QoS provision in a dynamic channel allocation based on admission control decisions

CLAUDINA RATTARO, Facultad de Ingeniería, Universidad de la República, Uruguay
LAURA ASPIROT, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay
ERNESTO MORDECKI, Facultad de Ciencias, Universidad de la República, Uruguay
PABLO BELZARENA, Facultad de Ingeniería, Universidad de la República, Uruguay

Cognitive Radio Networks have emerged in the last decades as a solution of two problems: spectrum underutilization and spectrum scarcity. In this work we propose a dynamic spectrum sharing mechanism, where primary users have strict priority over secondary ones, in order to improve the mean spectrum utilization with the objective of providing to secondary users a satisfactory grade of service with a small interruption probability. We study a stochastic model for Cognitive Radio Networks with fluid limits techniques. Our main findings consist in a Gaussian limit theorem in the sub-critical case, and a non-Gaussian limit theorem, under a different scaling scheme, in the critical case. These results provide us practical QoS criteria for sharing policies. We support our analysis with representative simulated examples in both scenarios.

CCS Concepts: • **Mathematics of computing** → **Probability and statistics**; • **Networks** → **Network performance modeling**; **Network dynamics**; *Network performance analysis.*

Additional Key Words and Phrases: Cognitive radio networks, QoS, fluid limits, asymptotic distribution

ACM Reference Format:
Claudina Rattaro, Laura Aspirot, Ernesto Mordecki, and Pablo Belzarena. 2019. QoS provision in a dynamic channel allocation based on admission control decisions. *ACM Trans. Model. Perform. Eval. Comput. Syst.* 1, 1, Article 1 (January 2019), 30 pages. https://doi.org/10.1145/3372786

1 INTRODUCTION

During the last two decades we have seen an explosive development of wireless networks which is reflected on the widely extended use of wireless technologies in our everyday lives (e.g. mobile phones, sensors, laptops). Consequently the demand for electromagnetic spectrum has increased to unprecedented levels resulting in the spectrum scarcity problem. In spite of this, spectrum utilization measurements have shown that licensed bands are vastly underutilized while unlicensed bands are too crowded [6, 21, 39]. The significant underutilization occurs even in densely populated urban areas. Therefore, it is vital to move forward with a means for better utilization of the spectrum.

Nowadays industrial and academic communities are focusing much of their efforts to define the main characteristics of the fifth-generation of mobile networks. It is expected that 5G connections will appear on the scene in 2020 and will grow more than a thousand percent from 2.3 million in 2020 to over 25 million in 2021 [11]. With this in mind, how to optimize the spectrum capacity becomes

Authors' addresses: Claudina Rattaro, Facultad de Ingeniería, Universidad de la República, Julio Herrera y Reissig 565, Montevideo, 11200, Uruguay, crattaro@fing.edu.uy; Laura Aspirot, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Montevideo, Uruguay, laspirot@ccee.edu.uy; Ernesto Mordecki, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, mordecki@cmat.edu.uy; Pablo Belzarena, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, belza@fing.edu.uy.

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2376-3639/2019/1-ART1 $15.00
https://doi.org/10.1145/3372786

a key ingredient in 5G system design. In this regard Cognitive Radio represents an attractive candidate, especially for 5G networs [1, 17, 19, 28, 34].

The concept of Cognitive Radio (CR) was introduced by Mitola [26, 27]. CR represents a promising technology which, based on dynamic spectrum access, strives at solving the important problems that we mentioned before: spectrum underutilization and spectrum scarcity. In this paradigm we can identify two classes of users: primary and secondary. Primary users (PUs) are those for which a certain portion of the spectrum has been allocated to (often in the form of a paid contract). Secondary users (SUs) are devices which are capable of detecting unused licensed bands and adapt their transmission parameters for using them. The key requirement in this context is that the PUs ought to be as little affected as possible by the presence of SUs. One challenge then is to distribute the spectrum holes efficiently and fairly. Another goal is to provide quality of service (QoS) to PUs and also to SUs.

In what follows we focus on the analysis and characterization of a dynamic spectrum sharing mechanism where PUs have strict priority over secondary ones. We present tools and criteria that can be used in order to improve the mean spectrum utilization with the commitment of providing to SUs a satisfactory grade of service and a small interruption probability. We are interested in SUs whose service cannot be interrupted with high probability (like a phone call or other interactive services). For these services it is preferable to be rejected and to avoid the situation where the connection is established and then interrupted. These decisions (start service or not) represent a mechanism that can be adopted by the SUs as a sort of admission control policy. We analyze two indicators of these types of systems: the mean spectrum utilization and the probability that the system is working at its capacity limit which is related with the interruption probability for SUs services. Associated with this last issue we analyze a possible admission control policy in order to reduce this probability.

In more detail we consider a scenario with C subchannels to be distributed between SUs and PUs, and where PUs have strict priority. That is to say, we assume a scenario without spatial reuse of subchannels where if a PU arrives when all the resources are in use, one of the SUs will be deallocated immediately. For instance, consider a cellular network that employs frequency division duplexing where the operator has C frequency bands (subchannels) to be assigned to its users (PUs). If the primary network is LTE, we can assume a subchannel as one or more resource blocks. Another example is given by the digital TV spectrum bands. In both scenarios, if there are free resources, the SUs could use them with the constraint that their communications can be interrupted at any time. However, there are some differences between both examples: in a cellular network PUs can use any of the C subchannels, but in the case of digital TV each TV channel has its own frequency band. Then we assume in our model that when a PU arrives while a SU is using its subchannel, and there are free subchannels, this SU can be moved instantly to another unused band, without any consequence to its service. If there is not a free subchannel (all of them are busy), the SUs communication will be interrupted with consequences to its QoS. As our model takes into account only the number of subchannels that are being used by PUs and SUs, if there are free subchannels, the case of a PU that arrives to its subchannel and a SU must be moved instantly to another free one will be modeled as if the PU arrives to a free subchannel.

There are some previous works which contribute in this direction, with the most representative examples being [2, 20, 23, 40]. In [20] the authors study a spectrum sharing allocation for PUs and SUs where preemption can occur. Similar to our model, they consider that a SU call is immediately dropped when all resources are in use and a PU call arrives. However, they only analyze the queuing system (as two $M/M/C/C$ queues) but do not address the possibility of providing certain level of QoS to the users. Therefore they do not investigate the behavior of the system when admission control mechanisms are applied. In [23] the authors study preemptive and non-preemptive priority

queues. The main difference to our model is that the affected SU (which is deallocated when a PU needs a subchannel and are all in use) has to wait in the system until a subchannel is available again (in the paper it is assumed the existence of a buffer). In our case we consider a total interruption of the communication, with no buffer. Our assumption is more reasonable in services like a phone call or other interactive services. On the other hand, in [40] the authors study the stationary regime for a multiclass queue with preemption, similar to ours, obtaining a recursive expression for the probability of interruption. Finally in [2], although the authors study an admission control mechanism over SUs, they do no obtain an analytic expression of QoS metrics and only evaluate them through simulations.

In this work we model the cognitive radio network as a two-dimensional continuous time Markov chain (CTMC). In our approach the Markovian structure allows us to analyze its asymptotic behavior by means of a deterministic approximation obtained as a solution of a differential equation, i.e. its fluid limit. Under some regularity hypotheses on transition rates, the fluid limit is the solution of an ordinary differential equation (ODE). In cases as ours, with discontinuous transition rates, the fluid limit is a piecewise smooth dynamical system (PWSDS). We characterize the system behavior studying the PWSDS fixed points and we show that, in many cases, a SUs admission control mechanism is required in order to ensure a low probability of service interruption. Some preliminary results were published in our previous article [30].

In the present paper we extend the analysis to a probabilistic admission control mechanism and most importantly, we incorporate aspects such as the description of the asymptotic distribution associated with the fluid limit. This distribution depends strongly on the fixed points of the deterministic approximation, both with and without admission control, and we find, depending on the parameters of the model, Gaussian and non-Gaussian asymptotic distributions. The asymptotic distribution permits the analysis of the interruption probability for SUs giving confidence bounds valid when the number of users is large. In the case of non-Gaussian distribution, the stationary regime and its limit are described explicitly in a simplified case. Using our results we also present some practical network design criteria. In particular, the asymptotic distribution that was only mentioned in [30] is developed here.

Applications of fluid models to telecommunications appeared in the literature and were widely developed in the last decade. Some recent examples include for instance: peer-to-peer systems and mobile networks (see for example [3, 5, 14, 29, 31, 32, 37] and references therein). In particular concerning applications to CR networks there are some related works that we would like to highlight. In [38] the authors use a fluid model to study SU's queuing delay performance. In [41] they study the coexistence of two wireless networks with different priorities and compare throughput and delay obtained in both networks. On the other hand, [42] is focused on the collaborative sensing within the SUs and its impact in QoS. In these papers the authors analyze the delay and throughput in different CR scenarios, so their results can be complementary to ours where we analyze the interruption probability.

The rest of the paper is structured as follows. After this introductory section, in Section 2, we describe our model of spectrum sharing in CR networks. Section 3 presents the fluid model and in the following subsections we show our analysis and characterization of the behavior of the system under different admission control policies. We study the system for three different admission control policies: free, deterministic and probabilistic. For all cases we introduce a fluid limit, and study the system through it , analyzing fixed points and the asymptotic distribution around the limit. Finally, we conclude and discuss future work in Section 4.

2 MODEL DESCRIPTION

In this section we introduce our stochastic model for the number of primary and secondary users in the system as a two-dimensional CTMC. This is a comprehensive and although simple model for the system under study, however, it does not have an analytical solution. Therefore, we introduce in Section 3 a scaled version of the CTMC, in order to find a fluid limit that allows us to study the system analytically.

We model the arrival processes for both type of users as independent Poisson processes, and the service durations are also independent and exponentially distributed random variables. We also model the possibility of admission control decisions when a SU arrives to the system (SUs shall decide, depending on the state of the system, whether to enter or not). We associate one user with one channel.

In this context a general model and its variations due to the admission control policy assumptions are stated in the following definitions.

Definition 2.1 (General model). Consider a CTMC (X_1, X_2) defined as follows.

- $X_1(t)$, $X_2(t)$: number of PUs and SUs at time t respectively,
- C: total number of identical subchannels, therefore, the state space is the subset:

$$E = \{(x_1, x_2) \in \mathbb{N}^2 : 0 \leq x_1, 0 \leq x_2, x_1 + x_2 \leq C\}, \tag{1}$$

- λ_1, λ_2: arrival rates for PUs and SUs respectively (independent Poisson arrivals),
- μ_1, μ_2: service rates for PUs and SUs respectively (independent exponentially distributed service times),
- $a(x_1, x_2)\colon E \longrightarrow \mathbb{R}$, admission decision function for SUs, and represents the probability that a SU that arrives starts being served when there are x_1 PUs and x_2 SUs.

Thus the stochastic process $(X_1(t), X_2(t))$ has transition rates $q((x_1, x_2), (x_1', x_2'))$, from state (x_1, x_2) to state (x_1', x_2'), defined by:

- $q((x_1, x_2), (x_1 + 1, x_2)) = \lambda_1$, if $x_1 + x_2 < C$,
- $q((x_1, x_2), (x_1 - 1, x_2)) = \mu_1 x_1$,
- $q((x_1, x_2), (x_1, x_2 + 1)) = a(x_1, x_2)\lambda_2$, if $x_1 + x_2 < C$,
- $q((x_1, x_2), (x_1, x_2 - 1)) = \mu_2 x_2$,
- $q((x_1, x_2), (x_1 + 1, x_2 - 1)) = \lambda_1$, if $x_1 + x_2 = C$ and $x_2 \neq 0$.

Definition 2.2 (Free admission control policy). We call free admission control model when in the previous definition we consider no admission control policy, then $a(x_1, x_2) = 1$ if $x_1 + x_2 < C$ and $a(x_1, x_2) = 0$ if $x_1 + x_2 = C$.

Definition 2.3 (Deterministic admission control policy). In the deterministic case $a(x_1, x_2) \in \{0, 1\}$; if $a(x_1, x_2) = 1$ and a SU arrives, it will start being served, and when $a(x_1, x_2) = 0$, it will not. In this work we assume $x_1 + x_2 = \delta$, with $0 < \delta < C$, as an admission control boundary. That is to say, $a(x_1, x_2) = 1$ if $x_1 + x_2 < \delta$ and $a(x_1, x_2) = 0$ if $x_1 + x_2 \geq \delta$.

Definition 2.4 (Probabilistic admission control policy). In the probabilistic admission control SUs can access the system with a probability related with the number of users in the system. Let us assume that $a(x_1, x_2)$ is a continuous function that vanishes close to the border $\gamma = \{(x_1, x_2) : x_1 + x_2 = C\}$. In this work we consider $a(x_1, x_2) = 1 - (x_1 + x_2)/C$.

Concerning the above formulation we may make some remarks. First notice that PUs behave as a $M/M/C/C$ queue, independent of the behavior of the SUs and of the admission policy. Secondly, in this work we consider two approaches for the admission control mechanisms: a deterministic and a probabilistic one. Thirdly, it is important to highlight that in the case when $\mu_1 = \mu_2$, for some

type of admission policies $a(x_1, x_2)$, the process is a one-dimensional CTMC and the stationary distribution can be computed explicitly. However, when $\mu_1 \neq \mu_2$ (which represents the natural situation in cognitive radio networks) it is not possible to obtain a closed form expression of its stationary distribution (see for example [43] and the references therein). Finally, although in the general case the stationary distribution can be computed numerically, our approach consists in formulating the corresponding fluid limit in order to characterize the system behavior and study the influence of the admission control decisions in a more feasible and efficient way. We infer properties of the stochastic system from the study of fixed points of the deterministic fluid limit and the asymptotic distribution around it, and we define practical QoS criteria for sharing spectrum policies.

3 FLUID MODEL

The Markovian structure of a process allows us to analyze its asymptotic behavior by means of a simpler deterministic approximation: the fluid limit (generally obtained as a solution of an ODE). Fluid limits have been used for particle systems, biology, epidemics, as well as in the study of telecommunication networks. Generally speaking, starting from a stochastic model, the objective is to find a deterministic approximation for the original process. Whereas the stochastic process is a microscopic description of the system, we can say that the corresponding differential equation gives a macroscopic description that captures the main characteristics of the system. See for instance classical results on convergence of Markov processes in [10, 12, 13, 33, 36].

As a very simplified description of the method, the proof of this approximation result is generally based on a martingale decomposition of the Markov process, which shows that the average behavior of the stochastic process is captured by the drift part while the stochastic fluctuation of second order (corresponding to the martingale) vanishes with the scaling and limit procedure. More specifically, consider a Markov process $\widetilde{X}^N(t)$ parametric in N and its martingale decomposition:

$$\widetilde{X}^N(t) = \widetilde{X}^N(0) + \int_0^t \widetilde{Q}^N(\widetilde{X}^N(s))ds + \widetilde{M}^N(t),$$

where $\widetilde{Q}^N(l)$ is the so-called drift of the process at state l, which is calculated as $\sum_{m\in S}(l-m)q(l,m)$, being $q(l,m)$ the transition rate from state l to m, S the state space, and where $\widetilde{M}^N(t)$ is a martingale. Consider now the scaled process $X^N(t) = \widetilde{X}^N(t)/N$, then:

$$X^N(t) = X^N(0) + \frac{1}{N}\int_0^t \widetilde{Q}^N(\widetilde{X}^N(s))ds + \frac{\widetilde{M}^N(t)}{N}.$$

If there exists a Lipschitz function Q such that:

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \left\| \frac{\widetilde{Q}^N(\widetilde{X}^N(t))}{N} - Q(X^N(t)) \right\| = 0 \tag{2}$$

in probability, where $\|\cdot\|$ is the Euclidean norm, and $\widetilde{M}^N(t)/N$ converges to zero in probability, then $X^N(t)$ converges in probability over compact time intervals to a deterministic process $x(t)$, described by the ODE:

$$x'(t) = Q(x(t)). \tag{3}$$

The drift Q may be interpreted as the expected rate of change of the stochastic process (see Figure 1 for an illustration of this convergence in a toy example).

This introduces the problem of finding the suitable scale for the approximation. A typical scaling procedure consists in dividing the process by N and considering transition rates multiples of N; jumps are of order $1/N$ and transition rates are of order N, which means that the product remains or

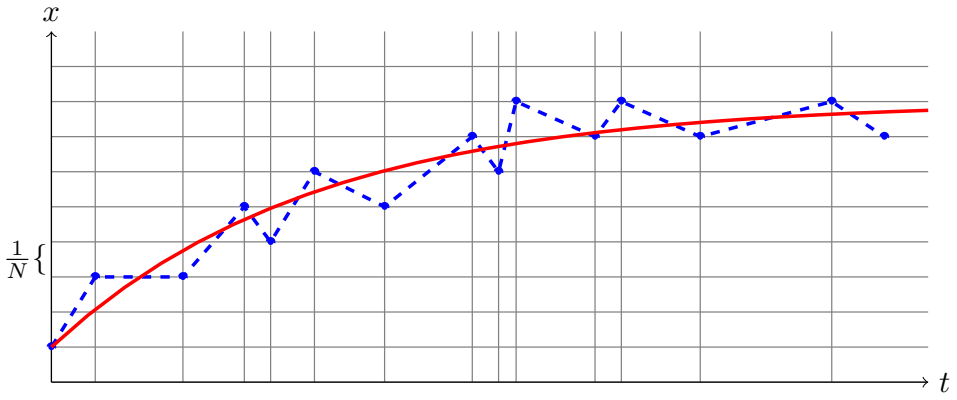


Fig. 1. Illustration of a Markov process convergence. The continuous red line represents the solution of the ODE and the dashed blue one is a trajectory of $X^N(t)$.

tends to a constant as N increases. This is the scaling considered in this work, we refer to [18, 25, 33] as references for this and other scaling regimes, under which there are different limit results.

Let N be the scaling factor, and we define, as for the original model presented in Definition 2.1, the following sequence of CTMCs $(\widetilde{X}_1^N, \widetilde{X}_2^N)$ and whose state space is:

$$\widetilde{E}^N = \{(Ni, Nj) : (i, j) \in E\}$$

where E is defined in Equation (1).

Definition 3.1. Consider the sequence of CTMCs $(\widetilde{X}_1^N, \widetilde{X}_2^N)$ defined as follows.

- $\widetilde{X}_1^N(t)$, $\widetilde{X}_2^N(t)$: number of PUs and SUs at time t, respectively,
- CN: total number of subchannels,
- $\lambda_1 N$, $\lambda_2 N$: arrival rates for PUs and SUs, respectively,
- μ_1, μ_2: service rates for PUs and SUs, respectively,
- $\widetilde{a}^N(x_1, x_2)$: admission decision for SUs in each state.

The admission decision in each state should verify:

$$\lim_{N \to +\infty} \widetilde{a}^N(Nx_1, Nx_2) = a(x_1, x_2).$$

This scaling concerns both admission control schemes presented in Section 2. In the case of a deterministic admission control $\widetilde{a}^N(x_1, x_2) = 1$ if $x_1 + x_2 < N\delta$ and $\widetilde{a}^N(x_1, x_2) = 0$ if $x_1 + x_2 \geq N\delta$. For the probabilistic admission control let $\widetilde{a}^N(x_1, x_2) = 1 - (x_1 + x_2)/CN$.

The sequence of scaled stochastic process $(\widetilde{X}_1^N(t), \widetilde{X}_2^N(t))$ has transition rates $\widetilde{q}^N((x_1, x_2), (x_1', x_2'))$, from state (x_1, x_2) to state (x_1', x_2'), defined by:

- $\widetilde{q}^N((x_1, x_2), (x_1 + 1, x_2)) = \lambda_1 N$, if $x_1 + x_2 < CN$,
- $\widetilde{q}^N((x_1, x_2), (x_1 - 1, x_2)) = \mu_1 x_1$,
- $\widetilde{q}^N((x_1, x_2), (x_1, x_2 + 1)) = \widetilde{a}^N(x, x_2)\lambda_2 N$, if $x_1 + x_2 < CN$,
- $\widetilde{q}^N((x_1, x_2), (x_1, x_2 - 1)) = \mu_2 x_2$,
- $\widetilde{q}^N((x_1, x_2), (x_1 + 1, x_2 - 1)) = \lambda_1 N$, if $x_1 + x_2 = CN$ and $x_2 \neq 0$.

In Table 1 we summarize the scaled parameters and their relationship with the original ones.

In this scaling scheme we go from our original system in Section 2 to a system where arrival rates and capacity are multiplied by N, that can be interpreted as a large network, with many users (both PUs and SUs) and large capacity. On the other hand, service rates in each channel are not scaled,

Table 1. Original and scaled parameters for the two-dimensional process (X_1, X_2) and $(\widetilde{X}_1^N, \widetilde{X}_2^N)$.

(X_1, X_2)	λ_1	λ_2	μ_1	μ_2	C	a
$(\widetilde{X}_1^N, \widetilde{X}_2^N)$	$\lambda_1 N$	$\lambda_2 N$	μ_1	μ_2	CN	$\widetilde{a}^N$

as for instance they depend on the service type, and they do not increases individually, despite the total service time increases with the number of users as $\mu_1 \widetilde{X}_1^N(t)$ for PUs and $\mu_2 \widetilde{X}_2^N(t)$ for SUs. Finally the admission control in the large system depends only on the proportion of resources occupied.

We consider now the scaled process (X_1^N, X_2^N), defined by

$$(X_1^N, X_2^N) = (\widetilde{X}_1^N, \widetilde{X}_2^N)/N \tag{4}$$

and whose state space is:

$$E^N = \{(i/N, j/N) : (i, j) \in E\}$$

where E is defined in Equation (1).

This scaled process will converge to the deterministic fluid limit. In order to state the fluid limit result we verify that our process satisfies equation (2). We compute the drift in the following proposition.

Proposition 3.2. *The drift for the process stated in Definition 2.1 is*

$$\begin{aligned}
Q(x_1, x_2) &= \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ \lambda_2 a(x_1, x_2) - \mu_2 x_2 \end{pmatrix}, \ \text{if } x_1 + x_2 > C; \\
Q(x_1, x_2) &= \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ -\lambda_1 - \mu_2 x_2 \end{pmatrix}, \ \text{if } x_1 + x_2 = C, \ x_2 > 0; \\
Q(x_1, x_2) &= \begin{pmatrix} -\mu_1 x_1 \\ 0 \end{pmatrix}, \ \text{if } (x_1, x_2) = (C, 0).
\end{aligned}$$

The drift $\widetilde{Q}^N(x_1, x_2)$ for the corresponding scaled process presented in Definition 3.1 verifies

$$\lim_{N \to +\infty} \frac{1}{N} \widetilde{Q}^N \left(\widetilde{X}_1^N, \widetilde{X}_2^N \right) = Q \left(X_1^N, X_2^N \right).$$

Proof. We compute the drift of (X_1, X_2) defined as:

$$Q(x_1, x_2) = \sum_{(x_1', x_2') \in E} q((x_1, x_2), (x_1', x_2')) \left[(x_1', x_2') - (x_1, x_2) \right].$$

$$\begin{aligned}
Q(x_1, x_2) &= \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a(x_1, x_2) \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \mu_1 x_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_2 x_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \ \text{if } x_1 + x_2 < NC, \\
Q(x_1, x_2) &= \lambda_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \mu_1 x_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_2 x_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \ \text{if } x_1 + x_2 = C \text{ and } x_2 > 0, \\
Q(x_1, 0) &= \mu_1 x_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \ \text{if } x_1 = C \text{ and } x_2 = 0.
\end{aligned}$$

We compute the drift of $(\widetilde{X}_1^N, \widetilde{X}_2^N)$ defined as:

$$\widetilde{Q}^N(x_1, x_2) = \sum_{(x_1', x_2') \in \widetilde{E}^N} \widetilde{q}^N((x_1, x_2), (x_1', x_2')) \left[(x_1', x_2') - (x_1, x_2) \right].$$

$$\widetilde{Q}^N(x_1,x_2) = N\lambda_1\begin{pmatrix}1\\0\end{pmatrix} + \widetilde{a}^N(x_1,x_2)N\lambda_2\begin{pmatrix}0\\1\end{pmatrix} + \mu_1 x_1\begin{pmatrix}-1\\0\end{pmatrix} + \mu_2 x_2\begin{pmatrix}0\\-1\end{pmatrix}, \text{ if } x_1+x_2<NC,$$

$$\widetilde{Q}^N(x_1,x_2) = N\lambda_1\begin{pmatrix}1\\-1\end{pmatrix} + \mu_1 x_1\begin{pmatrix}-1\\0\end{pmatrix} + \mu_2 x_2\begin{pmatrix}0\\-1\end{pmatrix}, \text{ if } x_1+x_2=NC \text{ and } x_2>0,$$

$$\widetilde{Q}^N(x_1,0) = \mu_1 x_1\begin{pmatrix}-1\\0\end{pmatrix}, \text{ if } x_1=NC \text{ and } x_2=0.$$

Then we replace (x_1, x_2) by (Nx_1, Nx_2) in the previous equations and divide by N. We need that $\lim_{N\to+\infty} \widetilde{a}^N(Nx_1, Nx_2) = a(x_1, x_2)$, in fact we have that $\widetilde{a}^N(Nx_1, Nx_2) = a(x_1, x_2)$ for the three different admission control schemes presented in Section 2. According to this we can conclude that

$$\lim_{N\to+\infty} \frac{1}{N}\widetilde{Q}^N\left(\widetilde{X}_1^N, \widetilde{X}_2^N\right) = Q\left(X_1^N, X_2^N\right).$$

□

Classical results on convergence of Markov processes assume some regularity properties of the fluid ODE, i.e. the vector field $Q(x_1, x_2)$ defining the ODE must be a Lipschitz continuous function in the domain of interest. It is a sufficient condition for existence and uniqueness of solutions given initial conditions. In our system this regularity condition does not always hold. In this context, using results obtained by Bortolussi in [7, 9], it is possible to determine a PWSDS that is the fluid limit.

The rest of the paper is structured as follows. In Subsection 3.1 we introduce some results from [7, 9] and other references. Then, as a first step we analyze in Subsection 3.2 the system without an admission control policy (free admission control policy). In Subsection 3.3 we continue with the characterization of the system considering a deterministic admission control and finally in Subsection 3.4 we study a probabilistic admission control policy. In all cases we analyze, together with the fluid limit, the fixed points for the deterministic approximation and the asymptotic distribution, all with the objective of designing practical admission control algorithms.

3.1 Piecewise smooth approximation

In the case where the drift is discontinuous the fluid limit can be obtained in the framework of differential equations with discontinuous right-hand side [15]. In this context the differential equation is replaced by what is called a differential inclusion, this means that Equation (3) is replaced by

$$x'(t) \in \overline{Q}(x(t)) \tag{5}$$

where $\overline{Q}$ is a set-valued mapping known as Filippov extension of Q defined as the convex hull of the accumulation points of the drift. We define a Filippov solution as an absolutely continuous function $x(t)$ such that $x(0) = x_0$ and $x'(t) \in \overline{Q}(x(t))$ almost everywhere. We refer to [7, 16] and references therein for a more detailed exposition.

Concretely, considering $x'(t) = Q(x), Q : E \to \mathbb{R}^2, E \subseteq \mathbb{R}^2, \bigcup R_i \supseteq E$ (R_i $i = 1, 2$ is a set of disjoint regions), where Q is smooth on R_i and can be discontinuous only on the boundaries of R_i, and restricting our attention to our two different problems we have:

- Free admission control and probabilistic admission control policies: we only have one region (R_1) and its border γ. We have $Q_1(x)$ and $Q_2(x)$ the velocity vectors, both smooth in $R_1 = \{x : h(x) < 0\}$ and $\gamma = \{x : h(x) = 0\}$ respectively with $h(x)$ smooth such that $\nabla h(x) \neq 0$ in γ.

Then $x'(t) \in \overline{Q}(x(t))$ with:

$$\overline{Q}(x) = \begin{cases} Q_1(x) & \text{if } x \in R_1 \\ conv(Q_1, Q_2) & \text{if } x \in \gamma \end{cases}.$$

- Deterministic admission control policies: we have a system with two regions (R_1, R_2) and the boundary between them (γ'). Let $Q_1(x)$ and $Q_2(x)$ be continuous in $R_1 = \{x : h(x) < 0\}$ and $R_2 = \{x : h(x) > 0\}$ respectively and $\gamma' = \{x : h(x) = 0\}$ with $h(x)$ a smooth function with $\nabla h(x) \neq 0$ in γ'.) Then $x'(t) \in \overline{Q}(x(t))$ with:

$$\overline{Q}(x) = \begin{cases} Q_1(x) & \text{if } x \in R_1 \\ conv(Q_1, Q_2) & \text{if } x \in \gamma' \\ Q_2(x) & \text{if } x \in R_2 \end{cases},$$

where $conv(Q_1, Q_2)$ is the convex hull of the vectors $\{Q_1, Q_2\}$. Notice that for each continuity point x of Q, $\overline{Q}(x) = \{Q(x)\}$, so that we have a proper differential inclusion only in the discontinuity region (the borders γ or γ').

Filippov proved results about existence and uniqueness of solutions [15]. If Q_1 and Q_2 are C^1, $Q_1 - Q_2$ is C^1 in γ (or γ' as appropriate), h is C^2 in γ (or γ') and at least one of $n^T(x)Q_1(x) > 0$ or $n^T(x)Q_2(x) < 0$ holds, with x on the border γ (or γ') and $n(x)$ the normal vector to the border, then there exists a unique Filippov solution from each initial condition. Considering x on the border γ' (or γ), there are different behaviors of a solution starting in x depending on the value of $n^T(x)Q_1(x)$ and $n^T(x)Q_2(x)$:

- transversal crossing: if $n^T(x)Q_1(x)$ and $n^T(x)Q_2(x)$ have the same sign, e.g. if $n^T(x)Q_1(x) > 0$ and $n^T(x)Q_2(x) > 0$, a solution starting in R_1 crosses the border and stays in R_2;
- sliding motion: if $(n^T(x)Q_1(x))(n^T(x)Q_2(x)) < 0$ there is sliding motion, there are two cases: unstable sliding motion when $n^T(x)Q_1(x) < 0$ and $n^T(x)Q_2(x) > 0$ (in this case there is no uniqueness for solutions) and stable sliding motion when $n^T(x)Q_1(x) > 0$ and $n^T(x)Q_2(x) < 0$. In this last case, that is ours, the system cannot escape from the border, then the solution follows a vector field obtained as convex combination of Q_1 and Q_2, obtaining a new vector field $g(x) = (1 - \alpha(x))Q_1 + \alpha(x)Q_2$ with $\alpha(x) \in [0, 1]$ that verifies $n^T(x)g(x) = 0$;
- tangential crossing: if $n^T(x)Q_1(x) = 0$ (or $n^T(x)Q_2(x) = 0$), then the trajectory continues in the region pointed by the non-zero vector field.

In our context, in all cases (free, deterministic and probabilistic admission control) we have a discontinuous drift that leads to a differential equations with discontinuous right-hand side and to the presence of sliding motion, as we show in the following subsections.

3.2 Free admission control policy

In this subsection we assume that $a(x_1, x_2) = 1$ for all $(x_1, x_2) \in R_1$ such that $R_1 = \{(x_1, x_2) : x_1 + x_2 - C < 0\}$, 0 otherwise. The goal is to study the behavior of the system without any intervention: if a SU arrives and there is at least one idle subchannel, the SU will be served. Here we have discontinuous transition rates, as in the border of the state space $\gamma = \{(x_1, x_2) : x_1 + x_2 - C = 0\}$ we have $a(x_1, x_2) = 0$. As we explained, these discontinuities lead to a deterministic limit whose trajectories are continuous but not differentiable. Because of this, some trajectories present sliding motion.

Proposition 3.3. *Let $Q_1(x_1, x_2)$ and $Q_2(x_1, x_2)$ be vector fields, both smooth in R_1 and γ respectively such that*

$$Q_1(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ \lambda_2 - \mu_2 x_2 \end{pmatrix}, \qquad Q_2(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ -\lambda_1 - \mu_2 x_2 \end{pmatrix},$$

and let $n(x_1, x_2)$ be the normal vector to the boundary γ ($n^T(x_1, x_2) = (1, 1)$ for all $(x_1, x_2) \in \gamma$) then we have a PWSDS driven by the following equations.

If $x_1 + x_2 < C$ or $\lambda_1 + \lambda_2 \leq \mu_1 x_1 + \mu_2 x_2$:

$$\begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= \lambda_2 - \mu_2 x_2. \end{cases},$$

if $x_1 + x_2 = C$ and $\lambda_1 + \lambda_2 > \mu_1 x_1 + \mu_2 x_2$:

$$\begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= -\lambda_1 + \mu_1 x_1. \end{cases}.$$

and if $x_1 = C$:

$$\begin{cases} x_1' &= -\mu_1 x_1, \\ x_2' &= 0. \end{cases}.$$

Sketch of the proof. We have that $n^T(x_1, x_2)Q_1(x_1, x_2) = 0 \Leftrightarrow \lambda_1 + \lambda_2 - \mu_1 x_1 - \mu_2 x_2 = 0$ and $n^T(x_1, x_2)Q_2(x_1, x_2) = 0 \Leftrightarrow -\mu_1 x_1 - \mu_2 x_2 = 0$. Then, for studying $n^T(x_1, x_2)Q_i(x_1, x_2)$ we have several cases depending on the position of the line $\lambda_1 + \lambda_2 - \mu_1 x_1 - \mu_2 x_2 = 0$. It is clear that it depends on the values of λ_1, λ_2, μ_1 and μ_2. In particular, we have that $n^T(x_1, x_2)Q_1(x_1, x_2) > 0$ if $\lambda_1 + \lambda_2 - \mu_1 x_1 - \mu_2 x_2 > 0$, Q_1 and n are tangent in the points over the line $\lambda_1 + \lambda_2 - \mu_1 x_1 - \mu_2 x_2 = 0$ and $n^T(x_1, x_2)Q_1(x_1, x_2) < 0$ if $\lambda_1 + \lambda_2 - \mu_1 x_1 - \mu_2 x_2 < 0$. On the other hand, $n^T(x_1, x_2)Q_2(x_1, x_2) < 0$ in γ independently of the parameters λ_i and μ_i. Because of this we are in the presence of sliding motion when $-\mu_1 x_1 + \mu_2 x_2 < \lambda_1 + \lambda_2$ and $x_1 + x_2 = C$. In that case we define the differential inclusion by a convex combination $g(x_1, x_2) = (1 - \alpha(x_1, x_2))Q_1 + \alpha(x_1, x_2)Q_2$ verifying $n^T(x_1, x_2)g(x_1, x_2) = 0$ (the solution cannot escape from the border). Computing $\alpha(x_1, x_2)$ we obtain

$$\alpha(x_1, x_2) = \frac{\mu_1 x_1 + \mu_2 x_2}{\lambda_1 + \lambda_2}$$

then substituting in $g(x_1, x_2)$ the result is proved. □

Let

$$(x_1(t), x_2(t)) \tag{6}$$

be the PWSDS that is the solution to the previous equations with initial condition $(x_1(0), x_2(0))$, we describe the trajectories for different system parameters and initial conditions. Complementary to this, in Prop. 3.5 we study the fixed points.

The discussion depends on the existence or not of times where trajectories hit or leave the border γ. Those times are the solutions of the equations $x_1(t) + x_2(t) - C = 0$ and $\lambda_1 + \lambda_2 = \mu_1 x_1(t) + \mu_2 x_2(t)$ and can not always be computed explicitly.

- Case 1: If the initial condition $(x_1(0), x_2(0))$ satisfies $x_1(0) + x_2(0) < C$ we have two subcases:
  - if $\nexists\, t^+ > 0$ such that $x_1(t^+) + x_2(t^+) - C = 0$, then the solution $(x_1(t), x_2(t))$ evolves according to $Q_1(x_1, x_2)$ remaining in R_1 and converges to its attractor fixed point (x_1^*, x_2^*). See Fig. 2 first row left. Note that the fixed point may also be on γ.
  - if $\exists\, t^+ > 0$ such that $x_1(t^+) + x_2(t^+) - C = 0$, then the solution $(x_1(t), x_2(t))$ evolves according to $Q_1(x_1, x_2)$ remaining in R_1 until it hits the border at t^+. Afterwards, the solution evolves according to

    $$g(x_1, x_2) = \begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= -\lambda_1 + \mu_1 x_1. \end{cases}.$$

    and depending on the fixed point we have:

491 ∗ if the fixed point $(x_1^*, x_2^*) \in R_1$, always $\exists t^{++}$, an exit point, such that $t^{++} > t^+ : \mu_1 x_1(t^{++}) + \mu_2 x_2(t^{++}) = \lambda_1 + \lambda_2$. Then for $t > t^{++}$ the solution evolves according to $Q_1(x_1, x_2)$ and converges to (x_1^*, x_2^*). See Fig. 2 second row left.
494 ∗ if the fixed point $(x_1^*, x_2^*) \in \gamma$, the solution remains in γ $\forall t > t^+$. See Fig. 2 second row right.
496 • Case 2: If the initial condition $(x_1(0), x_2(0))$ satisfies $x_1(0) + x_2(0) = C$ (with $x_2(0) > 0$)
497 – if $\exists\, t^* > 0$ such that $\mu_1 x_1(t^*) + \mu_2 x_2(t^*) = \lambda_1 + \lambda_2$, the solution $(x_1(t), x_2(t))$ evolves according to $g(x_1, x_2)$ (defined before) until t^* (an exit point) and then it evolves according to $Q_1(x_1, x_2)$. It converges to the fixed point $(x_1^*, x_2^*) \in R_1$. See Fig. 2 first row right.
500 – if $\nexists\, t^* > 0$ such that $\mu_1 x_1(t^*) + \mu_2 x_2(t^*) = \lambda_1 + \lambda_2$, then
501 ∗ if $(x_1^*, x_2^*) \in R_1$, $\forall t > 0$ the solution $(x_1(t), x_2(t))$ evolves according to $Q_1(x_1, x_2)$ remaining in R_1 and converges to (x_1^*, x_2^*). See Fig. 2 third row left.
503 ∗ if $(x_1^*, x_2^*) \in \gamma$, $\forall t > 0$ the solution $(x_1(t), x_2(t))$ evolves according to $g(x_1, x_2)$ remaining in γ and converges to (x_1^*, x_2^*). See Fig. 2 third row right.
505 • Case 3: If the initial condition $(x_1(0), x_2(0))$ satisfies $x_1(0) = C$:
506 – if $\lambda_1/\mu_1 > C$, the solution remains at $(C, 0)$ being the fixed point.
507 – if $\lambda_1/\mu_1 < C$, the solution evolves like Case 2.

508 Next we present the fluid limits results.
509
510 THEOREM 3.4. *Consider the process* $(\widetilde{X}_1^N, \widetilde{X}_2^N)$ *with transition rates defined in Table 1, define:*

$$(X_1^N(t), X_2^N(t)) = (\widetilde{X}_1^N(t), \widetilde{X}_2^N(t))/N$$

512
513 *and let* (x_1, x_2) *be the PWSDS defined in Equation* (6) *with initial condition* $(x_1(0), x_2(0))$. *If*

$$\lim_{N\to\infty} (X_1^N(0), X_2^N(0)) = (x_1(0), x_2(0))$$

515
516 *then, for all* $T > 0$,

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \left\| (X_1^N(t), X_2^N(t)) - (x_1(t), x_2(t)) \right\| = 0$$

518
519 *in probability.*

520 SKETCH OF THE PROOF. The proof follows straightforward from Theorem IV.2 in [7]. The hypotheses of this Theorem are verified in our case; i.e. the scaling scheme of the process, as we have defined in Section 3 and sumarized in Table 1, and the existence of a unique PWSDS with regular trajectories, presented Subsection 3.2 in order to define Equation (6). Theorem IV.2 in [7] considers two different regions R_1 and R_2 with a border between them, where the process and the PWSDS may change many times from one region to the other and there may be several different pieces of sliding motion at the border. In our case we have only one region and the border, but the proof in [7] is suitable for this case. More specifically, the proof there consists in splitting the whole PWSDS trajectory in pieces in each region (where classical results, for example from [13], hold) and in sliding trajectories at the border. For that case the proof in [7] consists in replacing the discontinuous drift by the sliding vector field and prove convergence by an uniformization procedure. The proof of Lemma 3 in [8] that considers two different regions is valid for the case with a region and its border.

533 In our case, as we in fact have a switched linear system we only have one piece of sliding motion, and as our initial condition can only be in R_1 we have three cases. In the first case the deterministic trajectory stays all the time in R_1 and classical results hold. In the second case the trajectory starts in R_1 and then presents sliding motion and stays at the border for $t \to \infty$. In the last case the trajectory PSWDS starts on R_1 or at the border, then presents sliding motion and exits at the border, and then remains in R_1. In that case we need to check the exit conditions for the sliding motion,
539

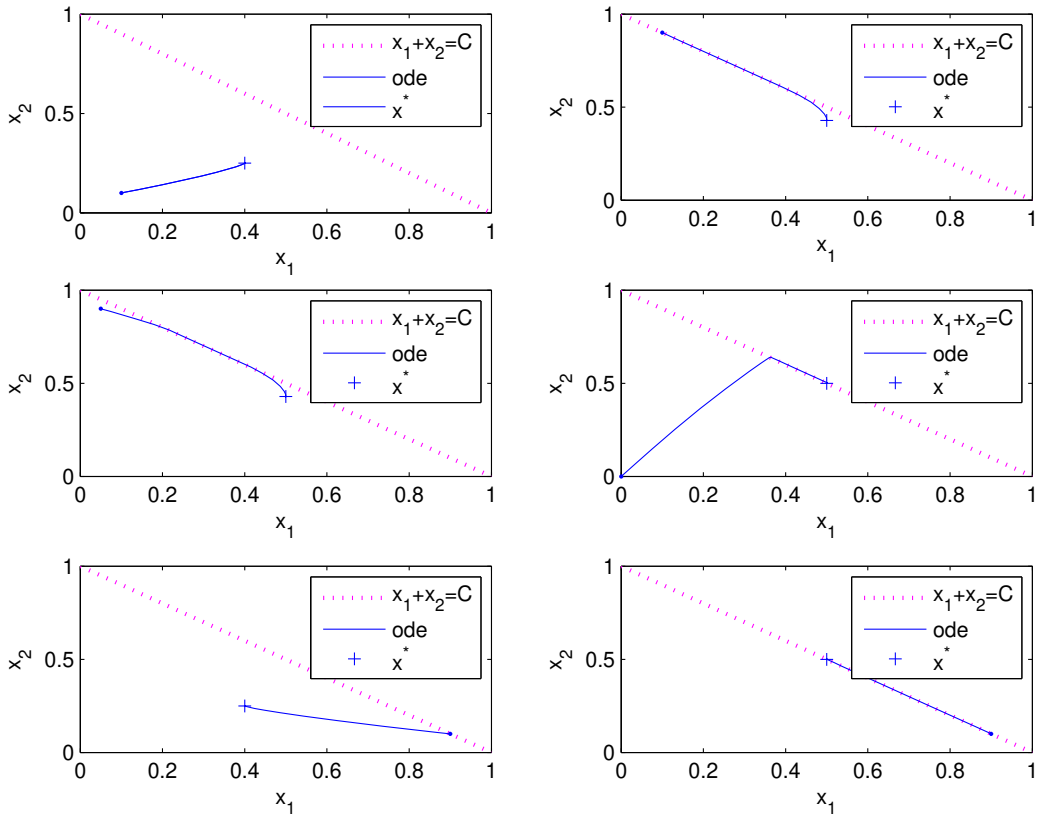


Fig. 2. Different trajectories for the PWSDS defined in Proposition 3.3 with $C = 1$. First row (left) shows a trajectory in R_1, starting at $(0.1, 0.1)$ ($\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$, $\mu_2 = 4$). First row (right) shows a trajectory with initial condition $(0.1, 0.9)$ at γ, with an exit point and that converges to the fixed point in R_1 ($\lambda_1 = 3$, $\lambda_2 = 1.5$, $\mu_1 = 6$, $\mu_2 = 3.5$). The second row (left) shows a trajectory with initial condition $(0.05, 0.9)$ at R_1, that stays for a while on γ and finally converges to the fixed point ($\lambda_1 = 3$, $\lambda_2 = 1.5$, $\mu_1 = 6$, $\mu_2 = 3.5$) and (right) a trajectory with initial condition $(0, 0)$ in R_1 which finally remains on γ ($\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$, $\mu_2 = 5$). The third row shows trajectories with initial condition $(0.9, 0.1)$ at γ: (left) the whole trajectory for $t > 0$ in R_1 ($\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$, $\mu_2 = 4$) and (right) trajectory on γ $\forall t > 0$ ($\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$, $\mu_2 = 5$.)

that are guaranteed because both vector fields are not tangential to the border at the same time (let us recall that Q_2 always points towards R_1). □

An alternative approach can be done following [36]. In Chapter 8 the authors define what they call flat boundary process, that is a Markov process where transitions at the border are different from transitions in the interior of the state space, and transition rates are discontinuous at the border. Under some regularity conditions the authors of [36] obtained very similar results than [7]: they prove existence and uniqueness for the solution of the PWSDS and they prove convergence to such systems. In their presentation the coefficients of the convex combination defining the sliding vector field have a probabilistic interpretation. They consider a process with three coordinates, where the first two correspond to PUs and SUs as ours, and the third, valued 0 or 1 indicates if the system is at the border or not. With the same scaling as ours the third coordinate tends to 0 when

scaled. In addition, as transition rates in the interior are smooth, while process stays in the interior the scaled process (dividing by N) changes a little. On the other hand, the number of jumps from the interior to the border per time unit increases with N, so the third coordinate that indicates where the process is reaches its invariant distribution before the number of PUs and SUs. In that interpretation the coefficient α is the probability or the proportion of time that the process spends at the border and $(1-\alpha)$ the the proportion of time that the process spends in the interior.

For further work concerning fluid limits, including systems with discontinuous rates we also refer to [9, 16, 22].

In the context of fluid limits it is usual to infer from the fixed point analysis of the deterministic system the behavior of the stochastic one in the stationary regime. If there is a unique fixed point that is a global attractor, the stochastic invariant distributions converges in probability to this fixed point [4, 24, 37]. In what follows we will exploit this general result.

Proposition 3.5. *Considering $a(x_1, x_2) = 1$ for all $(x_1, x_2) : x_1 + x_2 < C$, and 0 otherwise, letting R_1 and γ be the above defined zone and border and setting (x_1^*, x_2^*) as the PWSDS fixed point, then:*

a. *If $\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} < C$, then the fixed point $(x_1^*, x_2^*) = \left(\frac{\lambda_1}{\mu_1}, \frac{\lambda_2}{\mu_2}\right) \in R_1$ and the mean system utilization is $\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2}$ (sub-critical case).*
b. *If $\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} \geq C$ and $\frac{\lambda_1}{\mu_1} < C$, then the fixed point $(x_1^*, x_2^*) = \left(\frac{\lambda_1}{\mu_1}, C - \frac{\lambda_1}{\mu_1}\right) \in \gamma$ and the mean system utilization is C (critical case).*
c. *If $\frac{\lambda_1}{\mu_1} \geq C$, then the fixed point $(x_1^*, x_2^*) = (C, 0) \in \gamma$ and the mean system utilization is C (critical case).*

Sketch of the proof. Let Q_1 and Q_2 be the velocity vectors, both continuous in R_1 and γ respectively:

$$Q_1(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ \lambda_2 - \mu_2 x_2 \end{pmatrix}, Q_2(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ -\lambda_1 - \mu_2 x_2 \end{pmatrix},$$

and let $n(x_1, x_2)$ be a normal vector to the line $\gamma = \{(x_1, x_2) : x_1 + x_2 = C\}$, so that $n^T(x_1, x_2)$ is collinear with $(1, 1)$ for all $(x_1, x_2) \in \gamma$.

As it is explained in Subsection 3.1, when summarizing the different possible behaviors of both vector fields we have that all possible cases, for different values of the parameters, can be categorized into two groups represented by Case 1 and Case 2 from Figure 3. In that figure the solid line represents γ, the dotted line is $\lambda_1 + \lambda_2 - \mu_1 x_1 - \mu_2 x_2 = 0$ and the vectors are Q_1 and Q_2. In Case 1 the PWSDS fixed point is in R_1 (Proposition 3.5.a); we call it sub-critical case. It is easy to note that $(x_1^*, x_2^*) = (\lambda_1/\mu_1, \lambda_2/\mu_2)$. On the other hand, in Case 2, the fixed point is on γ and its value is $(x_1^*, x_2^*) = (\lambda_1/\mu_1, C - \lambda_1/\mu_1)$ representing a critical case, (Proposition 3.5.b). When $\lambda_1/\mu_1 + \lambda_2/\mu_2 \geq C$ we can identify a sliding motion behavior near the fixed point, more precisely we can affirm that the equation solution will live on γ most of the time. Both examples of Figure 3 consider $\lambda_1/\mu_1 < C$. Finally, when the system is saturated by PUs ($\lambda_1/\mu_1 \geq C$), as a corollary from Proposition 3.5.b the fixed point is $(x_1^*, x_2^*) = (C, 0)$ (Proposition 3.5.c). □

In Figures 4 and 5 we show the deterministic approximation and a trajectory of the stochastic processes for the Cases 1 and 2 of Figure 3. In each one, in the left graphic we show the simulation of one trajectory of the scaled Markov process and the PWSDS. In the right we show for the same simulation the evolution on the plane of the Markov chain and the corresponding PWSDS. In Figure 4 the fixed point is in R_1 and in Figure 5 it is on γ. It is important to note that in both cases,

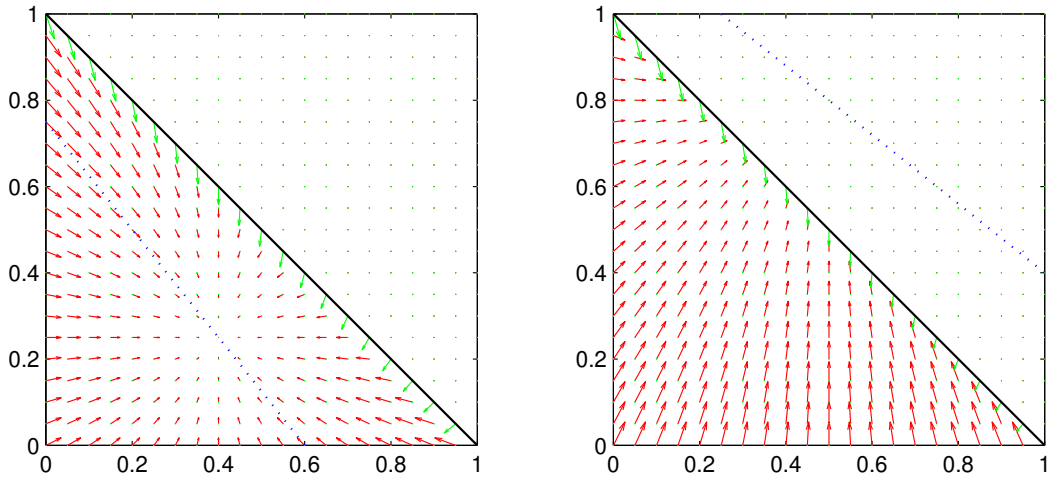


Fig. 3. Vector field for Case 1 (left): $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$, $\mu_2 = 4$ and Case 2 (right): $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$, $\mu_2 = 5$. The solid line represents γ and the dotted line is $\lambda_1 + \lambda_2 - \mu_1 x_1 - \mu_2 x_2 = 0$. The region R_1 is defined by $R_1 = \{(x_1, x_2) : x_1 + x_2 < 1\}$.

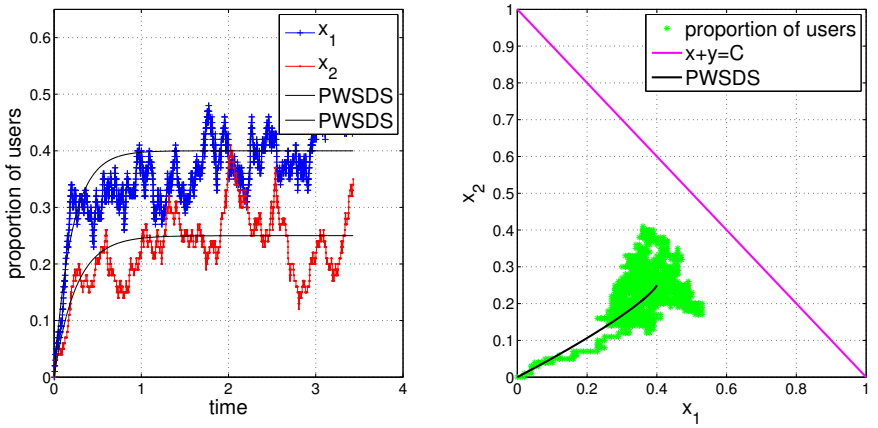


Fig. 4. Case 1 with parameters: $N = 100$, $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$ and $\mu_2 = 4$. The PWSDS fixed point is $(x_1^*, x_2^*) = (\lambda_1/\mu_1, \lambda_2/\mu_2) = (2/5, 1/4)$.

for large time values, the scaled number of users in the stochastic process is around the PWSDS fixed point (x_1^*, x_2^*).

Recalling the description of the system, if $x_1 + x_2 = C$ and a PU arrives, a SU will be immediately deallocated giving the subchannel to the new PU. In this case, the QoS perceived by the SU will be affected because of the interruption of its communication. We are interested in SUs whose service cannot be interrupted with high probability (like a phone call or other interactive services). With this in mind, we can relate the interruption probability with the probability that the process lives on γ. We can conclude that the PWSDS fixed point has to be far enough from γ to avoid a strong impact on secondary communications. However, it has to be as close as possible to γ to permit

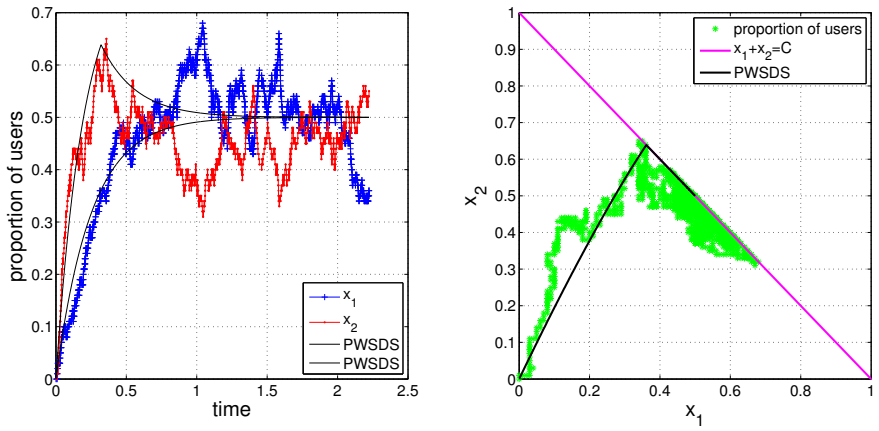


Fig. 5. Case 2 with parameters: $N = 100$, $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$. The PWSDS fixed point is $(x_1^*, x_2^*) = (\lambda_1/\mu_1, C - \lambda_1/\mu_1) = (1/2, 1/2)$.

more spectrum utilization and a high access probability for SUs. According to Proposition 3.5 and observing Figures 4 and 5 we can identify two cases: when the PWSDS fixed point is in R_1 and when it is on γ. For the last one, the system in stationary regime works near γ, so the probability of service interruption is too large. Then, the question in this case is how can we move the fixed-point? The analysis in the subsections 3.3 and 3.4 will be concentrated on answering that question. In particular we move the fixed point using admission control decisions. On the other hand, in the next two subsections we concentrate in cases like Case 1 (when the PWSDS fixed point is in R_1). More specifically, we concentrate our efforts on answering the question: is the fixed point far enough from γ to assure a small interruption probability?

We study the interruption probability by means of the probability that the system is full $P(X_1 + X_2 = C)$. Both probabilities are highly related by

$$P(\text{deallocate a SU}) + P(\text{block a PU}) = P(X_1 + X_2 = C|\text{PU arrives}),$$

by PASTA property (Poisson Arrivals See Time Averages) the last probability is $P(X_1 + X_2 = C)$ and $P(\text{block a PU})$ is the blocking probability for a $M/M/C/C$ queue.

3.2.1 Gaussian asymptotic distribution in sub-critical cases. Another issue studied in the context of fluid limits is the velocity of this convergence by looking at the fluctuations of the process around the limit. We refer to [13, 18, 33, 36] for the analysis of different scaling regimes and limit theorems in that sense, concerning different kinds of limit distributions.

If we consider cases when $\lambda_1/\mu_1 + \lambda_2/\mu_2 < C$ and the PWSDS trajectory remains all the time in R_1, it is possible to apply known results (see Theorem 2.3 of Chapter 11 in [13]) in order to obtain the asymptotic distribution. Let (x_1, x_2) be the trajectory of the PWSDS with initial condition $(x_1(0), x_2(0))$. If:

$$\lim_{N\to+\infty} \sqrt{N}[(X_1^N(0), X_2^N(0)) - (x_1(0), x_2(0))] = \chi(0)$$

with $\chi(0)$ deterministic, then:

$$\sqrt{N}[(X_1^N, X_2^N) - (x_1, x_2)] \Rightarrow \chi,$$

where $(\chi(t))$ is a two-dimensional Gaussian process and $\Rightarrow$ means convergence in distribution. $(\chi(t))$ has a covariance matrix determined explicitly by:

$$Cov(\chi(t), \chi(r)) = \int_0^{t\wedge r} e^{A(t-s)} G(x_1(s), x_2(s)) e^{A(r-s)} ds$$

where $A = \begin{pmatrix} -\mu_1 & 0 \\ 0 & -\mu_2 \end{pmatrix}$ and $G(x_1(s), x_2(s)) = \begin{pmatrix} \lambda_1 + \mu_1 x_1(s) & 0 \\ 0 & \lambda_2 + \mu_2 x_2(s) \end{pmatrix}$.

In sub-critical cases, we can conclude that $\lim_{N\to+\infty} P(X_1^N(t) + X_2^N(t) = C) = 0$ for all t. However, considering the defined Gaussian process and a finite large N we can present a practical criterion to analyze if the PWSDS fixed point is far enough from γ. In particular, we can obtain confidence bounds and also infer an adequate number of subchannels in order to avoid a high interruption probability for SUs.

3.2.2 Practical QoS design criterion in sub-critical cases. As we have shown in the simulated examples of Figures 4 and 5, the fluid limit is an accurate approximation when N is large. Then, in practice, a possible criterion to determine whether the PWSDS fixed point is far enough from γ would be to consider a certain confidence region of $(X_1^N(t), X_2^N(t))$ assuming a large value of t. If the resulting confidence ellipse is entirely inside R_1, high probability of non-interruption is guaranteed. Otherwise, we should try to move the fixed point. In particular we consider $\lim_{t\to+\infty} G(x_1(s), x_2(s)) = G(x_1^*, x_2^*)$ and define this limit matrix as $G(\infty) = G(x_1^*, x_2^*)$. Then, following the development in [29] we can obtain the covariance matrix $\Sigma(\infty)$ by solving:

$$A\Sigma(\infty) + \Sigma(\infty)A^T = -G(\infty). \tag{7}$$

Considering a fixed relation between both classes $\frac{\lambda_1/\mu_1}{\lambda_2/\mu_2} = constant$, using the deterministic confidence ellipse we can infer which is the ideal scaling parameter. In other words, we can obtain an idea of the optimal number of resources (subchannels) necessary to guarantee a small interruption probability for SUs. In Figure 6 we show the theoretical 95% confidence ellipses considering different values of N for two different parameter sets (Case A and Case B). In particular, we have considered $C = 1$, then N represents the number of channels of the system $(\widetilde{X}_1^N(t), \widetilde{X}_2^N(t))$. For Case A, (with $C = 1$, $\lambda_1 = 9$, $\lambda_2 = 13$, $\mu_1 = 50$ and $\mu_2 = 20$) we have that the ellipse is tangent to γ when $N = 180$. Therefore, we can conclude that an admission control does not make sense in the system $(\widetilde{X}_1^N(t), \widetilde{X}_2^N(t))$ when $N \geq 180$. In Table 2 we confirm that the interruption probability of SUs can be analyzed studying the probability that the system is full of users (blocking probability of PUs can be approximated by 0). For Case B (with $C = 1$, $\lambda_1 = 10$, $\lambda_2 = 12$, $\mu_1 = 25$ and $\mu_2 = 30$), we have an analogous conclusion when $N \geq 120$.

Table 2. Values of full system probability and blocking PU probability for Case A of Figure 6. $P(\widetilde{X}_1^N(t) + \widetilde{X}_2^N(t) > C)$ is approximated by a Gaussian distribution.

N	100	110	120	130	140	150
$P(\widetilde{X}_1^N(t) + \widetilde{X}_2^N(t) > C)$	0.031	0.0252	0.0205	0.0167	0.0136	0.0111
P(block a PU)	0.5×10^{-40}	0.7×10^{-45}	0.8×10^{-48}	0.1×10^{-53}	0.1×10^{-55}	0

3.3 Deterministic admission control policy

In this subsection we analyze the system when a deterministic admission control is applied. In this situation we consider $\gamma' = \{(x_1, x_2) : x_1 + x_2 = \delta\}$ as the admission control border, then the question is: what is a reasonable value of δ to guarantee certain level of QoS to SUs?

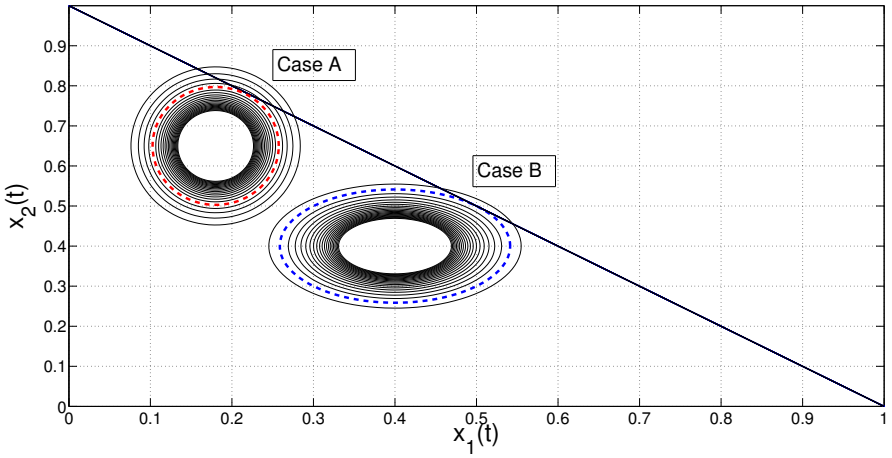


Fig. 6. Theoretical 95% confidence ellipses of $(X_1^N(t), X_2^N(t))$, considering different N values ($N = 100, 120, 140, 160, \ldots, 500$). Parameters of Case A: $C = 1$, $\lambda_1 = 9$, $\lambda_2 = 13$, $\mu_1 = 50$ and $\mu_2 = 20$. Parameters of Case B: $C = 1$, $\lambda_1 = 10$, $\lambda_2 = 12$, $\mu_1 = 25$ and $\mu_2 = 30$.

In this case the fluid limit is obtained in the same way as in Subsection 3.2, but the sliding motion occurs in γ'.

Proposition 3.6. *Let $Q_1(x_1, x_2)$ and $Q_2(x_1, x_2)$ be vector fields, both smooth in $R_1 = \{(x_1, x_2) : x_1 + x_2 - \delta < 0\}$ and $R_2 = \{(x_1, x_2) : x_1 + x_2 - \delta > 0\}$ respectively such that*

$$Q_1(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ \lambda_2 - \mu_2 x_2 \end{pmatrix}, \qquad Q_2(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ -\mu_2 x_2 \end{pmatrix},$$

and let $n(x_1, x_2)$ be the normal vector to the boundary γ' ($n^T(x_1, x_2) = (1, 1)$ for all $(x_1, x_2) \in \gamma'$) then we have a PWSDS driven by the following equations.

If $x_1 + x_2 < \delta$:

$$\begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= \lambda_2 - \mu_2 x_2. \end{cases},$$

else, if $x_1 + x_2 - \delta = 0$ (γ'):

$$\begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= -\lambda_1 + \mu_1 x_1. \end{cases},$$

and if $x_1 + x_2 > \delta$:

$$\begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= -\mu_2 x_2. \end{cases}.$$

Notice that we consider the case $\frac{\lambda_1}{\mu_1} < C$ because its practical importance.

The behavior is very similar as in the free admission control case, and we can think in this case as a sort of translation of the free case where the border that presents sliding motion is shifted from γ to γ'. Then, the proof is totally analogous to Proposition 3.3.

A convergence theorem analogous to Theorem 3.4 holds, in this case using directly Theorem IV.2 in [7], as for the deterministic policy there are two regions $R_1' = \{(x_1, x_2) : x_1 + x_2 < \delta\}$, $R_2' = \{(x_1, x_2) : x_1 + x_2 > \delta\}$, and the border $\gamma' = \{(x_1, x_2) : x_1 + x_2 = \delta\}$, where the drift is discontinuous.

In addition we can study fixed points, by an analysis similar to Proposition 3.5. However, the fixed point of the PWSDS in the most interesting cases lies in γ'. This fact hinders the development of a design criterion like in Subsection 3.2.2, as we do not have the same results for the asymptotic distribution.

In the deterministic policy scheme the hypotheses of Theorem 2.3 of Chapter 11 in [13] are not verified, due to the non-smoothness of the PWSDS at the border γ'.

We have convergence of the stationary regime to the PWSDS fixed point but there is not a general framework that allows to state a Gaussian asymptotic distribution. Therefore, we proceed to compute the asymptotic stationary distribution of the total number of users in the particular case when service rates are the same for primary and secondary users.

The limitation to a particular case comes from our proof method that depends on the explicit stationary distribution, that can only be explicitly obtained for equal service rates. However, we will give some insight on the asymptotic distribution in this case, for different service rates, based on simulations. Further study of the general case is object of future work.

Using the limit of the invariant distribution for the particular case, we can build some design criteria in order to find an estimation of the optimal δ value for certain maximum level of interruption probability. We will show that the criteria can be extended to the case $\mu_1 \neq \mu_2$ with good results in 3.3.2.

3.3.1 Non-Gaussian invariant distribution. We consider the original system (when $\mu_1 = \mu_2$) and compute its invariant distribution, both for the original system and the scaled one. Then for the scaled system that depends on N we find the limit of the invariant distribution when N goes to infinity. For practical purposes as in this case, the invariant distribution can be computed explicitly. There is no need to compute the asymptotic distribution, but theoretically we want to show that the asymptotic distribution in the case where the equilibrium lies in the admission control border is not Gaussian; with a different scaling, we find a geometric asymptotic distribution. The proof is only in the particular case when service rates are the same for primary and secondary users, that is when the problem is one-dimensional. The scaling, different from $\sqrt{N}$, and the geometric distribution are not frequent in related works.

Consider the Markov chain defined in Section 2 with $\mu_1 = \mu_2$ where the access control is defined considering γ'. In this case $X = X_1 + X_2$ is a one dimensional Markov chain with state space $E = \{0, 1, \ldots, C\}$, and non-zero transition rates from i to j, $q(i, j)$, given by:

$$q(i, i+1) = \begin{cases} \lambda_1 + \lambda_2 & \text{for } 0 \leq i < \delta \\ \lambda_1 & \text{for } \delta \leq i < C \end{cases}, \qquad q(i, i-1) = i\mu \text{ for } 0 < i \leq C,$$

where $\delta \in E$ denotes the border of the admission control, that is when we have δ or more users we prohibit the access of new secondary users. Let us observe that in the general case described in Section 2 the total number of users $X = X_1 + X_2$ is Markovian only if $\mu_1 = \mu_2$, so the reduction to a one dimensional Markov chain follows only in this case. In order to simplify notation in what follows let us call $\nu_2 = \lambda_1 + \lambda_2$, and $\nu_1 = \lambda_1$. We then have the following transition rates, where $\nu_1 < \nu_2$:

$$q(i, i+1) = \begin{cases} \nu_2 & \text{for } 0 \leq i < \delta \\ \nu_1 & \text{for } \delta \leq i < C \end{cases}, \qquad q(i, i-1) = i\mu \text{ for } 0 < i \leq C,$$

We consider the scaled process $\widetilde{X}^N = \widetilde{X}_1^N + \widetilde{X}_2^N$ with the scaling in Table 3. Theorem 3.4 holds in this case and we also have that $X^N = \widetilde{X}^N/N$ converges in probability, uniformly over compact time intervals to x, given by the following equations, for $0 \leq \delta \leq C$.

Table 3. Scaling for the one dimensional processes (X) and $(\widetilde{X}^N)$.

(X)	ν_1	ν_2	C	δ
$(\widetilde{X}^N)$	$\nu_1 N$	$\nu_2 N$	CN	δN

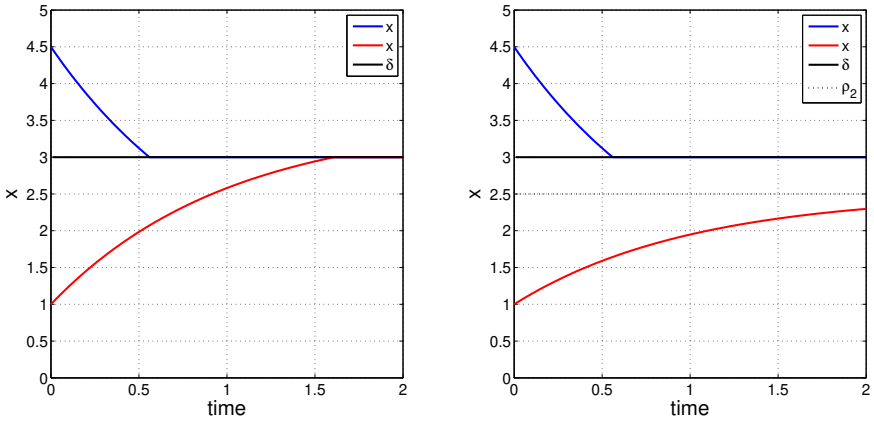


Fig. 7. PWSDS solution in the one dimensional case for two different initial conditions. Parameters: $C = 5$, $\delta = 3$, $\rho_1 = 1$, $\rho_2 = 3.5$ (left) and $C = 5$, $\delta = 3$, $\rho_1 = 1$, $\rho_2 = 2.5$ (right).

If the initial condition is $x(0) < \delta$:

$$x(t) = \begin{cases} \rho_2 + (x(0) - \rho_2)e^{-\mu t} & \text{if } t < \tau_2 \\ \delta & \text{if } t \geq \tau_2. \end{cases}$$

If the initial condition is $x(0) > \delta$:

$$x(t) = \begin{cases} \rho_1 + (x(0) - \rho_1)e^{-\mu t} & \text{if } t < \tau_1 \\ \delta & \text{if } t \geq \tau_1, \end{cases}$$

where $\rho_1 = \dfrac{\nu_1}{\mu}$, $\rho_2 = \dfrac{\nu_2}{\mu}$, $\tau_1 = \dfrac{1}{\mu}\log\left(\dfrac{x(0)-\rho_1}{\delta-\rho_1}\right)$ and $\tau_2 = \dfrac{1}{\mu}\log\left(\dfrac{x(0)-\rho_2}{\delta-\rho_2}\right)$. (See Figure 7 for different behaviors of these solutions.)

Now we come back to the original system. As the problem now is one-dimensional we will explicitly compute the stationary distribution of X, and then for the scaled system $\widetilde{X}^N$ for fixed N and obtain its limit when $N \to \infty$. X is a modification of a $M/M/C/C$ queue, where the arrivals changes their rate depending on the number of clients in the queue. We will analyze our system as in [33] for the $M/M/C/C$ queue. Computing the stationary distribution π for this Markov chain:

$$\begin{aligned} \pi(i)\nu_2 &= \pi(i+1)(i+1)\mu \text{ for } 0 \leq i < \delta, \\ \pi(i)\nu_1 &= \pi(i+1)(i+1)\mu \text{ for } \delta \leq i < C. \end{aligned}$$

Then,

$$\pi(i) = \frac{\rho_2^i}{i!}\pi(0) \quad \text{for} \quad 0 \le i \le \delta,$$
$$\pi(i) = \left(\frac{\rho_2}{\rho_1}\right)^{\delta} \frac{\rho_1^i}{i!}\pi(0) \quad \text{for} \quad \delta < i \le C,$$

where:

$$\pi(0)^{-1} = \sum_{i=0}^{\delta} \frac{\rho_2^i}{i!} + \left(\frac{\rho_2}{\rho_1}\right)^{\delta} \sum_{i=\delta+1}^{C} \frac{\rho_1^i}{i!}.$$

The stationary distribution $\widetilde{\pi}^N$ for the process $\widetilde{X}^N$ is:

$$\widetilde{\pi}^N(i) = \frac{(N\rho_2)^i}{i!}\widetilde{\pi}^N(0) \quad \text{for} \quad 0 \le i \le N\delta,$$
$$\widetilde{\pi}^N(i) = \left(\frac{\rho_2}{\rho_1}\right)^{N\delta} \frac{(N\rho_1)^i}{i!}\widetilde{\pi}^N(0) \quad \text{for} \quad N\delta < i \le C,$$

with:

$$\widetilde{\pi}^N(0)^{-1} = \sum_{i=0}^{N\delta} \frac{(N\rho_2)^i}{i!} + \left(\frac{\rho_2}{\rho_1}\right)^{N\delta} \sum_{i=N\delta+1}^{NC} \frac{(N\rho_1)^i}{i!}.$$

Theorem 3.7. *Consider the original processes with equal service rates $\mu_1 = \mu_2$ and X^N defined as before. Then, the stationary distribution of $\widetilde{X}^N - N\delta$ converges to the distribution of an integer variable Z given by*

$$P(Z = j) = \begin{cases} \rho \left(\dfrac{\rho_2}{\delta}\right)^j & \text{if } j < 0, \\ \rho \left(\dfrac{\rho_1}{\delta}\right)^j & \text{if } j \ge 0, \end{cases}$$

where $\rho_1 = \dfrac{\nu_1}{\mu}$, $\rho_2 = \dfrac{\nu_2}{\mu}$, $\rho = \left(\dfrac{\rho_1}{\delta - \rho_1} + \dfrac{\rho_2}{\rho_2 - \delta}\right)^{-1}$.

Proof. Let us compute, for $0 \le k < N\delta$ the stationary distribution $\widetilde{\pi}^N(N\delta - k)$. Computing its inverse:

$$\begin{aligned}
\widetilde{\pi}^N(N\delta - k)^{-1} &= \frac{(N\delta - k)!}{(N\rho_2)^{N\delta - k}}\widetilde{\pi}^N(0)^{-1} \\
&= \frac{(N\delta - k)!}{(N\rho_2)^{N\delta - k}} \left(\sum_{i=0}^{N\delta} \frac{(N\rho_2)^i}{i!} + \left(\frac{\rho_2}{\rho_1}\right)^{N\delta} \sum_{i=N\delta+1}^{NC} \frac{(N\rho_1)^i}{i!} \right) \\
&= \frac{(N\delta - k)!}{(N\rho_2)^{N\delta - k}} \left(\sum_{j=0}^{N\delta} \frac{(N\rho_2)^{N\delta - j}}{(N\delta - j)!} + \left(\frac{\rho_2}{\rho_1}\right)^{N\delta} \sum_{i=N\delta+1}^{NC} \frac{(N\rho_1)^i}{i!} \right) \\
&= \sum_{j=0}^{N\delta} \frac{(N\rho_2)^k (N\delta - k)!}{(N\rho_2)^j (N\delta - j)!} + \sum_{j=1}^{N(C-\delta)} \frac{(N\rho_1)^j (N\rho_2)^k (N\delta - k)!}{(N\delta + j)!} \\
&= \rho_2^k \sum_{j=0}^{N\delta} \rho_2^{-j} \frac{N^k (N\delta - k)!}{N^j (N\delta - j)!} + \sum_{j=1}^{N(C-\delta)} \frac{(N\rho_1)^j (N\rho_2)^k (N\delta - k)!}{(N\delta + j)!}.
\end{aligned}$$

Using Stirling's formula we obtain:

$$\lim_{N\to\infty} \frac{N^k(N\delta-k)!}{N^j(N\delta-j)!} = \frac{\delta^j}{\delta^k}$$

and then for the first term, using dominated convergence we have:

$$\lim_{N\to\infty} \rho_2^k \sum_{j=0}^{N\delta} \rho_2^{-j} \frac{N^k(N\delta-k)!}{N^j(N\delta-j)!} = \left(\frac{\rho_2}{\delta}\right)^k \frac{1}{1-\dfrac{\delta}{\rho_2}}$$

For the second term we have using Stirling's formula that:

$$\lim_{N\to\infty} \frac{(N\rho_1)^j(N\rho_2)^k(N\delta-k)!}{(N\delta+j)!} = \left(\frac{\rho_1}{\delta}\right)^j \left(\frac{\rho_2}{\delta}\right)^k,$$

then, with dominated convergence:

$$\lim_{N\to\infty} \sum_{j=1}^{N(C-\delta)} \frac{(N\rho_1)^j(N\rho_2)^k(N\delta-k)!}{(N\delta+j)!} = \frac{\rho_1}{\delta-\rho_1}\left(\frac{\rho_2}{\delta}\right)^k$$

and, for $0 \le k < N\delta$, we obtain:

$$\lim_{N\to\infty} \widetilde{\pi}^N(N\delta-k) = \rho\left(\frac{\delta}{\rho_2}\right)^k.$$

If $\widetilde{Z}^N = \widetilde{X}^N - N\delta$ we have that the stationary distribution $\widetilde{\mu}^N$ of $\widetilde{Z}^N$ verifies, for $j < 0$, that:

$$\lim_{N\to\infty} \widetilde{\mu}^N(j) = \lim_{N\to\infty} \widetilde{\pi}^N(N\delta-(-j)) = \rho\left(\frac{\rho_2}{\delta}\right)^j.$$

In the same way we compute, for $0 \le k \le NC - N\delta$, $\widetilde{\pi}^N(N\delta+k)$. □

In Figure 8 we show the limit distribution and the stationary distribution for different values of N. We see that the limit distribution represents a good estimation of the stationary one for all N values.

3.3.2 Practical QoS criterion for δ selection. Using Theorem 3.7 we can improve the average utilization of the spectrum while ensuring a small probability of interruption to the secondary users. Please note that the result is valid for $\mu_1 = \mu_2$. For the general case, we cannot obtain an analytical expression of the distribution but we can estimate δ considering an equivalent system. If μ_1 and μ_2 are of the same order we can assume $\mu = \min\{\mu_1, \mu_2\}$ (the worst case). Another approach considers a μ−scaled system ($\mu = \mu_1$, $\lambda_1 = \lambda_1$ and $\lambda_2 = \lambda_2\mu/\mu_2$) [35]. Then, with the equivalent system, in the same way as in the Gaussian approximation, our proposal consists in relating the value of δ with the probability that the process lives on γ, then

$$P(\widetilde{X}_1^N + \widetilde{X}_2^N = CN) = P(\widetilde{X}^N = CN) = P(\widetilde{X}^N - N\delta = CN - N\delta).$$

According to Theorem 3.7 we can assume that:

$$P(\widetilde{X}_1^N + \widetilde{X}_2^N = CN) \approx \rho\left(\frac{\rho_1}{\delta}\right)^{N(C-\delta)} \tag{8}$$

for large values of N.

Assuming a threshold on the probability of interruption which can be represented by $P(\widetilde{X}_1^N + \widetilde{X}_2^N = CN) \le \epsilon$ for certain ϵ and considering NC as the total number of subchannels in the system, using equation (8) it is possible to obtain an upper bound of $\hat{\delta} = \delta(N, C, \epsilon)$.

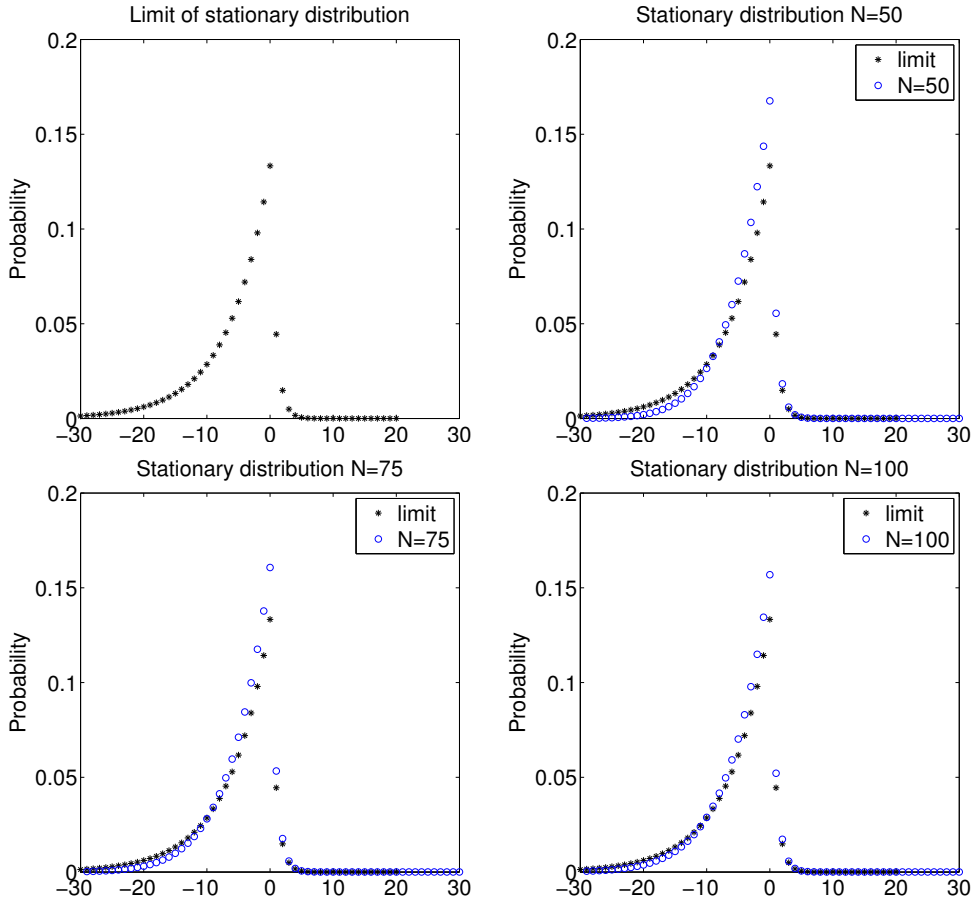


Fig. 8. Limit ($N \to \infty$) of stationary distribution and stationary distributions for different N values. Parameters: $C = 5$, $\delta = 3$, $\rho_1 = 1$, $\rho_2 = 3.5$.

As an example we made different sets of simulations changing the value of N considering the parameters $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$. Figure 9 shows the probability that the system is full for different values of N. In this case we have used $\mu = 4$ (the most critical case). Considering $\epsilon = 0.1$ we can conclude that if we need $P(\widetilde{X}_1^N + \widetilde{X}_2^N = CN) \leq 0.1$, we obtain an upper bound of δ ($\hat{\delta}$) depending on N (see Table 4).

Table 4. Upper bounds to δ in order to ensure $P(\widetilde{X}_1^N + \widetilde{X}_2^N = CN) \leq 0.1$.

N	100	80	50	20	10
$\hat{\delta}$	0.986	0.982	0.951	0.920	0.791
P(block a PU)	0.0163×10^{-8}	0.8675×10^{-8}	0.3602×10^{-5}	0.0019	0.0184

3.3.3 Simulations for the general case with different service rates. The asymptotic distribution obtained in Theorem 3.7 has the drawback that assumes equal services rates for PUs and SUs. This

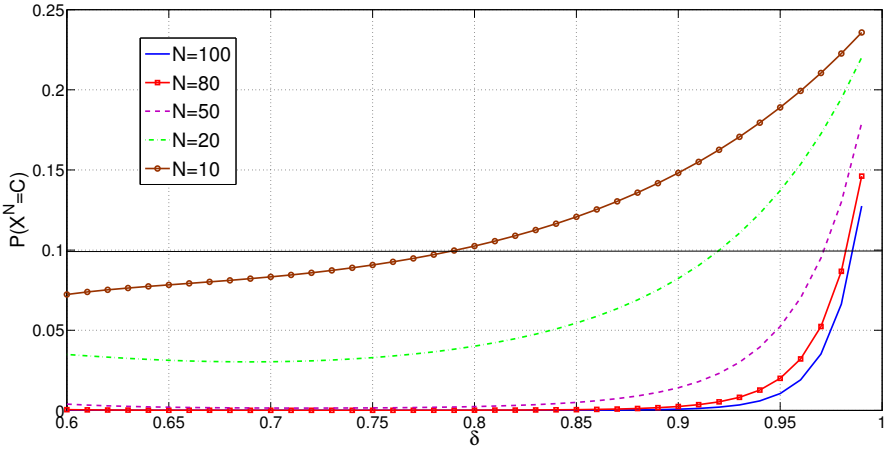


Fig. 9. Interruption probability vs δ. Parameters: $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$.

hypothesis is needed because of the technique that we used to prove Theorem 3.7. Our proof, based on [33], relies on the explicit computation of the stationary distribution for the original system and for the scaled one, and in the calculus of the limit when the scaling factor N goes to infinity. This approach is not suitable for different service rates as there is not a closed formula for the stationary distribution. Then, a different approach would be necessary to obtain a non-Gaussian asymptotic distribution in the general case.

Despite of this restriction, in 3.3.2 we have applied our results about geometric distribution to the case of different services rates in order to design an admission control policy. In addition, for different service rates, when the fixed point is in the admission control border, the asymptotic distribution is non-Gaussian.

Here we analyze with simulations the case with different service rates. We consider a deterministic admission control where the fixed point (x_1^*, x_2^*) of the PWSDS is at the admission control border γ'. In this case we cannot obtain a Gaussian distribution as it is obtained when the fixed point is in the interior the region $\{(x_1, x_2) : x_1 + x_2 < \delta\}$. In this last case it actually does not matter if there is a deterministic admission control at $x_1 + x_2 = \delta$ or we consider the free admission control policy. In both cases, if the fixed point is in the interior the behavior is similar. On the other hand, when the fixed point lies in γ', fluctuations around the fixed point are asymmetric and much smaller than in the case of the interior fixed point, so that scaled by $\sqrt{N}$ as in the Gaussian case we would not obtain such limit.

We simulate two cases, with different parameters and the same deterministic admission control, one with the fixed point in the interior an the other at γ'. In both cases the starting point is the fixed point. For both cases we show the trajectories in Figure 10. The case at the right is more asymmetrical and shows less dispersion around the fixed point than the left one. In Figure 11 we show the kernel density estimations of the normalized fluctuations of $X_1^N + X_2^N$ around the mean system utilization $x_1^* + x_2^*$ and in Figure 12 the corresponding QQ-plots. In both figures it is more visible the asymmetry but specially that the distribution at the right is concentrated around zero, and cannot be approximated by a Gaussian.

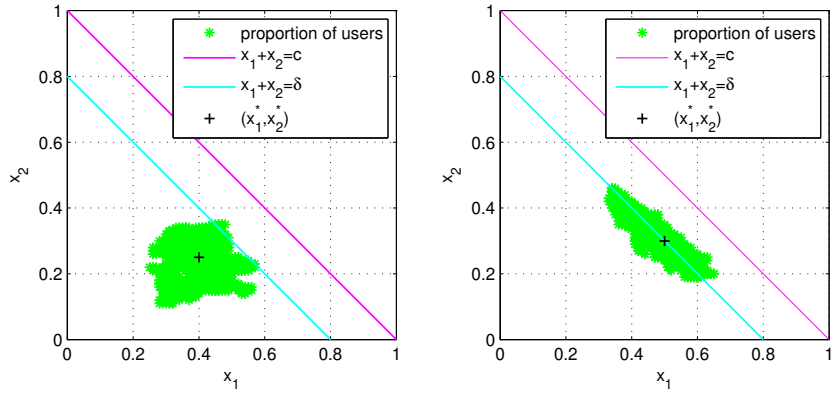


Fig. 10. Left: $N = 100$, $C = 1$, $\delta = 0.8$ $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$ and $\mu_2 = 4$. The PWSDS fixed point is $(2/5, 1/4)$. Right: $N = 100$, $C = 1$, $\delta = 0.8$ $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$. The PWSDS fixed point is $(1/2, 3/10)$.

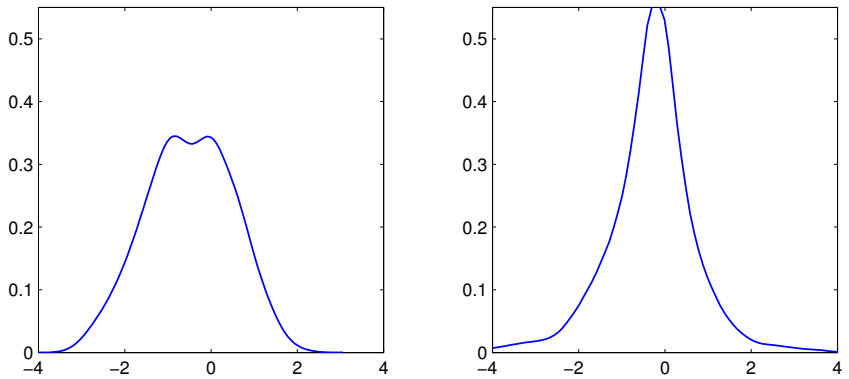


Fig. 11. Left: $N = 100$, $C = 1$, $\delta = 0.8$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$ and $\mu_2 = 4$, kernel density estimation for the normalized variable $X_1^N + X_2^N - 0.65$. Right: $N = 100$, $C = 1$, $\delta = 0.8$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$, kernel density estimation for the normalized variable $X_1^N + X_2^N - 0.8$.

3.4 Probabilistic admission control policy

Let us consider another class of admission control: a probabilistic admission mechanism where secondary users can access the system with a probability related to the number of users in the system. Let $a(x_1, x_2)$ be the probability that a secondary user that arrives enters the system when there are x_1 primary users and x_2 secondary users. Then, when computing the entry rates for the whole system, the arrival rates of secondary users appear multiplied by this probability. Let us assume that $a(x_1, x_2)$ is a Lipschitz function that vanishes close to the border $\{(x_1, x_2) : x_1 + x_2 = C\}$ (as an example we could choose $a(x_1, x_2) = 1 - (x_1 + x_2)/C$), that is:

$$\lim_{x_1+x_2 \to C} a(x_1, x_2) = 0.$$

In this case we also have discontinuous transition rates in the border of the state space. Following the same lines as for the free admission case in Subsection 3.2 Proposition 3.3 we have:

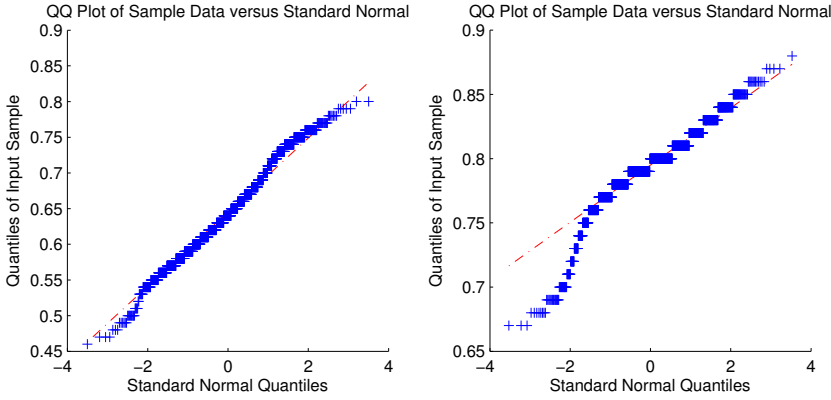


Fig. 12. Left: $N = 100$, $C = 1$, $\delta = 0.8$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$ and $\mu_2 = 4$. Right: $N = 100$, $C = 1$, $\delta = 0.8$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$.

Proposition 3.8. *Let $Q_1(x_1, x_2)$ and $Q_2(x_1, x_2)$ be vector fields, both smooth in $R_1 = \{(x_1, x_2) : x_1 + x_2 - C < 0\}$ and γ respectively such that*

$$Q_1(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ \lambda_2 a(x_1, x_2) - \mu_2 x_2 \end{pmatrix}, \qquad Q_2(x_1, x_2) = \begin{pmatrix} \lambda_1 - \mu_1 x_1 \\ -\lambda_1 - \mu_2 x_2 \end{pmatrix},$$

and let $n(x_1, x_2)$ be the normal vector to the boundary γ ($n^T(x_1, x_2) = (1, 1)$ for all $(x_1, x_2) \in \gamma$) then we have a PWSDS driven by the following equations.

If $x_1 + x_2 < C$ or $\lambda_1 \leq \mu_1 x_1 + \mu_2 x_2$:

$$\begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= \lambda_2 a(x_1, x_2) - \mu_2 x_2. \end{cases}$$

If $x_1 + x_2 = C$ and $\lambda_1 > \mu_1 x_1 + \mu_2 x_2$:

$$\begin{cases} x_1' &= \lambda_1 - \mu_1 x_1, \\ x_2' &= -\lambda_1 + \mu_1 x_1. \end{cases}$$

This system has a unique solution with initial condition $(x_1(0), x_2(0))$. Let:

$$(x_1(t), x_2(t)) \tag{9}$$

be the piecewise smooth dynamical system that is the solution to the previous equations with initial condition $(x_1(0), x_2(0))$. Theorem 3.4 holds in this context.

This probabilistic admission control is different from the deterministic one. The solution presents sliding motion when $\lambda_1/\mu_1 \geq C$ or when $\lambda_1/\mu_2 \geq C$. In the first case the system in stationary regime is always saturated by PUs, despite of the admission control. In the second case sliding motion depends on the initial condition but it does not influence the stationary regime.

Let us consider the fixed point (x_1^*, x_2^*) for the processes defined by (9). For $\theta_1 = \lambda_1/\mu_1 < C$ and $\theta_2 = \lambda_2/\mu_2$, we have $x_1^* = \theta_1$ and x_2^* verifies the equation $x_2^* = \theta_2 a(\theta_1, x_2^*)$. Therefore, x_2^* is unique if we assume that the equation $\theta_2 a(\theta_1, x) - x = 0$ has an unique solution. In the example, with $a(x_1, x_2) = 1 - x_1 + x_2/C$, x_2^* can be obtained explicitly.

Proposition 3.9. *Considering $a(x_1, x_2)$ continuous with $\lim_{x_1+x_2 \to C} a(x_1, x_2) = 0$ and defining (x_1^*, x_2^*) as the PWSDS fixed point, then:*

a. *If* $\theta_1 < C$, *then* $x_1^* = \frac{\lambda_1}{\mu_1}$ *and* $x_2^* = \theta_2 a(\theta_1, x_2^*)$ *where* $(x_1^*, x_2^*) \in R_1$ *and the mean system utilization will be* $\theta_1 + \theta_2 a(\theta_1, x_2^*)$.
b. *If* $\theta_1 \geq C$, *then* $(x_1^*, x_2^*) = (C, 0)$ *and the mean system utilization will be* C.

Sketch of the proof. Results follows from similar arguments than in Proposition 3.5. In addition we could assume hypotheses about the probabilistic admission control function $a(x_1, x_2)$ that ensures asymptotic stability of the solutions (for example negative real part of eigenvalues for the linearized system). This condition follows in the example $a(x_1, x_2) = 1 - x_1 + x_2/C$, where we have a linear ODE. (See Figure 13 in order to analyze the vector field behavior in two different cases.) □

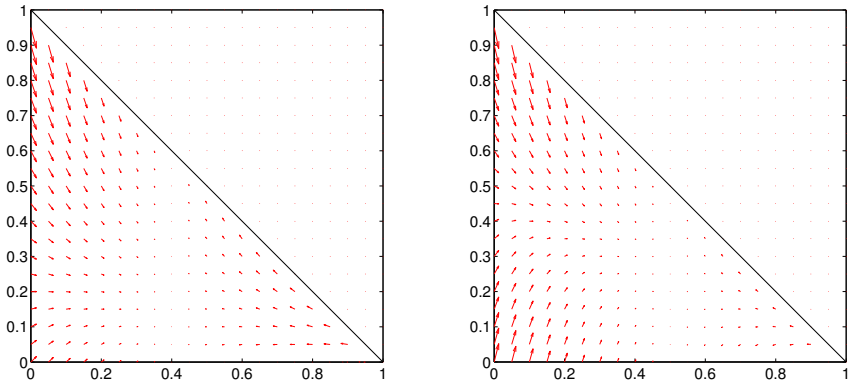


Fig. 13. Vector field for Case 1 (left): $N = 100$, $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\mu_1 = 5$, $\mu_2 = 4$ and Case 2 (right): $N = 100$, $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$, $\mu_2 = 5$. The continuous line represents γ.

Remark. *For case a. of Proposition 3.9, for initial conditions where there is not sliding motion, it is possible to apply the same results about Gaussian limits. Let* (x_1, x_2) *be the trajectory of the solution with initial condition* $(x_1(0), x_2(0))$. *If:*

$$\lim_{N \to +\infty} \sqrt{N}[(X_1^N(0), X_2^N(0)) - (x_1(0), x_2(0))] = \chi(0)$$

with $\chi(0)$ *deterministic, then* $\sqrt{N}[(X_1^N, X_2^N) - (x_1, x_2)] \Rightarrow \chi$, *where* $(\chi(t))$ *is a Gaussian process whose matrix can be determined explicitly by:*

$$Cov(\chi(t), \chi(r)) = \int_0^{t \wedge r} e^{A(t-s)} G(x_1(s), x_2(s)) e^{A(r-s)} ds$$

where

$$A = \begin{pmatrix} -\mu_1 & 0 \\ 0 & -\mu_2 \end{pmatrix},$$

$$G(x_1(s), x_2(s)) = \begin{pmatrix} \lambda_1 + \mu_1 x_1(s) & 0 \\ 0 & \lambda_2 a(x_1(s), x_2(s)) + \mu_2 x_2(s) \end{pmatrix}.$$

In particular in Figure 14 we show the simulation of one trajectory of the scaled Markov process and the trajectory of the ODE. In the same way as in Section 3.2, considering a large finite value of N we can calculate the covariance matrix $\Sigma(\infty)$ solving Equation (7). For instance in Figure 15 we illustrate different confidence ellipses (for different N values) for the case of Figure 14.

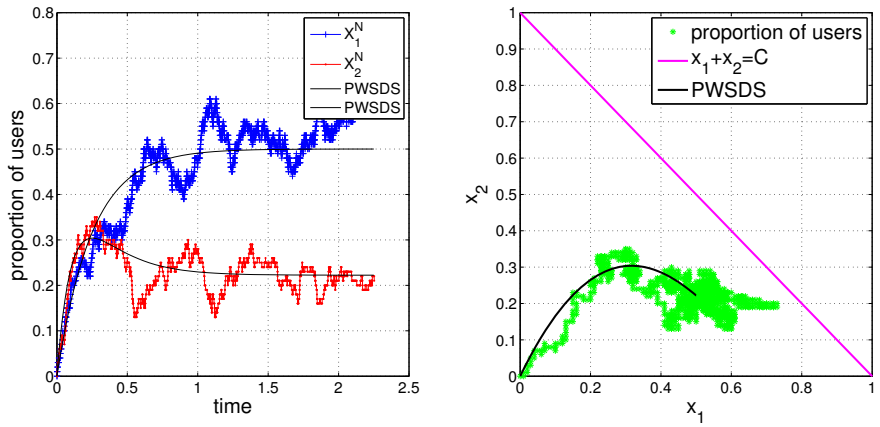


Fig. 14. Probabilistic control ($a(x_1, x_2) = 1 - x_1 - x_2$), parameters: $N = 100$, $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$. The PWSDS fixed point, with $\theta_1 = \frac{\lambda_1}{\mu_1}$ and $\theta_2 = \frac{\lambda_2}{\mu_2}$, is $(x_1^*, x_2^*) = \left(\theta_1, \theta_2 \left(\frac{C-\theta_1}{C+\theta_2}\right)\right) = (\frac{1}{2}, \frac{2}{9})$.

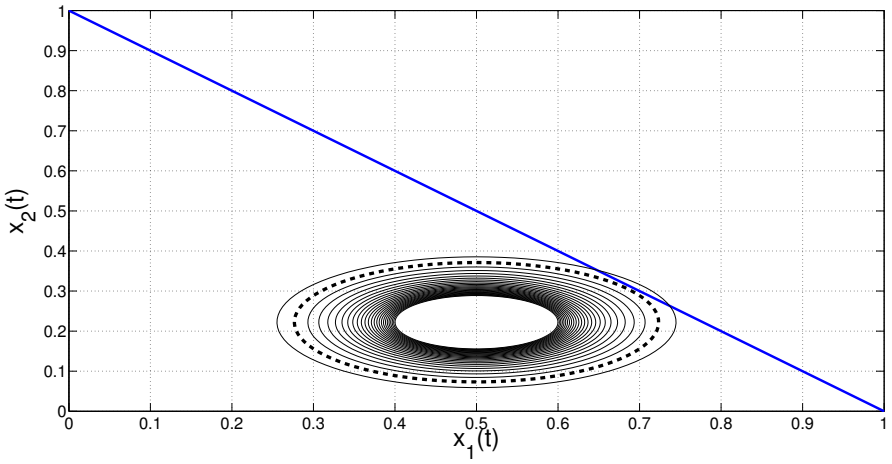


Fig. 15. Theoretical 95% confidence ellipses of $(X_1^N(t), X_2^N(t))$, considering different N values ($N = 50, 60, 70, \ldots, 300$). Parameters: $C = 1$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\mu_1 = 4$ and $\mu_2 = 5$.

Remark. *For case b. of Proposition 3.9, where there is sliding motion, it is possible to apply a similar approach as in Theorem 3.7.*

4 CONCLUSIONS AND FUTURE WORK

The main contributions of this work are the analysis and characterization of a possible model of spectrum sharing in cognitive radio networks that improve the average utilization of the spectrum while ensuring a small probability of interruption to the secondary users.

We considered a continuous time Markov chain that represents the population of the different types of users in the system. We formulated the associated fluid model and we studied its solutions. We studied the behavior of the system considering deterministic and probabilistic admission control

mechanisms by means of the study of fixed points and the related asymptotic distributions. For the deterministic case we obtained non-Gaussian and Gaussian asymptotic distributions being the non-Gaussian result one of the main contributions of the paper. It is an input for future work our preliminary results about asymptotic distribution in the non-Gaussian case. We aim to find theoretical results in the general case with different service rates.

For systems with a large number of users we presented practical design criteria to guarantee with high probability that secondary users in the system will not have service interruptions. These design criteria were suggested by a theoretical analysis and supported by simulations.

ACKNOWLEDGMENTS

This work was financially supported by CSIC (Grupo I+D Artes ID 129, Proyecto I+D "Límites fluidos, aproximación por difusión y grandes desvíos en sistemas de comunicación de gran dimensión.") and STIC-AmSud 2018 ("Algorithms for the Capacity Crunch problem in Optical Networks", ACCON).

REFERENCES

[1] Fayaz Akhtar, Mubashir Husain Rehmani, and Martin Reisslein. 2016. White space: Definitional perspectives and their role in exploiting spectrum opportunities. *Telecommunications Policy* 40, 4 (2016), 319 – 331. https://doi.org/10.1016/j.telpol.2016.01.003

[2] S. A. AlQahtani and H. Ahmed. 2014. An admission control scheme for secondary users in cognitive radio networks. In *Proceedings of the 11th IEEE International Conference on Networking, Sensing and Control*. 245–250. https://doi.org/10.1109/ICNSC.2014.6819633

[3] F. Baccelli, F. Mathieu, I. Norros, and R. Varloot. 2013. Can P2P networks be super-scalable?. In *2013 Proceedings IEEE INFOCOM*. 1753–1761. https://doi.org/10.1109/INFCOM.2013.6566973

[4] Michel Benaïm and Jean-Yves Le Boudec. 2008. A class of mean field interaction models for computer and communication systems. *Perform. Eval.* 65, 11-12 (2008), 823–838.

[5] Paola Bermolen, Matthieu Jonckheere, Federico Larroca, and Pascal Moyal. 2016. Estimating the transmission probability in wireless networks with configuration models. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems* 1, 2 (apr 2016), 1–23. https://iie.fing.edu.uy/publicaciones/2016/BJLM16

[6] V. Blaschke, H. Jaekel, T. Renk, C. Kloeck, and F.K. Jondral. 2007. Occupation Measurements Supporting Dynamic Spectrum Allocation for Cognitive Radio Design. In *Cognitive Radio Oriented Wireless Networks and Communications, 2007. CrownCom 2007. 2nd International Conference on*. 50–57. https://doi.org/10.1109/CROWNCOM.2007.4549771

[7] Luca Bortolussi. 2011. Hybrid Limits of Continuous Time Markov Chains. In *Proceedings of Eighth International Conference on Quantitative Evaluation of Systems (QEST), 2011*. 3 –12. https://doi.org/10.1109/QEST.2011.10

[8] Luca Bortolussi. 2011. Supplementary material of Hybrid Limits of Continuous Time Markov Chains. www.dmi.units.it/~bortolu/files/qest2011supp.pdf [Online; accesed 22-Feb-2019].

[9] Luca Bortolussi. 2016. Hybrid behaviour of Markov population models. *Information and Computation* 247, Supplement C (2016), 37 – 86. https://doi.org/10.1016/j.ic.2015.12.001

[10] H. Chen and D.D. Yao. 2001. *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*. Springer.

[11] Cisco. 2017. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*. Technical Report. Cisco.

[12] R. W. R. Darling and J. R. Norris. 2008. Differential equation approximations for Markov chains. *Probab. Surv.* 5 (2008), 37–79.

[13] Stewart N. Ethier and Thomas G. Kurtz. 1986. *Markov processes: Characterization and convergence*. John Wiley & Sons Inc., New York. x+534 pages.

[14] A. Ferragut and F. Paganini. 2016. Fluid Models of Population and Download Progress in P2P Networks. *IEEE Transactions on Control of Network Systems* 3, 1 (March 2016), 34–45. https://doi.org/10.1109/TCNS.2015.2434092

[15] Aleksej Fedorovic Filippov and F. M. Arscott. 1988. *Differential equations with discontinuous righthand sides*. Kluwer Academic Publ, Dordrecht.

[16] Nicolas Gast and Bruno Gaujal. 2012. Markov Chains with Discontinuous Drifts Have Differential Inclusion Limits. *Perform. Eval.* 69, 12 (Dec. 2012), 623–642. https://doi.org/10.1016/j.peva.2012.07.003

[17] Z. Haider, R. Hussain, I. L. Khan, A. Shakeel, B. Ijaz, and S. A. Malik. 2017. Evaluation of capabilities of open source Cognitive radio network simulators. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. 1814–1817. https://doi.org/10.1109/IWCMC.2017.7986559

[18] Shlomo Halfin and Ward Whitt. 1981. Heavy-Traffic Limits for Queues with Many Exponential Servers. *Operations Research* 29 (06 1981), 567–588. https://doi.org/10.1287/opre.29.3.567

[19] M. R. Hassan, G. Karmakar, J. Kamruzzaman, and B. Srinivasan. 2017. Exclusive Use Spectrum Access Trading Models in Cognitive Radio Networks: A Survey. *IEEE Communications Surveys Tutorials* PP, 99 (2017), 1–1. https://doi.org/10.1109/COMST.2017.2725960

[20] Peter J. Smith, Abdulla Firag, Pawel Dmochowski, and Mansoor Shafi. 2012. Analysis of the M/M/N/N Queue with Two Types of Arrival Process: Applications to Future Mobile Radio Systems. 2012 (01 2012).

[21] J. Jacob and B. R. Jose. 2016. Spectrum occupancy measurement and analysis in Kochi-India from a cognitive radio perspective. In *2016 Sixth International Symposium on Embedded Computing and System Design (ISED)*. 328–333. https://doi.org/10.1109/ISED.2016.7977107

[22] Matthieu Jonckheere and Seva Shneer. 2014. Stability of Multi-Dimensional Birth-and-Death Processes with State-Dependent 0-Homogeneous Jumps. *Advances in Applied Probability* 46, 1 (2014), 59–75. https://doi.org/10.1239/aap/1396360103

[23] N. Khedun and V. Bassoo. 2015. Analysis of priority queueing with multichannel in cognitive radio network. In *IEEE EUROCON 2015 - International Conference on Computer as a Tool (EUROCON)*. 1–6. https://doi.org/10.1109/EUROCON.2015.7313750

[24] Jean-Yves Le Boudec. 2010. *The Stationary Behaviour of Fluid Limits of Reversible Processes is Concentrated on Stationary Points*. Technical Report. http://infoscience.epfl.ch/record/152160

[25] Avi Mandelbaum, William A. Massey, and Martin I. Reiman. 1998. Strong Approximations for Markovian Service Networks. *Queueing Syst. Theory Appl.* 30, 1/2 (June 1998), 149–201. https://doi.org/10.1023/A:1019112920622

[26] Joseph Mitola. 2000. *Cognitive radio*. Ph.D. Dissertation. KTH, Teleinformatics. http://www.diva-portal.org/smash/get/diva2:8730/FULLTEXT01.pdf NR 20140805.

[27] J. Mitola and G. Q. Maguire. 1999. Cognitive radio: making software radios more personal. *IEEE Personal Communications* 6, 4 (Aug 1999), 13–18. https://doi.org/10.1109/98.788210

[28] Hua Mu and Teck Hu. 2017. *Cognitive Radio and the New Spectrum Paradigm for 5G*. Springer Singapore, Singapore, 265–286. https://doi.org/10.1007/978.981.10.2254.8.10

[29] Dongyu Qiu and R. Srikant. 2004. Modeling and performance analysis of BitTorrent-like peer-to-peer networks. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*. ACM, New York, NY, USA, 367–378. https://doi.org/10.1145/1015467.1015508

[30] C. Rattaro, L. Aspirot, and P. Belzarena. 2015. Analysis and characterization of dynamic spectrum sharing in Cognitive Radio Networks. In *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*. 166–171. https://doi.org/10.1109/IWCMC.2015.7289076

[31] Claudina Rattaro, Paola Bermolen, and Pablo Belzarena. 2018. Multi-resource allocation : Analysis of a paid spectrum sharing approach based on fluid models. *IEEE Transactions on Cognitive Communications and Networking* (2018), 1–14. https://iie.fing.edu.uy/publicaciones/2018/RBB18

[32] Claudina Rattaro, Federico Larroca, Paola Bermolen, and Pablo Belzarena. 2017. Estimating the medium access probability in large cognitive radio networks. *Ad Hoc Networks* 63 (aug 2017), 1–13. https://iie.fing.edu.uy/publicaciones/2017/RLBB17

[33] Philippe Robert. 2003. *Stochastic Networks and Queues*. Springer-Verlag, New York. xvii+398 pages.

[34] S. Sasipriya and R. Vigneshram. 2016. An overview of cognitive radio in 5G wireless communications. In *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. 1–5. https://doi.org/10.1109/ICCIC.2016.7919725

[35] Sergei V. Savin, Morris A. Cohen, Noah Gans, and Ziv Katalan. 2005. Capacity Management in Rental Businesses with Two Customer Bases. *Operations Research* 53, 4 (2005), 617–631. https://doi.org/10.1287/opre.1040.0193

[36] Adam Shwartz and Alan Weiss. 1995. *Large deviations for performance analysis*. Chapman & Hall, London. x+556 pages.

[37] Max Tschaikowski and Mirco Tribastone. 2017. Spatial fluid limits for stochastic mobile networks. *Performance Evaluation* 109 (2017), 52 – 76. https://doi.org/10.1016/j.peva.2016.12.005

[38] Shanshan Wang, Junshan Zhang, and Lang Tong. 2010. Delay analysis for cognitive radio networks with random access: A fluid queue view. In *INFOCOM, 2010 Proceedings IEEE*. IEEE, 1–9.

[39] J. Xue, Z. Feng, and K. Chen. 2013. Beijing Spectrum Survey for Cognitive Radio Applications. In *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*. 1–5. https://doi.org/10.1109/VTCFall.2013.6692114

[40] Shuna Yang and Norvald Stol. 2014. Performance Modeling in Multi-Service Communications Systems with Preemptive Scheduling. *JCM* 9, 6 (2014), 448–460. https://doi.org/10.12720/jcm.9.6.448-460

[41] Changchuan Yin, Long Gao, and Shuguang Cui. 2010. Scaling Laws for Overlaid Wireless Networks: A Cognitive Radio Network Versus a Primary Network. *IEEE/ACM Trans. Netw.* 18, 4 (Aug. 2010), 1317–1329. https://doi.org/10.1109/TNET.2010.2041467

[42] Caoxie Zhang, Xinbing Wang, Xinping Guan, and Hsiao-Hwa Chen. 2009. Quality-of-Service in Cognitive Radio Networks with Collaborative Sensing. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*. 1–5. https://doi.org/10.1109/GLOCOM.2009.5425517

[43] Zhen Zhao. 2012. A preemptive connection pool manager for web-based application collaboration. In *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2012, Pittsburgh, PA, USA, October 14-17, 2012*. 326–334. https://doi.org/10.4108/icst.collaboratecom.2012.250465