A performance analysis of channel fragmentation in dynamic spectrum access systems

**Ed Coffman · Philippe Robert · Florian Simatos ·
Shuzo Tarumi · Gil Zussman**

Abstract Dynamic Spectrum Access systems offer temporarily available spectrum to opportunistic users capable of spreading transmissions over a number of non-contiguous subchannels. Such methods can be highly beneficial in terms of spectrum utilization, but excessive fragmentation degrades performance and hence off-sets the benefits. To get some insight into acceptable levels of fragmentation, we present experimental and analytical results derived from a mathematical model. According to the model, a system operates at capacity serving requests for bandwidth by assigning a collection of one or more gaps of unused bandwidth to each request as bandwidth becomes available. Our main result is a proof that, even if fragments can be arbitrarily small, the system remains stable in the sense that the average total number of fragments remains bounded. Within the class of dynamic fragmentation models, includ-

---

This paper is a revision for journal publication of the conference paper [3]. All of the results herein can be found in [3].

E. Coffman · S. Tarumi · G. Zussman
Electrical Engineering, Columbia University, New York, NY, USA

E. Coffman
e-mail: egc@ee.columbia.edu

S. Tarumi
e-mail: shuzo@ee.columbia.edu

G. Zussman
e-mail: gil@ee.columbia.edu

P. Robert
INRIA Paris-Rocquencourt, Paris, France
e-mail: Philippe.Robert@inria.fr

F. Simatos (✉)
CWI, Amsterdam, The Netherlands
e-mail: florian.simatos@cwi.nl

ing models of dynamic storage allocation that have been around for many decades, this result appears to be the first of its kind.

In addition, we provide extensive experimental results that describe behavior, at times unexpected, of fragmentation as parameter values are varied. Different scanning rules for searching gaps of available spectrum, all covered by the above stability result, are also studied. Our model applies to dynamic linked-list storage allocation, and provides a novel analysis in that domain. We prove that, interestingly, a version of the 50 % rule of the classical, non-fragmented allocation model holds for the new model as well. Overall, the paper provides insights into the behavior of practical fragmentation algorithms.

1 Introduction

This paper focuses on the analysis of dynamic resource allocation algorithms and is motivated by applications of these algorithms to Dynamic Spectrum Access Networks (also known as Cognitive Radio Networks). Cognitive Radios can adapt their transmitter parameters to the environment in which they operate and are viewed as key enablers of efficient use of the underutilized wireless spectrum [1, 7, 8, 17]. Under the basic model of Cognitive Radio Networks [1], Secondary Users (SUs) are allowed to use *white spaces* (also known as *spectrum holes*) that are not used by the Primary Users but must avoid interfering with active Primary Users (for instance, see Fig. 1).

An SU's channel allocation may be broken down into a number of *subchannels*, each allocated to a distinct hole of unused bandwidth. As shown in Fig. 1, this can be realized, for example, by employing a variant of Orthogonal Frequency-Division Multiplexing (OFDM) [10, 15, 18–20, 22]. Although the physical-layer aspects of OFDM-based Dynamic Spectrum Access have been extensively studied recently, allowing channel fragmentation introduces several new problems [12, 20, 23] that significantly differ from classical Medium Access Control (MAC).[1]
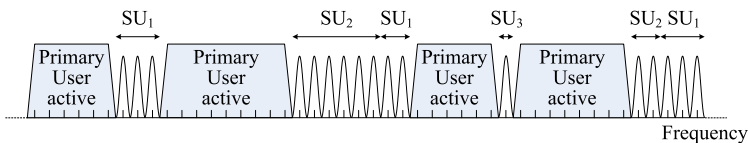
Fig. 1 Non-contiguous OFDM with Primary and Secondary Users (SUs), where the Secondary Users use non-contiguous channels that do not overlap with the Primary Users' channels

---

[1]Further details on Dynamic Spectrum Access Networks can be found in [3], which also contains an extensive review of related work.

In this paper, we study a baseline *theoretical model* in which the spectrum is shared by SUs only. Those users have to transmit and receive data, and accordingly need some bandwidth for given amounts of time. Hence, bandwidth requests of SUs are characterized by a desired total bandwidth and the duration of a time interval over which it is needed. The data transmission can take place over a non-contiguous channel (i.e., a number of subchannels). Once a transmission terminates, some fragments (subchannels) are vacated, and therefore, gaps (spectrum holes/white spaces) develop randomly in both size and position. When allocating a channel to a new SU request, it is fragmented (in the frequency domain) into available gaps until the full requested bandwidth is provided. This process repeats itself, until the next request fails to fit into the available fragments (more details are provided in the model subsection below).

The main goal of the paper is to investigate the phenomena of fragmentation induced by spectrum allocation algorithms. For this purpose, we will ignore the particularities of techniques such as OFDM and make a couple of assumptions (for a complete system description, see Sect. 2): (i) the system operates at capacity and there is always a waiting bandwidth request; and (ii) the fragment size is not bounded from below. Making the first assumption allows us to study the effect of fragmentation in the worst case. Clearly, if there are idle periods, when there are no waiting requests, only departures occur and the fragmentation level of the system decreases during these periods. Similarly, the latter assumption allows us to study the system performance when artificial lower bounds on the fragment size are not imposed. This differs from OFDM-based systems in which a subcarrier has a given minimal bandwidth.

In the application of these spectrum allocation algorithms, intriguing new problems in dynamic allocation arise. For example, because of the dynamic use of the spectrum, the bandwidth available to a new user will usually be distributed among a number of gaps in the spectrum. In principle, the random arrival and departure of users might cause the system to evolve in such a way that sequences of many, very small gaps are often allocated to user requests. In such cases, the system performance will deteriorate, since the algorithms maintaining the state of a highly fragmented spectrum become more time-consuming. To put these remarks on a proper footing, we first need to introduce the details of our model.

1.1 The model

A continuous frequency band of given width is made available to users. Each such user makes a request composed of a required total bandwidth and the duration of a residence time for the exchange of data over a channel of this bandwidth. Channels are allocated to requests on a first-come-first-served basis, subject to available bandwidth. A channel must remain fixed while active, and on departure it returns its allocation to the pool of available bandwidth. The channel of an allocated request is allowed to consist of multiple, disjoint bandwidth fragments, each being accommodated by a gap of unused spectrum that was available at the time of allocation.

For convenience, the model normalizes the spectrum to the interval [0, 1], so the sizes of all bandwidth requests are numbers in [0, 1]. Our goal is to characterize the fragmentation of requests when the system is operating at capacity, so we assume that

there is effectively an infinite queue of waiting users. For the purposes of the example to be given shortly, we denote the sizes of their bandwidth requests by $u_1, u_2, \ldots$. We will also use u_i as the name of the ith user. An initial, non-fragmented state is constructed as follows. Starting at 0, consecutive subintervals of sizes $u_1, u_2, \ldots$ are assigned as the channels of waiting users until a request size u_i, $i > 1$, is encountered which exceeds available bandwidth, i.e.,

$$u_1 + \cdots + u_{i-1} \leq 1 < u_1 + \cdots + u_i$$

At this point, all $i - 1$ of the channels in this initial state begin their residence times. Subsequent state transitions take place at departure epochs when the residence times of currently allocated channels expire. Suppose all requests up to u_j have been allocated channels and a request u_i, $i \leq j$, departs, releasing its allocated channel. If there is still not enough bandwidth for u_{j+1}, then it must wait for one or more additional departures. Otherwise, $u_{j+1}, u_{j+2}, \ldots$ are allocated their requested bandwidths until, once again, a request is encountered that asks for more bandwidth than is available. All channels then begin or continue their residence times as before until the next departure.

The *linear-scan* (LS) rule sets up a channel by scanning the spectrum in order of increasing frequency, allocating gaps of available bandwidth in partial fulfillment of the request until enough bandwidth has been allocated to satisfy the entire request. In general, the last gap is only partially used, in which case the last fragment is left-justified in the last gap. An example is shown in Fig. 2. Allocations are shown for the first eight user requests, assuming that, at time 0, the spectrum is not in use. The requests with sizes $u_1, u_2,$ and u_3 are the first to be allocated channels; u_4 must wait for a departure, since the first four request sizes sum to more than 1. The variables t_i give the sequence of departure times of allocated requests. We see that the first occurrence of fragmentation takes place at the departure of u_2 and the subsequent admission of u_4; an initial fragment of u_4 is placed in the gap left by u_2 and a final fragment is placed after u_3. Note also that, even after u_2 and u_4 have departed, there is still not enough bandwidth for u_5. After the additional departure of u_1, both u_5 and u_6, but not u_7, can be allocated bandwidth.
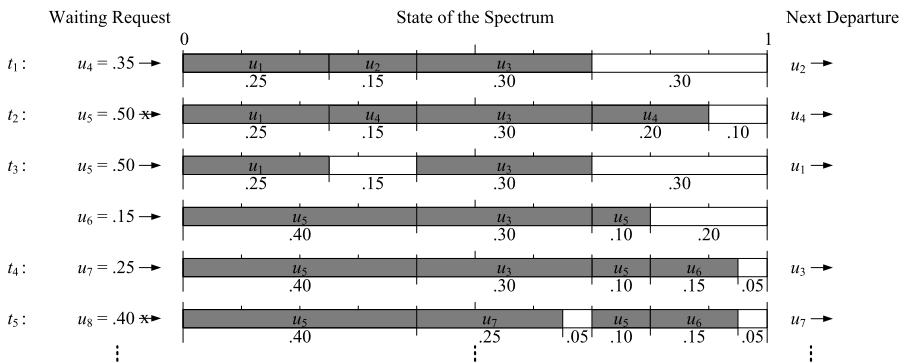
Fig. 2 An example of the admission and departure processes and the resulting spectrum fragmentation

We complete the model description by giving the probability laws governing the residence times and sizes of the waiting requests. With $\alpha \in (0, 1]$ a parameter of the model, the sizes $u_{j+1}, u_{j+2}, \ldots$ are drawn independently from the uniform distribution on $(0, \alpha]$. The residence times of allocated requests are independent of request sizes and form a sequence of i.i.d. exponentially distributed random variables. For simplicity, the mean is normalized to 1.

1.2 Main results

At first glance, since request sizes can be arbitrarily small, the fragments used by bandwidth requests might well become progressively narrower as time passes, in which case the number of fragments could grow without bound. Clearly, such an operational model cannot be supported by a realistic system, so it is of great interest to know whether this can actually happen. The principal theoretical contribution of this paper is a proof that in fact, with high probability, the number of fragments in the system remains acceptably low. A precise statement is deferred to the next section where additional notation and concepts are introduced.

In addition to mathematical results, and in order to gain further insight into the performance of the system, we present the results of extensive experiments. These show that although for a given maximum request size α, the number of fragments has a finite expected value, there is a linear relationship between $1/\alpha$ and the expected number of fragments into which a request is divided. This indicates that when the requests are constrained to be small, they are also fragmented into a relatively large number of fragments. From a practical point of view, this implies that, if we can impose a lower bound on the fragment size by rejecting gaps that are too small, then we have a useful control on the ill effects of excessive fragmentation.

Our experiments show that different bandwidth allocation algorithms exhibit significantly different performance in terms of the fragmentation occurring in the system. As one example, an algorithm that scans the gaps in decreasing order of their sizes reduces the fragmentation by almost an order of magnitude compared to LS. Interestingly, the experiments also show that the number of fragments is distributed according to a normal law with a relatively low mean value. Finally, the experiments for small α led to the interesting observation that the ratio of the average number of gaps to the average number of requests being served was approximately 1/2. Although we were unable to prove this ratio-of-averages result, we were able to prove a corresponding 50 % limit law for the expected value of the ratio of the number of gaps to the number of requests being served.

The spectrum allocation problems described here can be characterized in the context of dynamic storage allocation problems of computers; see Knuth [14]. In that context, the term "spectrum" refers to the storage unit, "bandwidth requirements" refer to storage requirements, and "channel" refers to a region in storage containing a file. In the original dynamic storage model, fragmentation referred only to the gaps of unoccupied storage interspersed with intervals of occupied storage: files were not fragmented. The reader familiar with that model may have found the 50 % rule observed in our experiments to be reminiscent of Knuth's 50 % rule, which asserts that the number of gaps is approximately half the number of files when exact

fits of files into gaps are rare. Our notion of fragmentation applies also to the files themselves, so our discovery that a similar 50 % rule continued to apply was a surprising one at first. Note that, in terms of the storage application, our model corresponds precisely to *linked-list* allocation of files—channels allocated to requests are sets of linked, disjoint segments of storage allocated to file fragments. Our results provide a novel analysis for such systems, and have implications for the garbage-collection/defragmentation process in linked-list systems. We note that studies of dynamic storage allocation have been around for some 40 years, and widely recognized as posing very challenging problems to both combinatorial and stochastic modeling and analysis. In particular, results of the type found in this paper, rigorous within stochastic models, seem to be quite new.

The remainder of the paper begins in the next section with the introduction of notation, a formalization of the spectrum state space, and the stochastic processes of interest in later sections. Section 3 presents experimental results that bring out the effects of fragmentation, particularly as a function of the maximum request size α. In Sect. 4 relations between variables describing the configuration of gaps and fragments are proved in preparation for (1) a *50 % limit law* for the relation between the number of active channels and the number of gaps in the spectrum at departure times, and (2) our main stability result, which shows that the expected value of the total number of fragments and gaps is bounded. These results are proved in Sects. 4 and 5, respectively. Section 5 also proves that, for α large enough, an equilibrium distribution exists. Section 6 discusses algorithmic issues, such as changes in performance resulting from alternative allocation algorithms, and Sect. 7 discusses experimental results that exhibit normal approximations for the total number of fragments and gaps. Section 8 concludes the paper with a discussion of the results and future research directions.

2 Preliminaries

The state of a fragmentation process must carry the information given by a sequence of gaps alternating with sequences of contiguous fragments. Accordingly, we formalize a state space $\mathcal{S}$ for our model as follows.

Definition 1 A state $x \in \mathcal{S}$ is given by

$$x = (L_1, \ldots, L_r; u)$$

where u is the size of the request waiting at the head of the queue, $r \geq 1$ is the number of currently active channels, and L_i is the list of open subintervals of [0, 1] occupied by the fragments of the ith channel. For x to be admissible, the open intervals in $\bigcup_i L_i$ must be mutually disjoint, and, since the size of u exceeds the bandwidth available, $u > 1 - \sum_i s_i$ has to hold, where s_i is the cumulative size of the fragments in L_i.

Since channel residence times are i.i.d. exponentially distributed random variables, the process $(X(t))$ on $\mathcal{S}$ is a Markov process. The Markov chain embedded at departure times t_n will be of special interest.

When treated as a random variable, the size of the request waiting to be allocated bandwidth at time t is denoted by $U(t)$. Similarly, $U_i(t)$ denotes a random variable giving the size of the ith request behind the head of the queue. Except for Sect. 5, the dependence on time will usually be omitted. We denote by $R(t)$ the number of requests with an active channel at time t. $F(t)$ and $G(t)$ denote, respectively, the numbers of fragments and gaps at time t.

With this notation, we are in position to state our main contribution to mathematical foundations. We prove that the total number of gaps and fragments at departure epochs has bounded exponential moments, as follows: For some $\eta > 0$ and any initial state $x \in \mathcal{S}$,

$$\sup_{n \geq 1} \mathbb{E}_x\big(e^{\eta(F(t_n)+G(t_n))}\big) < +\infty$$

where $\mathbb{E}_x$ denotes expectation for the process X started at $X(0) = x$. While our choice of a uniform distribution for request sizes is a useful one, it is not essential to this result; it holds for more general distributions. The result shows that the sum $F(t_n) + G(t_n)$ is strongly concentrated near the origin. This fact provides strong support for the informal assertion made in Sect. 1 to the effect that, with high probability, the level of fragmentation stays reasonably low. Our analysis has several basic ingredients: relations between the number of fragments and gaps in the spectrum (Lemma 2); a drift relation of Lyapunov type for the total number of gaps, fragments, and requests (Proposition 1); a general inequality for Markov chains (Theorem 4); and a stability result of [13].

The last of these results refers to the early work of Kipnis and Robert [13] on a non-fragmented version of our model. Channel allocations can be moved as needed in order to put all available bandwidth together in one block, so fragmentation is avoided entirely. They consider a system with arrivals, but with arrival rates sufficiently high, their analysis of maximum throughput gives us an analysis of $(R(t_n))$, assuming that the same probability laws for request sizes and residence times apply. This follows simply from the fact that the admission criterion for waiting requests is the same in both models. A major result in [13] asserts the existence and uniqueness of an invariant measure for $(R(t_n))$. Explicit formulas are hard to come by, but those in [13] for the maximal departure rate in special cases have provided useful checks for our experiments.

In Sect. 6, we shall also evaluate two scanning alternatives to LS. The first is *circular scan* (CS), which is still a linear scan of the gaps, but each scan starts where the previous one left off; after the last gap in [0, 1], CS cycles back to the first gap in [0, 1]. The second is *largest-first-scan* (LFS), which is intended to further reduce fragmentation by assigning gaps in order of decreasing size. Although these algorithms make different scans of the gap sequence, they are all alike in their treatment of the last gap occupied: *the last fragment is left-justified in the last gap*. This is a key assumption, and it is very likely to hold in practice. In our probability model, it follows that, with probability 1, the last gap used in a channel allocation will be changed to a fragment and a smaller residual gap.

3 Experimental results

The experimental study reported in this section serves two related roles. First, it brings out characteristics of the fragmentation process that need to be borne in mind in implementations, particularly where these characteristics show parameter values that must be avoided, if a system with fragmentation is to operate efficiently. The second role is that of experimental mathematics, in which results indicate where behavior might well be formalized and rigorously proved as a contribution to mathematical foundations. In the latter role, this section leads up to the next two sections, which formalize and prove the stability of the fragmentation process.

The experiments were conducted with a discrete-event simulator written in C that also includes a stochastic arrival process, a capability that we intend to explore in future research. The maximum request size α is the single parameter surviving the normalizations of the simpler mathematical model of this paper. The simulations were most demanding, of course, for small α, when large numbers of departure events were needed to ensure behavior near the stationary regime. For every α value chosen in the interval [0.01, 1], 20 million departure events were simulated starting in an empty state, with data collected for the last 10 million events. For every choice of α in [0.001, 0.01], 100 million departure events were processed and data collection was performed during the last 50 million events. The excellent accuracy of this tool was established in tests against exact results for special classes of queueing systems, and against the maximum throughput results derived in [13]. Examples of the test results can be found in the appendix and in [21].

The results for the average number of channels, the average number of gaps, and the average number of fragments per channel are shown in Fig. 3. The curves are nearly linear in $1/\alpha$ for small values of α; indeed, the errors in the linear fits are within the thickness of the printed lines. In particular, the asymptotic (as $\alpha \to 0$) average number of channels in the spectrum is $2/\alpha$. When requests are large relative to the spectrum (i.e., for $\alpha > 1/3$), the behavior is not given by functions quite so simple. As such cases are of less practical interest, we omit the relevant data.

The asymptotic linear growth of the average number of channels as a function of channel size is obvious, but the linearity of the other two measures is not so obvious. A closer look shows that the average number of gaps is almost exactly one half the

Fig. 3 Average numbers of channels, of gaps, and of fragments per channel
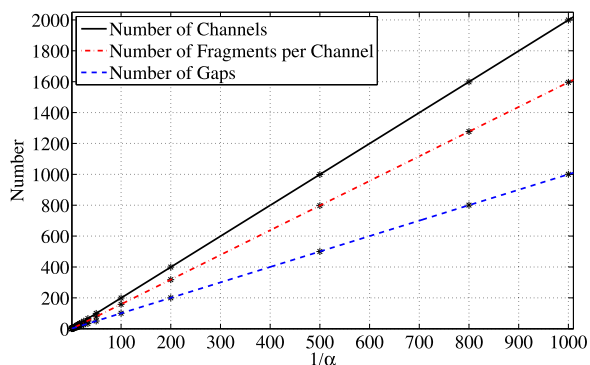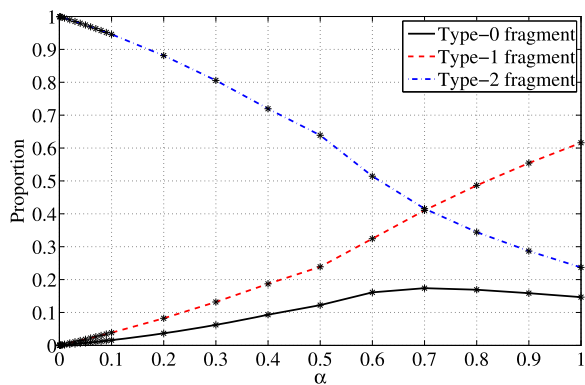
Fig. 4 Average total number of fragments vs.$1/\alpha$: a quadratic fit
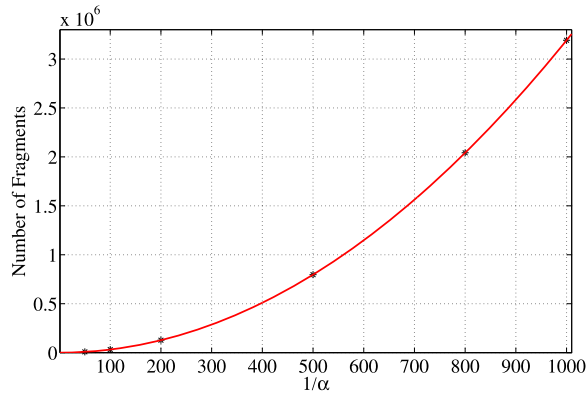

Fig. 5 Percentage of type-i fragments

average number of channels for even relatively small $1/\alpha$. This is the unexpected version of Knuth's 50 % rule that we mentioned in Sect. 1. We return to this behavior in the next section, where we prove a 50 % limit law. The linear growth of the average number of fragments per channel may also be unexpected at first glance: the fragmentation of channels *increases* as the average channel size *decreases*. This linear growth implies the quadratic growth of the average total number of fragments plotted in Fig. 4 (the accuracy of the fit is as before: the error is within the thickness of the printed lines).

The analysis in later sections will focus largely on tracking fragment *types* defined as follows: a fragment is of type i, if it is adjacent to i other fragments, where $i = 0$, 1, or 2. It can be seen in Fig. 5 that for small α, more than 90 % of the fragments are type-2 fragments. In addition, clearly, the number of type-0 and type-1 fragments is a function of the number of gaps. These observations and the results illustrated in Figures 3 and 4 indicate that, even for relatively small $1/\alpha$, the average total number of type-0 and type-1 fragments grows linearly in $1/\alpha$, but the average number of type-2 fragments grows quadratically.

Figure 6 compares the average gap and fragment sizes. As might be expected, for relatively small α, they are close to each other. The relation holds even for moderately large α, although for α rather close to 1, the difference amounts to about a factor of 2.

Fig. 6 Average sizes of
fragments and gaps
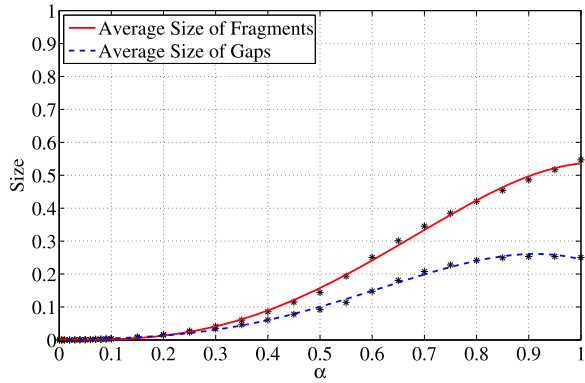
(a) $0 \leq \alpha \leq 1$
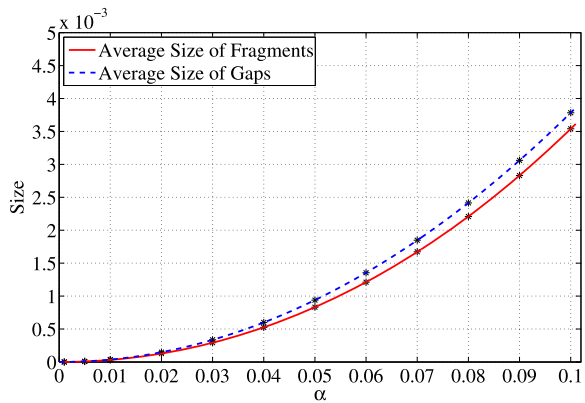
(b) $0 \leq \alpha \leq 0.1$

With this property and the 50 % rule suggested by Fig. 3, the linear growth in the number of fragments per channel (shown in Fig. 3) is easily explained for moderately small α in the following way.

As mentioned above, for moderately small α, the number of channels is approximately $2/\alpha$ (i.e., the spectrum size divided by the average request size). By the 50 % rule, the number of gaps is roughly $1/\alpha$. At any time, the total size of the gaps is at most α, since there is a request waiting for departures whose requested bandwidth exceeds the total size of the gaps. Therefore, at most α available bandwidth is spread among $1/\alpha$ gaps, giving an average gap size on the order of α^2. The average fragment size is at most (and indeed very close to) the average gap size. The fragments must occupy at least $1 - \alpha$ of the spectrum since, as mentioned above, at most a fraction α of the spectrum is devoted to gaps. Thus, the number of fragments must be on the order of $1/\alpha^2$, and so the average number of fragments per channel must be on the order of (in particular, linear in) $1/\alpha$. As $\alpha \to 0$, the asymptotics of these estimates become more precise.

4 Numbers of fragments and gaps

This section presents analytical results relating the numbers of fragments and gaps under the fragmentation process $(X(t))$. Recall the definition of fragment types: For $i = 0$, 1 or 2, a fragment is said to be of type i if it touches exactly i other fragments. $N_i(t)$ denotes the number of type i fragments at time t, so that $F(t) = N_0(t) + N_1(t) + N_2(t)$ is the total number of fragments.

Let $\sigma(t)$ denote the sum of the numbers of fragments and gaps,

$$\sigma(t) = F(t) + G(t). \tag{1}$$

The number of gaps and the numbers of fragment types are related as follows.

Lemma 1 *With probability* 1,

$$G(t) = N_0(t) + \frac{1}{2}N_1(t) + I(t) \tag{2}$$

for any $t \geq 0$, *where* $I(t) = 1$, *if there is a gap starting at the origin, and* 0 *otherwise.*

Proof Each gap, except for boundary gaps starting at 0 or ending at 1, separates two fragments. Two gaps surround a type-0 fragment not touching the origin, only one touches a type-1 fragment not touching the origin, and none touch a type-2 fragment, so the gaps strictly inside $(0, 1)$ are double-counted in $2N_0(t) + N_1(t)$. Gaps at the boundaries are counted only once in this expression, so if there are gaps touching each boundary, 2 must be added to $2N_0(t) + N_1(t)$ to produce a double count of all gaps. Then $N_0(t) + N_1(t)/2 + 1$ counts the gaps as called for by the lemma.

With probability 1, a gap always touches the boundary at 1, so the only case left to consider is the absence of a gap touching the origin. In this case, there is a type-0 or type-1 fragment touching the origin, and so a nonexistent gap has been counted in $2N_0(t) + N_1(t)$. This over-count cancels the under-count of the gap touching 1, and so no correction term is needed, i.e., $N_0(t) + N_1(t)/2$ counts all gaps as stated in the lemma. □

Definition 2 Let (t_k) denote the sequence of departure times, let $D_i(t_k)$ denote the number of type i fragments in the channel leaving at time t_k, and let $A(t_k)$ denote the number of requests admitted to the spectrum at time t_k. Finally, define the *drift* in the total number of fragments and gaps:

$$\Delta\sigma(t_k) \stackrel{\text{def}}{=} \sigma(t_k) - \sigma(t_{k-1}) \tag{3}$$

with the convention $t_0 = D_i(t_0) = A_i(t_0) = 0$.

The following lemma is the basis of the stability analysis of $\sigma(t)$ in Sect. 5, and the 50 % rule proved later in this section.

Lemma 2 *With probability* 1, *the departure at* t_k *creates the following change in the total number of fragments and gaps*:

$$\Delta\sigma(t_k) = A(t_k) - 2D_0(t_k) - D_1(t_k) + J(t_k), \quad k \geq 1 \tag{4}$$

with $t_0 = 0$, *and* $J(t_k) = 1$, *if a fragment starting at the origin is in the departing channel, and* 0 *otherwise*.

Proof With probability 1, each new channel allocation covers completely every gap it is allocated, except for the last one, which is only partially covered. Thus, with probability 1, each new channel allocation changes gaps to fragments, except for the last gap which is changed to a fragment plus a gap; this adds one to $\sigma(t_{k-1})$ for each admission, which accounts for the total of $A(t_k)$ in (4).

Two fragments of the same channel cannot be contiguous, so it is correct to add up the changes created by departing fragments, with each being treated separately. Suppose first that there is no fragment $(0, b)$ against the origin. Then for every type-0 fragment in the departing channel, two gaps and a fragment are replaced by a single gap for a net decrease of two, and for every departing type-1 fragment, a gap and a fragment are replaced by a single gap for a net reduction of one. This gives the reduction of $2D_0(t_k) + D_1(t_k)$ appearing in (4). If there is a fragment $(0, b)$, it must be of type 0 or 1; if it is of type 0, then its departure gives a decrease of one; if it is of type 1, its departure has no effect. Each of these contributions is one less than it would be were the fragment not touching the origin. There can only be one such fragment, so the correction shown in $J(t_k)$ for a fragment $(0, b)$ follows. □

We will denote by $G^-(t_k)$ the total number of gaps just after the kth departure, but before new admissions, if any, are made. Note that if we remove $A(t_k)$ from the right-hand side of (4) and add back the total number of departing fragments at t_k, i.e., $D_0(t_k) + D_1(t_k) + D_2(t_k)$, we get the number of gaps available to admissions at the kth departure:

$$G^-(t_k) = G(t_{k-1}) - D_0(t_k) + D_2(t_k) + J(t_k) \tag{5}$$

with $J(t_k) = 1$ as in Lemma 2.

Knuth's widely known *50 % rule* appears in a very different context than the model here, so it is difficult to anticipate the apparent fact that it also holds for our fragmentation model. However, one can argue a similar result assuming that the fragmentation process has a stationary distribution. The result is given below as an expected value of a ratio, rather than a ratio of expected values.

Theorem 1 *Assume that for each* $\alpha > 0$, *the fragmentation process at departure epochs* $(X(t_k), k \geq 0)$ *with request sizes uniformly distributed in* $(0, \alpha)$ *admits at least one stationary distribution. For each* $\alpha > 0$ *pick one stationary distribution* π_α. *Then*

$$\lim_{\alpha \to 0} \mathbb{E}_{\pi_\alpha}\left(\frac{G(0)}{R(0)}\right) = \frac{1}{2}.$$

Proof Any stationary distribution π_α must satisfy $\mathbb{E}_{\pi_\alpha}[\sigma(t_1)] = \mathbb{E}_{\pi_\alpha}[\sigma(t_0)]$ and $\mathbb{E}_{\pi_\alpha}[A(t_1)] = 1$. Thus, using Lemma 2 one gets

$$\mathbb{E}_{\pi_\alpha}\big[2D_0(t_1) + D_1(t_1) - J(t_1)\big] = 1. \tag{6}$$

Now for a state x having $R(0) > 0$ channels and $N_i(0)$ type-i $(i = 0, 1, 2)$ fragments, we have from Lemma 1

$$\mathbb{E}_x\big[2D_0(t_1) + D_1(t_1) - J(t_1)\big] = \big(2N_0(0) + N_1(0)\big)/R(0) - \mathbb{E}_x\big[J(t_1)\big]$$
$$= 2\big[G(0) - I(0)\big]/R(0) - \mathbb{E}_x\big[J(t_1)\big].$$

The term $\mathbb{E}_x[J(t_1)]$ is equal to the probability that a fragment starting at the origin leaves. This is certainly smaller than the probability that the channel with the fragment closest to the origin leaves, which by symmetry is equal to $1/R(0)$. Using in addition $I(0) = 0$ or 1 one gets

$$0 \leq \frac{1}{2} - \mathbb{E}_{\pi_\alpha}\big(G(0)/R(0)\big) \leq 2\mathbb{E}_{\pi_\alpha}\big(1/R(0)\big).$$

Now for any stationary distribution π_α the random variable $R(0)$ under π_α must be distributed according to the stationary distribution of the non-fragmented model of Kipnis and Robert [13], and one easily checks that $\mathbb{E}_{\pi_\alpha}(1/R(0))$ goes to 0 as α goes to 0, hence the result. □

Note that, because we have a Lyapunov function (see Proposition 1), the existence of a stationary distribution would be guaranteed if we could prove that the process $(X(t_n))$ has the Feller property, see for instance Proposition 12.1.3 in Meyn and Tweedie [16].

5 Stability results

This section establishes that the average total number of fragments and gaps remains bounded and that, for certain distributions of request sizes, ergodicity holds. The analysis leads to the following two results. Recall that (t_n) is the sequence of departure times $(t_0 = 0)$ and $\sigma(t) = F(t) + G(t)$, defined in (1), is the total number of fragments and gaps.

Theorem 2 *There exists some $\eta > 0$ such that, for any initial state $x \in \mathcal{S}$,*

$$\sup_{n \geq 1} \mathbb{E}_x\big(e^{\eta\sigma(t_n)}\big) < +\infty. \tag{7}$$

Clearly, this implies that for any initial state $x \in \mathcal{S}$, the sequence $(\mathbb{E}_x(\sigma(t_n)), n \geq 0)$ is bounded. With an additional assumption on the distribution of the request size, a stronger stability result can be proved.

Theorem 3 *When $\alpha > 1/2$, the process $(X(t))$ is positive Harris recurrent; in particular, it has a unique stationary distribution.*

A criterion for finite exponential moments using a Lyapunov function is established next. Then, we provide some estimates of the drift of the number of fragments between departures which will show us how to construct such a Lyapunov function. The proofs of Theorems 2 and 3 will finally be given in Sects. 5.4 and 5.5.

5.1 A criterion for finite exponential moments

Before stating the main result, some results on Markov chains are needed. In the sequel, $\leq_{st}$ refers to stochastic ordering, i.e., $V \leq_{st} Z$ means that $\mathbb{E}(f(V)) \leq \mathbb{E}(f(Z))$ for any increasing function f. For reasons that will become clear in Lemma 5, the following lemma focuses on admissions at four consecutive departure times. Recall that $(U_i, i \geq 1)$ are the sizes of the requests waiting to be allocated bandwidth after U, the first one, and that they are assumed to be i.i.d.

Lemma 3 *The random variable $A(t_1) + \cdots + A(t_4)$ is stochastically dominated by a random variable Z such that $\mathbb{E}(e^{\lambda Z}) < +\infty$ for some $\lambda > 0$.*

Proof It is clear that $A(t_1) \leq Z + 1$ where

$$Z = 1 + \inf\{n \geq 1 : U_1 + \cdots + U_n \geq 1\}.$$

Markov's inequality shows that for any $z \geq 0$,

$$\mathbb{P}(Z \geq z + 1) = \mathbb{P}(U_1 + \cdots + U_z \leq 1) \leq e\big(\mathbb{E}\big(e^{-U_1}\big)\big)^z$$

and so $\mathbb{E}(e^{\eta Z})$ is finite for $\eta > 0$ small enough. From this observation, it is not difficult to extend the result to $A(t_1) + \cdots + A(t_4)$ instead of just $A(t_1)$. $\qquad\square$

This lemma shows in particular that

$$\xi \stackrel{\text{def}}{=} \sup_{i \geq 1} \sup_{x \in \mathcal{S}} \mathbb{E}_x\big(A(t_i)\big) < +\infty.$$

is well-defined; this constant will be used repeatedly throughout the rest of the analysis. The proof of the following lemma is standard, and therefore, omitted.

Lemma 4 *Let $Z \geq 0$ be a positive, real-valued random variable such that $\mathbb{E}(e^{\lambda Z}) < +\infty$ for some $\lambda > 0$, and define $c = \lambda^{-2}\mathbb{E}(e^{\lambda Z} - 1 - \lambda Z)$. Then for any $0 \leq \varepsilon \leq \lambda$ and any real-valued random variable V such that $V \leq_{st} Z$, we have $\mathbb{E}(e^{\varepsilon V}) \leq 1 + \varepsilon \mathbb{E}(V) + \varepsilon^2 c$.*

The following result is closely related to a result of Hajek [11].

Theorem 4 *Let (Y_k) be a discrete-time, continuous state-space Markov chain such that for some function $f \geq 0$, there exist $K, \gamma > 0$ such that for any initial state y with $f(y) > K$, the inequality $\mathbb{E}_y(f(Y_1) - f(Y_0)) \leq -\gamma$ holds. Assume that there exists a random variable Z such that for any initial state y, Z dominates stochastically the*

random variable $f(Y_1) - f(Y_0)$ under $\mathbb{P}_y$. Assume finally that $\mathbb{E}(e^{\lambda Z}) < +\infty$ for some $\lambda > 0$. Then there exist $\eta > 0$ and $0 \leq C < +\infty$ such that for any initial state y,

$$\sup_{n \geq 1} \mathbb{E}_y\big(e^{\eta f(Y_n)}\big) \leq e^{\eta f(y)} + C.$$

Proof For any $0 < \varepsilon \leq \lambda$,

$$\mathbb{E}_y\big(e^{\varepsilon f(Y_{n+1})}\big) = \mathbb{E}_y\big(\mathbb{E}_{Y_n}\big(e^{\varepsilon(f(Y_1)-f(Y_0))}\big)e^{\varepsilon f(Y_n)}\mathbb{1}_{\{f(Y_n) \leq K\}}\big)$$
$$+ \mathbb{E}_y\big(\mathbb{E}_{Y_n}\big(e^{\varepsilon(f(Y_1)-f(Y_0))}\big)e^{\varepsilon f(Y_n)}\mathbb{1}_{\{f(Y_n) > K\}}\big). \qquad (8)$$

Since by assumption, $f(Y_1) - f(Y_0)$ under $\mathbb{P}_y$ is stochastically dominated by Z for every y, one gets

$$\mathbb{E}_{Y_n}\big(e^{\varepsilon(f(Y_1)-f(Y_0))}\big) \leq \mathbb{E}\big(e^{\varepsilon Z}\big) = \mathbb{E}\big(e^{\lambda Z}\big) \overset{\text{def}}{=} D$$

and therefore

$$\mathbb{E}_y\big(\mathbb{E}_{Y_n}\big(e^{\varepsilon(f(Y_1)-f(Y_0))}\big)e^{\varepsilon f(Y_n)}\mathbb{1}_{\{f(Y_n) \leq K\}}\big) \leq De^{\varepsilon K}. \qquad (9)$$

For the second term, we apply Lemma 4 to the random variable $f(Y_1) - f(Y_0)$ under $\mathbb{P}_{Y_n}$:

$$\mathbb{E}_{Y_n}\big(e^{\varepsilon(f(Y_1)-f(Y_0))}\big) \leq 1 + \varepsilon \mathbb{E}_{Y_n}\big(f(Y_1) - f(Y_0)\big) + c\varepsilon^2.$$

Thus, on the event $\{f(Y_n) > K\}$, one gets

$$\mathbb{E}_{Y_n}\big(e^{\varepsilon(f(Y_1)-f(Y_0))}\big) \leq 1 - \gamma\varepsilon + c\varepsilon^2 \overset{\text{def}}{=} \rho_\varepsilon \qquad (10)$$

and finally

$$\mathbb{E}_y\big(\mathbb{E}_{Y_n}\big(e^{\varepsilon(f(Y_1)-f(Y_0))}\big)e^{\varepsilon f(Y_n)}\mathbb{1}_{\{f(Y_n) > K\}}\big) \leq \rho_\varepsilon \mathbb{E}_y\big(e^{\varepsilon f(Y_n)}\big). \qquad (11)$$

Gathering the two bounds (9) and (11) in (8) finally gives

$$\mathbb{E}_y\big(e^{\varepsilon f(Y_{n+1})}\big) \leq \rho_\varepsilon \mathbb{E}_y\big(e^{\varepsilon f(Y_n)}\big) + De^{\varepsilon K}$$

which leads by induction to

$$\mathbb{E}_y\big(e^{\varepsilon f(Y_n)}\big) \leq \rho_\varepsilon^n e^{\varepsilon f(y)} + \frac{1 - \rho_\varepsilon^{n+1}}{1 - \rho_\varepsilon} De^{\varepsilon K}.$$

Let $\eta = (\varepsilon/(2\gamma)) \wedge \lambda$. From the definition of ρ_ε in (10), one can easily check that $0 < \rho_\eta < 1$. Therefore, choosing $\varepsilon = \eta$ gives the result. $\qquad \square$

Theorem 4 will be applied to the Markov chain $(X(t_{4n}))$ with a function f of the form $\sigma_\kappa \overset{\text{def}}{=} \sigma + \kappa R$ for some $\kappa > 0$ suitably chosen. $(X(t_{4n}))$ is not the most natural choice at first glance, but it appears to be needed because of the complexity of the state space.

By Lemma 2, $\sigma(t_1) - \sigma(t_0) \leq A(t_1) + 1$ and clearly $R(t_1) - R(0) = A(t_1) - 1$, so that

$$\sigma_\kappa(t_4) - \sigma_\kappa(t_0) \leq (\kappa + 1)\big(A(t_1) + \cdots + A(t_4)\big) + 4$$

and therefore, by Lemma 3, $\sigma_\kappa(t_4) - \sigma_\kappa(t_0)$ is stochastically dominated by some random variable Z with an exponential moment. Therefore, one has to establish a negative drift relation for $\sigma_\kappa(t_4) - \sigma_\kappa(t_0)$ in order to apply Theorem 4 to the fragmentation process $(X(t_{4n}))$. This is the purpose of the following two subsections. The most delicate part is to control the term $\sigma(t_4) - \sigma(t_0)$, which is the objective of the following section.

5.2 Evolution of the number of fragments

Let $x \in \mathcal{S}$, the initial state of the system, have r active channels, and define the total available gap size $h = 1 - (s_1 + \cdots + s_r)$. Time 0, referring to the initial state x, will usually be omitted; e.g., $\sigma(0), F(0), G(0), \ldots$ will be simplified to $\sigma, F, G, \ldots$. The quantity $\Delta\sigma(t_n)$ is defined in (3) as $\sigma(t_n) - \sigma(t_{n-1})$.

Lemma 5 *Fix $0 < \varepsilon < 1$ and $0 < \eta < 1/2$, and let $x \in \mathcal{S}$ be an initial state such that $\sigma = G + F \geq 2K + 1$ for some fixed $K \geq 0$.*
  Then $F = N_0 + N_1 + N_2 \geq K$, and

(1) *If $r = 1$, then $\mathbb{E}_x(\Delta\sigma(t_1)) \leq \xi - K$.*
(2) *If $r > 1$ and $N_0 + N_1 \geq \varepsilon K$, then*

$$\mathbb{E}_x\big(\Delta\sigma(t_1)\big) \leq \xi + 1 - \frac{\varepsilon K}{r}.$$

Assume in the remaining cases that $r > 1$, define $K' = K((1 - \varepsilon)/r - \varepsilon)^+$, and let $i^* \in \{1, \ldots, r\}$ index a channel L_{i^*} in x with the most type-2 fragments.

(3) *If $N_0 + N_1 \leq \varepsilon K$ and $u > h + s_{i^*}$, then*

$$\mathbb{E}_x\big(\Delta\sigma(t_2)\big) \leq \xi + 2 - \frac{K'}{r(r-1)}. \tag{12}$$

(4) *If $N_0 + N_1 \leq \varepsilon K$, $u < h + s_{i^*}$ and $h + s_{i^*} < \eta\alpha$, then*

$$\mathbb{E}_x\big(\Delta\sigma(t_3)\big) \leq \xi + 2 - \frac{(1-\eta)K'}{r^2(r-1)}. \tag{13}$$

(5) *If $N_0 + N_1 \leq \varepsilon K$, $u < h + s_{i^*}$ and $\eta\alpha < h + s_{i^*}$, then there exists a $\gamma(\eta) > 0$ such that*

$$\mathbb{E}_x\big(\Delta\sigma(t_4)\big) \leq \xi + 2 - \frac{\gamma(\eta)K'}{r^5}. \tag{14}$$

It follows that there exists a $\overline{\xi} > 0$ and a function $\psi(r) > 0$ such that for any x with $\sigma \geq 2K + 1$,

$$\mathbb{E}_x\big(\sigma(t_4) - \sigma\big) \leq \overline{\xi} - K\psi(r). \tag{15}$$

Proof As is readily verified, $G \leq F+1$, so $2K+1 \leq \sigma = F+G \leq 2F+1$, and hence $F \geq K$ as claimed. In what follows, we use repeatedly the two following simple facts:

$$\mathbb{E}_x\big(D_0(t_1) + D_1(t_1)\big) = (N_0 + N_1)/r, \tag{16}$$

and by Lemma 1,

$$G \geq K \quad \Rightarrow \quad N_0 + N_1 \geq K - 1. \tag{17}$$

– *First case*: $r = 1$. In this case, right after the only channel initially present leaves, there is no channel allocated bandwidth, and therefore, $\sigma(t_1) = A(t_1)$. Note that $r = 1$ is only possible when $\alpha > 1/2$, and in this case the possibility for a channel to be alone is crucial in the proof of the Harris recurrence stated in Theorem 3.

– *Second case*: $r > 1$, $N_0 + N_1 \geq \varepsilon K$. Under these assumptions, the inequality follows from (4):

$$\mathbb{E}_x\big(\Delta\sigma(t_1)\big) \leq \xi + 1 - \mathbb{E}_x\big(D_0(t_1) + D_1(t_1)\big) = \xi + 1 - \frac{N_0 + N_1}{r} \leq \xi + 1 - \frac{\varepsilon K}{r}.$$

In the 3 remaining cases, let N_j^* denote the number of type-j fragments in any channel i^* which has the most type-2 fragments. If $N_0 + N_1 \leq \varepsilon K$, then since $F \geq K$, necessarily $N_2 \geq (1-\varepsilon)K$ and $N_2^* \geq (1-\varepsilon)K/r$. Define the event $D^* = \{$channel L_{i^*} leaves at $t_1\}$ and recall that G^- denotes the number of gaps right after L_{i^*} leaves but before new admissions, if any, are made. It follows from (5) that $G^- \geq K'$ in the event D^*, since

$$G^- = G - N_0^* + N_1^* + J(t_1) \geq \big(-\varepsilon K + (1-\varepsilon)K/r\big)^+ = K'.$$

The remaining analysis tacitly assumes that $r > 1$, that $N_0 + N_1 \leq \varepsilon K$, and that the channel L_{i^*} leaves at t_1.

– *Third case*: $u > h + s_{i^*}$. Under this condition, $A(t_1) = 0$, since when L_{i^*} leaves it does not provide enough additional bandwidth for U. In particular, $R(t_1) = r - 1$ and $G(t_1) = G^- \geq K'$, and so

$$\mathbb{E}_x\big(\Delta\sigma(t_2)\big) \leq \xi + 1 - \mathbb{E}_x\big(D_0(t_2) + D_1(t_2); D^*\big).$$

The strong Markov property makes it possible to lower-bound this last term.

$$\mathbb{E}_x\big(D_0(t_2) + D_1(t_2); D^*\big) = \mathbb{E}_x\big(\mathbb{E}_{X(t_1)}\big(D_0(t_1) + D_1(t_1)\big); D^*\big)$$
$$= \mathbb{E}_x\left(\frac{(N_1 + N_2)(t_1)}{R(t_1)}; D^*\right)$$
$$\geq \frac{K' - 1}{r - 1}\mathbb{P}_x\big(D^*\big) = \frac{K' - 1}{r(r - 1)}$$

and therefore, $\mathbb{E}_x(\Delta\sigma(t_2)) \leq \xi + 2 - K'/(r(r-1))$.

– *Fourth case*: $u < h + s_{i^*} < \eta\alpha$. In this case U is admitted at t_1. Thus it makes sense to define the event

$$E_4 = D^* \cap \{U \text{ leaves at } t_2 \text{ and } U_1 > \eta\alpha\}.$$

Then as before

$$\mathbb{E}_x\big(\Delta\sigma(t_3)\big) \leq \xi + 1 - \mathbb{E}_x\big(D_0(t_3) + D_1(t_3); E_4\big).$$

In the event E_4, U is admitted at t_1 and leaves at t_2, while U_1 stays blocked at t_1 and t_2, so that $G(t_2) = G^- \geq K'$ and $R(t_2) = r - 1$. Hence as in the second case,

$$\mathbb{E}_x\big(D_0(t_3) + D_1(t_3); E_4\big) \geq \frac{K' - 1}{r - 1}\mathbb{P}_x(E_4) \geq \frac{(1 - \eta)K'}{r^2(r - 1)} - 1$$

since $\mathbb{P}_x(E_4) = (1 - \eta)/r^2$. Thus (13) holds.

   – *Fifth case*: $u < h + s_{i*}$ and $\eta\alpha < h + s_{i*}$. Again, U is admitted at t_1. Letting U_i denote the sizes of the requests behind U, define the event

$$B = \{U_i < \eta\alpha, \ i = 1, \dots, \tau \ \text{and} \ U_{\tau+1} > 2\eta\alpha\}$$

with $\tau = \inf\{n \geq 0 : U_1 + \cdots + U_n > h + s_{i*} - \eta\alpha\}$ and $E_5' = D^* \cap B \cap \{U \text{ leaves at } t_2\}$. It is readily verified that $1 \leq \tau < +\infty$ almost surely. Moreover, one has in E_5'

$$0 < h^* \overset{\text{def}}{=} h + s_{i*} - (U_1 + \cdots + U_\tau) < \eta\alpha < U_{\tau+1}.$$

This means that at t_2, exactly τ new requests $U_1, \dots, U_\tau$ have been admitted, and $U_{\tau+1}$ is blocked. Moreover, for any $i \in \{1, \dots, \tau\}$, one has $h^* + U_i < 2\eta\alpha < U_{\tau+1}$, so that if one of the τ channels allocated to the (U_i) leaves, $U_{\tau+1}$ remains blocked.

   When L_{i*} left, there were $G^- \geq K'$ gaps; in the remainder of the analysis, we call an *initial gap* a gap present right after L_{i*} left. After L_{i*} left, U and $A(t_1) - 1$ new requests were admitted, and then U left and $A(t_2)$ new requests were admitted at t_2. Thus, at t_2, each initial gap is in either of two states: either it is completely filled, or it is still a gap, i.e., it has not been filled completely. Let k be the number of initial gaps completely filled at t_2, and let $k' = G^- - k$: then $k + k' = G^- \geq K'$. In each initial gap completely covered at t_2, there is at least one type-2 fragment of one of the τ new channels. Therefore, $N_{1,2} + N_{2,2} + \cdots + N_{\tau,2} \geq k$ with $N_{i,2}$ the number of type-2 fragments of the channel corresponding to U. In particular there is a channel L_{j*}, $j* \in \{1, \dots, \tau\}$ with at least the average k/τ of type-2 fragments: $N_{j*,2} \geq k/\tau$. Define finally the event $E_5 = E_5' \cap \{L_{j*} \text{ leaves at } t_3\}$. Since $h^* + U_{j*} < U_{\tau+1}$, then $U_{\tau+1}$ remains blocked at t_3 when E_5 occurs, and therefore (note that when $j*$ leaves, some gaps may merge, but not two initial gaps),

$$G(t_3) \geq N_{j*,2} + k' \geq k/\tau + k' \geq (k + k')/\tau \geq K'/\tau.$$

Now we proceed as before to obtain

$$\mathbb{E}_x\big(\Delta\sigma(t_4)\big) \leq \xi + 1 - \mathbb{E}_x\big(D_0(t_4) + D_1(t_4); E_5\big)$$

and, using the Markov property at time t_3,

$$\mathbb{E}_x\big(D_0(t_4) + D_1(t_4); E_5\big) = \mathbb{E}_x\big(\mathbb{E}_{X(t_3)}\big(D_0(t_1) + D_1(t_1)\big); E_5\big)$$

$$= \mathbb{E}_x\left(\frac{(N_0 + N_2)(t_3)}{R(t_3)}; E_5\right)$$

$$\geq \mathbb{E}_x\left(\frac{(K'/\tau - 1)^+}{r + \tau - 2}; E_5\right)$$

since $R(t_3) = r + \tau - 2$ in E_5. The same kind of reasoning as before then leads to

$$\mathbb{E}_x\left(\frac{(K'/\tau - 1)^+}{r + \tau - 2}; E_5\right) \geq \frac{K'}{r^5} f(\eta, h + s_{i*} - \alpha\eta) - 1$$

with the function $f(\eta, \cdot)$ defined for $y > 0$ by

$$f(\eta, y) = \mathbb{E}\big((1 + \tau(y))^{-5}; B(\eta, y)\big)$$

with $\tau(y) = \inf\{n \geq 1 : U_1 + \cdots + U_n \geq y\}$ and

$$B(\eta, y) = \big\{U_i < \eta\alpha, i = 1, \ldots, \tau(y) \text{ and } U_{\tau(y)+1} > 2\eta\alpha\big\}.$$

It is not difficult to show that $\gamma(\eta) = \inf_{0 < y < 1} f(\eta, y) > 0$, which then gives the result.

It remains to prove (15). One only needs to assemble the various bounds, taking into account that $\mathbb{E}_x(\Delta\sigma(t_i)) \leq \xi + 1$ for any $x \in \mathcal{S}$ and $i \geq 0$, to arrive at 4 separate bounds on $\mathbb{E}_x(\sigma(t_4) - \sigma)$. For example, using the former bound for the first two terms and the last term of $\mathbb{E}_x(\sigma(t_4) - \sigma) = \sum_{1 \leq i \leq 4} \mathbb{E}_x \Delta\sigma(t_i)$ and then the bound in (13) for the third term, we find that one of the four bounds, which applies when x satisfies the inequalities of the fourth case, is

$$\mathbb{E}_x\big(\sigma(t_4) - \sigma\big) \leq 4\xi + 5 - \frac{(1 - \eta)K'}{r^5}$$

Computing the minimum over these bounds with $\eta = 1/4$ and $\varepsilon = 1/r^2$, one obtains (15) after setting $\bar{\xi} = 4\xi + 5$ and

$$\psi(r) = \frac{\varphi(r)}{r^6} \times \big((1 - \eta) \wedge \gamma\big)$$

with $\varphi(r) = 1 - 2r^{-2}$. This concludes the proof. □

The bound (15) gives a negative drift for σ when the term $K\psi(R(0))$ is large enough. But $\psi(r)$ vanishes when r goes to infinity, and so this bound cannot yield a drift uniformly negative, the problem occurring when $R(0)$ is large. For this reason σ is not a Lyapunov function but the simple modification $\sigma_\kappa = \sigma + \kappa R$ introduced earlier is. The purpose of the additional term κR is precisely to give a negative drift when $R(0)$ is large.

5.3 Construction of a Lyapunov function

Since the variation $\Delta R(t_k)$ in the number of channels at a departure is exactly equal to $A(t_k) - 1$, one readily gets that

$$\sigma_\kappa(t_4) - \sigma_\kappa = (\sigma(t_4) - \sigma) + \kappa(A(t_1) + \cdots + A(t_4) - 4)$$

with $\sigma_\kappa = \sigma + \kappa R$. In particular, if $x \in S$ is such that $\sigma \geq 2K + 1$ and $r = R(0) \leq K_1$ for some $K_1 \geq 0$, then (from now on, we assume without loss of generality that the function ψ given by (15) in Lemma 5 is decreasing)

$$\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \leq \overline{\xi} - K\psi(K_1) + 4\kappa(\xi - 1)$$

whereas if $r \geq K_1$,

$$\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \leq 4(\xi + 1) + \kappa\mathbb{E}_x(r(t_4) - r)$$

and so we see that we only need to control $\mathbb{E}_x(R(t_4) - r)$ for r large. In this case, the negative drift comes from the fact that, except perhaps at t_1, with high probability there is no admission at a departure, since the channel that leaves is small with high probability.

Lemma 6 *There exist $K_1, \gamma_1 > 0$ such that if $x \in S$ is such that $r = R(0) \geq K_1$, then $\mathbb{E}_x(R(t_4) - r) \leq -\gamma_1$.*

Sketch of Proof The proof of this inequality is easier than the proof of Lemma 5, we only give a sketch of it. From $s_1 + \cdots + s_r = 1 - h \leq 1$ one gets $\#\{i : s_i \geq \gamma\} \leq 1/\gamma$, and therefore, $\mathbb{P}_x(s_{i_1} \geq 1/\sqrt{r}) \leq 1/\sqrt{r}$ with L_{i_1}, $i_1 \in \{1, \ldots, r\}$, the channel that leaves at t_1. Thus, when r is large, with high probability a small channel leaves.

If $h - u$ is away from α, then the event $\{u_1 > h - u + s_{i_1} + \cdots + s_{i_4}\}$ (with i_k defined similarly) has high probability, and in this event $A(t_1) + \cdots + A(t_4) \leq 1$. If in contrast $h - u$ is large, then with high probability U_1 is admitted and with high probability $h - u - U_1$ is away from α; hence we can do the same again, and get that, with high probability, $A(t_1) + \cdots + A(t_4) \leq 2$. □

Gathering Lemma 6 and the bound (15) we can now prove that for κ suitably chosen, the function $\sigma_\kappa = \sigma + \kappa R$ is indeed a Lyapunov function.

Proposition 1 (Lyapunov function inequality) *There exist κ and $K > 0$ such that if $x \in S$ is such that $\sigma_\kappa \geq K$, then $\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \leq -1$.*

Proof Let K_1 and γ_1 be given by Lemma 6, and take κ and K as follows:

$$\kappa = \frac{4\xi + 5}{\gamma_1} \quad \text{and} \quad K = \frac{8\xi + 2\overline{\xi} + 2}{\psi(K_1)} + \kappa K_1 + 1.$$

Assume that $\sigma_\kappa \geq K$. If $r = R(0) \geq K_1$, then

$$\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \leq 4(\xi + 1) - \kappa\gamma_1 = -1.$$

Otherwise, $r \leq K_1$, and since $\sigma_\kappa \geq K$, this necessarily gives $\sigma \geq K - \kappa K_1 = 2\hat{K} + 1$ with $\hat{K} = (K - \kappa K_1 - 1)/2$. Thus,

$$\mathbb{E}_x\big(\sigma_\kappa(t_4) - \sigma_\kappa\big) \leq \overline{\xi} - \hat{K}\psi(K_1) + 4\xi = -1$$

and the proposition follows. □

5.4 Proof of Theorem 2

Theorem 4 and Lemma 3 applied to the Markov chain $(X(t_{4n}), n \geq 0)$, and the function σ_κ show that for some $\eta > 0$ and some constant $0 \leq C < +\infty$,

$$\sup_{n \geq 0} \mathbb{E}_x\big(e^{\eta\sigma_\kappa(t_{4n})}\big) \leq e^{\eta\sigma_\kappa} + C.$$

Then the Markov property gives for any $i \geq 0$

$$\sup_{n \geq 0} \mathbb{E}_x\big(e^{\eta\sigma_\kappa(t_{4n+i})}\big) \leq \mathbb{E}_x\big(e^{\eta\sigma_\kappa(t_i)}\big) + C < +\infty$$

from which (7) follows readily.

5.5 Proof of Theorem 3

In the following discussion, no conceptual argument is missing, only some formalism needed to handle the continuous state space $\mathcal{S}$. These details are routine and left to the interested reader. In the analysis below, requests are said to be *big* if their size exceeds $1/2$.

We argue that $(X(t))$ visits infinitely often a state in which there are no fragmented channels, and such that all size distributions remain the same at all visits. This is enough to show Harris recurrence; see for instance Asmussen [2]. For this purpose, it is convenient to pick a simple regeneration set $E \subset \mathcal{S}$ in which (i) the spectrum is being used by a big request, alone and with an unfragmented channel of the form $(0, b)$, and (ii) the request U waiting at the head of the queue is also big. Each of these has the conditional request-size distribution given that its size is larger than $1/2$, i.e., the uniform distribution on $(1/2, \alpha)$, see [13].

To verify that E is visited infinitely often, consider the process $(R(t), U(t))$ with $R(t)$ the number of requests allocated a channel at time t and $U(t)$ the size of the request at the head of the queue; the process $(R(t), U(t))$ is simply the process $(X(t))$ when the data on fragmentation is ignored. This process is positive Harris recurrent, as shown in Kipnis and Robert [13]. In particular it visits infinitely often states with $R(t) = 1$ and $U(t) > 1/2$; one can add $U_1(t) > 1/2$ as well (i.e., the first request in line behind the head of the queue request is also big), since this happens with a geometric probability. Then when the only channel leaves, the process $(X(t))$ enters E, since there is exactly one channel, it is big, it is necessarily unfragmented and of the form $(0, b)$, and a big request is waiting at the head of the queue. Moreover, this argument shows that the time between visits to E is integrable, which in turn establishes positive Harris recurrence. This completes the proof.
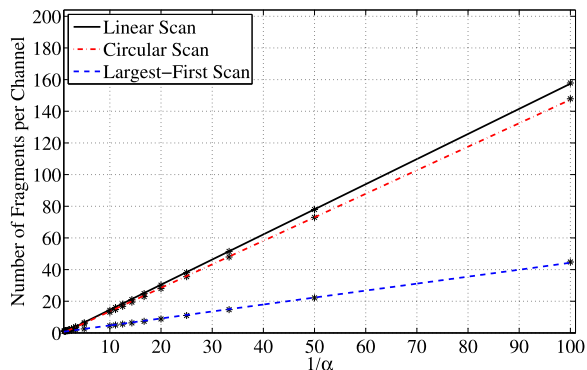
6 Algorithms

Although the focus so far has been on measures of fragmentation as a function of α, algorithmic issues are also of obvious interest. For example, more uniform patterns of gaps might be an advantage. The linear scan LS discussed in the previous sections, tends to push the gaps towards the end of the spectrum, particularly when the spectrum is viewed at random times in steady state. Interestingly, our experiments have shown that, for all $\alpha < 1/3$, the starting position of the first gap in the spectrum remains very close to 0.64.

To uniformize gap locations, an alternative gap scan resembles the circular-scan sequences of dynamic storage allocation [14]. In our case, circular scan (CS) uses a gap list, in which the successor to the last gap in [0, 1] is the first gap in [0, 1]. The scan is still linear like LS, but the starting gap of the scan moves as follows: if the last fragment of a channel is placed in gap g, then the residual gap of g is the first gap scanned in constructing the next channel. Clearly, although CS will tend to uniformize gap sizes as a function of position, boundary effects will persist so long as the spectrum itself is not circular, i.e., gaps and fragments are not allowed to overlap the end of the spectrum, a restriction that would likely be dictated in practice.

The average number of fragments per channel is a direct measure of gap-search times, and one that we use here. For values of α expected to be of interest in applications, the effects of a circular scan on gap-search times are only within a few percent relative to LS, as can be seen in Fig. 7. The figure also shows the average number of fragments per channel for LFS. This algorithm is designed to speed up the process of finding a set of gaps sufficient to create a new channel. It selects available gaps in a decreasing order of their sizes and allocates them to a request, thereby greedily minimizing the number of gaps needed to fulfill a request. The extra mechanism needed for such a search will of course tend to reduce overall performance gains. The results in Fig. 7 for LFS show a surprisingly large improvement in the average number of fragments per channel—as can be seen, a reduction by a factor more than 3 is achieved for even moderately small α.

The Probability Mass Functions (pmf's) of the number of fragments per channel are shown in Fig. 8. Notice that while all probabilities are small under LS and CS, the largest applies to the case of no fragmentation at all. The much more peaked distribution for LFS has both much smaller mean and variance: The standard deviation under
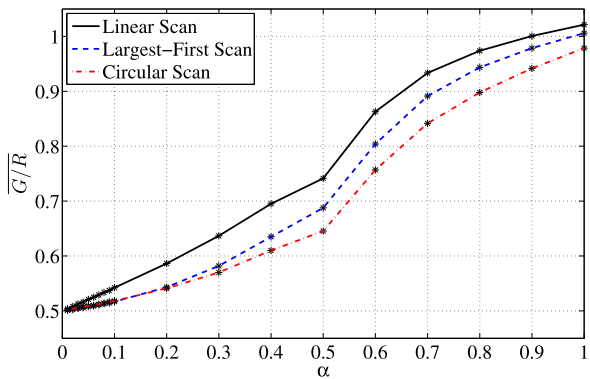

Fig. 7 Average number of fragments per channel

Fig. 8 Distribution of the
number of fragments per
channel for $\alpha = 0.1$
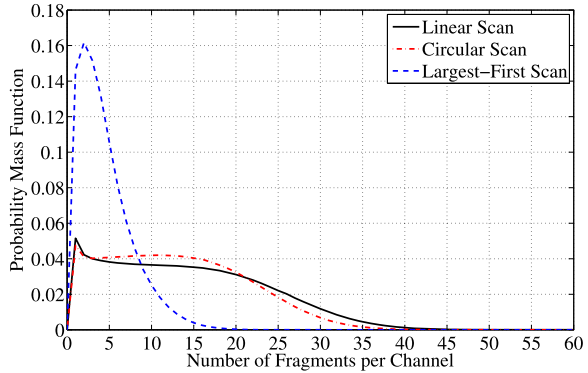
Fig. 9 $\overline{G/R} \to 1/2$ as $\alpha \to 0$
under LS, LFS, and CS

the linear algorithms is approximately 1.5 to 2.0 times that under LFS. The same
limiting behavior called for by the 50 % law holds for CS and LFS, which was to be
expected, as the arguments supporting the 50 % law did not depend on the sequence
in which gaps were scanned. But an interesting result of our experiments with CS is
that the 50 % approximation to expectation of the ratios is within a couple of percent
even when the maximum request size is as much as one fifth of the spectrum size. This
is easily seen in Fig. 9. The convergence rate of LFS to 50 % is intermediate between
LS and CS.

7 Normal approximations

After the usual scaling (i.e., first centering then normalizing by the standard devi-
ation), the scaled version of the number of channels, R, tends in probability to the
standard Normal, $\mathcal{N}(0, 1)$, as $M \to \infty$, where $M = \lfloor 1/\alpha \rfloor$ (it is convenient to ex-
press the asymptotics in this section in terms of M). This result follows easily from
the corresponding heavy-traffic limits in [4, 5], and the ergodicity of $(R(t))$.

Interestingly, it was discovered in the experiments that a normal-limit law also
appears to hold for the total number F of fragments as $M \to \infty$ (see for example
Fig. 10). This is not surprising, since F is the sum over all requests in system of

Fig. 10 Distributions of the total number of fragments and corresponding normal fits

(a) $\alpha = 0.5$

(b) $\alpha = 0.1$
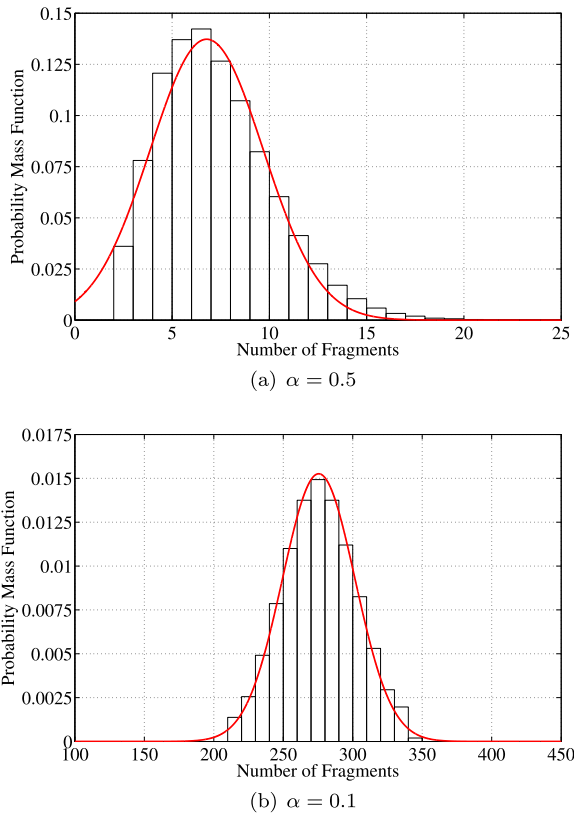
the numbers F_i, $1 \leq i \leq R$, of fragments allocated to requests. The requests have a mutual dependence, but one whose effect can be expected to weaken for large M. As illustrated in Fig. 7, the mean of F_i is proportional to M. Our experiment indicate that the standard deviation of F_i is also approximately proportional to M. Hence, if $\xrightarrow{p}$ denotes convergence in probability as $M \to \infty$, then

$$\frac{F - \beta M^2}{\theta M^{3/2}} \xrightarrow{p} \mathcal{N}(0, 1) \tag{18}$$

with $\beta \approx 1.5$ and $\theta \approx 0.9$ formalizes the normal-limit law suggested by our experiments.

Note the two departures from the standard Central Limit Theorem set-up. First, consistent with the linear fragmentation observation, individual request fragmentation scales linearly in M, so that the total number of fragments scales as M^2. Restating the limit law in terms of the sums of random variables (F_i/M) clearly eliminates this discrepancy. Second, the number of channels (R) is random and satisfies the normal-limit law discussed above. Thus, the plausibility of (18) requires an appeal to Central Limit Theorems for random sums (e.g., see [9, p. 258], for an appropriate version). Finally, Fig. 10 gives some idea of the convergence of the fits to the normal density; as can be seen, fits for $\alpha \leq 0.1$ are indeed close to the simulation data.

Fig. 11 Distributions of the total number of fragments for $\alpha = 0.1$ and the corresponding normal fits
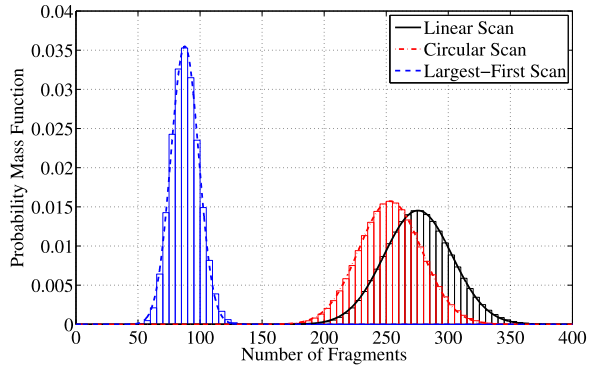
Fig. 12 Distributions of the number of gaps at different epochs for $\alpha = 0.01$
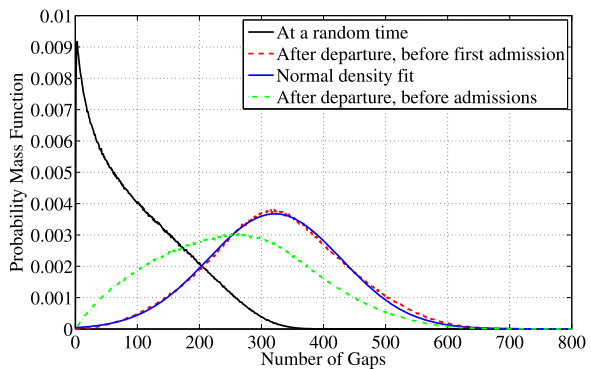
We note that the normal approximations shown for the total number of fragments under LS were also found to hold under CS and LFS. This is illustrated in Fig. 11.

The distributions of the number of gaps at different epochs are shown in Fig. 12. The distribution at a random time shows a decreasing pmf. What appeared to be yet another normal approximation was discovered when looking at the *first-admission* pmf in Fig. 12; this is the distribution as seen by the first admission immediately after a departure at just those epochs when there is at least one admission. The third curve is the pmf of the number, G^-, of gaps as seen right after departures, but before determining whether or not the head of the queue fits in the total available bandwidth. The fit to the Normal density is also illustrated in the figure. The proof of a similar limit law for Renyi's space-filling problem can be found in [6], but extension of known techniques once again faces the difficult challenges posed by our more difficult fragmentation problem.

8 Conclusions

The results of this paper prepare the ground for further research on several fronts. Before listing a number of the more important ones, we review what we have learned. Our experiments brought out first an unexpected reappearance of a *50 % rule* relating

the expected numbers of gaps and channels in the limit of small request sizes relative to the spectrum size. In our case, we were able to prove the limit law. Also, experimental results described a linear relationship between the inverse of the maximum request size, α, and the expected number of fragments into which a request was divided at the time of allocation. Interestingly, the smaller α was taken, the greater was the resulting fragmentation of requests.

Our stability results established the beginning of a mathematical foundation of fragmentation processes. Particularly, we showed that for $\alpha > 1/2$, the fragmentation process is Harris recurrent. For general α, we proved that the total number of fragments is bounded in expected value. We examined alternative algorithms for sequencing through the available gaps and showed that using the largest-first-scan algorithm leads to significantly less fragmentation than using the linear-scan and circular-scan algorithms. Finally, we exhibited experimentally a limiting, small-α behavior in which, with appropriate scaling, distributions tend to the normal.

A broad direction for further research is in extending the parameters of our mathematical model. For instance, while Uniform distributions are generally the assumption of choice in fragmentation models, it would be interesting to see what new effects are created by other distributions of request size, e.g., by varying a in the generalized uniform distributions on $[0, \alpha]$, with densities x^a/α^{a+1}. The exponential residence-time assumption is likely to yield simplifications to analysis, but changes in behavior resulting from other distributions are worth investigating. Moreover, instead of a system operating at capacity, in which there is always a request waiting, one could adopt an underlying, fully stochastic model of demand; e.g., a Poisson arrival process, as found in [13].

More realistic, but in all likelihood significantly more difficult models, would relax the independence assumptions. A prime example appropriate for Dynamic Spectrum Access applications would be allowing residence times to depend on fragmentation, the greater the fragmentation of a request, the longer its residence time.

The results regarding the performance of the different algorithms imply that the algorithms' design should also be considered carefully. Some examples of algorithms that come to mind will aim to better fit the fragments into the available gaps. A more challenging objective would be to develop algorithms that take into account spectrum sensing capabilities during the gap allocation process.

Finally, another broad and very important avenue of research that introduces more realistic models discretizes request sizes and the bandwidth allocation process (as is being done while allocating OFDM subcarriers). As in other models of fragmentation, the continuous limit represented in this paper may conceal important effects, or, conversely, it may introduce effects not present in discrete models. We are actively pursuing this avenue of research.

Appendix: Tool validation

In the first test of the simulation tool, we allowed arrivals, discretized the spectrum size, and fixed the requested bandwidth, in such a way that the system reduced to a conventional $M/M/k$ queue. We then matched simulation results with those obtained from classical formulas. In all cases checked, the error compared to the exact results was negligible.

Keeping with our continuous model operating at capacity with random bandwidth requests, one finds that there are very few explicit results for measures of interest, even for the process $R(t)$ for the number of channels in the system. Those that do exist can be found in [13] along with the elegant formula below, which unfortunately requires calculations that are rarely tractable.

Theorem 5 (Kipnis and Robert [13]) *Let S_n denote the sum of n i.i.d. request sizes. Then, the maximum throughput (expected number of requests in the system) is*

$$\mathbb{E}(R) = \frac{1}{\sum_{n=1}^{+\infty}(1/n)\mathbb{P}(S_n \leq 1 < S_{n+1})}$$

$$= \frac{1}{\sum_{n=2}^{+\infty}\big(1/n(n-1)\big)\mathbb{P}(S_n > 1)}.$$

A system with requests that are uniformly distributed on $[0, 1]$ is one case that admits of a simple formula. Here, the density of S_n on $[0, 1]$ is $z^{n-1}/(n-1)!$, and so an easy calculation gives

$$\mathbb{E}(R) = \frac{1}{e-2} = 1.392\ldots.$$

Transform methods are used in [13] to evaluate $\mathbb{E}(R)$ for other values of α. In particular, computations are based on the (numerical) inversion of a Fourier–Laplace

Table 1 Averages $\overline{R}$, computed from our experiments, vs. the corresponding values $\mathbb{E}(R)$, obtained from the numerical computations in [13], for various values of α. Except for the case $\alpha = 1$, the latter results were given to two decimal digits only

α	$\mathbb{E}(R)$	$\overline{R}$	α	$\mathbb{E}(R)$	$\overline{R}$
0.05	39.51	39.325	0.55	2.9	2.892
0.1	19.4	19.317	0.6	2.54	2.541
0.15	12.69	12.653	0.65	2.26	2.261
0.2	9.34	9.306	0.7	2.04	2.039
0.25	7.32	7.303	0.75	1.87	1.868
0.3	5.98	5.961	0.8	1.73	1.731
0.35	5.01	5.003	0.85	1.62	1.621
0.4	4.28	4.275	0.9	1.53	1.531
0.45	3.71	3.709	0.95	1.45	1.456
0.5	3.29	3.281	1	1.392	1.392

transform. To evaluate the accuracy of our simulator, we checked our experimental results against the computations for the values of α shown in Table 1. As can be seen, the numerics agree well with simulations.

References

1. Akyildiz, I.F., Lee, W.Y., Vuran, M.C., Mohanty, S.: NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**(13), 2127–2159 (2006)
2. Asmussen, S.: Applied Probability and Queues, 2nd edn. Springer, New York (1987)
3. Coffman, E., Robert, P., Simatos, F., Tarumi, S., Zussman, G.: Channel fragmentation in dynamic spectrum access systems—a theoretical study. In: Proc. ACM SIGMETRICS'10 (2010)
4. Coffman, E.G. Jr., Puhalskii, A.A., Reiman, M.I.: Storage limited queues in heavy traffic. Probab. Eng. Inf. Sci. **5**(4), 499–522 (1991)
5. Coffman, E.G. Jr., Reiman, M.I.: Diffusion approximations for storage processes in computer systems. In: Proc. ACM SIGMETRICS'83 (1983)
6. Dvoretsky, A., Robbins, H.: On the 'parking' problem. Publ. Math. Inst. Hung. Acad. Sci. **9**, 209–226 (1964)
7. FCC: 03-222. Notice of proposed rulemaking (2003)
8. FCC: 08-260. Second report and order, ET Docket No. 04-186, unlicensed operation in the TV broadcast bands (2008)
9. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. II, 2nd edn. Wiley, New York (1966)
10. Ghasemi, A., Sousa, E.: Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. IEEE Commun. **46**(4), 32–39 (2008)
11. Hajek, B.: Hitting-time and occupation-time bounds implied by drift analysis with applications. Adv. Appl. Probab. **14**(3), 502–525 (1982)
12. Jia, J., Zhang, Q., Shen, X.: HC-MAC: a hardware-constrained cognitive MAC for efficient spectrum management. IEEE J. Sel. Areas Commun. **26**(1), 106–117 (2008)
13. Kipnis, C., Robert, P.: A dynamic storage process. Stoch. Process. Appl. **34**(1), 155–169 (1990)
14. Knuth, D.E.: The Art of Computer Programming. Fundamental Algorithms, vol. 1, 3rd edn. Addison Wesley Longman Publishing Co., Redwood City (1997)
15. Mahmoud, H., Yucek, T., Arslan, H.: OFDM for cognitive radio: merits and challenges. IEEE Wirel. Commun. **16**(2), 6–15 (2009)
16. Meyn, S., Tweedie, R.L.: Markov Chains and Stochastic Stability, 2nd edn. Cambridge University Press, Cambridge (2009)
17. Mitola, J. III: Cognitive radio: an integrated agent architecture for software defined radio. Ph.D. thesis, Doctor of Technology, Royal Inst. Technol. (KTH), Stockholm, Sweden (2000)
18. Poston, J.D., Horne, W.D.: Discontiguous OFDM considerations for dynamic spectrum access in idle TV channels. In: Proc. IEEE DySPAN'05 (2005)
19. Rajbanshi, R., Wyglinski, A.M., Minden, G.J.: OFDM-based cognitive radios for dynamic spectrum access networks. In: Bhargava, V.K., Hossain, E. (eds.) Cognitive Wireless Communication Networks, pp. 165–188. Springer, Berlin (2007)
20. Shukla, A., Willamson, B., Burns, J., Burbidge, E., Taylor, A., Robinson, D.: A study for the provision of aggregation of frequency to provide wider bandwidth services. Tech. rep., QinetiQ (2006)
21. Tarumi, S.: Analysis of channel fragmentation in dynamic spectrum access networks. Ph.D. thesis, Columbia Universtiy, Electrical Engineering (2010)
22. Weiss, T.A., Jondral, F.K.: Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency. IEEE Commun. **42**(3), S8–14 (2004)
23. Yuan, Y., Bahl, P., Chandra, R., Moscibroda, T., Wu, Y.: Allocating dynamic time-spectrum blocks in cognitive radio networks. In: Proc. ACM MobiHoc'07 (2007)