

# Economics of User-in-the-Loop Demand Control with Differentiated QoS in Cellular Networks

Rainer Schoenen<sup>1,2</sup>, Halim Yanikomeroglu<sup>1</sup>

<sup>1</sup>Department of Systems and Computer Engineering, Carleton University, Canada

<sup>2</sup>Communication Networks (ComNets), Faculty 6, RWTH Aachen University, Germany

**Abstract**—Increasing cellular traffic is the driving force for innovations in wireless communications. While voice traffic is not expected to increase much and does not require 4G systems, traffic for video and data applications is expected to grow with a rate of 100% per year. Smart mobile devices, tablets and laptop dongles will certainly make this a reality. On the other hand the supply side cannot grow with the same rate. Base stations, eNB, pico- and femtocells will bring more heterogeneity in space and new applications will bring more heterogeneity in demand over time. Designing for over-provisioning capacity has been the standard approach to stabilize traffic, but it will be harder and harder, with more congestion situations in time (busy hour) and space (crowded cell) which will break application traffic and give bad quality-of-experience of users. Furthermore, over-provisioning comes with more power consumption and higher financial expenditures for infrastructure and operating costs.

The user-in-the-loop (UIL) approach offers a solution orthogonal to the traditional supply-only view. In addition to technical improvements, having a temporal demand control can alleviate the severity of busy-hour situations which formerly caused congestion and connection failures. Demand shaping is implemented by a dynamic usage-based tariff and adaptive rates depending on the load condition. The users in a cell are part of a closed control loop which reacts in cases of severe demand overload. In this paper three different service classes are controlled individually and results from analysis and simulation show the performance in stationary and dynamic scenarios. The economics of tariffs and dynamic prices and the resulting operator revenue on one side is compared to the dissatisfaction of rejected users and this gives decision indicators for the investment into new infrastructure. Overall this saves money, energy and turns situations of hard congestion into an elastic stationarity which is in the interest of both users and operators.

**Index Terms**—User-in-the-loop (UIL); demand shaping; congestion; tariff; QoS; sustainability; cross-layer; green; economics.

## I. INTRODUCTION

**I**N cellular networks the trend is towards increasing data rates. On the supply side new IMT-Advanced generations are evolving using sophisticated PHY and MAC layer improvements and increasing the antenna density with pico- and femtocells will also provide better capacity. On the demand side the data rates are increasing with an estimated growth rate of almost 100% over the next 10 years [1]–[4] due to the popularity of smart mobile devices, tablets and laptop dongles which run more and more data hungry applications, software updates, etc. While voice traffic will not contribute with more than 10% to the total traffic (and does not require

new wireless generations) the killer applications are streaming video, app data and software downloads. The demand is fueled by successful marketing of user terminal (UT) devices and flat rate tariffs which the user believes he needs to make all this work. There are no good tariff alternatives on offer which really address the user needs:

- reliable service (“it works”), anytime, anywhere,
- no locked services (e.g., VoIP, IMAP),
- affordable prices (less than a few ct per MB),
- a precise overview about the currently accumulated costs,
- no bill shock (surprises of \$20..\$50 more on the bill),
- ability to set limits to expenses (incl. parental control).

On the other side, the operators fear heavy users, which are currently a minority of 20% but are responsible for 80% of the traffic. For this reason they introduced caps on flat rates and a varying set of blocked application ports which on the other hand also annoys serious users. Flat rate caps are not a good solution (open loop control only) as users still have to monitor their traffic (“when do I reach the cap?”) and some are intentionally using up the remaining quota at the end of the month. This is obviously not helping the business and cannot avoid congestion on the wireless links. Indeed, congestion is more and more likely to happen in the future as the capacity (supply) cannot scale with the huge increase of demand. Congestion will happen during the busy hours first and at some point the links can be assumed to be 100% utilized all the time, while the real demand is higher than the capacity. Application traffic fails under such circumstances due to huge delays, TCP timeouts and bursty effects due to TCP rate control running over hundreds of parallel connections. In effect, the user satisfaction (quality of experience, QoE) will drop and the problem is blamed on the wireless carrier.

It is therefore important to avoid overload and congestion situations in all networks. In this paper, the solution approach is user-in-the-loop (UIL). Previous work [5], [6] has shown that influencing the user behavior is a promising approach with substantial gains in spectral efficiency and successful congestion mitigation. Dynamic usage-based tariffs are controlled in a closed loop including the user. The rate controller [7] is revisited in short form before we study three different service classes as well as the economics of dynamic prices in equilibrium. The revenue of operators is compared in situations with and without UIL, taking infrastructure costs into account.

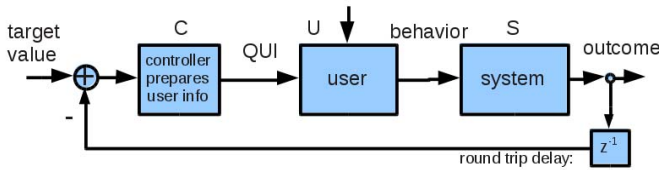


Fig. 1. User-in-the-loop (UIL): control of user and system [5].

Finally, the user dissatisfaction by either congestion or too high service prices is integrated into the picture.

The paper organization is as follows. The UIL concept is introduced first, followed by the control loop analysis and user behavior study. Next, the economics from operator perspective is treated and related to the user QoE.

## II. USER IN THE LOOP IN SPACE AND TIME

The closed loop control of user behavior for either a spatial or a temporal user reaction is called user in the loop (UIL) [5]. The user is part of a closed loop control system (Figure 1). The user receives hints and incentives or penalties (quantified user information, QUI) in order to convince him to deviate from the default behavior (which is unconstrained, open loop). The outcome is the user moving from one location to a better one [5] or postponing session traffic out of the busy hours [6].

The user receives this QUI on the graphical user interface (GUI), e.g., by the proposal of a better time to start his session (out of the busy hour). The system block is the wireless channel usage by user-initiated traffic. The outcome is a measurement telling the current status of the system output. This can be spectral efficiency or data rate produced/consumed. After comparison with the target value the controller calculates how much control to exert and changes QUI accordingly. Due to the closed loop, the target value can be achieved within a short settling time.

Spatial UIL control has been treated in [5], [8], [9], providing substantial gains in spectral efficiency  $\bar{\gamma}$ . Recent survey results support the assumptions [10], [11].

In this paper the focus is on temporal control. The long term spiral of growth leads to a fast exponential demand increase [1], [2] which is shown in Figure 2 together with the result of UIL control for three traffic classes. Note that the exponential graph is unconstrained demand. In practice there are external bounds by technology and (open loop) tariff caps. Either way, there will be a congestion problem at some point in the future because wireless capacity cannot grow adequately with 100% per year. Figure 2 shows the solution and outcome of the UIL control with three service classes. Voice, video and data have different QoS requirements and are therefore treated separately.

Temporal UIL control in Figure 1 means the target value is the sustainable rate (close to the capacity of a radio cell), e.g., load  $\rho^{(t)} = 0.95$ , at which all applications still work with reasonable QoS. A priority scheme is always assumed in MAC layer, so that real-time (RT) traffic is preferred and observes a much lower load and therefore lower queueing delay. The signals inside the control loop are vectorized for the traffic

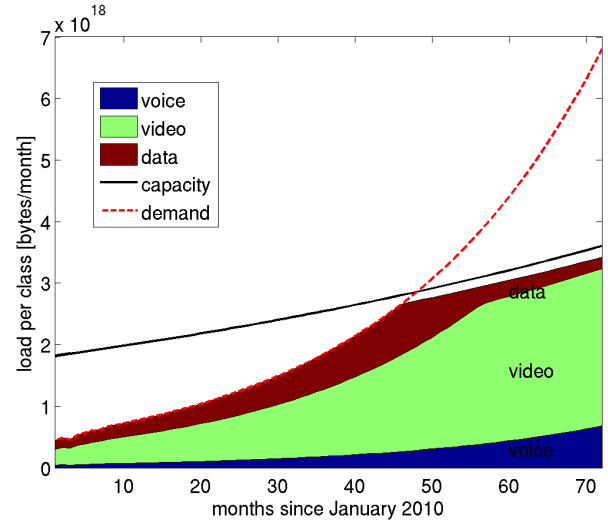


Fig. 2. This figure shows the exponential growth of demand for mobile traffic  $r^{(u)}(t)$  ( $u$ =unconstrained), the limited capacity  $\hat{R}(t) = r^{(t)}(t)$  and the effect of temporal UIL control to different traffic classes.

classes voice (S), video (V) and data (D) assumed in this paper. The controller determines a dynamic price rate [6] per class - valid for all users and all new sessions - and updates this information to the user terminal (UT). The handling on the UT is up to the operating system. Either the price (rate) is displayed before each transaction when it is above a user-defined threshold or an agent or manager software on the UT will act on behalf of the user. This must be fully transparent to the user but can be muted for convenience. The user or a software agent decides to use or not to use the service at the current time, location and price. The output of the user black box is a reduced traffic demand when the session price goes up. The pricing method will change the user behavior and the traffic similar to the system of dynamic electricity tariffs and smart-grid applications.

All users in the same cell see the same specific price  $\pi_C$  per class of service  $C$  ( $\pi_C$  is in \$/bit). For the example later we assume nominal prices  $\pi_C^{(N)}$  for voice: 8 ct/MB, video: 4 ct/MB, data: 2 ct/MB. Results are only affected by the ordering, not the absolute amount.

We define  $\chi_C$  as normalized price increase, per class  $C$ , with a default of 0, i.e.,  $\chi_C = (\pi_C - \pi_C^{(N)})/\pi_C^{(N)}$ .  $\chi_C \geq 0$ .

The user reaction to  $\chi_C$  is the control ratio, defined as;  $p = r^{(c)}/r^{(u)}$ .  $r^{(c)}$  is the controlled rate (output) and  $r^{(u)}$  is the uncontrolled rate (where we assume nominal demand at nominal price level). The control ratio  $p$  is the average reduction per user. It can either come from single users reducing their demand or a split of users into a set of users not changing their original demand ( $p$ ), and the set of users not triggering traffic  $(1 - p)$ . The controller determines  $\chi_C$ . This stays on 0 if the system is not in congestion, but acts when  $r^{(u)} > r^{(t)}$ .

The user reacts stochastically but in total over all users the reaction per class  $C$ ,  $p_C = f(\chi_C)$  leads to the decreased demand  $r^{(c)} = p_C \cdot r^{(u)}$ . In a real system the user behavior is estimated (Kalman filter) by gathering statistics on his

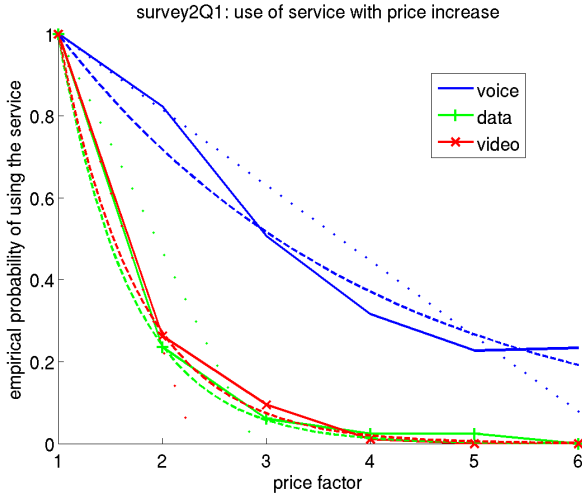


Fig. 3. For three different service classes this shows the user reaction [11] to an increase in price  $p(\chi)$ . The dotted and dashed lines show the linear and exponential fits (Eq. 1). Obviously data and video are more elastic than voice.

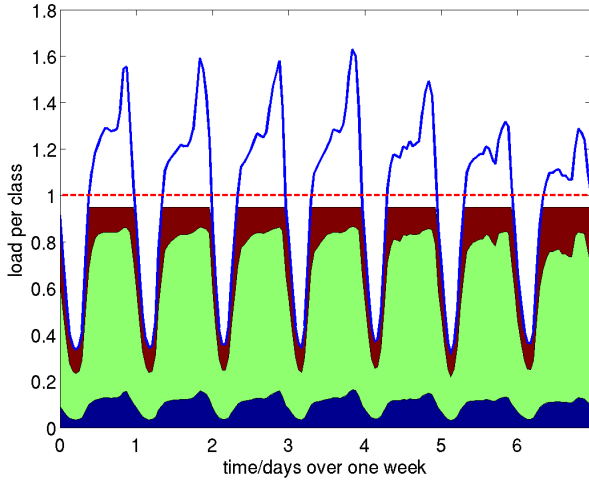


Fig. 4. Simulated congestion during busy hours (7 days). UIL reduces the aggregate traffic below the threshold. The three classes voice (blue), video (green) and data (red) are put together by the ratio which results from individual controlled price rates (method M3). Voice is rarely changed.

conditional accept/deny pattern.

The user behavior cannot be precisely known, but only the aggregate behavior of all users in a cell is required and a recent survey [11] provides some working assumptions for the classes S, V, D, which improves upon the early UIL work [6] assuming a linear elastic model. In this paper the user response is assumed following these equations, displayed in Figure 3.

$$p_C(\chi) = e^{-\eta_C \chi_C}, \quad (1)$$

$$p_C^{(lin)}(\chi) = \max(0, 1 - e_C \cdot \chi_C). \quad (2)$$

Equations 1 and 2 hold for  $\chi_C \geq 0$ . The constants determined by survey data [11] are  $\eta_S = 0.330$ ;  $\eta_D = 1.429$ ;  $\eta_V = 1.304$  and  $e_S = 0.18$ ;  $e_D = 0.73$ ;  $e_V = 0.52$ .

To reduce the uncontrolled traffic load  $r^{(u)}$  to  $r^{(t)}$ , the overall control ratio  $p$  must be chosen as  $p = \min(r^{(t)}/r^{(u)}, 1)$ , independent of the traffic class. The next task of the controller is the determination of the  $\chi_C$ , depending on the pricing

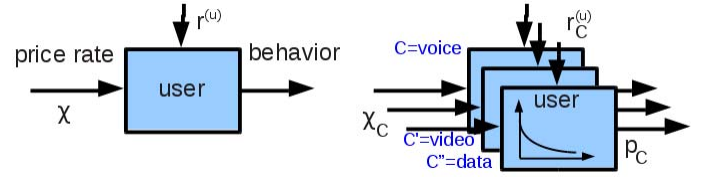


Fig. 5. I/O model of the user box. Input is the unconstrained demand  $r_C^{(u)}$  per traffic class  $C$  and the dynamic price indicator  $\chi_C$  from the controller. Output is the traffic intensity value  $r_C^{(c)} = p_C \cdot r_C^{(u)}$ . During the length of a session the price is assumed constant. Thus there is a temporal system response function.

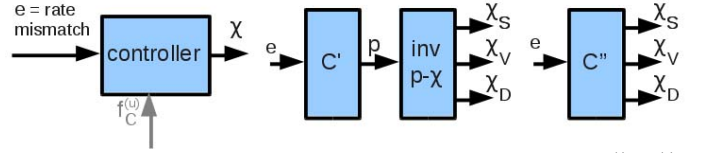


Fig. 6. I/O model of the controller. Input is the rate mismatch  $e = r^{(t)} - r^{(c)}$ . The traffic class proportions  $f_C^{(u)}$  are known optionally as side information. Three different methods M1, M2, M3 are shown from left to right.

model [6]. In this paper, a pricing proportional to the usage volume  $v$  is assumed using  $\pi$  as the dynamic price=tariff rate.

$$\pi_C = \pi_C^{(N)} \cdot (1 + \chi_C). \quad (3)$$

Figures 5 and 6 show the model of user and controller box. In stationarity, the I/O response from  $\chi$  is a control ratio  $p$  (estimated by Eq. 1). For numeric results we assume a traffic class mix of

$$f_S^{(u)} = 10\%, \quad f_V^{(u)} = 59\%, \quad f_D^{(u)} = 31\%, \quad (4)$$

which are the projected proportions for 2020 [2] and assumed constant for the whole time span (Figure 2). The unconstrained traffic demand  $r_C^{(u)}$  is assumed to be split into traffic these classes  $C$  with  $r_C^{(u)} = f_C^{(u)} \cdot r^{(u)}$ .

Dynamic simulation results are shown in Figure 4 and Figure 7 shows the relative price  $\chi_C + 1$  and control ratio  $p_C$  for different control methods explained now.

For the following control discussion, three ways of control are assumed (Figure 6), different in the way they treat the ratio of traffic classes.

- M1="same relative price" (Figure 6 left),
- M2="same control ratio" (Figure 6 middle),
- M3="prioritized" (Figure 6 right).

If service class distinction is not a concern, then method M1 solves the control by providing the same normalized price  $\chi$  for all classes. Figure 7 left shows that it has a different effect on the traffic of each class because the elasticity is different

In stationarity  $e = 0$  ( $r^{(t)} = r^{(c)}$ ) can be assumed due to the I-component of the controller. Thus we can derive the stationary solution here by solving for  $\Theta := r^{(u)}/r^{(t)}$  in Eq. 6:

$$r^{(t)} = r^{(c)} = r^{(u)} \cdot \sum_{C=1}^3 f_C^{(u)} \cdot p_C(\chi_C), \quad (5)$$

$$0 = \Theta \cdot \sum_{C=1}^3 f_C^{(u)} \cdot e^{-\eta_C \chi} - 1. \quad (6)$$

The value  $\chi$  is the zero of  $f(\Theta)$  in Eq. 6. When  $\Theta \geq 0$  the function decreases monotonically from  $+\Theta$  to  $-1$  and there is

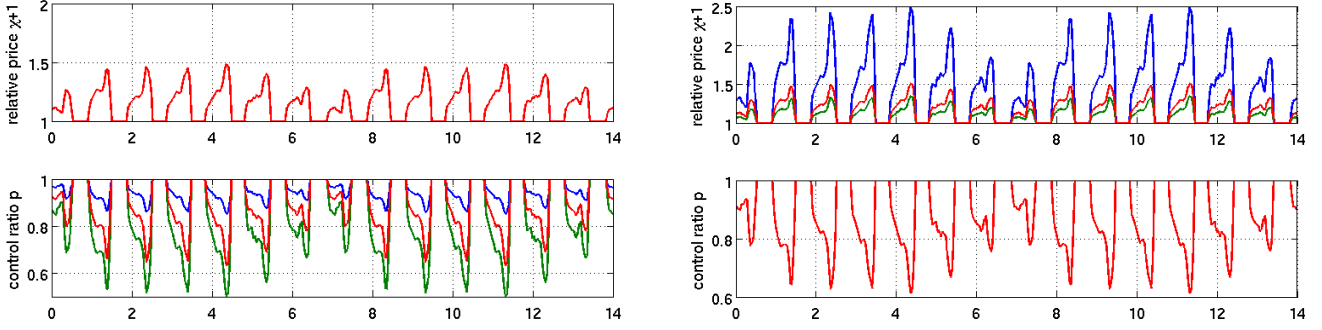


Fig. 7. UIL temporal control over a 14 days period for all three service classes (blue=voice, red=video, green=data) for the control method M1="same relative price" ( $\chi$ ) left and M2="same control ratio" ( $p$ ) right. The sum rate is controlled below the target capacity. Shown is the relative price  $\chi_C + 1$  above and the control ratio  $p_C$  below. Graphs show that M1 reduces data most (same  $\chi$ ) while M2 raises the tariff for voice most (same  $p$ ).

exactly one zero. If the user response model is assumed linear (Eq. 2), then  $\chi$  can be explicitly calculated:

$$0 = \Theta \cdot \sum_{C=1}^3 f_C^{(u)} \cdot (1 - \eta'_C \cdot \chi) - 1, \quad (7)$$

$$\chi = (1 - 1/\Theta) / \sum_{C=1}^3 f_C^{(u)} \cdot \eta'_C. \quad (8)$$

A more elaborate method, M2, is designed to reduce each traffic class by the same degree  $p$ . The controller needs to calculate  $p$  first, assuming  $K > 0$ , and then translates it into different  $\chi_C$ . The latter step is straightforward by inverting the user response functions of Eq. 1:

$$\begin{aligned} \chi_{\text{voice}}(p) &= -\ln(p)/0.330 \\ \chi_{\text{data}}(p) &= -\ln(p)/1.429 \\ \chi_{\text{video}}(p) &= -\ln(p)/1.304. \end{aligned} \quad (9)$$

The last method, M3, calculates  $\chi_C$  (and  $p_C$  internally) directly and different for each class. According to an assumed priority inside the scheduler and corresponding stacked prices, M3 controls the cheapest traffic first, as long as its absolute price  $\pi_D$  (data) is below the next level of  $\pi_V$  (video). From this load (operating) point on, both are controlled proportionally. A relative gap of  $\beta$  between them is optional. The next switching point is when  $\pi_V$  touches  $\pi_S$ . beyond that point, all of them are controlled proportionally. Figure 8 graph 2 shows how the prices are aligned. The order of QoS classes (by absolute price) is preserved. Voice traffic would be modified only minimally in extreme demand overload, because in this exotic case users are willing to pay the same price also for video and data. This assumption is driven by the surveyed data. The reality might give different price tags, but the principle works for a wide range of tariffs with our algorithm explained here.

In Figure 8 the uncontrolled load  $\rho^{(u)} = 5$  is not serious for the QoS of voice traffic, since the partial load for voice is only  $\rho_S \approx 0.3$  at  $\rho^{(u)} = 5$ . Thus the system looks like a queueing system in underload for the prioritized voice traffic.

Note that in Figure 8, the operating point (x-axis) of  $\rho = \lambda/\mu = r^{(u)}/\hat{R}$  ("deep demand overload situation") is beyond values usually used in traffic and queueing theory. There  $\rho < 1$  is required for stability. By using UIL, traffic is controlled back to a level close to  $r^{(t)}$ . Here  $\rho^{(t)} = 0.95$ . An interesting

interpretation of this is as load-dependent "soft" CAC. In this CAC, sessions are accepted only if the user has agreed to paying the negotiated price for it.

The calculation of  $\chi_C$  in model M3 works as follows. Eq. 5 has to be solved like for M1, but the  $\chi_C$  are different per class. The numeric solution requires a distinction of cases depending on which classes are affected by the constraint in Eq. 10. 1) No class, 2) only D, 3) D+V, 4) all classes:

$$\pi_S \cdot (1 + \chi_S) \geq \beta \pi_V \cdot (1 + \chi_V) \geq \beta^2 \pi_D \cdot (1 + \chi_D). \quad (10)$$

For the solution we substitute  $\chi_V = (1 + \chi_D) \cdot \beta \pi_D / \pi_V - 1$  and  $\chi_S = (1 + \chi_D) \cdot \beta^2 \pi_D / \pi_S - 1$  into Eq. 5. The resulting (huge) term has the form of  $f(\chi_D) = 0$ . The solution is obtained by finding the zero. This analysis provided all results of Figure 8.

As Figure 8 shows, traffic of the cheaper class is affected before it touches the price of the next higher traffic class. As soon as two classes touched, they are treated the same way by UIL from that point on. A priority scheduler in the MAC layer must still be present in order to guarantee class specific QoS, e.g., delay bounds.

### III. ECONOMICS OF UIL

While the UIL approach in this paper provides a technical solution for congestion mitigation, there are interesting questions about the economics on the user side and operator side. The revenue rate  $R$  (unit \$/second) for the operator at a moment of time  $t$  (with  $\pi_C^{(N)}$  = nominal tariff for class  $C$ ) is

$$R(t) = \sum_{\forall C} \pi_C^{(N)} \cdot (1 + \chi_C(t)) \cdot r_C^{(c)}(t). \quad (11)$$

Without loss of generality a constant rate  $r_C^{(u)} \propto \rho^{(u)}$  per class  $C$  is assumed, representing an average only depending on  $\rho^{(u)}$ .

Assuming usage-based pricing, the revenue increases with the carried traffic volume, which is proportional to the load of the system (wireless cell) as long as it is not congested ( $\rho^{(u)} < 1$ ). If the system could carry more traffic, the revenue would continue to increase (uncontrolled). But in the best case assumed, the carried traffic is constant equal to the capacity (limited). Figure 9 shows these two boundary cases in black dashed lines. With UIL control as defined in section II, price rates will increase in congestion with no further increase in traffic. This has some advantages:

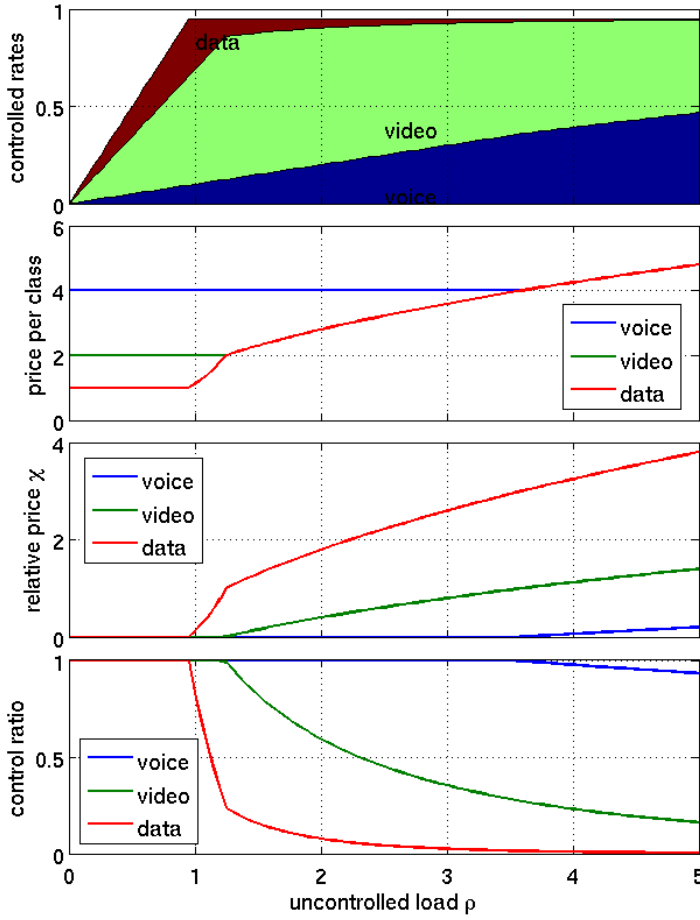


Fig. 8. This figure shows the controlled traffic for the “priority price” rule (M3, with  $\beta = 1$ ) with x-axis parameter  $0 \leq \rho^{(u)} \leq 5$ . The top graph shows the proportions  $\rho_C^{(u)}$  of each traffic class. The sum never exceeds 100%. Graph 2 shows the prices  $\pi_C$  for each class. Graph 3 shows the relative price  $\chi_C$ , which affects the user to react with  $p_C$  as shown in Graph 4.

- 1) The real traffic is in the stable region  $\rho^{(c)} < 1$ ;
- 2) short peaks in demand have a soft consequence (price up, demand down) instead of congestion collapse and application failures;
- 3) the region  $1 \leq \rho^{(u)} < 2$  can even produce more revenue in some cases (see M2 results in Figure 9);
- 4) Incentive to offload some downloads to WiFi or cable;
- 5) capital expenditures (CAPEX) for infrastructure do not need to be spent too early in order to over-provision. Instead, the moment to invest can be postponed.

While this is advantageous, there is a catch in this picture. The user satisfaction with UIL will drop to zero depending on a) the price increase and b) the frustration for traffic not carried. On the other hand, without UIL, the satisfaction will drop when the system in overload breaks application performance. In both cases the dissatisfaction has an economic impact in mid-term, if there is better competition in the market.

To quantify the user side, a model called “virtual refund” is proposed. A virtual refund is not paid back in real money, neither does the customer know about it, but it is a variable subtracted from the pure usage-based revenue to account for

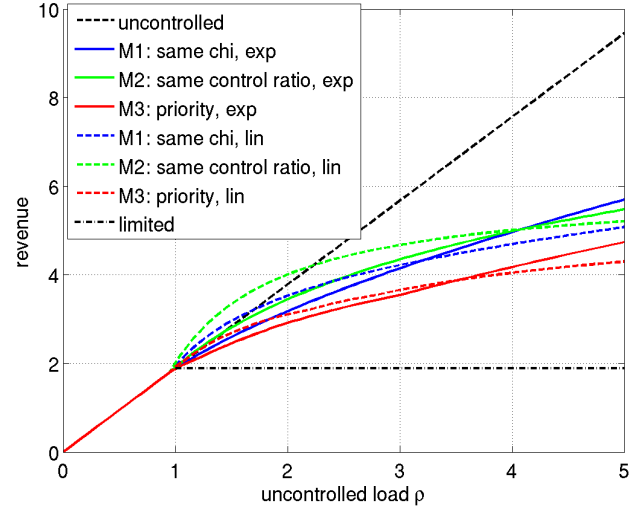


Fig. 9. Comparison of revenues in the full range of (over)load situations (unconstrained demand overload  $\rho^{(u)}$ ). Different UIL assumptions: Price rates are multiplied with controlled data rates. Thus this graph shows the total revenue rates  $[\$/time]$ . Revenue is shown for the three methods M1, M2 and M3. M3 (the priority scheme) imposes the least financial burden on the customer, but this corresponds to the lowest revenue for the operator. It can be observed that the revenue can be increased further even by operating in virtual overload  $\rho^{(u)} \leq 1$ . For most methods and in the range between  $1 \leq \rho^{(u)} \leq 2$ , the revenue is similar or even above the regular uncontrolled revenue (where we assume the system has infinite capacity and all traffic was carried).

unsatisfied users. We assume that with satisfaction  $S = 0$ , dissatisfaction  $D = 1 - S$  gives  $D = 1$  and the total price will be imaginarily refunded. With full satisfaction  $S = 1$ ,  $D = 0$ , the revenue stays as in Eq. 11. Therefore we get:

$$R(t) = \sum_{\forall C} \pi_C^{(N)} \cdot (1 + \chi_C) \cdot r_C^{(c)} \cdot \bar{S}_C. \quad (12)$$

$\bar{S}_C \in [0, 1]$  is the average satisfaction with service  $C$  over all users. The choice of  $S$  depending on QoS is arbitrary, but a mapping like QoE is recommended. For simplicity we assume in this paper that without UIL and in overload  $\rho^{(u)} > \rho^{(t)}$  all applications show failures, except voice due to its priority in the scheduler. Thus, satisfaction is binary as

$$\bar{S}_{C \in \{V, D\}}^{(noUIL)}(\rho^{(u)}) = 1 \cdot (\rho^{(u)} \leq \rho^{(t)}). \quad (13)$$

With UIL, the satisfaction is assumed to drop with the control rate  $p_C$  to reflect unhappiness of those users not accepting the current price rate:

$$\bar{S}_C^{(UIL)}(\rho^{(u)}) = p_C(\rho^{(u)}). \quad (14)$$

Figure 10 shows the virtual revenue using the virtual refund approach for situations with and without UIL. It can be observed that without UIL only voice traffic is handled adequately in overload due to its priority handling and low effective sum rate. With UIL, despite dissatisfaction with the price level, there is a comfortable soft crossover region around  $\rho^{(u)} = 1$  which makes it a realistic strategy for regular overload conditions, like those happening during the busy hours.

The difference between the uncontrolled revenue in Figure 10 and the virtual revenue with dissatisfied customers is

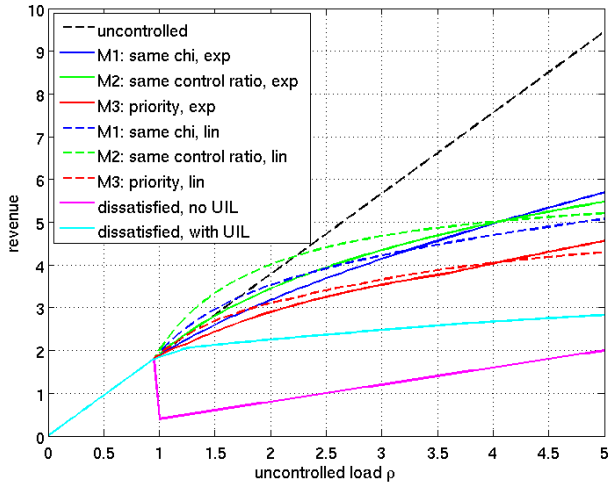


Fig. 10. Revenue comparison considering user satisfaction: Compared to Figure 9 the virtual revenue (revenue reduced by virtual refunds) is still useful for  $\rho^{(u)} \geq 1$  while it shows failure to perform without UIL. Without UIL the satisfaction shown here is  $> 0$  only because voice traffic is still carried without disruptions.

proportional to the incentive of an operator to invest into new infrastructure. The comparison is performed over a reasonable time span, e.g. one month, taking the  $PDF(\rho^{(u)})$  of the load situations over time into account, e.g., from Figure 4.

A practical problem these days is the dominance of flat-rate plans. Technically, a differentiated QoS is not meaningful without separate specific prices per class, and UIL is not possible without a (financial) motivation. Figure 11 shows how any flat-rate plan can be converted into a usage-based plan by providing refunds to the user if the included volume  $c$  (cap) is not utilized. This way it allows a smooth transition strategy in a sensitive market environment. Alternative usage-based functions other than linear (e.g. logarithmic tariff [6]) are possible with UIL by simply counting an “effective volume”, proportional to  $(1 + \chi)$  times the real (measured) volume. At the end of the month, higher-priced intervals appear as a higher volume, thus the effective volume can be multiplied with the standard nominal price rate  $\pi_C^{(N)}$  again. To the users, the interface optionally just shows the severity factor  $(1 + \chi)$ , e.g., in colors green, yellow and red, based on user preferences.

#### IV. CONCLUSION

In this paper the control and economic aspect of the user-in-the-loop (UIL) concept is investigated in order to find relevant business cases for the operator and the wireless subscribers. The closed loop allows for differentiated services, dynamic prices and respects quantitative user response data. The proposed temporal UIL control resolves congestion situations in the busy hour by providing dynamic prices individually to each service class. Numeric results show the feasibility of controlling traffic into a stable operating point this way. In simulations this has turned out to be robust against inaccuracies of user behavior and nonlinear effects. The method has been applied to the traffic classes voice, video and data.

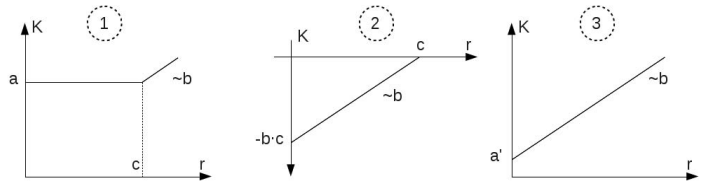


Fig. 11. Any flat rate plan ① can be converted to a linear usage-based tariff ③ by providing a refund ② for unused capacity ( $r < c$ ).  $r$  is monthly volume [GB],  $K$  is cost or revenue [\$/month].  $c$  is the monthly cap,  $a$  the flat price,  $b$  the linear price [\$/MB]. Add ① and ② to get ③.

The main contribution of this paper is an economic interpretation of the UIL price situation. The revenue of the operator is compared between the case without and with UIL. UIL helps to operate the system safely in the demand overload condition during busy hours which makes such operating conditions softer compared to the hard application failures which would happen without UIL in real overload. The economic analysis can also be used to support operators to find the point at which an investment into new infrastructure is recommended. On the user side, UIL can also lead to dissatisfaction, but to a much lower degree than the anger about broken applications in congestion.

Avoiding excess traffic at certain times can postpone infrastructure investments, therefore also saving energy and not only operators' money. Therefore it supports the idea of green or sustainable networks. Having a price for each transaction (like watching a video for \$1) trains users to be green-aware, which would be impossible with flat-rate tariffs. Please follow [12].

#### REFERENCES

- [1] “2010 mobile internet phenomena report,” Sandvine Inc., White Paper. [Online]. Available: <http://www.sandvine.com/downloads/documents/>
- [2] “Mobile broadband capacity constraints and the need for optimization,” Rysavy Inc., White Paper, February 2010. [Online]. Available: <http://www.rysavvy.com/Articles/>
- [3] “UMTS Forum report 44 - mobile traffic forecasts 2010-2020,” <http://www.umts-forum.org/>, UMTS Forum, Tech. Rep., Jan 2011.
- [4] “Cisco visual networking index: Global mobile data traffic forecast update, 2010–2015,” Cisco Systems Inc., White Paper, February 1, 2011. [Online]. Available: <http://www.cisco.com/en/US/solutions/>
- [5] R. Schoenen, H. Yanikomeroglu, and B. Walke, “User-in-the-loop: Mobility aware users substantially boost spectral efficiency of cellular OFDMA systems,” *IEEE Communications Letters*, vol. 15, no. 5, pp. 488–490, May 2011.
- [6] R. Schoenen, G. Bulu, A. Mirtaheri, and H. Yanikomeroglu, “Green communications by demand shaping and User-in-the-Loop tariff-based control,” in *Proceedings of the 2011 IEEE Online Green Communications Conference (IEEE GreenCom'11)*, Online, 2011.
- [7] R. Schoenen and H. Yanikomeroglu, “Dynamic demand control with differentiated qos in user-in-the-loop controlled cellular networks,” in *submitted to 2012 IEEE Globecom*.
- [8] R. Schoenen, “On increasing the spectral efficiency more than 100% by user-in-the-control-loop,” in *Proceedings of the 16th Asia-Pacific Conference on Communications (APCC)*, Auckland, October 2010.
- [9] T. Beitelmal, R. Schoenen, and H. Yanikomeroglu, “On the impact of correlated shadowing on the performance of user-in-the-loop for mobility,” in *Proc. UNet-Workshop at the IEEE International Conference on Communications (ICC)*, Ottawa, Canada, June 2012.
- [10] R. Schoenen, G. Bulu, A. Mirtaheri, T. Beitelmal, and H. Yanikomeroglu, “First survey results of quantified user behavior in user-in-the-loop scenarios for sustainable wireless networks,” in *currently under review*, 2012.
- [11] —, “Quantified user behavior in user-in-the-loop spatially and demand controlled cellular systems,” in *Proceedings of the European Wireless*, Poznan, 2012.
- [12] R. Schoenen, “User-in-the-loop project,” <http://userinthe-loop.org>.