Impatience in mobile networks and its application to data pricing

Fabrice Guillemin, Salah Eddine Elayoubi
Orange Labs, France
{fabrice.guillemin,salah.elayoubi}@orange.com

Philippe Robert, Christine Fricker, Bruno Sericola
INRIA, France
{philippe.robert,christine.fricker,bruno.sericola}@inria.fr

***Abstract*—We consider in this paper an important Quality of Experience (QoE) indicator in mobile networks that is reneging of users due to impatience. We specifically consider a cell under heavy load conditions and compute the reneging probability by using a fluid limit analysis. By solving the fixed point equation, we obtain a new QoE perturbation metric quantifying the impact of reneging on the performance of the system. This metric is then used to devise a new pricing scheme accounting of reneging. We specifically propose several flavors of this pricing around the idea of having a flat rate for accessing the network and an elastic price related to the level of QoE perturbation induced by communications.**

I. Introduction

An important aspect of the quality of experience (QoE) in cellular networks, which is in general a complex mixture of many parameters, is the reneging by users. Reneging results from impatience when users feel that their communications last for an excessive amount of time. Such a phenomenon is negative for both the user and the network as radio resources are uselessly consumed by impatient users before they abort their communications.

Impatience has been the object of several works dealing with fixed networks. In [1], a new version of Erlang formula has been derived and is applicable to the case of streaming-like flows, where the service duration is independent of the quantity of resources obtained by the user. This is a key difference with the case of data traffic considered in this paper, where service duration is proportional to the quantity of resources obtained by the user. In [2], the authors model impatience using the deterministic service curves approach and used simulations to quantify its impact on the system performance for several bandwidth sharing disciplines.

In [3], the authors analyze data traffic at the flow level and consider impatience of users in the overload regime, when the mean arrival intensity is larger than the mean service rate. Along the same line of investigations, we develop in this paper a fluid flow analysis of impatience in cellular networks. We notably establish a fixed point formulation for the computation of the reneging probability for users sharing bandwidth of a cell under overload conditions. These reneging probabilities are then used to introduce a new metric, namely QoE perturbation, expressing how much a particular flow impacts the reneging probability in the system. We then use this QoE perturbation metric to design a new pricing framework.

The main contributions of this paper are as follows:

- We develop a tractable analytic model for impatience of users in mobile networks, from both network and user perspectives (steady-state probabilities and probability of reneging);
- We develop a QoE perturbation model based on the impact of a new communication on the reneging probability of users in the cell;
- We discuss how this QoE perturbation scheme can be exploited for pricing.

In Section II, we describe the system under consideration and introduce impatience from a network point of view. In Section III, we develop a fluid limit of the system. On the basis of this analysis, we devise in Section IV new pricing schemes accounting of impatience. Concluding remarks are presented in Section V.

II. System model with impatience

A. Basic model with no impatience

We deal with one cell of a cellular network and consider a general model, applicable to 3G and 4G networks, where resources (time slots in HSDPA and Resource Blocks in LTE) are equally divided among users. Because of propagation and interference, the capacity at the edge of a cell is lower than that at its center, as illustrated in Figure 1 where the throughput distribution obtained from drive test measurements is represented.

In this paper, we represent the radio cell as a concatenation of K concentric rings. The bit rate in one ring is assumed to be constant and is denoted by c_k for ring k with $c_1 > c_2 > \ldots > c_K$. From a modeling point of view, this corresponds to a multi-class system where users at different positions belong to different classes. Ring k has a weight in the total traffic demand equal to p_k such that $\sum_{k=1}^{K} p_k = 1$.

We assume that users present in the cell download data, thus giving rise to data flows (typically TCP connections). In the following, we assume that flows appear according to a Poisson process with rate λ. The cell capacity is shared among the various data flows according to the processor sharing discipline. If a data flow originates in ring k, the instantaneous service rate for this flow is equal to c_k/n if there are globally n active flows.

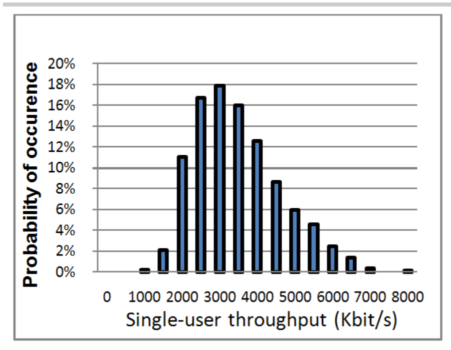

Fig. 1. Probability distribution function of throughput obtained from drive test measurements in an HSDPA network in a large European city (category 14 devices with LMMSE receivers).

Let us assume that the volume σ of data per flow has a general distribution with mean $\mathrm{E}(\sigma)$. Given the heterogeneity of the bit rates available to users according to their location, we are led to consider a multiclass processor sharing system, where a customer of class k has a requested service time equal to $T_k = \sigma/c_k$. In the following we define $\mu_k = 1/\mathrm{E}(T_k) = c_k/\mathrm{E}(\sigma)$.

The steady state probability of the multiclass processor sharing queue is known in the literature to have a product form [4]. Specifically, the steady state probability that there are n_k flows in progress in class k for $k = 1, \ldots, K$ is

$$P(\mathbf{n}) = (1-\rho)\frac{|\mathbf{n}|!}{\mathbf{n}!}\prod_{k=1}^{K}\rho_k^{n_k},$$

where $\mathbf{n} = (n_1, \ldots, n_K)$, $\mathbf{n}! = n_1! \ldots n_K!$, $|\mathbf{n}| = n_1 + \ldots + n_K$, $\rho_k = \lambda p_k/\mu_k$ and $\rho = \sum_{k=1}^{K}\rho_k$. The stability condition of the system reads $\rho < 1$.

In the following, we consider for $n \geq 0$ the set $\mathcal{S}_n$ composed of those K integer tuples $(n_1, \ldots, n_K) \in \mathrm{N}^K$ such that $n_1 + \ldots + n_K = n$. The cardinal of $\mathcal{S}_n$ is $|\mathcal{S}_n| = \binom{n+K-1}{n}$.

B. Modeling impatience

We suppose that users may abort their download if the transmission of data takes too much time. We specifically assume that users renege at the expiration of some timer with duration τ independent of the transmission time. We further assume that the random variable τ is exponential with mean $1/\mu_0$ and that $\mu_0 < \mu_K$. This assumption means that a user is ready to wait more than the time needed for the completion of the download if he were alone in the cell in the worst radio conditions.

Let $n_k(t)$ denote the number of customers (i.e., the number of active flows) of class k in the system at time t. Assuming that the amount σ of data to transmit is exponentially distributed, the process $((n_1(t), \ldots, n_K(t)))$ is a Markov process.

Let $q(\mathbf{n}, \mathbf{m})$, where $\mathbf{n}, \mathbf{m} \in \mathrm{N}^K$, denote the transition rate from state $\mathbf{n}$ to state $\mathbf{m}$. The non-null transition rates are

$$q(\mathbf{n}, \mathbf{n}+\mathbf{e}_k) = \lambda p_k, \quad \text{and} \quad q(\mathbf{n}, \mathbf{n}-\mathbf{e}_k) = \frac{n_k\mu_k}{|\mathbf{n}|} + n_k\mu_0,$$

for $k = 1, \ldots, K$, where $\mathbf{e}_k$ is the row vector with all components equal to 0 except the kth one equal to 1. Furthermore, we set $q(\mathbf{n}, \mathbf{n}) = -\sum_{\mathbf{m}\neq\mathbf{n}} q(\mathbf{n}, \mathbf{m})$.

The number of customers in the system is less than the number of customers in an $M/M/\infty$ queue with arrival λ and service rate μ_0. This implies that the system under consideration is stable even if $\rho > 1$. There exists a unique invariant probability distribution given by the row vector $(\pi(\mathbf{n}), \mathbf{n} \in \mathrm{N}^K)$ which satisfies

$$\pi Q = 0 \quad \text{and} \quad \sum_{\mathbf{n}\in\mathrm{N}^K}\pi(\mathbf{n}) = 1. \tag{1}$$

C. Probability of reneging

Let $P_k(\mathbf{n})$ be the probability that a customer of class k reneges while there are n_ℓ customers of class $\ell = 1, \ldots, K$ in the system upon its arrival so that $\mathbf{n} = (n_1, \ldots, n_K)$. By using the memoryless property of the exponential distribution, we can easily prove the following result.

Lemma 1: The probabilities $P_k(\mathbf{n})$ for $\mathbf{n} \in \mathrm{N}^K$ satisfy the recurrence relations

$$P_k(\mathbf{n}) = \frac{\mu_0}{\Lambda_k(\mathbf{n})} + \sum_{\ell=1}^{K}\frac{\lambda p_\ell}{\Lambda_k(\mathbf{n})}P_k(\mathbf{n}+\mathbf{e}_\ell) + \sum_{\ell=1}^{K}\frac{1}{\Lambda_k(\mathbf{n})}\left(\frac{n_\ell\mu_\ell}{|\mathbf{n}|+1} + n_\ell\mu_0\right)P_k(\mathbf{n}-\mathbf{e}_\ell), \tag{2}$$

where

$$\Lambda_k(\mathbf{n}) = \lambda + \sum_{\ell=1}^{K}\frac{(n_\ell + \delta_{k,\ell})\mu_\ell}{|\mathbf{n}|+1} + (|\mathbf{n}|+1)\,\mu_0$$

with $\delta_{k,\ell}$ denoting the Kronecker symbol.

The recurrence relation (2) can be rewritten in matrix form as

$$(\mathrm{I} - M_k)\mathbf{P}_k = \mathbf{u}_k, \tag{3}$$

where I is the identity matrix, $\mathbf{P}_k$ (resp. $\mathbf{u}_k$) is the column vector with components $P_k(\mathbf{n})$ (resp. $\mu_0/\Lambda_k(\mathbf{n})$), $\mathbf{n} \in \mathrm{N}^K$, and the matrix M_k is given by

$$M_k = \begin{pmatrix} 0 & A_{k,0} & & & \\ B_{k,1} & 0 & A_{k,1} & & \\ & B_{k,2} & 0 & A_{k,2} & \\ & & \cdot & \cdot & \cdot \end{pmatrix} \tag{4}$$

with matrices $A_{k,n}$ and $B_{k,n}$ being defined as follows:

- The non-null entries of matrix $A_{k,n}$ are defined for $n \geq 0$ and $\mathbf{n} = (n_1, \ldots, n_K) \in \mathcal{S}_n$ by for $j = 1, \ldots, K$
$$a_{k,n}(\mathbf{n}, \mathbf{n}+\mathbf{e}_j) = \frac{\lambda p_j}{\Lambda_k(\mathbf{n})};$$
- The non-null entries of matrix $B_{k,n}$ are defined for $n \geq 0$ and $\mathbf{n} = (n_1, \ldots, n_K) \in \mathcal{S}_n$ by for $j = 1, \ldots, K$
$$b_{k,n}(\mathbf{n}, \mathbf{n}-\mathbf{e}_j) = \frac{1}{\Lambda_k(\mathbf{n})}\left(\frac{n_j\mu_j}{|\mathbf{n}|+1} + n_j\mu_0\right).$$

By using the fact that $0 \leq P_k(\mathbf{n}) \leq 1$ for all $\mathbf{n} \in \mathrm{N}^K$ (see [5] for details).

Proposition 1: The vector $\mathbf{P}_k$ is given by

$$\mathbf{P}_k = \sum_{r=0}^{\infty} (M_k)^r \mathbf{u}_k, \tag{5}$$

where $\mathbf{u}_k$ is introduced in Equation (3).

The reneging probability $\mathcal{P}_k$ that a class k customer reneges is eventually given by

$$\mathcal{P}_k = \pi.\mathbf{P}_k, \tag{6}$$

where π is the row vector satisfying Equation (1).

Since Equation (3) may be very difficult to solve since we have to deal with the infinite matrix M_k, we develop in the next section an approximating model when the load $\rho > 1$. With regard to impatience, this case is the most interesting as many customers may renege.

III. Approximating model

We have seen in the previous section that the computation of the reneging probability implies intricate matrix computations. In this section, we consider the case when the load ρ of the cell is greater than one ($\rho > 1$).

Define $C_k(\mathbf{n}(t)) \stackrel{\text{def.}}{=} n_k(t)/(n_1(t) + \cdots + n_K(t))$, where $\mathbf{n}(t) = (n_1(t), \ldots, n_K(t))$ is the number of customers in the various rings of the cell at time t. Additionally, let us recall that due to impatience, a class k customer can leave the system at rate $\mu_0 > 0$.

For $\xi > 0$, let $\mathcal{N}_\xi(\mathrm{d}t)$ denote a Poisson process on R_+ with rate ξ and $(\mathcal{N}_{\xi,i}(\mathrm{d}t))$ is an i.i.d. sequence of such processes. All Poisson processes are assumed to be independent.

Provided that $\mathbf{n}(t)$ is not 0, the process $(\mathbf{n}(t))$ can then be expressed as the solution of the following SDE (Stochastic Differential Equation)

$$\mathrm{d}n_k(t) = \mathcal{N}_{\lambda_k}(\mathrm{d}t) - \mathcal{N}_{\mu_k C_k(\mathbf{n}(t-))}(\mathrm{d}t) - \sum_{i=1}^{n_k(t-)} \mathcal{N}_{\mu_0,i}(\mathrm{d}t) \tag{7}$$

with initial condition $(n_k(0))$ for $k = 1, \ldots, K$.

A. Scaled Version

In order to qualitatively analyze the system, we consider a scaling similar to the one used in Gromoll *et al.* [6] for a processor-sharing queue with impatience with a single class of jobs but with general assumptions on the distribution of service duration and impatience. The average impatience of jobs is assumed to be of the order of a large factor N as follows: the parameter μ_0 is replaced by μ_0/N. The corresponding Markov process will be denoted as $(\mathbf{n}^N(t))$. The SDE (7) then reads

$$\mathrm{d}n_k^N(t) = \mathcal{N}_{\lambda_k}(\mathrm{d}t) - \mathcal{N}_{\mu_k C_k(\mathbf{n}^N(t-))}(\mathrm{d}t) - \sum_{i=1}^{n_k^N(t-)} \mathcal{N}_{\mu_0/N,i}(\mathrm{d}t).$$

Let $\bar{n}_k^N(t) = n_k^N(Nt)/N$, the corresponding fluid scaling of the process. By integrating the above SDE, $(\bar{n}_k^N(t))$ can be expressed as, for $1 \le k \le K$,

$$\bar{n}_k^N(t) = \bar{n}_k^N(0) + \lambda_k t - \int_0^t \frac{\mu_k \bar{n}_k^N(u)}{\bar{n}_1^N(u) + \cdots + \bar{n}_K^N(u)}\,\mathrm{d}u - \mu_0 \int_0^t \bar{n}_k^N(u)\,\mathrm{d}u + M_k^N(t), \tag{8}$$

for $t < T_0^N \stackrel{\text{def.}}{=} \inf\{s \ge 0 : \bar{\mathbf{n}}^N(s) = 0\}$, where $\mathbf{M}^N(t) = (M_k^N(t))$ is a martingale whose predictable increasing process is given by

$$\left\langle M_k^N \right\rangle(t) = \frac{1}{N}\left(\lambda_k t + \int_0^t \frac{\mu_k \bar{n}_k^N(u)}{\bar{n}_1^N(u) + \cdots + \bar{n}_K^N(u)}\,\mathrm{d}u + \mu_0 \int_0^t \bar{n}_k^N(u)\,\mathrm{d}u\right).$$

B. Convergence results

Assume that the initial conditions are such that $(\bar{\mathbf{n}}^N(0))$ converges to a non-zero vector $\ell(0) = (\ell_k)$. By using classical methods (see [5] for details), we can show the martingale $(\mathbf{M}^N(t))$ converges in distribution to 0 for the uniform convergence on compact sets when N tends to infinity. As a consequence one gets that the system does not empty when $\rho > 1$.

Lemma 2: Under the condition $\rho > 1$ and if $\mathbf{n}^N(0)/N$ converges to a non-zero limit, the hitting time of 0 by process $(\bar{\mathbf{n}}^N(t))$ converges in distribution to infinity, i.e. for any $t > 0$, $\lim_{N\to+\infty} P\left(T_0^N < t\right) = 0$.

Proof: By introducing the process $(\widetilde{n}^N(t))$ such that $\widetilde{n}^N(t) = e^{\mu_0 t}\sum_{k=1}^K \frac{\bar{n}_k^N(t)}{\mu_k}$, we have

$$\mathrm{d}\widetilde{n}^N(t) = e^{\mu_0 t}(\rho - 1)\,\mathrm{d}t + e^{\mu_0 t}\sum_{k=1}^K \frac{\mathrm{d}M_k^N(t)}{\mu_k}.$$

By using the convergence of the martingale term to 0 and the above SDE, for $\varepsilon > 0$ sufficiently small, one gets

$$\lim_{N\to+\infty} \mathrm{P}\left(\inf_{0\le s\le t} \widetilde{n}^N(s) > \varepsilon\right) = \mathrm{P}\left(\inf_{0\le s\le t} \mathbf{n}(s) > \varepsilon\right) = 1.$$

the result easily follows. ■

We can now state the main result of this section (the proof is omitted and is given in [5]).

Theorem 1 (Fluid Limits): If $\mathbf{n}^N(0)/N$ converges to a non-zero limit $(\ell_k(0))$, then the process $(\mathbf{n}^N(Nt)/N)$ converges in distribution to $(\ell_k(t))$, the solution of the differential equation

$$\dot{\ell}_k(t) = \lambda_k - \frac{\mu_k \ell_k(t)}{\ell_1(t) + \cdots + \ell_K(t)} - \mu_0 \ell_k(t) \tag{9}$$

with the prescribed initial condition $(\ell_k(0))$.

As an easy consequence of the above result, we have the convergence of distributions.

Corollary 1: If $\rho > 1$ and $(n_k^N(\infty))$ denotes a random variable with the same distribution as the invariant probability of $(n_k^N(t))$ then, for the convergence in distribution $\left(\frac{n_k^N(\infty)}{N}\right) \to (\ell_k)$ when $N \uparrow \infty$, with, for $1 \le k \le K$,

$$\ell_k = \frac{\frac{\lambda_k}{\mu_0} S}{\mu_k + S}, \tag{10}$$

where S is the unique non-negative solution to the equation

$$\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k + S} = 1. \quad (11)$$

The reneging probability for class k customers is then approximated by the quantity

$$\widetilde{P}_k = \frac{S}{\mu_k + S}. \quad (12)$$

The global reneging probability is thus computed by

$$\widetilde{P} = \sum_{k=1}^{K} \frac{\lambda_k}{\Lambda} \widetilde{P}_k = \frac{S}{\Lambda} \sum_{k=1}^{K} \frac{\lambda_k}{\mu_k + S} = \frac{S}{\Lambda} \quad (13)$$

where $\Lambda = \sum_{k=1}^{K} \lambda_k$ and we have used Equation (11). It is worth noting that the reneging probabilities do not depend on μ_0 but only on the quantity S, which itself does not depend on μ_0 (see Equation (11)). We can further prove the following result for second order asymptotics [5].

Theorem 2: Provided that $\left((n_k^N(0) - N\ell_k(0))/\sqrt{N}\right)$ converges to (δ_k), then the sequence of processes $\left((n_k^N(Nt) - N\ell_k(t))/\sqrt{N}\right)$ converges in distribution to $(Z_k(t))$, the solution of the following SDE

$$\mathrm{d}Z_k(t) = \sqrt{2\lambda_k - \dot{\ell}_k(t)}\, \mathrm{d}B_k(t)$$
$$- Z_k(t)\left(\mu_0 + \frac{\mu_k}{\sum_{i=1}^{K} \ell_i(t)}\right) \mathrm{d}t + \frac{\mu_k \ell_k(t)}{\sum_{i=1}^{K} \ell_i(t)} \sum_{i=1}^{K} Z_i(t)\, \mathrm{d}t,$$

such that $(Z_k(0)) = (\delta_k)$, where $(B_k(t))$ is a standard K-dimensional Brownian motion and $(\ell_k(t))$ is the solution of the ODE (9) with initial condition $(\ell_k(0))$.

C. Numerical example

For illustrating the previous results, we consider the case when $K = 2$. The elements of the matrix Q and the states of the Markov chain $(n_1(t), n_2(t))$ are indexed by the natural integers. For $i \in \{n(n+1)/2, \ldots, (n+1)(n+2)/2 - 1\}$ with $n \geq 0$, state i means that there are i customers of class i and $n - i$ customers of class 2.

To compute the stationary probability π satisfying Equation (1), we truncate the matrix Q to rank $(M+1)(M+2)/2 - 1$ for some integer $M > 0$ such that $\sum_{i=0}^{(M+1)(M+2)/2-1} \pi_i \geq 1 - \varepsilon$ for some $\varepsilon \ll 1$, where π_i is the stationary probability of being in state i. This leads us to consider the matrix Q^M and then the matrix $\widetilde{Q}^M$ obtained from matrix Q^M by deleting the first line and the first row. We solve the equation $\pi^M \widetilde{Q}^M = -\pi_0^M(\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2)$. The probability π_0^M is then determined by the normalizing condition

$$\pi_0^M = 1/(1 - (\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2)(\widetilde{Q}^M)^{-1}) \cdot \mathbf{e}),$$

where $\mathbf{e}$ is the column vector with all entries equal to 1.

To illustrate Corollary 1 and Theorem 2, we have considered the case when $\lambda_1 = \lambda_2 = .6$ and $\mu_1 = 1.0$, $\mu_2 = .5$ (meaning that cell center users (class 1) have twice more physical capacity than cell edge users (class 2)), $\mu_0 = .1$ and $N = 10$. The stationary density functions of the number of customers in each class, numerically obtained for $M = 100$, is depicted in Figure 2 and exhibit a Gaussian shape as stated in Theorem 2. The mean values are equal to 20.04 for class 1 and 28.9 for class 2. The respective values from Corollary 1 (equal to $N\ell_k = \frac{N\frac{\lambda_k}{\mu_0}S}{\mu_k+S}$ where S is the unique positive solution to Equation (11)) are 20 and 30. This and simulations show good agreement with the values obtained by solving Equation (1).

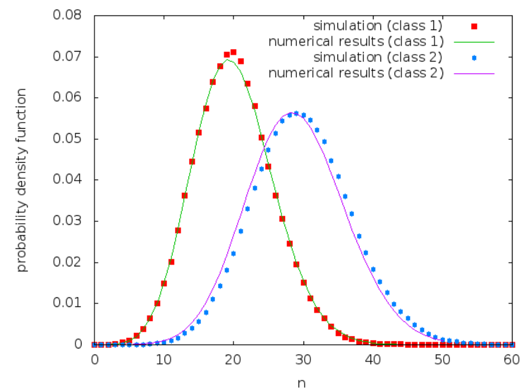

Fig. 2. Stationary distributions of the number of customers in each class for $\lambda_1 = \lambda_2 = .6$ and $\mu_1 = 1.$, $\mu_2 = .5$, $\mu_0 = .1$ and $N = 10$ - class 1 with squares and class 2 in line with dots.

Now by using Equation (12), we can approximate the reneging probability for each class. Equation (11) is a quadratic equation with a unique positive solution given by

$$S = \frac{-s + \sqrt{s^2 + 4\mu_1\mu_2(\rho - 1)}}{2}$$

with $s = \mu_1 + \mu_2 - \lambda_1 - \lambda_2$. In Figure 3, we have plotted the reneging probability for class 1 and 2 customers as a function of the arrival rate λ_1 for a fixed load $\rho = 1.5$ (with $\mu_1 = 1.$ and $\mu_2 = .5$). The global reneging probability $\widetilde{\mathcal{P}} = \frac{S}{\lambda_1+\lambda_2}$ (Equation (13)) is essentially due to class 1 when λ_1 increases. The reneging probability of class 2 customers is very high but their contribution is small.

We now move to a more realistic numerical example with a large number of radio conditions issued from Figure 1. We plot in Figure 4 the reneging probability of the different classes of users and the average reneging probability when the cell load increases. It can be observed that there is a large difference between cell center and edge users.

IV. QoE-perturbation based pricing scheme

We develop in this section a practical application of the model presented in this paper that consists of introducing a QoE perturbation metric and its usage in the design of a pricing scheme.

A. Classical pricing schemes and their shortcomings

Pricing of data communications in cellular networks is usually based on capped offers. The user can download up

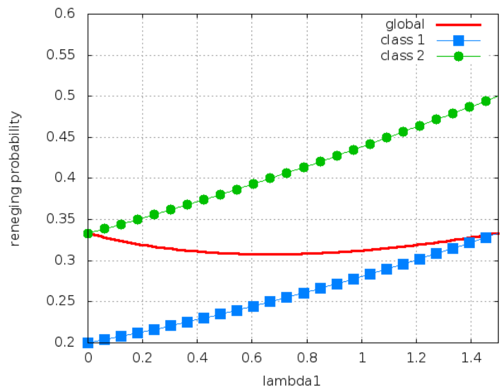

Fig. 3. Reneging probabilities as a function of λ_1 for $\rho = 1.5$ - line with squares (resp. dots) for class 1 (resp. class 2) and solid line for the global reneging probability.

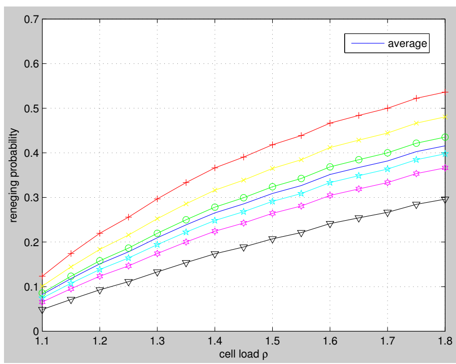

Fig. 4. Reneging probabilities of the different radio classes function of the cell load. The upper curve corresponds to cell edge users (peak throughput of 2 Mbps) and the lower one corresponds to cell center users (peak throughput of 7 Mbps). The dashed curve corresponds to the average reneging probability over the cell.

to a certain threshold and is then throttled independently of the state of the network when the volume of transmitted data exceeds the prescribed threshold. The price that the user has to pay precisely depends on this threshold.

With the rapid development of mobile Internet and the deployment of 4G technology, capped offers will however not be well suited to the new usage of wireless communications, which offer a performance comparable to that in fixed networks. This is all the more true as users do not care so much of the way they are accessing the Internet, via cellular or WiFi/fixed access. The major drawback of capped offers is that they do not account of network state and they require the deployment of a monitoring infrastructure, which can be very costly and leads to the concentration of traffic in gateways, namely Packet Data Network (PDN) gateways in the 4G architecture.

In this paper, we investigate new ways of pricing data communications by going beyond capped offers. We instead exploit the fact that the capacity of a cell is shared between all active users thanks to the scheduling algorithm implemented in the base station. A first idea in this context would be to charge the user according to the average bit rate he obtains for his communications (bit rate pricing). This is however unfair in several respects. For instance, the achieved bit rate depends on the location of the user in the cell, which is a parameter not completely under control by the user. Moreover, the higher the bit rate for a communication, the shorter is the communication and hence the smaller is the impact of this user on other users.

An alternative approach would be to introduce a social welfare function for the various customers present in a cell and use Vickrey-Clarke-Groves pricing model [7], [8]. As shown in [9], when customers share a common transmission capacity c under the fair share policy (i.e., the welfare function for a customer is proportional to $\log(x)$ where x is the achieved bit rate) then the price to pay is an increasing and concave function of the achieved rate x. This scheme has hence the same shortcomings as bit rate pricing. Other pricing schemes based on rate allocation are also investigated in [9].

Congestion could also be used for pricing as in Conex scheme that supposes that traffic sources (using for instance TCP) adapt to congestion signals [10]. However, in today's networks, TCP is more and more bursty and may create backlog even if the network is not congested (in terms of average offered traffic). In addition to this, the bottleneck in wireless networks is often the radio access so that congestion based on buffer levels as in Conex is difficult to assess because of fluctuations of the available bandwidth. Moreover, customers in bad radio conditions may create backlog in the base station but this is not a sign of congestion; packets accumulate only because the bit rate achievable by the customer is small.

We propose in this section a alternative pricing scheme based on the notion of social cost of a user, or the QoE perturbation induced by the arrival of this user into the cell. We begin by describing the metric of QoE perturbation before proposing the pricing scheme.

B. *QoE perturbation metric*

QoE perturbation is related to the impact of the presence of a given flow on the QoE of other users. Recall that $P_k(\mathbf{n})$ is the probability of reneging for a class k customer when there are $\mathbf{n}$ customers in the system. The extra perturbation introduced by a class k customer when the system is in state $\mathbf{n}$ can be computed by:

$$\Gamma_k(\mathbf{n}) = \sum_{j=1}^{K} p_j (P_j(\mathbf{n} + \mathbf{e_k}) - P_j(\mathbf{n})). \tag{14}$$

This means that the QoE perturbation for class k users is the average additional impatience observed by users when there is an additional user of class k entering the system. Note that this QoE perturbation depends on the state of the system, leading to a pricing scheme that depends on the instantaneous state of the system. It is thus preferable to define an average QoE perturbation metric as follows:

$$\hat{\Gamma}_k(\lambda_1, ..., \lambda_K) = \sum_{\mathbf{n}} \Gamma_k(\mathbf{n}) \pi(\mathbf{n}). \tag{15}$$

Note that this metric may be complex to compute, as it involves the computation of the reneging probabilities. We then make use of the approximation of the previous section and define the QoE perturbation function for class k users as follows:

$$\tilde{\Gamma}_k(\lambda_1, ..., \lambda_K) = \frac{\partial \tilde{R}}{\partial \lambda_k}, \quad (16)$$

where the global impatience rate is defined as follows:

$$\tilde{R}(\lambda_1, ..., \lambda_K) = \sum_{j=1}^{K} \lambda_j \tilde{\mathcal{P}}_j(\lambda_1, ..., \lambda_K) = S(\lambda_1, ..., \lambda_K).$$

In the general case of K radio conditions, it is sufficient to derive the fixed point equation (11) in order to obtain the QoE perturbation metric:

$$\tilde{\Gamma}_k(\lambda_1, \ldots, \lambda_K) = \frac{\partial S}{\partial \lambda_k} = \left((\mu_k + S) \sum_{j=1}^{K} \frac{\lambda_j}{(\mu_j + S)^2} \right)^{-1} \quad (17)$$

Figure 5 illustrates the QoE perturbation caused by the different classes of radio conditions (taken from Figure 1). The following properties can be observed:

- QoE perturbation is larger for cell edge users than for cell center users. Cell edge users contribute more to reneging than cell center users.
- QoE perturbation is almost a flat function of the cell load, even if the impact of cell edge users reduces at large loads (when the impact of radio conditions reduces compared to the impact of increasing number of active users). This is an important property as the operator can set a fixed approximate QoE perturbation metric function of the radio condition and independent of the traffic load.
- QoE perturbation does not depend on the impatience rate (this is not observed from the figure but directly from the mathematical formula (17) as S is independent of μ_0). This is an important property as the operator does not have to estimate the impatience rate, difficult to be assessed from field measurements, but only the distribution of radio conditions.

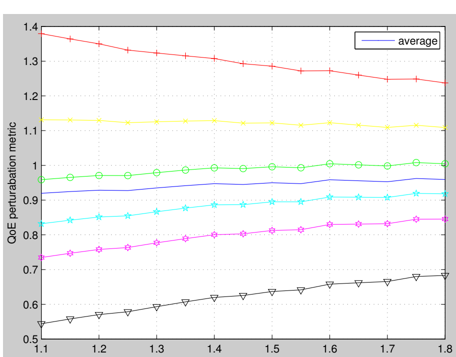

Fig. 5. QoE perturbation metric for users belonging to different classes.

C. Proposed pricing scheme

Based on this QoE perturbation metric and for a convergence of fixed and cellular networks in terms of pricing, we propose a pricing scheme comprising two components:

- A flat rate to enable a customer to attach to the network (authentication, roaming, etc.);
- An elastic component reflecting the "social" cost of a user, i.e. depending on the QoE perturbation metric defined above.

Different possibilities exist for the elastic component. A first possibility is to define a per-Mbyte price for the different radio conditions, proportional to the QoE perturbation metric. A second possibility, more inline with the current capped offers and the Conex strategy, would be to define a "congestion right limit" that is decremented proportionally to the QoE perturbation associated with the radio condition. A pricing notification mechanism could also be implemented, where users receive, depending on their position and the network state, their price/congestion information.

V. Conclusion

By using a fluid limit approximation, we have studied in this paper the reneging probability of customers sharing the radio resources of a cell of a mobile network. We showed that, under heavy load, reneging probability can easily be derived using a simple fixed point equation. We then introduced a QoE perturbation metric that corresponds to the impact a particular communication has on the QoE of other users. We then devised new pricing policies that comprise a fixed component (flat rate as in fixed networks) and an elastic one depending on the QoE degradation metric.

References

[1] F. Baccelli and G. Hebuterne, "On queues with impatient customers," *Performance*, vol. 81, pp. 159–179, 1981.

[2] S.-C. Yang and G. de Veciana, "Bandwidth sharing: The role of user impatience," in *IEEE Globecom*, 2001.

[3] J. R. T. Bonald, "Congestion at flow level and the impact of user behaviour," *Computer networks*, vol. 42, pp. 521–536, 2003.

[4] K. W. Ross, *Multiservice loss models for broadband telecommunication networks*. Springer, 1995.

[5] F. Guillemin, S. Elayoubi, P. Robert, C. Fricker, and B. Sericola, "Controlling impatience in cellular networks using QoE-aware radio resource allocation," submitted for publication.

[6] C. Gromoll, P. Robert, and B. Zwart, "Fluid limits for processor sharing queues with impatience," *Mathematics of Operations Research*, vol. 33, no. 2, pp. 375–402, May 2008. [Online]. Available: http://arxiv.org/abs/math/0604215

[7] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of Finance*, vol. 16, pp. 8–37, 1961.

[8] S. Yang and B. Hajek, "VCG-Kelly mechanisms for allocation divisible goods: Adapting VCG mechanisms to one-dimensional signals," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 6, pp. 1237–1243, 2007.

[9] S. Birmiwal, R. R. Mazumdar, and S. Sundaram, "Processor sharing and pricing implications," in *Proceedings of the ITC 24, Krakow, Poland*, 2012.

[10] http://datatracker.irtf.org/wg/conex, *Congestion exposure (conex)*, 2014. [Online]. Available: http://books.google.fr/books?id=g9xzlgEACAAJ