Revisiting the characterization and the modeling of user impatience in LTE networks

Bruno Baynat
Université Pierre et Marie Curie - LIP6
bruno.baynat@lip6.fr

Marion Vasseur
Université Pierre et Marie Curie - LIP6
marion.vasseur3@gmail.com

Thiago Abreu
Université Pierre et Marie Curie - LIP6
thiago.wanderley@lip6.fr

***Abstract*—In this paper we revisit the definition of user impatience in cellular networks, by proposing a new characterization that does not reduce impatience to a simple timeout, like most previous works do. Instead, we characterize impatience by decoupling the cause of it, i.e., the probability that a given user becomes impatient, and the consequence, i.e., the reduction of its download by a premature departure. From this new characterization, we develop two models based on M/G/1/PS queues that differ in the way impatient users decide to stop their download. We show by comparison with NS3 simulations, that these models have a high accuracy, and use them to investigate the influence of impatience on the performance of the cell and on the quality of experience of users. One of our main findings is that impatience does not significantly degrade the global satisfaction of users with regards to a system without impatience, and greatly extends the range of utilization of the cell where customers have an acceptable satisfaction.**

I. Introduction

Nowadays, communication networks based on 3G or 4G LTE (Long-Term Evolution) technologies must deal with an increasing number of mobile devices (smartphones and tablets) generating traffic. This constant increase together with the fact that users are more and more accustomed to using high speed connections, leads to the appearance of impatient users as soon as the network is subjected to occasional congestions. Impatience corresponds to the premature departure of a user, deciding to stop its download before it is completed, due to its own evaluation of the current network performance.

Impatience has first been studied in the context of streaming trafic in fixed networks. [1] developed Erlang-like formulas that take into account impatience as a maximum waiting time of users before receiving service. [4] studied a system where impatience also concerns customers being served, and is then defined as a maximum sojourn time of users before completing their service. In both cases, impatience is thus expressed as a random variable defining the maximum time after which impatience occurs and customers leave prematurely the system.

In the context of elastic trafic in cellular networks, impatience of users has a greater impact on performance. As a matter of fact, for elastic applications, the cell shares the total offered throughput among all active users, e.g., in a round-robin fashion. If the cell is temporarily congested, users receive a degraded throughput that extends their session duration. Users may react to this degradation by aborting their connection. This has the first consequence of decreasing the satisfaction of impatient users. But, on the other hand, impatient users alleviate the cell load and, as a result, leave more resources (time slots in HSDPA or resource blocks in LTE) to patient users, increasing their satisfaction. It is thus of high interest to correctly model impatience of users and analyze its impact on the performance of the cell and on the quality of experience of users.

Many works have analyzed wireless networks in which customers can be impatient (e.g., [1]–[7]). [2] provides a formalization of the problem of broadcast scheduling with impatient users in an overloaded wireless network. Authors assume that the duration of a flow cannot exceed a given patience duration. Different characterizations of the patience duration are studied. [5] also consider the broadcast scheduling problem taking user impatience into account, but simplify the analysis by considering an exponentially distributed patience duration. Authors of [7] propose a characterization of impatience based on the quantity of work already received by a user, and use simulations to quantify its impact on the system performance for several bandwidth sharing disciplines. In [6], authors develop an M/G/1/PS queue model to analyze the impact of user impatience on the performance of LTE networks. Impatience is still modeled as a maximum time after which users leave prematurely the system, but authors consider a more realistic case where this time is related to the flow size.

Most previous works have thus characterized user impatience by limiting the time a given user stays in the cell to some maximum value, making of impatience a timeout limit. This way of characterizing impatience has, according to us, several drawbacks. First, considering a fixed limiting sojourn time (as most previous works do), that does not depend on the actual size of the downloaded element, does not seem realistic. As a mater of fact, a user will more likely wait if he has a large file to download. And if some previous works have tried to relate the patience duration to the file size, the relation they propose does not seem very realistic. Second, this simple characterization does not take into account the fact that, whatever the conditions offered by cell are, some users are more likely to be impatient than others, and the proportion of impatient users cannot simply be reduced to a timeout.

In this paper we propose a new characterization of impatience that decouples the cause of it, i.e., the probability that a given user becomes impatient, and the consequence, i.e., the reduction of its download by a premature departure. Naturally, the impatience probability is related to the throughput a given user receives during its transfer. Indeed, the higher the load of the cell, the lower the throughput it receives, the more likely the user will be impatient. Conditioned by the fact that the user is impatient, we propose two different models that characterize the way an impatient user decides to stop his download. These two models are referred as to Model 1 and Model 2. The idea behind these models is to use the classical M/G/1/PS queue model of a cell without impatience, with an augmented service rate taking into account the fact that some customers are impatient and leave prematurely the cell, thus reducing the average service time of the queue. Model 1 and Model 2 differ in the way the size of the transfer of an impatient user is reduced

* This work has been carried out in the framework of IDEFIX project, funded by the ANR under the contract number ANR-13-INFR-0006.

with regard to a patient user.

The paper is organized as follows. Section II describes the system we consider and presents the new characterization of impatience. In Section III, we develop the two models associated with this new characterization. Section IV validates the models by comparing them to NS3 simulation. Finally Section V concludes this work.

II. System

A. A cell with impatient users

We consider a single LTE cell offering an average shared throughput of C Mbit/s to users, and focus on the downlink part of the radio transmission. We assume that the cell equitably shares the offered throughput among all active users in a round-robin fashion, i.e., when there are currently n active users in the cell, each one receives an instantaneous downloading throughput of $\frac{C}{n}$ Mbit/s. The cell thus acts as a Processor Sharing system.

We consider that some users in the cell can be impatient. An impatient user decides to stop its current download before it is completed, e.g., considering that it is too slow, and then leaves the system prematurely. Conversely, a patient user waits till the end of its download, whatever the conditions he obtains, and then leaves the system. The cell is shared by patient users and impatient users. The proportion of patient and impatient users is clearly related to the load of the cell, i.e., to the throughput they obtain during their download.

B. Characterizing Impatience

As stated in the introduction, most previous works have chosen to characterize impatience by limiting the time a given user stays in the cell to some maximum value. If this maximum sojourn time is reached before its download is completed, the considered user is declared to be impatient and leaves prematurely the system. This way of characterizing impatience takes into account neither the relation between the likelihood of impatience and the size of the download the user has to make, nor the diversity of users. Instead, we choose to decouple the characterization of impatience of users by first characterizing the probability of impatience, and then by defining how an impatient user behaves. Obviously, the lower the load of the system, the higher the throughput obtained by each user, the lower the probability for a given user to be impatient. We then propose to relate the probability that a given user is impatient, denoted as p, to the actual average throughput $\bar{\gamma}$ each user obtains, as follows:

$$p = \frac{1}{1+\alpha\bar{\gamma}} \tag{1}$$

where α is a coefficient that has to be estimated, e.g., from statistics (if we can estimate that half users are impatient for a given throughput $\bar{\gamma}_0$, then $\alpha = \frac{1}{\bar{\gamma}_0}$). With this characterization, when a new user arrives into the cell and obtains an average downloading throughput of $\bar{\gamma}$ Mbit/s, he has a probability p to be impatient and a probability $1-p$ to be patient. In the last case, as stated before, he will stay in the cell the time necessary to complete its download. We denote by σ the average size of an element a patient user has to download.

At this point, it is important to emphasize that any expression of p that would seem more appropriate can alternately be used in the modeling framework developed in Section III.

Conditioned by the fact that a user is impatient, we need to characterize when the user decides to stop its download. We have chosen to simply reduce the size of downloaded elements for impatient users, and propose two ways to do it. First, we assume that an impatient user decides to stop its download uniformly over all the downloading duration. This results in considering that the average size σ_I of an incomplete element an impatient user will download before leaving prematurely the system is just:

$$\sigma_I = \frac{\sigma}{2} \tag{2}$$

In the following, we will refer to this first characterization of impatience as "Model 1".

In what will be referred to as "Model 2", we consider that the size of an incomplete element an impatient user will download before leaving prematurely the system is itself a function of the actual average throughput it obtains during its download. This way, the average size of an incomplete downloaded element is related to $\bar{\gamma}$ as follows:

$$\sigma_I = \frac{\sigma}{1+\frac{1}{\beta\bar{\gamma}}} \tag{3}$$

where β is also a coefficient that has to be estimated from some statistics. With this second characterization, the higher the throughput obtained by the impatient user, the closer from the end the download is stopped. Conversely, when the average throughput obtained by the impatient user tends to zero, the actual size of the incomplete downloaded element also tends to zero. This seems to be a more realistic characterization of impatience.

Again, let us emphasize that one will be able to use any relevant alternative to these two specific characterizations, in the general modeling framework detailed below.

It may finally be noted that the probability p that a user is impatient is related to the average throughput $\bar{\gamma}$ offered to users (thanks to relation 1), which in turn depends on the proportion p of impatient users in the cell. In addition, in Model 2, the size σ_I of an incomplete download also depends on $\bar{\gamma}$, which in turns depends on σ_I. As a result the overall model will rely on the resolution of a fixed-point problem.

III. Models

A. PS queue model of a cell without impatience

The classical model associated with a single LTE cell without impatience is an M/M/1/PS queue. This model assumes that users globally generate requests for transmission in the cell according to a Poisson process with rate λ. All requests come with an identically distributed volume Σ of data to be transferred with an average size $\sigma = \mathbb{E}(\Sigma)$. The service rate of the queue is thus:

$$\mu = \frac{C}{\sigma} \tag{4}$$

It is worthwhile remembering that all average performance parameters of the processor sharing queue are insensitive to the actual distribution of the service time and thus to the actual distribution of Σ. The stability condition of the queue is: $\lambda < \mu$. If this condition is satisfied, the following average performance parameters can be obtained from the classical results of the M/M/1/PS queue. The average time $\bar{R}$ spent by a user in the system is:

$$\bar{R} = \frac{1}{\mu - \lambda} \tag{5}$$

And the average throughput $\bar{\gamma}$ obtained by a user for its download is:

$$\bar{\gamma} = \frac{\sigma}{\bar{R}} = C(1-\rho) \tag{6}$$

where

$$\rho = \frac{\lambda}{\mu} \tag{7}$$

B. Classical model where impatience is characterized by a time

As said before, most previous works have characterized impatience by a maximum sojourn time of active users in the cell. In the case the time Θ after which a user becomes impatient is exponentially distributed with rate θ, the resulting model becomes a Markovian birth-and-death process, where the rate from any state $n \geq 0$ to state $n+1$ is λ and corresponds to the arrival of a new request for transmission, and the rate from any state $n > 0$ to state $n-1$ is $\mu + n\theta$ and corresponds either to the end transmission of one of the n ongoing transfers, or to the impatience of one of the n current active users.

Because our new characterization of impatience is fundamentally different from the one used in this classical Markovian model, the results of our models (developed in the remaining subsections) cannot be compared to the results of this classical model. This is the reason why we don't give here the expressions of the performance parameters of this model.

C. Model 1: a new definition of impatience

The idea behind the new models we propose is to use the M/M/1/PS model of a cell without impatience, with an augmented service rate, denoted as μ_e, taking into account the fact that some customers are impatient and leave prematurely the cell (thus reducing the average service time of the queue). The service-time distribution of the new queueing model is described as an hyper-exponential distribution with 2 phases, one corresponding to patient users and the other corresponding to impatient users, as illustrated in Figure 1. With a probability $1-p$ a user is patient and has to complete a full download, corresponding to a rate $\mu = \frac{C}{\sigma}$, and with a probability p a user is impatient and has to complete a partial download, corresponding to a rate $\mu_I = \frac{C}{\sigma_I}$. Remember that the M/G/1/PS is insensitive to the service-time distribution of customers. As a result, our model is equivalent to an M/M/1/PS queue, thus having a global exponential service-time with a rate μ_e given by:

$$\frac{1}{\mu_e} = \frac{1-p}{\mu} + \frac{p}{\mu_I} \tag{8}$$

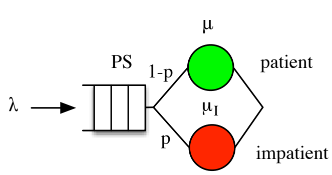


Fig. 1: PS queue model.

In Model 1, σ_I is simply given by relation 2, thus $\mu_I = 2\mu$ and $\mu_e = \frac{2\mu}{2-p}$. The stability condition of the associated PS queue is postpone twice as far (as the one corresponding to the case with no impatience): $\lambda < 2\mu$. Indeed, when we reach the limits of stability, almost all users become impatient (as they obtain a throughput that tends to zero). As a result, $p \to 1$ and $\mu_e \to \mu_I = 2\mu$. Provided this condition is satisfied, we can solve the system of equations given by relations 1, 6 and 8, with $\mu_I = 2\mu$ and $\rho = \frac{\lambda}{\mu_e}$. This system corresponds to a second degree equation with a single positive root giving the solution for p:

$$p = \frac{-\alpha\mu C - \mu + \alpha C\lambda + \mu\sqrt{(\alpha C + 1 - \frac{\alpha\lambda C}{\mu})^2 + 2\frac{\alpha\lambda C}{\mu}}}{\alpha C\lambda} \tag{9}$$

From p, we can derive the average throughput $\bar{\gamma}$ obtained by users, and the average time $\bar{R}$ spent by users in the cell as:

$$\bar{\gamma} = C(1 - \frac{\lambda(2-p)}{2\mu}) \tag{10}$$

$$\bar{R} = \frac{2-p}{2\mu - \lambda(2-p)} \tag{11}$$

D. Model 2: a more realistic characterization

As aforementioned, it is more realistic to consider that the size of an incomplete transfer an impatient user downloads before leaving prematurely the system is itself a function of the actual average throughput it obtains during its download, and is given by relation 3. It is important to note that Model 2 corresponds to a system that is always stable (whatever the load λ). Indeed the higher λ, the lower the throughput $\bar{\gamma}$, and the higher the proportion of impatient users in the system (like in Model 1). As a result, $p \to 1$, $\sigma_I \to 0$ and $\mu_e \to \mu_I \to \infty$. The performance of the corresponding model is still the solution of equations 1, 6 and 8, but with $\mu_I = (1 + \frac{1}{\beta\bar{\gamma}})\mu$ (and $\rho = \frac{\lambda}{\mu_e}$), leading to the following system:

$$\begin{cases} p = \frac{1}{1+\alpha\bar{\gamma}} \\ \mu_e = \mu\frac{1+\frac{1}{\beta\bar{\gamma}}}{(1-p)(1+\frac{1}{\beta\bar{\gamma}})+p} \\ \bar{\gamma} = C(1 - \frac{\lambda}{\mu_e}) \end{cases} \tag{12}$$

This system will be solved thanks to a fixed-point iterative technique, that will eventually provide the performance parameters of interest, namely the proportion p of impatient users, the throughput $\bar{\gamma}$ obtained by active users, and the average time $\bar{R}$ spent by users in the cell.

E. Customer Satisfaction

In addition to p, $\bar{\gamma}$ and $\bar{R}$, we can calculate for both models (1 and 2), another performance index measuring customer satisfaction. Let S denote this satisfaction index, ranging from 0 to 1, 1 corresponding to a fully satisfied user. If S_P is the satisfaction of patient users and S_I is the satisfaction of impatient users, the unconditioned satisfaction index can be obtained as:

$$S = (1-p)S_P + pS_I \tag{13}$$

To be consistent with our models, we consider that the satisfaction of patient users is an increasing function of the throughput they obtain. Indeed the higher the throughput, the faster their download. We choose to define S_P as:

$$S_P = \frac{\delta\bar{\gamma}}{1+\delta\bar{\gamma}} \tag{14}$$

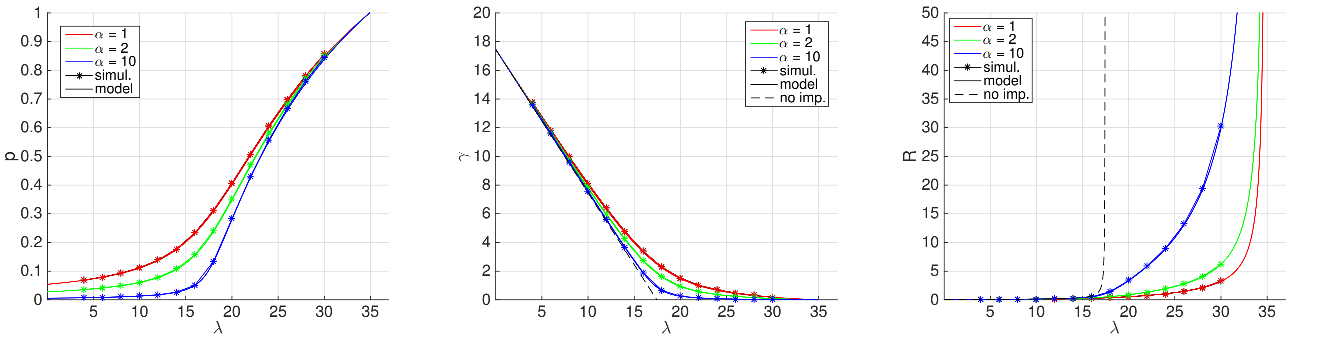
where δ is a tuning parameter. Any appropriate function starting from 0 when $\bar{\gamma} = 0$ and converging to 1 when $\bar{\gamma} \to \infty$ can alternately be used. We finally choose to consider that an impatient user has a nul satisfaction criterion:

$$S_I = 0 \tag{15}$$

IV. Numerical Results

A. Model 1

We first compare the performance of Model 1 with simulation results obtained with the NS3 simulator. The cell capacity C is equal to 17.45 Mbit/s and the average size of an element a patient user has to download is $\sigma = 1$ Mbit. We consider three different values of the parameter α used in relation 1 to calculate the probability of impatience: 1, 2 and 10. Table I gives for each α, the corresponding throughput $\bar{\gamma}_0$ leading to a 50% proportion of impatient users.



(a) Probability of impatience. (b) Average throughput obtained by users (Mbit/s). (c) Average time spent by users in the cell (sec.).

Fig. 2: Evolution of the performance parameters as a function of the load (Model 1).

α	1	2	10
$\bar{\gamma}_0$	1 Mbit/s	500 Kbit/s	100 Kbit/s

TABLE I: Throughput corresponding to 50% of impatient users.

It is important to understand how we simulate impatience of users. In order to reproduce our characterization of impatience, when a new user is generated in the simulation, we use the instantaneous throughput seen by this new user at the beginning of its transfer to calculate its probability p of impatience. In practice, if the new user starts its transfers in a cell with n users (including himself), the instantaneous throughput he obtains is $\frac{C}{n}$ and the simulation decides immediately if this user will be impatient or not according to a probability $p = 1/(1+\alpha\frac{C}{n})$. If the user is tagged as a patient user, the size of its download is drawn in an exponential distribution with mean σ, and if it is tagged as impatient, the size of its download is exponential with mean $\sigma_I = \frac{\sigma}{2}$. Note that, in the model, the probability of impatience is alternately derived from the average throughput a user obtains during its whole transfer (relation 1).

Figure 2 compares the performance parameters obtained from the model to those derived from simulation, when the load varies, for the three values of α (red, green and blue curves). It gives also the performance of a system with no impatience, i.e., where all users are assumed to be patient (dotted black curves). Figure 2a compares the probability p of impatience (relation 9), Figure 2b depicts the average throughput $\bar{\gamma}$ obtained by users (relation 10) and Figure 2c gives the average time $\bar{R}$ spent by users in the cell (relation 11). A first remark that can be globally made from the three curves of this figure concerns the stability condition. As stated in Section III-C, the system is stable if $\lambda < 2\mu = \frac{2C}{\sigma} = 34.9$. And we can see on the curves that when λ tends towards 34.9, $\bar{\gamma}$ logically tends to 0, p tends to 1, and $\bar{R}$ tends to infinity, for both model and simulation. Note that in the LTE standard (in thus in NS3), there is a physical limitation of 320 simultaneous ongoing transfers, that prevent us from testing by simulation loads close to saturation.

We can see on Figure 2a that, despite the fact that simulation calculates the probability of impatience using the instantaneous throughput seen by the arrival of a new user, whereas the model uses the average throughput, results of the model are very close to simulations. The relative error is, on average, 1.1%, with a maximum value less than 2%. The figure shows two regimes, the first corresponding to a system without impatience that is stable, i.e., $\lambda < \mu$, in which the probability of impatience increases slowly with the load, and the second corresponding to a system without impatience that is not stable, i.e., $\mu < \lambda < 2\mu$, in which the probability of impatience grows rapidly with the load.

As expected, it can be seen on Figure 2b that the average throughput obtained by users starts from the maximal value $C = 17.45$ Mbit/s at very low loads (where a user is very likely to be alone in the cell and obtains the total capacity of the cell), and decreases towards 0 when the load tends to the limit of stability. Again we can see two regimes. In the first one ($\lambda < \mu$) the evolution of the average throughput is close to the well-known linear decrease of a system without impatience (see Section III-A). The second one ($\mu < \lambda < 2\mu$) corresponds to a system without impatience that would not be stable, but the presence of impatient users stabilizes the system and offers to patient users enough throughput to finish their download. Again, we observe very small relative errors between the model and simulation (comparable to those made on parameter p).

Figure 2c depicts the average time spent by users in the cell, and clearly shows the difference between our system with impatience and the same system without impatience. The saturation limit is pushed up twice as far with the presence of impatient users. As a result, the cell benefits from impatience of users that increases its range of utilization.

Finally, Figure 3 represents the satisfaction criterion S_P of patient users as a function of the load. We used $\delta = 2$ in its expression (relation 14), corresponding to a satisfaction of 0.5 when patient users receive a throughput of 500 Kbit/s. Unsurprisingly, the criterion starts from a value close to 1 (corresponding to relation 14 with $\bar{\gamma} = C$) and decreases towards 0 when λ tends towards 2μ. It is interesting to note that, even in the region where the system without impatience is stable, i.e., $\lambda < \mu$, the satisfaction of patient users is always higher than the satisfaction of users in a system without impatience. Now, in the region where the system without impatience is not stable, i.e., $\mu < \lambda < 2\mu$, patient users can still have a pretty good satisfaction, benefiting from the early departure of impatient users. Now if we turn our attention to the global satisfaction S of users (relation 13), including the null satisfaction of impatient users, we can see on Figure 4 that impatience does not degrade significantly the satisfaction of users (patient and impatient) when the system without impatience is stable, and enables to largely extend the zone where users perceive an acceptable satisfaction. As an example, if we take $\alpha = 1$ and consider a lower limit of 0.2 for an acceptable satisfaction, we see from Figure 4 that the system with impatience can tolerate a maximum load of about 32 download requests by second if we just consider satisfaction of patient users (red curve of Figure 3), and of about 23 download requests by second if we consider the global satisfaction of all users (red curve of Figure 4), whereas the same system without impatience is limited to 17 requests by second (black dotted curve of both figures). This demonstrates that impatience of users is not only beneficial to the performance of the cell, but also to users themselves.

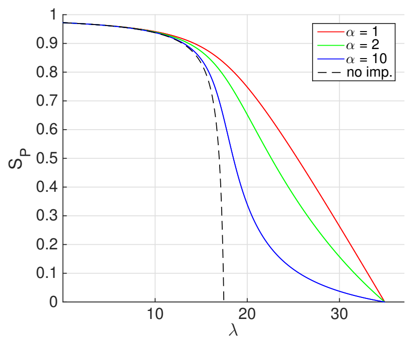


Fig. 3: Evolution of the satisfaction criterion for patient users as a function of the load (Model 1).

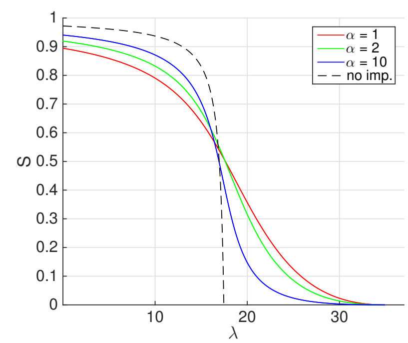


Fig. 4: Evolution of the global satisfaction criterion as a function of the load (Model 1).

B. Model 2

We solved iteratively the fixed-point equations of system 12 associated with Model 2, and compared the solution to NS3 simulations. In all cases, the iterative algorithm used to solve the model has converged very rapidly (in less than a few tens of iterations). As for the impatience probability, we use in simulations the instantaneous throughput obtained by an new arriving user, instead of the average throughput, in order to calculate from relation 3 the average size σ_I of the incomplete download of an impatient user. In all the results presented here, we chose the parameter β involved in this relation equal to α.

Figure 5 compares the performance parameters obtained from Model 2 (probability of impatience, average throughput obtained by users and average time spent by users in the cell) to simulations, for a varying load and three different values of $\alpha = \beta$ (1, 2 and 10). As can be seen on the figures, the accumulation points of the performance curves corresponding to Model 1 (Figure 2) have turned into asymptotes in Model 2. As a matter of fact, this system is stable for any value of the load (as explained in Section III-D), and when λ tends to infinity, the probability p obviously tends to 1 and the average throughput $\bar{\gamma}$ tends to 0. However, contrarily to Model 1, and as shown by Figure 5c, the average response time $\bar{R}$ in Model 2 converges towards a finite value. It can be easily proven from equations 12 that this asymptotic value is equal to $\sigma(\alpha + \beta)$.

Apart from that, the performance curves exhibit the same general behavior as those corresponding to Model 1, with two regimes, one corresponding to $\lambda < \mu$ (for which the system without impatience is stable), and the other corresponding to $\mu < \lambda$ (for which the system without impatience is unstable). The relative errors between the performance of Model 2 an simulation are similar to those of Model 1, and remain negligible.

The satisfaction criterion of patient users, S_P, and of all users, S, are depicted in Figures 6 and 7, respectively. Again, the accumulation points of the curves corresponding to Model 1 have turned into asymptotes in Model 2, where both criterions tend to 0 when λ tends to infinity. Conclusions made in Model 1 are even more true in Model 2: Impatience does not significantly degrade the global satisfaction criterion of users (with regards to a system without impatience), and greatly extend the range of the load where customers have an acceptable satisfaction.

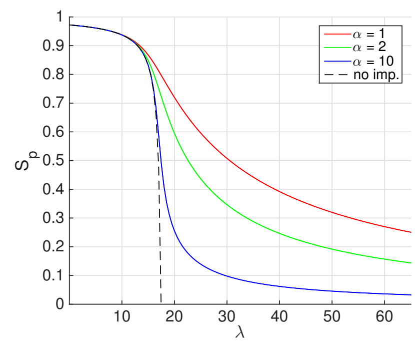


Fig. 6: Evolution of the satisfaction criterion for patient users as a function of the load (Model 2).

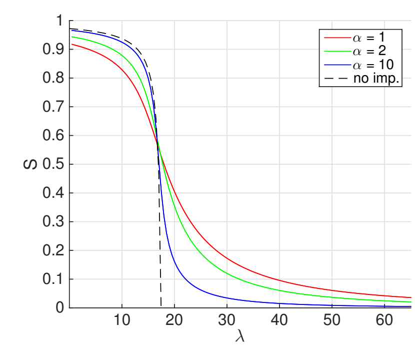
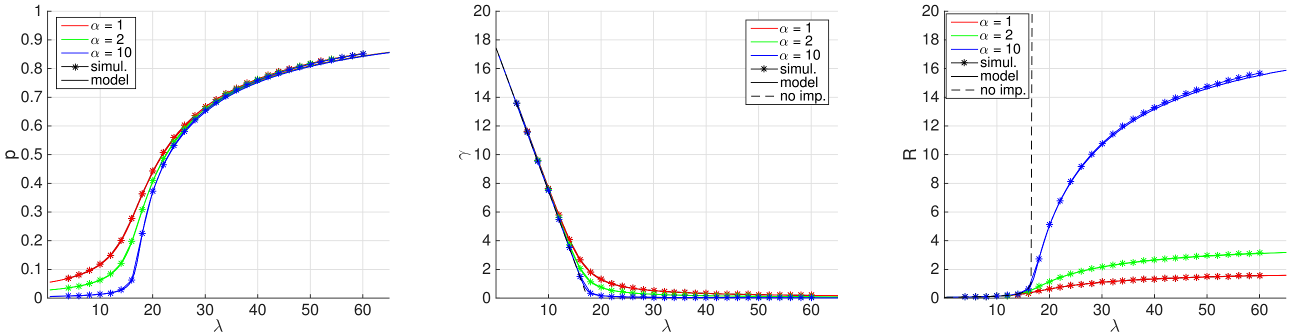


Fig. 7: Evolution of the global satisfaction criterion as a function of the load (Model 2).

V. Conclusion

We have proposed a new characterization of impatience of users in a LTE cell. This characterization differs from the classical one that assumes that impatience occurs when a predefined timeout expires, in defining the probability of impatience of a given user as a function of the throughput it receives during its transfer. This more realistic characterization takes into account the fact that users



(a) Probability of impatience. (b) Average throughput obtained by users (Mbit/s). (c) Average time spent by users in the cell (sec.).

Fig. 5: Evolution of the performance parameters as a function of the load (Model 2).

that have longer elements to download are more likely to wait. From this new definition, we developed two models that differ in the way an impatient user decides to stop its download. We compared the performance obtained by both models to simulations using NS3, and showed their accuracy. Based on the definition of a satisfaction criterion, we have shown that impatience of user in a LTE cell does not significantly degrade the global satisfaction of users with regards to a system without impatience, and greatly extends the range of utilization of the cell where customers have an acceptable satisfaction. As a conclusion, we can say that both the cell and users benefit from impatience.

References

[1] F. Baccelli and G. Hebuterne. On queues with impatient customers. Research Report RR-0094, 1981.

[2] T. Bonald and J. W. Roberts. Congestion at flow level and the impact of user behaviour. *Computer Networks*, 42(4):521–536, 2003.

[3] D. Collange, M. Hajji, J. Shaikh, M. Fiedler, and P. Arlos. User impatience and network performance. In *Next Generation Internet, 2012*, pages 141–148, 2012.

[4] D. J. Daley. General customer impatience in the queue gi/g/1. *Journal of Applied Probability*, 2(1):186–205, 1965.

[5] S. Jiang and N. H. Vaidya. Scheduling data broadcast to "impatient" users. In *International Workshop on Data Engineering for Wireless and Mobile Access*, pages 52–59, New York, NY, USA, 1999. ACM.

[6] C. Sanogo, T. Chahed, and S. E. Elayoubi. Modeling user impatience and its impact on performance in mobile networks. In *Personal Indoor and Mobile Radio Communications, 2013*, pages 2861–2865, 2013.

[7] S. Yang and G. de Veciana. Bandwidth sharing: the role of user impatience. In *Global Telecommunications Conference, 2001. IEEE*, volume 4, pages 2258–2262, 2001.