate the impact of the caps in both models.

For that purpose we use the following metric which we refer to as the RELATIVE TOTAL BALANCE, as the ratio between the total balance of the system with caps and the total balance of the system with no caps.

$$\text{RELATIVE TOTAL BALANCE}(x,y,z) = \frac{B(x,y,z)}{\sum_u b(u)}$$

We report on the RELATIVE TOTAL BALANCE for both models while varying the cap values (same number of free sms, MB of 3G data and on-net minutes) in Fig. 9. We also plot the absolute maxima of RELATIVE TOTAL BALANCE obtained by solving the optimization problems using a brute-force greedy approach.

We can infer the following from our results. First, carefully designed caps can significantly increase the total balance:

[7] By either discouraging them to join the network in the first place, or decreasing their satisfaction level by increasing their bill via caps to compensate for the cost they generate for the operator

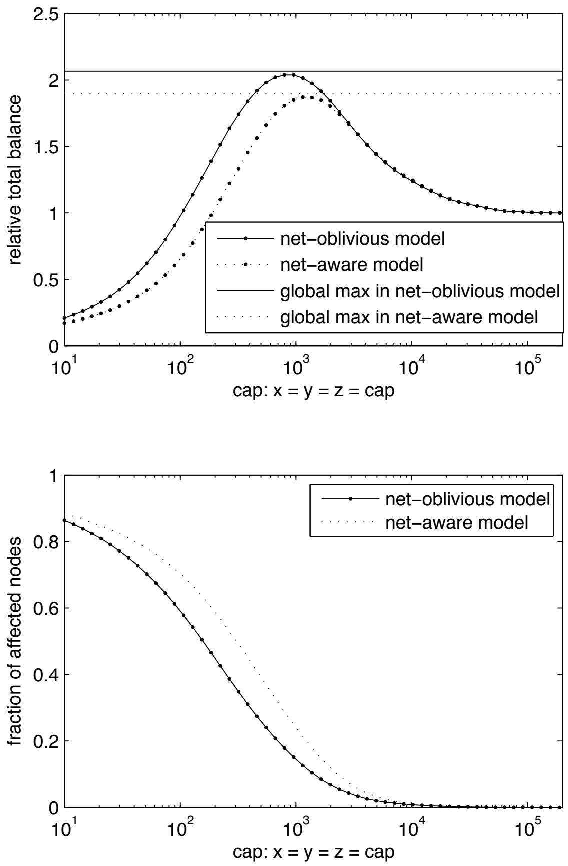

Figure 9: Top: The relative total balance(x,x,x), when the cap contains the same number of x free sms, x MB of 3G data and x on-net minutes. The straight lines represent the absolute maxima of relative total balance(x,y,z), which is equal to 2.07 in the net-oblivious, and 1.90 in the net-aware model. Bottom: The fraction of affected customers.

a factor of two increase can be expected in both models. Second, in the net-aware model one can expect lower total balance, though the impact of cap-initiated social pruning – users leaving due to network effects appears to be relatively small, and results in relatively small difference between the total balance in the two models of under 10%. This small difference can be in part attributed to the high-assortativity of the balance, shown in Fig. 6, where nodes with low balance are likely connected with the nodes with low balance too. Third, the cap in the simple form of x smses, x MB of 3G data and x free minutes of on-net calls, can recover most of the gains. Additionally, the optimal x that maximizes the total balance is around $\bar{x}_0 = 1150$ in the net-aware model and around $x_0 = 800$ in the net-oblivious model. Having slightly higher cap means that less nodes are directly affected by the cap and hence less nodes are affected indirectly by the social pruning. Overall around 16% of nodes are affected by caps at the optimal point in both models.

The absolute maxima of RELATIVE TOTAL BALANCE is achieved for:

$$(x_{opt}, y_{opt}, z_{opt}) = (931sms, 1240MB, 354min)$$

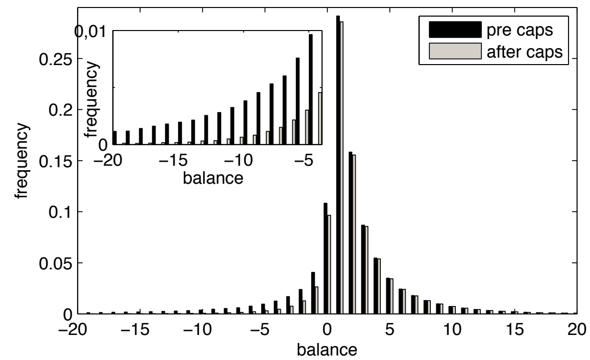

Figure 10: Histogram of per-user balance with and without caps. The net-aware model for computing the affected nodes used, with optimal cap $(1361sms, 1650MB, 471min)$. Most users with positive balance remain non-affected, while a large fraction of those with negative balance get removed due to caps.

in the net-oblivious model and

$$(\bar{x}_{opt}, \bar{y}_{opt}, \bar{z}_{opt}) = (1361sms, 1650MB, 471min)$$

in the net-aware model.

In Fig. 10 we depict the distribution of the balance per customer without caps and with caps in the optimal case of net-aware model. We observe that almost no users with balance >1 (= the mean) are affected by the cap, some customers are affected when the balance is close to zero, and almost all customers with large negative balance are affected by caps.

We finish this section by quantifying how capping individual services may affect the total balance. In Fig. 11 we plot the RELATIVE TOTAL BALANCE when one of the three originally-non-metered services is capped, while the other two are not, thus we look at the caps of form: (x, ∞, ∞), (∞, y, ∞) and (∞, ∞, z). From this figure, we can observe that capping one of the services can indeed provide certain improvement in the total balance, although the relative increment is much smaller than in the case of capping all three services. One can expect increase of up to around 40% when the 3G data or sms services are capped and around 25% increase when the on-net voice calls are capped. Hence to fully exploit the potential of the caps the operator should consider capping all three services.

6. RELATED WORK

If we consider the history of telecommunication services there is a distinctive trend of charging 'flat' fees for services: telegraph, post, fixed and mobile telephony, residential broadband, etc. This practice is justified by the customers' preference for convenient and simple tariffs as well as very low cost for the provider for accounting and delivering the service [17, 25]. In addition, users are willing to pay extra money for the convenience of not worrying about high bills that can result due to usage based pricing (UBP) [17, 25]. The focus of this work is *not* to study the historical precedents of flat-rate pricing nor is it to study the reasons why users may prefer flat-rate schemes. The focus is to understand, at large, the user behavior in terms of consumption

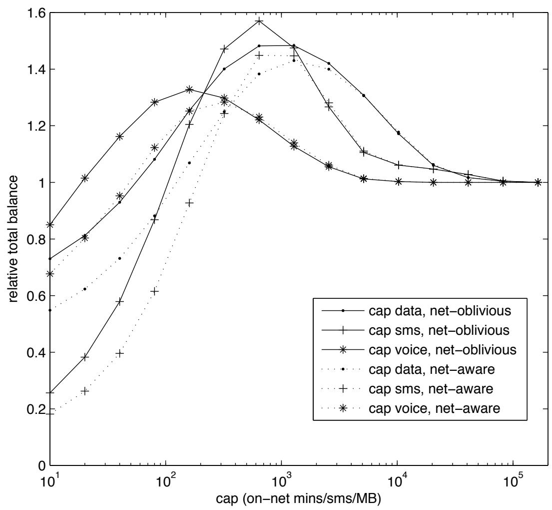

Figure 11: The relative total balance when only one service is capped. In the net-aware model the maxima are 1.43 for capping only sms, 1.42 for capping only 3G data, and 1.25 for capping only on-net voice.

under such schemes and how the behavior affects economics of a cellular provider, using real data.

There has been a lot of work on pricing issues in the Internet, including discussing relative merits and demerits of usage based pricing (UBP) in access and cellular networks [18, 16, 13]. The general consensus is UBP is used in access networks to raise revenues [18] while it is used in cellular networks to cope with congestion and runaway growth [18, 13]. It has also been suggested that UBP can be used to put an end to cross-subsidization [18, 13]. Our focus in this paper is not to study the underlying reasons to enforce UBP, but to quantitatively assess the effects of flat rate/UBP-tiered schemes using a large scale longitudinal dataset as well as investigate network effects in enforcing UBP schemes to end cross subsidization.

Most recently, the research community is starting to recognize the importance of pricing mobile broadband. Authors of TUBE [9] demonstrate that time-dependent pricing of mobile-broadband is powerful tool for reducing the peak-hour traffic in congested cells. While we do not investigate effects of temporal variability of demand here, we note that it plays important role in defining the traffic cost. Our time-independent approach in deriving the optimal caps, is driven by an implicit goal of simple tariffs. The design of more sophisticated, time-dependent and dynamic tariffs can benefit the efficiency of the provider, albeit it is not clear how large such benefits can be. On a more practical side, authors of [20, 21] investigate how accurate is the mobile broadband metering systems, isolate several sources of inaccuracy in such systems and describe various method for exploiting them. We note that the large variability of per-user demand has been documented in various domains in wired [6] and wireless [7] networks.

7. DISCUSSION

In this section we briefly discuss various factors that may affect the applicability of the observed results in other networks/operators. We also examine the implications of our findings on the end users, the operators and the mobile communications ecosystem in large.

Mobile network operators (MNO) that own radio spectrum and network infrastructure have a cost model that is not as simple and elastic as the one employed by MVNOs. In contrast to MVNOs that use and purchase the service on-demand, the MNO costs and revenues are not directly connected with the volumes of the usage of their customers. However, if the network has been planned optimally, the MNO capacity should closely match the demand generated by its customers. We believe that our results can offer insights on how the user behavior affects the revenues and usage of the cellular services and be instrumental in the process of network planning with the goal of maximizing the profits for the operator. We recall here that our dataset, that has tariffs with unlimited volume, lets us extract close to the ‘actual’ usage demands – usage demands that otherwise would be biased by self-regulation if volume caps had been in place.

The studied cost model is rather simple in that it treats each service unit equally in terms of cost, independent of the time or place of the request. The cost of delivering a service indeed depends whether the request was made in the peak-hour, the middle of the night, in an urban cell or a rural cell. The cost model we study averages all these conditions in order to simplify the billing of a service with millions requests (voice, sms, 3G data) per day. Other types of cost function, that take into account the spatial and/or temporal, variability of the traffic cost can easily be integrated in the studied model [27, 15, 9], though such considerations are out of scope of the present paper. Our analysis in this paper takes a rather static view on the user behavior, determined by their average usage and payment information. Understanding the dynamics of the studied phenomena would be an interesting direction for future research.

Additionally, we would like to stress that independently of whether the mobile provider is MNO or MVNO, targets low-income or high-income segments of the market, understanding the user behavior is a critical factor that drives not only the quarterly balance sheet, but also long-term strategy and network development. Therefore, the relationship between the users’ usage patterns on revenues and cost, studied holistically, is pivotal in determining how future networks will evolve.

Dealing with caps: user’s perspective

While the cost of delivery of mobile broadband is expected to decay due to the technological advancements, the decay is unlikely to be as quick as one may have wished; see Fig. 1. With the nontrivial cost of delivery of mobile broadband, some form of usage based pricing (UBP) with caps for network services appears inevitable as we have shown using our data. Given that majority of Internet users are accustomed to the non-metered Internet access, it will be interesting to observe their response to such UBP as well as the reaction from mobile app developers that would need to take a price-aware approach when designing apps that use mobile broadband and other network services.

Some early approaches to dealing with UBP, are reflected in several interesting projects regarding cap sharing that include Airmob [14] and Shair [10]. These projects are aimed at sharing available capacity across devices. Our results seem from Sec. 3 where we show that there is little correlation between usage of different services seems to indicate that it is indeed possible to share across different services. We have provided economic motivations for capacity sharing based solutions. From an app developer point of view, in addition to designing apps that judiciously consume data, they can themselves pay for part of the traffic the users generate [23]. Authors of [22] demonstrate that mobile broadband traffic has around 20% of redundant traffic that could be cached on the phone. Another method is to use existing network primitives when needed; a user who has exceeded the data caps and wants to use an over-the-top (OTP) app like Whatsapp can have some of her messages delivered via the sms channel – provided the user has sms capacity. Many existing operators have open APIs where this is feasible [31, 29].

8. CONCLUSIONS

This paper studies user behavior in terms of usage of different network services (voice calls, sms, data) and how the behavior affects revenues and costs of a cellular operator. In order to conduct this study, we use a large anonymized trace of over a million users spread over two years that consists of all usage information per user, the associated costs in delivering each service as well as their billing information for each user. Using the data, we characterize the individual user profitability and show that 20% of the users cost more to the operator than what they contribute as revenues. These heavy users are therefore being cross-subsidized by other users. We find that assigning value to a user without looking into her activity in the call graph, or without incorporating the role she can play in growing the network due to social network effects is myopic. We therefore quantify the social network effects and this enables us to more accurately assess the value of each user to the network. Armed with our knowledge, we study how volume caps affect the revenues and the costs of the operator and find that the difference between the revenues and the traffic costs can almost double by affecting just 16% of the user base.

To the best of our knowledge, this is the first work to provide insights on how user behavior impacts the economics of cellular operators, along with the viability of certain offerings (all-you-can buffet plans), quantitatively. We hope the findings of this work can help to plan new pricing mechanisms, do efficient network planning and provide motivation to develop new technologies to curb the increasing cellular traffic costs.

9. ACKNOWLEDGEMENTS

The authors would like to thank the reviewers of the paper for constructive comments, as well as our shepherd Vishnu Navda for his help.

10. REFERENCES

[1] Tellabs report, 2011. http://www.tellabs.com/news/2011/index.cfm/nr/142.cfm.

[2] Daniel Birke and G M Peter Swann. Network effects and the choice of mobile phone operator. *Journal of Evolutionary Economics*, 16(1-2):65–84, 2006.

[3] M. Chetty et al. You're capped: understanding the effects of bandwidth caps on broadband use in the home. In *Proceedings of ACM CHI*, 2012.

[4] Jon Crowcroft. Net neutrality: the technical side of the debate: a white paper. *ACM SIGCOMM CCR*, 37(1), 2007.

[5] Marcel D. et al. Glasnost: Enabling end users to detect traffic differentiation. In *NSDI*, pages 405–418, 2010.

[6] C. Estan and G. Varghese. New directions in traffic measurement and accounting. In *Proc. of ACM SIGCOMM*, 2002.

[7] H. Falaki et al. Diversity in smartphone usage. In *Proc. of ACM Mobisys*, 2010.

[8] M. Gjoka et al. Walking in facebook: A case study of unbiased sampling of osns. In *Proceedings IEEE INFOCOM*, 2010.

[9] S Ha et al. Tube: time-dependent pricing for mobile data. In *Proceedings of ACM SIGCOMM*, 2012.

[10] Pan Hui et al. Sharing airtime with shair avoids wasting time and money. In *Proceedings of ACM HotMobile 2009*.

[11] C. Kreibich et al. Netalyzr: illuminating the edge network. In *Proceedings of ACM IMC*, 2010.

[12] A. Lambrecht et al. Price discrimination in service industries. *Marketing Letters*, pages 1–16, 2012.

[13] Daniel A. Lyons. The impact of data caps and other forms of usage-based pricing for broadband access. In *Working Paper no 12-27, Mercatus Center, at George Washington Univ.*, 2012.

[14] MIT Media Labs. Airmobs Project, 2012. https://www.media.mit.edu/research/groups/viral-spaces.

[15] M. Motiwala et al. Towards a cost model for network traffic. *ACM SIGCOMM CCR*, 42(1):54–60, 2012.

[16] P. Nabipay, A. Odlyzko, and Z.L. Zhang. Flat versus metered rates, bundling, and "bandwidth hogs". In *NetEcon*, 2011.

[17] A. Odlyzko. Internet pricing and the history of communications. *Computer Networks*, 36:493–517, 2001.

[18] A. Odlyzko et al. Know your limits: Considering the role of data caps and usage based billing in internet access service. In *Public Knowledge*, 2012.

[19] J.-P. Onnela et al. Structure and tie strengths in mobile communication networks. *PNAS*, 104(18):7332–7336, 2007.

[20] C. Peng et al. Can we pay for what we get in 3g data access? In *Proceedings of ACM MOBICOM*, 2012.

[21] C. Peng et al. Mobile data charging: new attacks and countermeasures. In *Proceedings of ACM CCS*, 2012.

[22] F. Qian et al. Web caching on smartphones: Ideal vs. reality. In *Proceedings of the Mobisys*, 2012.

[23] H. Raj et al. Splitting the bill for mobile data with simlets. In *Proceedings of ACM HotMobile*, 2013.

[24] S. Keshav. Estimating the cost of SMS, 2008.

[25] S Soumya et al. Pricing data: A look at past proposals, current plans, and future trends. *CoRR*, abs/1201.4197, 2012.

[26] R. Stanojevic, V. Erramilli, and K. Papagiannaki. Cognitive bias in network services. In *Proc. of ACM HotNets*, 2012.

[27] R. Stanojevic, N. Laoutaris, and P. Rodriguez. On economic heavy hitters: shapley value analysis of 95th-percentile pricing. In *Proceedings of the ACM IMC*, 2010.

[28] Telco 2.0 Initiative. European Mobile: The Future's not Bright, it's Brutal, 2012. http://www.telco2research.com/articles/EB_European-mobile-future-brutal-telco_Summary.

[29] Telefonica. Bluevia Network API, 2013. http://bluevia.com/en/home/B.

[30] Upnext Mobile. White paper. How Iliad is shaping the future of mobile telecoms, 2012. http://upnextmobile.com/post/17945556081/freemobilefuturetelecoms.

[31] Verizon. Verizon Network API, 2013. http://developer.verizon.com/content/vdc/en/verizon-tools-apis/verizon_apis/network-api.html.

[32] Wikipedia. List of countries by mobile phones, 2012. http://en.wikipedia.org/wiki/List_of_countries_by_number_of_mobile_phones_in_use.

[33] Wired. Comcast bans Seattle man from Internet for his Cloudy Ways, 2011. http://www.wired.com/business/2011/07/seattle-comcast/.