
EURECOM
Department of Mobile Communications
Campus Sophia Tech
Les Templiers
450 route des Chappes
B.P. 193
06410 Biot
FRANCE

Research Report RR-14-296

# Underlay vs. Interweave: Which one is better?

November 07th, 2014
Last update December 19th, 2017

Fidan Mehmeti, Thrasyvoulos Spyropoulos

[1]EURECOM's research is partially supported by its industrial members: BMW Group, Cisco, Monaco Telecom, Orange, SAP, SFR, STEricsson, Swisscom, Symantec.

Tel : (+33) 4 93 00 81 00
Fax : (+33) 4 93 00 82 00
Email : {mehmeti,spyropou}@eurecom.fr

# Underlay vs. Interweave: Which one is better?

Fidan Mehmeti, Thrasyvoulos Spyropoulos

**Abstract**

Cognitive Networks have been proposed to opportunistically discover and exploit licensed spectrum bands, in which the secondary users' (SU) activity is subordinated to primary users (PU). Depending on the nature of the interaction between the SU and PU, there are two frequently encountered types of spectrum access: *underlay* and *interweave*. While a lot of research effort has been devoted to each mode, there is no clear consensus about which type of access performs better in different scenarios and for different metrics. To this end, in this paper we approach this question analytically, and provide closed-form expressions that allow one to compare the performance of the two types of access under a common network setup. This allows an SU to decide when one type of access technique or the other would provide better performance, as a function of the metric of interest and key network parameters. What is more, based on this analysis, we propose dynamic (hybrid) policies, that can decide at any point to switch from the one type of access to the other, offering up to another $50\%$ of additional performance improvement, compared to the optimal "static" policy in the scenario at hand. We provide extensive validation results using a wide range of realistic simulation scenarios.

**Index Terms**

Cognitive radio networks, Markov chains, Interweave, Underlay.

# Contents

# List of Figures

# 1 Introduction

Lately, we are witnessing a tremendous increase in the number of data-enabled wireless devices (smartphones, tablets, etc.) as well as in the applications and services that they provide. Coupled with the equally large market growth envisioned for the numerous small and large "things" requiring wireless connectivity [1], this creates a huge pressure on wireless network operators, and a resulting increase in spectrum demand.

Because of this, and due to the static spectrum allocation policies followed by authorities worldwide, spectrum scarcity has become a major problem in today's wireless industry. Nevertheless, measurements of the utilization of licensed wireless spectrum in fact reveal that the available spectrum is rather under-utilized, exhibiting high variability across space, frequency and time [2].

To address this issue, dynamic spectrum access techniques have recently been proposed, with cognitive radio (CR) as its key technology [3]. In a cognitive network, there exist licensed users, known as primary users (PU), which are provided the spectrum from the regulation authority, as well as unlicensed users that are known as cognitive or secondary users (SU) utilizing the spectrum opportunistically. Cognitive users are subordinated to primary users' activity. Hence, they have to adapt their transmission parameters, so that there are no impairments on PU Quality of Service (QoS).

One of the main functions of CRs is spectrum access [3]. Spectrum access is very important to prevent potential collisions between the SUs and PUs. Spectrum access techniques can be classified as: *underlay*, *interweave*, and *overlay*. In this paper, we are concerned only with the first two techniques. In the underlay mode (Fig. 1), the SU reduces the transmission power when a PU is utilizing a given channel such that the maximum interference level a PU can tolerate is not exceeded. In the interweave mode (Fig. 2), the cognitive user can transmit only when there is no PU, with the maximum power in accordance with the spectral mask. Whenever a PU claims a channel back, the SU must immediately cease its transmission and look for another *white space*, i.e. a part of the spectrum that is currently not utilized by its PU. In the overlay mode, the cognitive user serves as a relay to a licensed user and in turn the PU allows it to access to a portion of its spectrum. However, the necessity of complete channel knowledge from both PU and SU increases complexity and makes this mode less attractive.

A large number of works exist for both underlay and interweave access in CRNs [4, 5]. While some arguments for the one or the other exist (often related to the potential harm to PUs [6]), there is little consensus regarding which mode would offer the best performance to SUs.

While the possibility of transmitting without interruptions (usually causing issues to higher layer protocols) is certainly an advantage that underlay access offers, there are also some drawbacks associated with it. First, the user in this mode can transmit with maximum rate only when the PU is silent. These periods of time can be much shorter than the periods with PU, especially when dealing with high duty

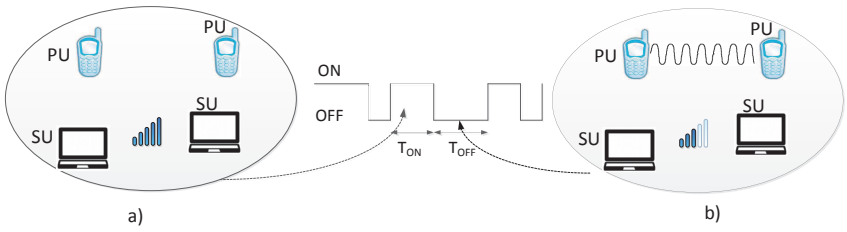


Figure 1: The illustration of the underlay mode with: a) idle PU (high periods), b) active PU (low periods).

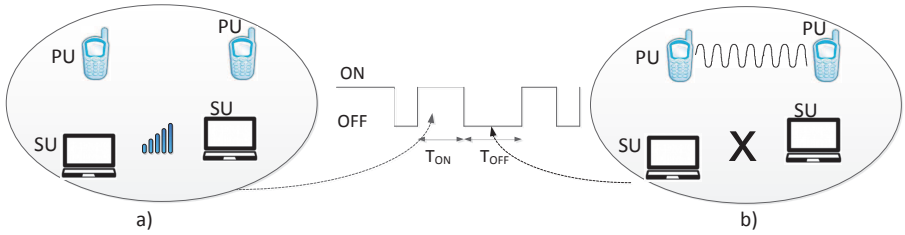


Figure 2: The illustration of the interweave mode with: a) idle PU (high periods), b) active PU (low periods).

cycle licensed users. Furthermore, if that PU is located in the vicinity of the SU, the SU's transmission rate might have to be significantly reduced. This could significantly reduce the effective (average) transmission rate and the resulting throughput.

Contrary to this, in interweave mode the SU can look for another idle channel when the PU arrives (possibly with a lower duty cycle) and start transmitting again at full power, possibly improving the average (long-term) throughput. Yet, the intermittent nature of communication relying on interweave access may delay some application flows (e.g. a request to transmit a short file, fetching a web page, etc.) significantly, if they happen to arrive while the SU is scanning for a new available channel. Such delays can be exacerbated not only if the SU resides in a relatively busy part of the spectrum (e.g. urban areas at "peak" hours), but even if the variability of this scanning time is high (e.g. sometimes a new channel is found quickly, but sometimes the SU might be stuck scanning for a long time).

Based on the above discussion, it is obvious that there are a number of tradeoffs involved, and it is not easy to say, a priori, which mode of spectrum access would perform better in a given scenario. The relative performance has a close dependence not only on specific network parameters (e.g. PU duty cycle, allowed transmission power, etc.), but also on the performance metric of interest, the type of SU traffic (sparse, frequent), size of requests, and even higher order statistics of key parameters, such as the time to find a new white space.

To this end, in this paper we approach this problem using an analytical framework to evaluate the individual performance of underlay and interweave access, as

well as to compare them in a range of settings. Our contributions can be summarized as follows:
(i) We derive closed-form expressions for the expected delay for underlay and interweave spectrum access as a function of key network parameters (average PU idle time, transmission rates, scanning time statistics), and user traffic statistics (traffic intensity, file size). This allows us to directly compare the performance of the two, and derive the conditions that would make the one or the other preferable. Finally, we also use these insights to propose a "hybrid" policy, that can switch between the two dynamically, in order to further improve performance. (ii) Using a wide range of realistic simulation scenarios, we validate our analytical predictions extensively, explore the conditions under which underlay or interweave policies perform better, and show when the dynamic policies can indeed offer additional performance improvements.

# 2 Performance modeling of spectrum access

We assume that traffic flows arrive randomly as a Poisson process with rate $\lambda$. The file sizes are assumed to be exponentially distributed. When a file arrives to find another file in the system, it will be queued. We consider First Come First Served (FCFS) order of service. The total time a file spends in the system is the sum of the service and queueing time, and is referred to as the *system time*. We use also the term *transmission delay* interchangeably with system time.

## 2.1 Problem setup for underlay access

In underlay access, the SU can transmit at full power when there is no PU communicating on that channel. When the PU resumes its transmission, the SU has to reduce its transmission power, so that there are no impairments on the PU transmission quality. Although the actual power allowed depends on the primary and the interfering channel quality (distance, LoS, etc.), we will assume for simplicity that the SU power can vary between two levels: "high" power when there is no PU, and "low" power when there is PU activity (e.g. perceived as an average value).

Consider a channel used by one or more PU. The occupancy of that channel can be modeled as an ON-OFF alternating renewal process [7] $\left(T_{ON}^{i}, T_{OFF}^{i}\right),\ i \geq 1$, as shown in Fig. 1. ON periods represent the absence of the PU on that channel, while the OFF periods denote the periods of time with active PU. $i$ denotes the number of ON-OFF cycles elapsed until time $t$. Unless otherwise stated, the duration of any ON period $T_{ON}$ is assumed to be exponentially distributed with parameter $\eta_H$, and is independent of the duration of any other ON or OFF period. Similarly, the duration of an OFF period is also assumed to be exponential, but with parameter $\eta_L$. While this assumption is necessary for the tractability of the delay analysis, as we shall see, we consider generic ON/OFF periods in the throughput analysis. Generic ON/OFF distributions could be introduced also in our delay

analysis by considering phase-type distributions and matrix analytic methods [8]. However, such methods only yield numerical solutions, that do not allow for direct analytical performance comparisons. What is more, simulation results (Section 6) suggest that, even for generic ON/OFF period distributions, the accuracy of our predictions is sufficient.

The data transmission rate during the ON periods is denoted with $c_H$, while during OFF periods the date rate is $c_L$. The actual values for these depend on technology, channel bandwidth, coding and modulation, etc. However, since the allowed transmission power during OFF periods is higher, it holds that $c_H > c_L$.

### 2.2 Problem setup for interweave access

In the interweave mode, the SU can transmit only when there is no PU activity (ON periods). It is again assumed that the periods with no PU activity are exponentially distributed with parameter $\eta_H$, and the data rate is $c_H$.

However, after the arrival of a PU in the channel (at the end of that ON period), the SU does not continue transmitting (at lower power), as in the underlay case, but starts looking for another available channel. As soon as it finds one, it resumes transmission at full power (i.e. with rate $c_H$). Consequently, we can again model this system with an alternating renewal process. However, OFF periods now correspond to scanning intervals during which no data can be transmitted, i.e. $c_L = 0$.[1] Hence, it changes its operation to the *scanning* mode. During the scanning mode (i.e. during an OFF period), the SU moves to a new channel and senses it for some time. If available, it resides there and goes back to the transmission mode. Otherwise, it switches to another frequency and senses another channel, and so on until finally an available channel is found[2], and the transmission process is resumed. Hence, the scanning time corresponding to one channel is actually the sum of the sensing time ($T_I$) and the switching delay ($T_{switch}$) introduced. So, the total scanning time can be expressed as

$$T_s = L\left(T_I + T_{switch}\right), \tag{1}$$

where $L$ is a random variable denoting the number of channels a SU has to sense until it finds the first available. We assume that sensing time per channel is much shorter than the ON and OFF periods.

[1] We assume that in both modes a single radio and antenna is used. Hence, a SU can only transmit or scan at any time, but not both. In contrast, to detect that the PU is back, we could just switch the radio periodically to receive mode, take a short time sample (in the order of $\mu$s) and do energy detection, to see if there is a PU signal [9]. Since we assume that the sensing time is much shorter than the actual durations of ON and OFF periods, we can safely ignore these sensing periods. However, the switching time is usually much higher than the sensing time (order of *ms* or even seconds) and cannot be ignored.

[2] We assume that the SU chooses channels sequentially from a list [10].

The switching delay while moving from the channel with frequency $f_s$ to the channel with frequency $f_d$ can be expressed as [10]

$$T_{switch} = \beta \frac{|f_d - f_s|}{\delta}, \tag{2}$$

where $\beta$ is the delay to move to the first contiguous channel, and is hardware dependent [10], while $\delta$ denotes the frequency separation between two neighboring channels.

By looking at Eq.(1), we can infer the following. If the probabilities of finding each channel available are independent and almost equal, we can say that the random variable $L$ is geometrically distributed. If we further assume that the switching time is the same when moving from one to another channel, and the well known fact that the geometric distribution can be obtained by rounding the exponential, we can infer that the scanning time can be approximated by an exponential distribution. For that purpose exponential scanning time distribution is assumed first. However, the nature of scanning time distribution depends heavily on the availability of the backup channels and on the frequency distance between them. Under most scenarios (high discrepancy on the availability probabilities between different channels, very low duty cycle of all the channels, the existence of available channels in the more remote parts of the spectrum etc.), the exponential assumption on the scanning time will not hold. For that reason, we also analyze the system for the scenarios when the switching time underlies high and low variance distributions.

Let's assume that the eligible channels have roughly the same duty cycles (% of time the PU is active on a channel) that are very low (sporadic activity of PUs), and that they are "neighbors" in the frequency context. Under this scenario, we would expect that the time needed to find an available channel is almost constant (most of the time only one channel needs to be sensed) and will not deviate much from the average value, $E[T_s]$. Hence, in this case the scanning time distribution would have a low variance (lower than the exponential distribution). A convenient and generic way to model such low variance distributions is by using a k-stage Erlang distribution [11].

As opposed to the previous scenario, there might be a number of channels with high duty cycles, and there might be one or few channels with much lower duty cycle located further away in the spectrum. Although there is a low probability, it may happen that all these channels located close to each other are busy, and the SU ends up searching the channel that is far away in the spectrum band. Since the switching time is proportional to the frequency difference, the scanning time will be considerably larger. So, there is a chance that some of the scanning time samples will deviate from the average to a considerable extent. Hence, in those cases the scanning time distribution can be considered as heavy-tailed, and the exponential assumption cannot capture that behavior. For that purpose, we model the scanning time with a hyperexponential distribution, in which with a probability $p_0$ the scanning time will be exponentially distributed with parameter $\eta_L$, and with

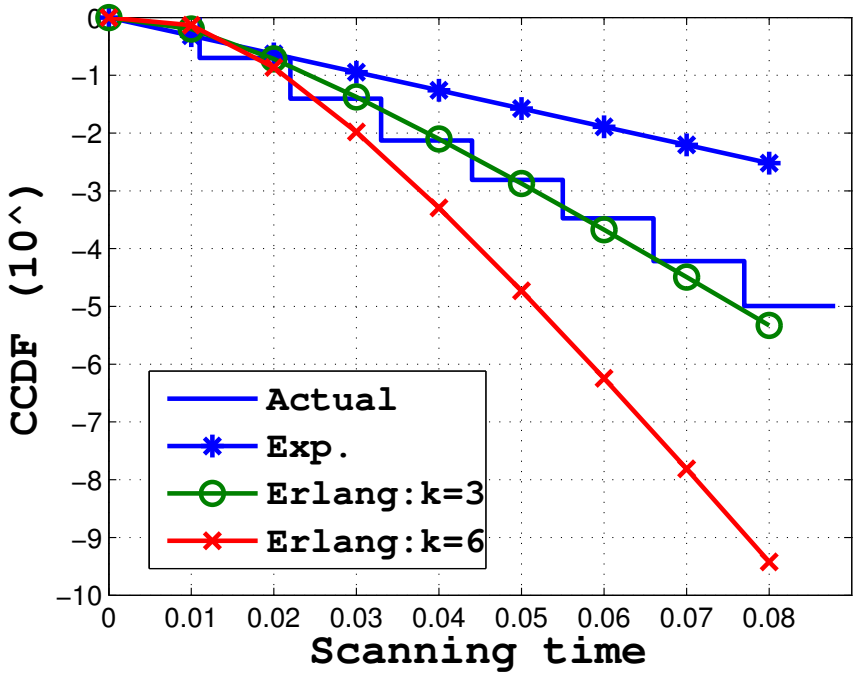


Figure 3: The distribution of the scanning time for low PU duty cycles.

a small probability $1 - p_0$ it will be exponentially distributed with parameter $\eta_V$. Note that $\eta_V << \eta_L$.

To support our aforementioned claims, we consider two scenarios. First, we assume that there is a group of 20 channels, which are close to each other in the spectrum. The duty cycles of the channels are all low (0.2), PU activities are i.i.d., $T_I = 1$ ms, and $T_{switch} = 10$ ms. Fig. 3 shows the complementary cumulative distribution function (CCDF) of the scanning time durations. On the same plot, the CCDFs of the exponential and Erlang distributions for $k = 3$ and $k = 6$, are also shown. The plot demonstrates that the exponential distribution cannot really capture the behavior of the system. Instead, an Erlang distribution needs to be used. Similar conclusion for the inability of the exponential distribution to capture the scanning time can be inferred from Fig. 4. As opposed to the previous scenario, in this case the channels have very high duty cycles (0.8), and the switching time between this group and the group with 2 channels (with a duty cycle of 0.2) is 0.5s. The hyperexponential distribution (shown also in the plot) has the following parameters: $p_0 = 0.2$, $\lambda_1 = 90$, $\lambda_2 = 3$. As can be seen, the hyperexponential distribution can capture to a better extent this scenario.

The above results highlight the need to consider more general distributions for scanning times directly in our analysis. Surprisingly, as we shall see, closed-form results for the system delay can still be found for such generic scanning times.

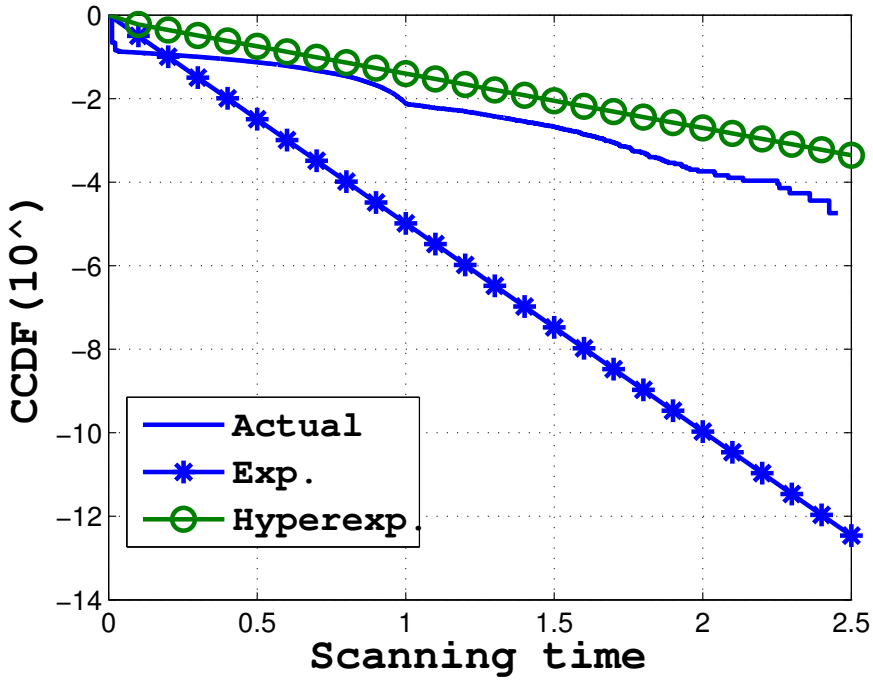


Figure 4: The distribution of the scanning time for two group of channels.

Before proceeding any further, we summarize in Table 1 some useful notation that will be used throughout the rest of the paper.

# 3 Delay analysis of underlay and interweave access

In this section we will derive the formulas for the average file delay for interweave and underlay access. For the former one, we perform the analysis over different scanning time distributions: exponential, k-stage Erlang and hyper exponential. For that purpose we use 2D Markov chain models, and the Probability Generating Functions (PGF) approach to derive the delay.

## 3.1 Delay analysis for interweave access

**Exponential scanning time.** Due to the assumptions made in Section 2.2 about the exponential distributions, we can model our system with a 2D Markov chain, shown in Fig. 5. Each state in this chain indicates the number of files present in the system and the presence (lower states) or absence (upper states) of the PU. $\pi_{i,L}$ ($\pi_{i,H}$) denotes the stationary probability of finding $i$ files when there is (not) a PU active on that channel. The transition rates $\eta_H$ and $\eta_L$ are the parameters of the exponentially distributed ON and OFF periods. While in the upper parts of the

Table 1: Variables and Shorthand Notation.

| Variable | Definition/Description |
|---|---|
| $T_{ON}$ | Duration of PU idle periods |
| $T_{OFF}$ | Duration of PU busy periods or scanning time |
| $\lambda$ | Average file arrival rate at the mobile user |
| $\pi_{i,L}$ | Probability of finding $i$ files in the OFF (low) period |
| $\pi_{i,H}$ | Probability of finding $i$ files in the ON (high) period |
| $\pi_{i,V}$ | Probability of finding $i$ files in the OFF (V-state) period |
| $\eta_H(\eta_L)$ | The rate of leaving the ON (OFF) state |
| $\eta_V$ | The rate of leaving the V-state |
| $\mu_H = \frac{c_H}{\Delta}$ | The service (transition) rate while being in a high state |
| $\mu_L = \frac{c_L}{\Delta}$ | The service (transition) rate while being in a low state |
| $\Delta$ | The average file size |
| $E[T]$ | The average system (transmission) time |

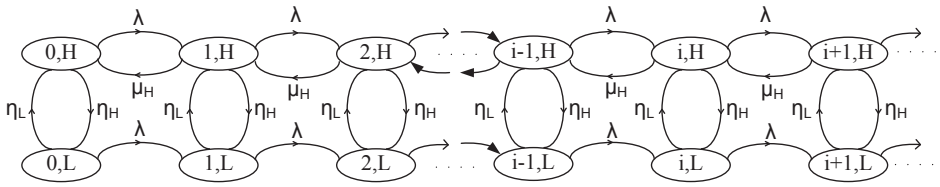


Figure 5: The 2D Markov chain for exponentially distributed scanning time.

chain there are transitions between states $(i, H)$ and $(i-1, H)$ with rates equal to $\mu_H = \frac{c_H}{\Delta}$, in the lower part (corresponding to the active PU), there is no transition going from state $(i, L)$ to $(i-1, L)$. This is a consequence of the inability of the SU to transmit while scanning. The transition rate from low to high periods is $\eta_L$, with exponential scanning time of average duration $E[T_s] = \frac{1}{\eta_L}$.

**Theorem 1.** *The average file delay in a cognitive radio network with interweave spectrum access and exponentially distributed scanning time is*

$$E[T_{exp}] = \frac{\eta_H(\eta_H + \mu_H)(E[T_s])^2 + 2\eta_H E[T_s] + 1}{(1 + \eta_H E[T_s])(\mu_H - \lambda - \lambda\eta_H E[T_s])}. \tag{3}$$

*Proof.* As is well known from the theory of Markov chains, a balance equation is an equation that describes the flow rate in and out of states. Basically, in equilibrium, the flow rate into a state is equal to the flow rate out of the same state. For the state $\{0, L\}$(Fig. 5), the possible transitions are to states $\{1, L\}$ and $\{0, H\}$, and these occur with rates $\lambda$ and $\eta_H$, respectively. So, the flow rate out of state $\{0, L\}$ is given by $\pi_{0,L}(\lambda + \eta_L)$. At the same time, the transitions into state $\{0, L\}$ can occur only from state $\{0, H\}$. So, the flow rate into state $\{0, L\}$ is $\eta_H \pi_{0,H}$. The first balance equation is:

$$\pi_{0,L}\,(\lambda + \eta_L) = \eta_H \pi_{0,H}. \tag{4}$$

The balance equations related to state $\{0, H\}$ are similar. There are two possible transitions out of this state: to states $\{1, H\}$ and $\{0, L\}$, with transition rates $\lambda$ and $\eta_H$, respectively. Hence, the flow rate out of this state is $\pi_{0,H}(\lambda + \eta_H)$. As opposed to state $\{0, L\}$, transitions into state $\{0, H\}$ can occur from two states: $\{0, L\}$ and $\{1, H\}$, with corresponding rates $\eta_L$ and $\mu_H$, respectively. As a result, the flow rate into this state is $\eta_L \pi_{0,L} + \mu_H \pi_{1,H}$. As the flow rates into and out of a state need to be equal, we obtain the second balance equation

$$\pi_{0,H} (\eta_H + \lambda) = \eta_L \pi_{0,L} + \mu_H \pi_{1,H}. \tag{5}$$

When it comes to states $\{i, H\}$ and $\{i, L\}$, for $i > 0$, it can be observed that the number of possible transitions is higher. There are two possible transitions out of state $\{i, L\}$, with corresponding rates of $\lambda$ and $\eta_L$, leading to a total flow rate out of this state of $\pi_{i,L}(\lambda+\eta_L)$. There are also two possible transitions into this state. The first one is from state $\{i, H\}$ with rate $\eta_H$, and the other is from state $\{i-1, L\}$ with rate $\lambda$, leading thus to a total flow rate (into the state) of $\lambda \pi_{i-1,L} + \eta_H \pi_{i,H}$. Equating the flow rates into and out of this state, we have

$$\pi_{i,L} (\lambda + \eta_L) = \lambda \pi_{i-1,L} + \eta_H \pi_{i,H},\ i \geq 1. \tag{6}$$

When it comes to state $\{i, H\}$, there are three possible transitions out of it. The corresponding rates are $\lambda$, $\eta_H$, and $\mu_H$. Hence, the total flow rate out of it is $\pi_{i,H}(\lambda + \eta_H + \mu_H)$. Transitions to state $\{i, H\}$ are possible from states $\{i, L\}$, $\{i-1, H\}$ and $\{i+1, H\}$ with rates $\eta_L$, $\lambda$, and $\mu_H$. So, the flow rate into state $\{i, H\}$ is $\eta_L \pi_{i,L} + \mu_H \pi_{i+1,H} + \lambda \pi_{i-1,H}$. Consequently, we have the following balance equation

$$\pi_{i,H} (\lambda + \mu_H + \eta_H) = \lambda \pi_{i-1,H} + \eta_L \pi_{i,L} + \mu_H \pi_{i+1,H},\ i \geq 1. \tag{7}$$

We define the PGFs for this chain as

$$G_L(z) = \sum_{i=0}^{\infty} \pi_{i,L} z^i, \text{and } G_H(z) = \sum_{i=0}^{\infty} \pi_{i,H} z^i, |z| \leq 1, z \in C.$$

Multiplying Eq.(6) with $z^i$ and adding it to Eq.(4), we obtain

$$(\lambda + \eta_L) \sum_{i=0}^{\infty} \pi_{i,L} z^i = \eta_H \sum_{i=0}^{\infty} \pi_{i,H} z^i + \lambda \sum_{i=1}^{\infty} \pi_{i-1,L} z^i, \tag{8}$$

that leads to

$$[\lambda(1 - z) + \eta_L] G_L(z) = \eta_H G_H(z). \tag{9}$$

Similarly, multiplying Eq.(7) with $z^i$ and summing with Eq.(5), we get

$$(\eta_H + \lambda)\sum_{i=0}^{\infty}\pi_{i,H}z^i + \mu_H\sum_{i=1}^{\infty}\pi_{i,H}z^i = \eta_L\sum_{i=0}^{\infty}\pi_{i,L}z^i + \lambda\sum_{i=1}^{\infty}\pi_{i-1,H}z^i + \mu_H\sum_{i=0}^{\infty}\mu_H\pi_{i+1,H}z^i. \tag{10}$$

Eq.(10) results in

$$\left[\lambda z(1-z) + \mu_H(z-1) + \eta_H z\right]G_H(z) - \eta_L z G_L(z) = \mu_H\pi_{0,H}(z-1). \tag{11}$$

Solving the system of equations (9) and (11) leads to

$$G_L(z) = \frac{\mu_H\pi_{0,H}(z-1)}{\frac{1}{\eta_H}\left[\lambda z(1-z) + \mu_H(z-1) + \eta_H z\right]\left[\lambda(1-z) + \eta_L\right] - z\eta_L}, \tag{12}$$

$$G_H(z) = \frac{1}{\eta_H}\left[\lambda(1-z) + \eta_L\right]G_L(z). \tag{13}$$

The only unknown in Eq.(12) is $\pi_{0,H}$ (the stationary probability of SU having zero files while there is no PU activity). To find it, we proceed as following. First, we write the balance equation across the vertical cut between states $(i, L)$ and $(i, H)$ on one side, and $(i, L+1)$ and $(i, H+1)$ on the other. This gives

$$\lambda\pi_{i,L} + \lambda\pi_{i,H} = \mu_H\pi_{i+1,H}. \tag{14}$$

After summing over all the values of $i$, we have

$$\lambda = \mu_H\left[G_H(1) - \pi_{0,H}\right]. \tag{15}$$

In Eq.(15), $G_H(1) = \sum_{i=0}^{\infty}\pi_{i,H}$ is the probability of finding the system in the high state. So, for the zero probability we have

$$\pi_{0,H} = \frac{\mu_H G_H(1) - \lambda}{\mu_H}. \tag{16}$$

Replacing $z = 1$ into Eq.(9) gives $G_L(1) = \frac{\eta_H}{\eta_L}G_H(1)$. It is also evident that $G_L(1) + G_H(1) = 1$, resulting in

$$G_H(1) = \frac{1}{1 + \frac{\eta_H}{\eta_L}}. \tag{17}$$

Replacing Eq.(17) into Eq.(16) enables us to find $\pi_{0,H}$:

$$\pi_{0,H} = \frac{1}{1 + \frac{\eta_H}{\eta_L}} - \frac{\lambda}{\mu_H}. \tag{18}$$

After finding $\pi_{0,H}$ and replacing it into Eq.(12), and the later into Eq.(13) we find $G_L(z)$ and $G_H(z)$ in closed form.

The next step is to find the average number of files in the system. It is the sum of the derivatives of partial PGFs at point $z = 1$, i.e.

$$E[N] = E[N_L] + E[N_H] = G_L^{'}(1) + G_H^{'}(1). \tag{19}$$

Differentiating Eq.(12) with respect to $z$ we have

$$G_L^{'}(z) = \frac{\mu_H \pi_{0,H} F(z) - \mu_H \pi_{0,H}(z-1)F^{'}(z)}{F^2(z)}. \tag{20}$$

In Eq.(20), $F(z) = A(z)B(z) - \eta_L z$, where $A(z) = \frac{\lambda z(1-z)+\mu_H(z-1)+\eta_H z}{\eta_H}$, and $B(z) = \lambda(1-z) + \eta_L$.

It can easily be proven that Eq.(20) is of the form $\frac{0}{0}$ at $z = 1$. After applying $L'H\hat{o}pital$'s rule twice, we get

$$G_L^{'}(z) = \frac{-\mu_H \pi_{0,H} F^{''}(z) + \mu_H \pi_{0,H} F^{'''}(z)(1-z)}{2F^{'}(z)^2 + 2F(z)F^{''}(z)}. \tag{21}$$

Based on Eq.(21), we have

$$E[N_L] = \lim_{z \to 1} G_L^{'}(z) = \frac{-\mu_H \pi_{0,H} F^{''}(1)}{2F^{'}(1)^2}. \tag{22}$$

After some algebra, we obtain

$$E[N_L] = \frac{\lambda \mu_H \pi_{0,H} \left(\eta_L + \mu_H + \eta_H - \lambda\right)}{\eta_H \left[\frac{1}{\eta_H}(\mu_H + \eta_H - \lambda)\eta_L - \lambda - \eta_L\right]^2}. \tag{23}$$

The next step is to find $E[N_H]$. For that purpose, Eq.(13) is differentiated, giving

$$G_H^{'}(z) = \frac{1}{\eta_H}\left\{-\lambda G_L(z) + \left[\lambda(1-z) + \eta_L\right] G_L^{'}(z)\right\}. \tag{24}$$

Since $E[N_H] = \lim_{z \to 1} G_H^{'}(z)$, substituting $z = 1$ into Eq.(24) results in

$$E[N_H] = \frac{1}{\eta_H}\left[-\lambda G_L(1) + \eta_L G_L^{'}(1)\right]. \tag{25}$$

In Eq.(25), $E[N_L] = G_L^{'}(1)$, and $G_L(1) = \frac{\eta_H E[T_s]}{1+\eta_H E[T_s]}$. So, Eq.(25) reduces to

$$E[N_H] = -\frac{\lambda E[T_s]}{1 + \eta_H E[T_s]} + \frac{1}{\eta_H E[T_s]} E[N_L]. \tag{26}$$

After some algebra, we can find that the average number of files in the system is $E[N] = E[N_L] + E[N_H]$.

Finally, using Little's law $E[N] = \lambda E[T]$ [7], we obtain the average file delay in the interweave access as in Eq.(3). □

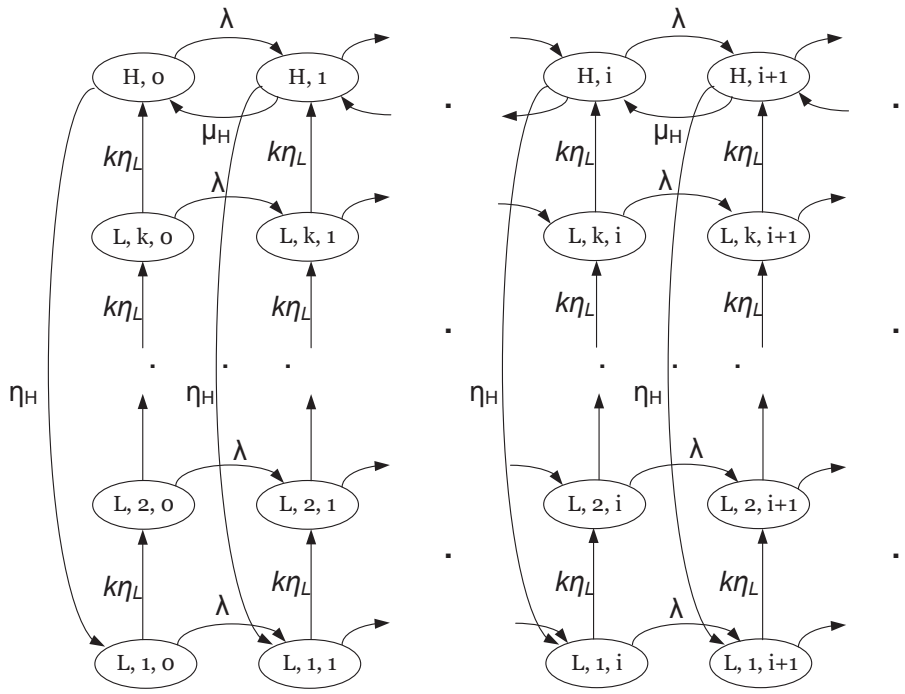


Figure 6: The 2D Markov chain for Erlang-distributed scanning time.

**Low-variability scanning time.** In the previous section we have derived the average file delay for exponentially distributed scanning times. However, as explained in Section 2.2, there exist some cases when the scanning time can have less variability than the exponential distribution. To capture this low variability an Erlang k-stage distribution is assumed. Our system can still be modeled with a 2D Markov chain, as depicted in Fig. 6. However, a transition from a low state (scanning) to a high state (finding and using a new available channel) would now have to go through an additional $k-1$ intermediate states (vertically), as opposed to going directly to the high state as in the exponential case (see Fig. 5). The transition rate between these states is $k\eta_L$. Since there are $k$ stages, the average scanning time is $E[T_s] = k\frac{1}{k\eta_L} = \frac{1}{\eta_L}$. It is not possible to make a transition backwards while in the scanning mode (no transmission).

In general, it is very difficult to solve these kind of Markov chains analytically, and one needs to use numerical, matrix-analytic methods [8]. However, numerical methods do not provide any insight on the nature of the solution and its dependency on certain parameters.

Interestingly, due to the particular structure of the MC at hand, we are nevertheless able to derive a closed form analytical expression. Although there are more than two states in the "vertical" direction, we can still write down the balance equations and follow our approach to solve a system of $k+1$ equations in the partial

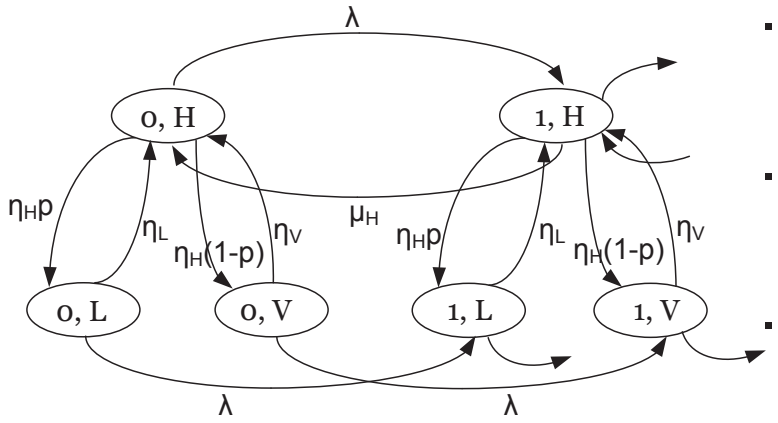


Figure 7: The 2D Markov chain for hyperexponentially distributed scanning time.

probability generating functions. This is the main difference with the scenario for scanning times that are exponential.

The following theorem gives the expected delay in this scenario.

**Theorem 2.** *The average file delay in the interweave access with Erlang distributed scanning time is given by*

$$E[T_{erl}] = \frac{\eta_H \left[\eta_H + \frac{(k+1)}{2k}\mu_H\right](E[T_s])^2 + 2\eta_H E[T_s] + 1}{(1+\eta_H E[T_s])(\mu_H - \lambda - \lambda\eta_H E[T_s])}. \tag{27}$$

*Proof.* Following the same procedure for the flow rates into and out of states of this chain (Fig. 6), as we did for the interweave mode with exponentially distributed scanning time, we can write the balance equations as follows

$$\begin{aligned}
\pi_{L,1,0}\left(\lambda + k\eta_L\right) &= \eta_H \pi_{H,0} && (28)\\
\pi_{L,1,i}\left(\lambda + k\eta_L\right) &= \eta_H \pi_{H,i} + \lambda \pi_{L,1,i-1},\ i \geq 1 && (29)\\
\pi_{L,j,0}\left(\lambda + k\eta_L\right) &= k\eta_L \pi_{L,j-1,0},\ 2 \leq j \leq k && (30)\\
\pi_{L,j,i}\left(\lambda + k\eta_L\right) &= k\eta_L \pi_{L,j-1,i} + \lambda \pi_{L,j,i-1},\ i \geq 1, 2 \leq j \leq k && (31)\\
\pi_{H,0}\left(\lambda + \eta_H\right) &= k\eta_L \pi_{L,k,0} + \mu_H \pi_{H,1} && (32)\\
\pi_{H,i}\left(\lambda + \eta_H + \mu_H\right) &= k\eta_L \pi_{L,k,i} + \lambda \pi_{H,i-1} + \mu_H \pi_{H,i+1},\ i \geq 1 && (33)
\end{aligned}$$

We define the partial probability generating functions for the scanning phases as

$$G_{L,j}(z) = \sum_{i=1}^{\infty} \pi_{j,i} z^i,\ j = 2, \ldots, k, |z| \leq 1, \tag{34}$$

and for the transmission phase

$$G_H(z) = \sum_{i=1}^{\infty} \pi_{H,i} z^i,\ |z| \leq 1. \tag{35}$$

Further, we continue with multiplying Eq.(29), Eq.(31), and Eq.(33) by $z^i$ and adding each one of them to Eq.(28), Eq.(30) and Eq.(32), respectively. After that, we obtain the following system of equations with partial probability generating functions as unknowns

$$[\lambda(1-z)+k\eta_L]\,G_{L,1}(z)=\eta_H G_H(z) \tag{36}$$

$$[\lambda(1-z)+k\eta_L]\,G_{L,j}(z)=\eta_L G_{L,j-1}(z) \tag{37}$$

$$[\lambda z(1-z)+\mu_H(z-1)+\eta_H z]\,G_H(z)=k\eta_L z G_{L,k}(z)+\mu_H\pi_{H,0}(z-1) \tag{38}$$

Eq.(36) can be expressed as

$$G_{L,1}(z)=\frac{\eta_H}{[\lambda(1-z)+k\eta_L]}G_H(z). \tag{39}$$

From Eq.(37), for the states $j=2,\ldots,k$ there are recursions involved. The partial PGF for the $j$th states can be written as

$$G_{L,j}(z)=\frac{k\eta_L}{[\lambda(1-z)+k\eta_L]}G_{L,j-1}(z),\ j\geq 2, \tag{40}$$

and after using this recursion successively we get

$$G_{L,j}(z)=\frac{(k\eta_L)^{j-1}\,\eta_H}{[\lambda(1-z)+k\eta_L]^j}G_H(z),\ j\geq 2. \tag{41}$$

Solving the system of equations Eq.(36)-Eq.(38), for $G_H(z)$ we obtain

$$G_H(z)=\frac{\mu_H\pi_{H,0}(z-1)}{\lambda z(1-z)+\mu_H(z-1)+\eta_H z-\frac{\eta_H z}{\left[1+\frac{\lambda}{k\eta_L}(1-z)\right]^k}}. \tag{42}$$

Replacing Eq.(42) into Eq.(39) and Eq.(41), we have the solutions for all the PGFs. However, there is an unknown component $\pi_{H,0}$ appearing in those expressions. We can find it by taking the vertical cut between the states corresponding to $i$ and $i+1$ files

$$\lambda\left(\sum_{j=1}^{k}\pi_{L,j,i}+\pi_{H,i}\right)=\mu_H\pi_{H,i+1}, \tag{43}$$

which after some rearrangements leads to

$$\pi_{H,0}=G_H(1)-\frac{\lambda}{\mu_H}. \tag{44}$$

The stationary probability of finding the system in the idle PU state, $G_H(1)$, can be found as follows. Eq.(39), for $z = 1$ reduces to $G_{L,1} = \frac{\eta_H}{k\eta_L} G_H(1)$. Similarly, for the other scanning phases the following result can be derived

$$G_{L,j}(1) = G_{L,j-1}(1),\ j \geq 2. \tag{45}$$

Obviously, it holds that

$$G_{L,1}(1) + \ldots + G_{L,k}(1) + G_H(1) = 1, \tag{46}$$

which after solving gives

$$G_H(1) = \frac{1}{1 + \frac{\eta_H}{\eta_L}}. \tag{47}$$

The term $\pi_{H,0}$ is obtained after replacing Eq.(47) into Eq.(44) as

$$\pi_{H,0} = \frac{1}{1 + \eta_H \eta_L} - \frac{\lambda}{\mu_H}. \tag{48}$$

After finding all the partial PGFs, we move further in finding the average file delay. Towards that direction, we need to find first the average number of files in the system, which is given by

$$E[N] = E[N_1] + \ldots + E[N_k] + E[N_H]. \tag{49}$$

We find the term $E[N_H] = G_H^{'}(1)$ as follows. Differentiating Eq.(42) with respect to $z$ leads to

$$G_H^{'}(z) = \frac{\mu_H \pi_{H,0} F(z) - \mu_H \pi_{H,0}(z-1)F^{'}(z)}{F(z)^2}, \tag{50}$$

with $F(z) = \lambda z(1-z) + \mu_H(z-1) + \eta_H z - \frac{\eta_H z}{[1+\lambda k \eta_L (1-z)]^k}$. It can be easily proven that $F(z) = 0$ making both the numerator and denominator equal to 0. Hence, we need to apply $L'H\hat{o}pital's$ rule twice. After that, we obtain

$$E[N_H] = \lim_{z \to 1} G_H^{'}(z) = -\frac{\mu_H \pi_{H,0} F^{''}(1)}{2F^{'}(1)^2}, \tag{51}$$

where $F^{'}(1) = \mu_H - \lambda - \lambda \frac{\eta_H}{\eta_L}$, and $F^{''}(1) = \lambda \left[2 + \frac{\eta_H}{\eta_L}\left(\frac{\lambda}{\eta_L} + \frac{\lambda}{k\eta_L} + 2\right)\right]$. Next, we need to determine the PGFs related to the scanning part. As shown earlier, the following relation holds

$$G_{L,j}(z) = \frac{(k\eta_L)^{j-1}\,\eta_H}{[\lambda(1-z) + k\eta_L]^j} G_H(z),\ j \geq 1. \tag{52}$$

Differentiating Eq.(52) with respect to $z$ yields

$$G_{L,j}^{'}(z) = (k\eta_L)^{j-1}\,\eta_H \cdot \frac{[\lambda(1-z) + k\eta_L]^j\, G_H^{'}(z) + j\lambda G_H(z)\,[\lambda(1-z) + k\eta_L]^{j-1}}{[\lambda(1-z) + k\eta_L]^{2j}}. \tag{53}$$

Since $E[N_{L,j}] = G'_{L,j}(1)$, for $j = 1, \ldots, k$, after some algebra we obtain

$$E[N_{L,j}] = \frac{\eta_H}{\eta_L} E[N_H] + \frac{\lambda \eta_H}{(k\eta_L)^2} \cdot j \cdot G_H(1),\ j \geq 1. \tag{54}$$

Replacing Eq.(51) and Eq.(54) into Eq.(49), we find the average number of files in the system as

$$E[N] = \left(1 + \frac{\eta_H}{\eta_L}\right) E[N_H] + \lambda \eta_H \frac{k+1}{2k} \frac{1}{\eta_L^2} G_H(1). \tag{55}$$

Replacing Eq.(47) and Eq.(51) into Eq.(55), we can find the average number of files in the system. Finally, using Little's law $E[N] = \lambda E[T]$, and expressing the scanning time as $E[T_s] = \frac{1}{k\eta_L}$, we obtain Eq.(27). □

By carefully comparing Eq.(3) and Eq.(27) one can notice that the average delay for exponential scanning time is always higher compared to the delay induced in the case with Erlang scanning time, since $\frac{k+1}{2k} < 1, \forall k > 1$. This is in accordance with queueing system experience, where higher variability usually reduces performance.

**High-variability scanning time.** Finally, we proceed with the case of high variability scanning time, which is modeled by an hyperexponential distribution with two branches that will be mapped into two separate states (denoted with the index *L* and *V*). The 2D Markov chain for this model is shown in Fig. 7. While being in the scanning phase, the SU can be either in one of the $(i, L)$ states (short time of finding an available channel), or in one of the $(i, V)$ states (long time until an available channel is found). The average scanning time in this setup is $E[T_s] = \frac{p_0}{\eta_L} + \frac{1-p_0}{\eta_V}$. In order to maintain the same $E[T_s]$ as before, but with much higher variability, we choose a very low value for $1 - p_0$ (e.g. lower than 0.05) and $\eta_V << \eta_L$.

Once more, the structure of this chain allows us to avoid numerical, matrix-analytic methods, and instead apply the methodology of PGFs to derive a closed form expression, given in the following theorem.

**Theorem 3.** *The average file delay in interweave access with hyperexponential scanning time is given by*

$$E[T_{hyp}] = \frac{(\eta_H E[T_s])^2 + \eta_H \mu_H \left(\frac{p_0}{\eta_L^2} + \frac{1-p_0}{\eta_V^2}\right) + 2\eta_H E[T_s] + 1}{(1 + \eta_H E[T_s])(\mu_H - \lambda - \lambda \eta_H E[T_s])}. \tag{56}$$

*Proof.* Following the same procedure for the flow rates into and out of states of this chain (Fig. 7), as we did for the interweave mode with exponentially and Erlang-

distributed scanning times, we can write the balance equations as follows

$$\begin{aligned}
\pi_{0,L}\left(\lambda+\eta_L\right) &= p_0\eta_H\pi_{0,H} && (57)\\
\pi_{i,L}\left(\lambda+\eta_L\right) &= \lambda\pi_{i-1,L}+p_0\eta_H\pi_{i,H},\ i\geq 1 && (58)\\
\pi_{0,V}\left(\lambda+\eta_V\right) &= (1-p_0)\eta_H\pi_{0,H} && (59)\\
\pi_{i,V}\left(\lambda+\eta_V\right) &= \lambda\pi_{i-1,V}+(1-p_0)\eta_H\pi_{i,H},\ i\geq 1 && (60)\\
\pi_{0,H}\left(\eta_H+\lambda\right) &= \eta_L\pi_{0,L}+\eta_V\pi_{0,V}+\mu_H\pi_{1,H} && (61)
\end{aligned}$$

$$\pi_{i,H}\left(\lambda+\mu_H+\eta_H\right)=\lambda\pi_{i-1,H}+\eta_L\pi_{i,L}+\eta_V\pi_{i,V}+\mu_H\pi_{i+1,H},\ i\geq 1. \quad (62)$$

The probability generating function for the V-state is defined as

$$G_V(z)=\sum_{i=0}^{\infty}\pi_{i,V}z^i, |z|\leq 1. \quad (63)$$

As in the previous subsections, we multiply Eq.(58), Eq.(60), and Eq.(62) with $z^i$ and add each one of them to Eq.(57), Eq.(59) and Eq.(62), respectively. After some simple algebra we get the following system of equations with partial PGFs as unknowns

$$\left[\lambda(1-z)+\eta_L\right]G_L(z)=p_0\eta_H G_H(z), \quad (64)$$

$$\left[\lambda(1-z)+\eta_V\right]G_V(z)=(1-p_0)\eta_H G_H(z), \quad (65)$$

$$\left[\lambda z(1-z)+\eta_H z+\mu_H(z-1)\right]G_H(z)=\eta_L zG_L(z)+\eta_V zG_V(z)+\mu_H\pi_{0,H}(z-1).$$

Expressing $G_L(z)$ and $G_V(z)$ through $G_H(z)$ in Eq.(64) and Eq.(65), and replacing them afterwards into Eq.(66) results in

$$G_H(z)=\frac{\mu_H\pi_{0,H}(z-1)}{\lambda z(1-z)+\eta_H z+\mu_H(z-1)-\frac{\eta_L z p_0\eta_H}{\lambda(1-z)+\eta_L}-\frac{\eta_V z(1-p_0)\eta_H}{\lambda(1-z)+\eta_V}}. \quad (66)$$

We replace the expression for $G_H(z)$ into Eq.(64) and Eq.(65) to get the $G_L(z)$ and $G_V(z)$, respectively. The only unknown in these equations is the zero probability for the high state, $\pi_{0,H}$. We derive it in the same way as before (using a vertical cut between neighboring triplets of states). We get

$$\pi_{0,H}=G_H(1)-\frac{\lambda}{\mu_H}. \quad (67)$$

A similar approach is used in finding $G_H(1)$. Replacing $z=1$ into Eq.(65), we have

$$G_V(1)=(1-p_0)\frac{\eta_H}{\eta_V}G_H(1). \quad (68)$$

In the same way, we replace $z=1$ into Eq.(64), which leads to

$$G_L(z)=p_0\frac{\eta_H}{\eta_L}G_H(1). \quad (69)$$

We next substitute Eq.(68) and Eq.(69) into $G_H(1)+G_L(1)+G_V(1)=1$. We get

$$G_H(1)=\left[\frac{p_0\eta_H}{\eta_L}+\frac{(1-p_0)\eta_H}{\eta_V}+1\right]^{-1}. \tag{70}$$

To get $E[N_H]=G_H^{'}(1)$, we need to differentiate Eq.(66). After doing that and using twice $L'H\hat{o}pital's$ rule, we obtain

$$E[N_H]=\frac{-\mu_H\pi_{0,H}F^{''}(1)}{2F^{'}(1)^2}, \tag{71}$$

where $F^{'}(1)=F_1^{'}(1)+F_2^{'}(1)+F_3^{'}(1)$, and $F^{''}(1)=-2\lambda-\frac{2\lambda p_0\eta_H(\lambda+\eta_L)}{\eta_L^2}-\frac{2\lambda(1-p_0)\eta_H(\lambda+\eta_V)}{\eta_V^2}$.

The "low component" of the average number of files is $E[N_L]=G_L^{'}(1)$. The first derivative of $G_L(z)$ is found from Eq.(64), and is

$$G_L^{'}(z)=\frac{p_0\eta_H G_H^{'}(1)\left[\lambda(1-z)+\eta_L\right]+p_0\eta_H\lambda G_H(z)}{\left[\lambda(1-z)+\eta_L\right]^2}. \tag{72}$$

So, for $E[N_L]$ we have

$$E[N_L]=\frac{p_0\eta_H\eta_L E[N_H]+p_0\eta_H\lambda G_H(1)}{\eta_L^2}, \tag{73}$$

where $E[N_H]$ is given by Eq.(71).

Next, we pursue the same procedure for the V-states. First, from

$$G_V(z)=\frac{(1-p_0)\eta_H}{\lambda(1-z)+\eta_V}\cdot G_H(z), \tag{74}$$

we get the first derivative as

$$G_V^{'}(z)=\frac{(1-p_0)\eta_H G_H^{'}(z)\left[\lambda(1-z)+\eta_V\right]+\lambda(1-p_0)\eta_H G_H(z)}{\left[\lambda(1-z)+\eta_V\right]^2}, \tag{75}$$

and at point $z=1$, $E[N_V]$ is given by

$$E[N_V]=G_V^{'}(1)=\frac{(1-p_0)\eta_H E[N_H]+\lambda(1-p_0)\eta_H G_H(1)}{\eta_V^2}. \tag{76}$$

The average number of files in the system is $E[N]=E[N_H]+E[N_L]+E[N_V]$. Applying the Little's law $E[N]=\lambda E[T]$, we obtain the average file delay as in Eq.(56). □

Let's consider now the term in brackets in the numerator of Eq.(56). It can be rewritten as

$$\frac{p_0}{\eta_L^2}+\frac{1-p_0}{\eta_V^2}=\frac{1}{\eta_V^2}+p_0\left(\frac{1}{\eta_L^2}-\frac{1}{\eta_V^2}\right)=\frac{1}{\eta_V^2}+\left(\frac{1}{\eta_L}+\frac{1}{\eta_V}\right)\cdot p_0\left(\frac{1}{\eta_L}-\frac{1}{\eta_V}\right). \tag{77}$$

The average scanning time for this scenario is

$$E[T_s]=\frac{p_0}{\eta_L}+\frac{1-p_0}{\eta_V}=\frac{1}{\eta_V}+p_0\left(\frac{1}{\eta_L}-\frac{1}{\eta_V}\right).$$

From the last equation, we have

$$p_0\left(\frac{1}{\eta_L}-\frac{1}{\eta_V}\right)=E[T_s]-\frac{1}{\eta_V}.$$

Replacing the last equation into Eq.(77), we obtain

$$\begin{aligned}\frac{p_0}{\eta_L^2}+\frac{1-p_0}{\eta_V^2}&=\frac{1}{\eta_V^2}+\left(\frac{1}{\eta_L}+\frac{1}{\eta_V}\right)\left(E[T_s]-\frac{1}{\eta_V}\right)>\frac{1}{\eta_V^2}+E[T_s]\left(E[T_s]-\frac{1}{\eta_V}\right)\\&=\left(E[T_s]\right)^2+\frac{1}{\eta_V}\left(\frac{1}{\eta_V}-E[T_s]\right)>\left(E[T_s]\right)^2.\end{aligned} \tag{78}$$

Since $p_0<1$ it holds that $\frac{1}{\eta_L}+\frac{1}{\eta_V}>\frac{p_0}{\eta_L}+\frac{1-p_0}{\eta_V}=E[T_s]$. Further, the average scanning time is much lower than the average (occasional) long periods, i.e., $\frac{1}{\eta_V}>E[T_s]$. This way we arrive at Eq.(78).

The corresponding term in the numerator of Eq.(3) is $(E[T_s])^2$, and the other terms in both the numerator and denominator are identical to the terms of Eq.(56). This way we have shown that $E[T_{hyp}]>E[T_{exp}]$, which is expected from queueing systems. To conclude, we have shown that the following relation holds $E[T_{hyp}]>E[T_{exp}]>E[T_{erl}]$. This means that the variability of the scanning time plays a crucial role on the average delay in the interweave access mode.

### 3.2 Delay analysis for underlay access

As we have already explained in Section 2.1, in the underlay CRN the SU can transmit all the time (both when in high and low states). We can again model the system with a 2D Markov chain, as shown in Fig. 8. Note the difference with Fig. 5. While in Fig. 5 there is no transition backwards in the low states, in the underlay CRN these transitions exist with rate $\mu_L=\frac{c_L}{\Delta}$.

We should mention that $\pi_{0,H}$ ($\pi_{0,L}$) denote, as before, the stationary probability of finding the SU with no files to transmit while being in a high (low) period.

**Theorem 4.** *The average file delay in the underlay access mode is given by*

$$E[T_u]=\frac{\eta_H+\eta_L+\mu_H(1-\pi_{0,H})+\mu_L(1-\pi_{0,L})-\lambda+\frac{\mu_L\mu_H}{\lambda}(\pi_{0,L}+\pi_{0,H}-1)}{\mu_H\eta_L+\mu_L\eta_H-\lambda(\eta_H+\eta_L)} \tag{79}$$

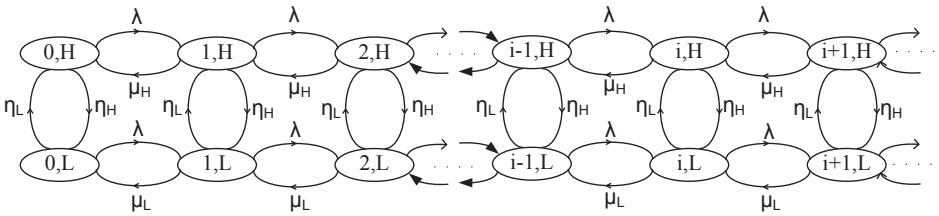


Figure 8: The 2D Markov chain for the underlay model.

*Proof.* There are two possible transitions out of state $\{0, L\}$ (Fig. 8), with rates $\eta_L$ and $\lambda$, respectively. Hence, the total flow rate out of this state is $\pi_{0,L}(\lambda + \eta_L)$. As far as the transitions into this state are concerned, there are two possibilities. The first one is from state $\{0, H\}$ with rate $\eta_H$, whereas the second one is from state $\{1, L\}$ with rate $\mu_L$. Hence, the flow rate into this state is $\pi_{0,H}\eta_H + \pi_{1,L}\mu_L$. Equating the flow rates into and out of state $\{0, L\}$, we get

$$\pi_{0,L}(\lambda + \eta_L) = \pi_{1,L}\mu_L + \pi_{0,H}\eta_H. \tag{80}$$

For state $\{i, L\}$, when $i > 0$, there are three possible transitions out of it, with rates $\lambda$, $\mu_L$, and $\eta_L$, respectively. The total flow rate out of this state is $\pi_{i,L}(\lambda + \eta_L + \mu_L)$. Transitions from the following three states into state $\{i, L\}$ are possible: $\{i + 1, L\}$, $\{i, H\}$, and $\{i - 1, L\}$. The rates for these transitions are $\mu_L$, $\eta_H$, and $\lambda$, respectively. Hence, the total flow rate into state $\{i, L\}$ is $\pi_{i-1,L}\lambda + \pi_{i+1,L}\mu_L + \pi_{i,H}\eta_H$, and the balance equation for states $\{i, L\}, i > 0$ is

$$\pi_{i,L}(\lambda + \eta_L + \mu_L) = \pi_{i-1,L}\lambda + \pi_{i+1,L}\mu_L + \pi_{i,H}\eta_H, (i > 0). \tag{81}$$

Obviously, the upper part of the chain is identical to the upper part of the chain corresponding to the interweave mode with exponential scanning time (Fig. 5). As a result, Eq.(5) and Eq.(7) from the previous analysis do not change. We rewrite those equations here:

$$\pi_{0,H}(\lambda + \eta_H) = \pi_{1,H}\mu_H + \pi_{0,L}\eta_L. \tag{82}$$

$$\pi_{i,H}(\lambda + \eta_H + \mu_H) = \pi_{i-1,H}\lambda + \pi_{i+1,H}\mu_H + \pi_{i,L}\eta_L, (i > 0). \tag{83}$$

Similarly as before, we define the probability generating functions for both the low and high states as

$$G_L(z) = \sum_{i=0}^{\infty} \pi_{i,L} z^i, \text{and } G_H(z) = \sum_{i=0}^{\infty} \pi_{i,H} z^i, |z| \leq 1.$$

After multiplying Eq.(81) by $z^i$, adding it to Eq.(80), and rearranging (in the same direction as we did for interweave CRN), we obtain

$$\begin{aligned}(\lambda+\eta_L+\mu_L)G_L(z) &= \lambda z G_L(z)+\eta_H G_H(z)\\ &+ \frac{\mu_L}{z}\left(G_L(z)-\pi_{0,L}\right)+\pi_{0,L}\mu_L. \end{aligned} \tag{84}$$

Multiplying Eq.(83) by $z^i$, adding it to Eq.(82), and rearranging, we get

$$\begin{aligned}(\lambda+\eta_H+\mu_H)G_H(z) &= \lambda z G_H(z)+\eta_L G_L(z)\\ &+ \frac{\mu_H}{z}\left(G_H(z)-\pi_{0,H}\right)+\pi_{0,H}\mu_H. \end{aligned} \tag{85}$$

Solving the system of equations Eq.(84)-(85) gives

$$f(z)G_L(z)=\pi_{0,H}\eta_H\mu_H z+\pi_{0,L}\mu_L\left[\eta_H z+(\lambda-z\mu_H)(1-z)\right], \tag{86}$$

where

$$\begin{aligned} f(z) &= \lambda^2 z^3-\lambda(\eta_L+\eta_H+\lambda+\mu_H+\mu_L)z^2\\ &+ (\eta_L\mu_H+\eta_H\mu_L+\mu_L\mu_H+\lambda\mu_H+\lambda\mu_L)z-\mu_L\mu_H. \end{aligned} \tag{87}$$

It can be proven that the polynomial in Eq.(87) has only one root in the open interval $(0,1)$ [12]. This root is denoted as $z_0$. Setting $z=z_0$ into Eq.(86), and after performing some simple algebra we get $\pi_{0,L}$ and $\pi_{0,H}$, as

$$\pi_{0,L} = \frac{\eta_H\left(\frac{\eta_H\mu_L+\eta_L\mu_H}{\eta_H+\eta_L}-\lambda\right)z_0}{\mu_L(1-z_0)(\mu_H-\lambda z_0)}, \tag{88}$$

$$\pi_{0,H} = \frac{\eta_L\left(\frac{\eta_H\mu_L+\eta_L\mu_H}{\eta_H+\eta_L}-\lambda\right)z_0}{\mu_H(1-z_0)(\mu_L-\lambda z_0)}. \tag{89}$$

Finally, using Little's law $E[N]=\lambda E[T]$ [7], we obtain Eq.(79). □

### 3.3 Underlay access with multiple power levels

In the underlay spectrum access technique, the assumption of two power levels for the SU might not be very realistic. Instead, depending on its position, an SU would be able to adjust its transmission power to multiple power levels, leading thus to a set of multiple data rates. Furthermore, if we assume that the time an SU spends being on a level with a given rate, then the corresponding Markov chain would have more states vertically. In the following, we derive the average system time for a file in such a scenario, considering that there are $M$ different data rates (power levels).

We model the SU time intervals with the multilevel scheme as shown in Fig. 9. The duration of each period is exponentially distributed with rate $\eta_i, i=1,\ldots,M$.

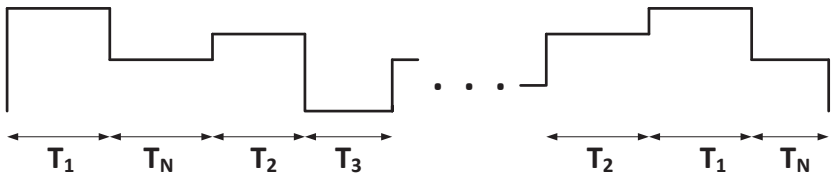


Figure 9: The multi-power underlay model.

Each period $T_i$ corresponds to the time during which the SU is having the same data rate. We call these periods *levels* or *phases*.

The durations of any level are mutually independent, and at the same time independent of durations of other levels. The data transmission rates during periods with different power levels are denoted by $\mu_i, i = 1, \ldots, M$. As before, the traffic arrival process is Poisson and file sizes are exponentially distributed. Again, we consider the FCFS scheduling discipline.

Based on the assumptions we have made, our system can be modeled with a 2D Markov chain that is bounded in one dimension (the dimension that represents the number of levels). This is shown in Fig. 10. The same Markov chain, but used in a different context is given in [13]. The interested reader can find more details there.

The possible level transitions are shown only for the state $(0, 1)$ and partially for $(j, 2)$ to avoid making the figure look more complex. $\pi_{k,i}$ denotes the stationary probability of being in level $i$ (where $i$ corresponds to the power level resulting in a data rate of $\mu_i$), and having $k$ files in the buffer. There are a number of possible transitions corresponding to the following events:

*new arrival*: From any state, the chain moves to the right (horizontally) with rate $\lambda$.

*flow finishes transmission*: From any state $\{k, i\}$ $(k > 0)$, the chain moves to the left (horizontally) with rate $\mu_i$.

*change in the power level:* The chain moves (vertically) from level $i$ to another level $j$ (transition to all the levels are possible with no exception) with rate $\eta_{i,j}, i \neq j$.

If we denote by $\eta_i$ the (total) rate at which a user leaves power level $i$, it holds that $\eta_i = \sum_{j=1}^{M} \eta_{i,j}$, for $i \neq j$.

We start by writing the balance equations for this Markov chain as

$$(\lambda + \eta_i)\, \pi_{0,i} = \mu_i \pi_{1,i} + \sum_{j=1}^{M} \eta_{j,i} \pi_{0,j}, \tag{90}$$

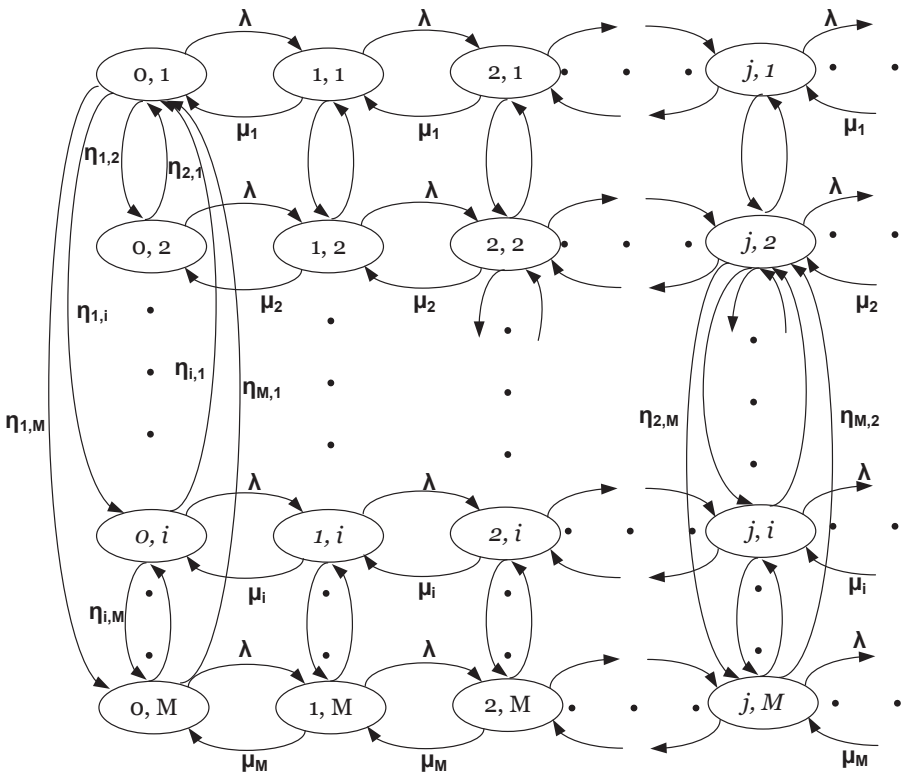


Figure 10: The 2D Markov chain for multilevel underlay model.

for $i = 1, \ldots, M$, $k = 0$, and

$$(\lambda + \mu_i + \eta_i)\, \pi_{k,i} = \lambda \pi_{k-1,i} + \mu_i \pi_{k+1,i} + \sum_{j=1}^{M} \eta_{j,i} \pi_{k,j}, \tag{91}$$

for $i = 1, \ldots, M$ and $k > 0$. Summing Eq.(90) and Eq.(91) multiplied by $z^k$, and then summing up over all $k$, we get

$$\begin{aligned} &\lambda \sum_{k=0}^{\infty} \pi_{k,i} z^k + \mu_i \sum_{k=1}^{\infty} \pi_{k,i} z^k + \sum_{k=0}^{\infty} \pi_{k,i} z^k \sum_{j=1}^{M} \eta_{i,j} \\ &= \lambda \sum_{k=1}^{\infty} \pi_{k-1,i} z^k + \mu_i \sum_{k=1}^{\infty} \pi_{k,i} z^{k-1} + \sum_{j=1}^{M} \eta_{j,i} \sum_{k=0}^{\infty} \pi_{k,j} z^k. \end{aligned} \tag{92}$$

We define the PGF for each level as

$$G_i(z) = \sum_{k=0}^{\infty} \pi_{k,i} z^k, \ |z| \leq 1, \ i = 1, \ldots, M. \tag{93}$$

Eq.(92) is now transformed into

$$\begin{aligned} &\lambda G_i(z) + \mu_i \left[ G_i(z) - \pi_{0,i} \right] + G_i(z) \sum_{j=1}^{M} \eta_{i,j} \\ &= \lambda z G_i(z) + \frac{\mu_i}{z} \left[ G_i(z) - \pi_{0,i} \right] + \sum_{j=1}^{M} \eta_{j,i} G_j(z). \end{aligned} \tag{94}$$

After performing some algebra, we have

$$\begin{aligned} &\left[ \lambda z(1-z) + \mu_i(z-1) + \eta_i z \right] G_i(z) - \sum_{j=1}^{M} \eta_{j,i} z G_j(z) \\ &\qquad = \mu_i (z-1) \pi_{0,i}, i = 1, \ldots, M. \end{aligned} \tag{95}$$

In Eq.(95), after introducing the substitution

$$f_i(z) = \lambda z(1-z) - \mu_i(1-z) + \eta_i z, \tag{96}$$

we obtain the following equation

$$\mathbf{F}(z)\mathbf{g}(z) = (z-1)\boldsymbol{\theta}, \tag{97}$$

where

$$\mathbf{F}(z) = \begin{bmatrix} f_1(z) & -\eta_{2,1} z & -\eta_{3,1} z & \ldots & -\eta_{M,1} z \\ -\eta_{1,2} z & f_2(z) & -\eta_{3,2} z & \ldots & -\eta_{M,2} z \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ -\eta_{1,M} z & -\eta_{2,M} z & -\eta_{3,M} z & \ldots & f_M(z) \end{bmatrix},$$

$$\mathbf{g}(z) = \begin{bmatrix} G_1(z) \\ G_2(z) \\ \vdots \\ G_M(z) \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \mu_1\pi_{0,1} \\ \mu_2\pi_{0,2} \\ \vdots \\ \mu_M\pi_{0,M} \end{bmatrix}.$$

Applying Cramer's rule to Eq.(97), we obtain

$$|\mathbf{F}(z)|\, G_i(z) = |\mathbf{F}_i(z)|\, (z-1). \tag{98}$$

$|\mathbf{F}_i(z)|$ is the determinant obtained after replacing the $i$th column of $|\mathbf{F}(z)|$ with $\boldsymbol{\theta}$. As can be observed from Eq.(96), at point $z = 1$, $f_1(1) = \eta_1$. Also, at this same point the sum of the elements in rows 2 to $M$ of the first column ($-\eta_{1,2}z - \eta_{1,3}z - \ldots - \eta_{1,M}z$) represents the sum of transition rates out of state 1 multiplied by $-1$. If we substract row 1 from the sum of the other rows, we have 0 at the element $\{1,1\}$ of the determinant $|\mathbf{F}_i(z)|$. Similar conclusions can be drawn for the other elements of the first row. Hence, we can obtain an equivalent determinant $|\mathbf{F}(z)|$ with all the elements of the first row equal to 0. So, $z = 1$ is one root of this determinant. Hence, we can write

$$|\mathbf{F}(z)| = (z-1)Q(z). \tag{99}$$

Replacing Eq.(99) into Eq.(98) we get

$$Q(z)G_i(z) = |\mathbf{F}_i(z)|\,. \tag{100}$$

In order to get the partial probability generating functions $G_i(z)$, we need first to find the zero probabilities $\pi_{0,1}, \pi_{0,2}, \ldots, \pi_{0,M}$. To do this, we proceed in the following way. First, we find the roots of $Q(z)$. Since our system is of order $M > 2$, these solutions can only be obtained numerically. The polynomial $Q(z)$ is of degree $2M-1$. However, only $M-1$ of its roots lie in the interval $(0,1)$ (which is our interval of interest)[3]. We denote these roots by $z_1, \ldots, z_{M-1}$. Since $G_i(z) \neq 0$ (all the probabilities $p_{k,i}$ are positive), then from Eq.(100), we have that $|\mathbf{F}_i(z_j)| = 0, i = 1, \ldots, M, j = 1, \ldots, M-1$. However, from Eq.(100) we can observe that, for each $z_j$, and any pair $1 \leq i, l \leq M$, $\frac{|\mathbf{F}_i(z_j)|}{|\mathbf{F}_l(z_j)|} = const$. This means that for each $z_j$ we have $M$ homogeneous linear equations that differ from each other only by a constant factor. Hence, $|\mathbf{F}_i(z_j)| = 0$ gives only one independent equation for each root $z_j$. Given that there are $M-1$ different roots $z_j$, it turns out that there are in total $M-1$ independent equations. Since we have $M$ unknown probabilities $\pi_{0,1}, \pi_{0,2}, \ldots, \pi_{0,M}$, and only $M-1$ equations, we cannot obtain unique solutions for these probabilities. So, we need another condition that relates these zero probabilities, and that is independent of the other $M-1$ equations.

[3]The proof to this claim is rather long and complicated, and we do not show it here. It was proven by Mitrani and Itzhak in [14], pp. 632-634.

Let's consider the vertical cut between states $k$ and $k+1$. The balance equation through this cut is

$$\lambda(\pi_{k,1} + \pi_{k,2} + \ldots + \pi_{k,M}) = \mu_1\pi_{k+1,1} + \ldots + \mu_M\pi_{k+1,M}. \quad (101)$$

Summing over all $k$ yields

$$\lambda \sum_{i=1}^{M} \pi_i = \mu_1(\pi_1 - \pi_{0,1}) + \ldots + \mu_M(\pi_M - \pi_{0,M}). \quad (102)$$

$\pi_i = \sum_{k=0}^{\infty} \pi_{k,i}$ denotes the percentage of time the system is in level $i$. Eq.(102) can be rewritten as

$$\mu - \lambda = \sum_{i=1}^{M} \mu_i\pi_{0,i}, \quad (103)$$

where $\mu = \sum_{i=1}^{M} \mu_i\pi_i$ is the average service rate of the system. Eq.(103) is the $M$th equation of the system we need to solve in order to get the zero probabilities. However, we do need to determine the probabilities $\pi_i$ first.

We can find $\pi_i$ by following a standard embedded MC approach for the (collapsed) chain with only $M$ states (corresponding to the $M$ levels). If we denote by $q_{i,j}$ the transition probabilities in the embedded chain, then $q_{i,j} = \frac{\eta_{i,j}}{\eta_i}$, and

$$\pi_i = \frac{\frac{r_i}{\eta_i}}{\sum_{i=1}^{M} \frac{r_i}{\eta_i}}, \quad (104)$$

where $r_i$ are the solutions to the global balance equations for the embedded Discrete Time Markov Chain (DTMC): $\sum_{i=1}^{M} r_i = 1$, and $r_j = \sum_{i=1}^{M} r_i q_{i,j}$.

Replacing Eq.(104) into Eq.(103), we have the $M$th equation of our system. Now, solving that system we get all the zero probabilities. The partial PGFs are found from Eq.(100) as

$$G_i(z) = \frac{|\mathbf{F}_i(z)|}{Q(z)}, i = 1, \ldots, M. \quad (105)$$

The average number of files in the system is

$$E[N] = \sum_{i=1}^{M} G_i'(1). \quad (106)$$

Using Little's law $E[N] = \lambda E[T]$, we get the following result:

**Result 5.** *The average file delay in a multilevel underlay spectrum access technique is given by*

$$E[T] = \frac{1}{\lambda} \sum_{i=1}^{M} \left( \frac{|\mathbf{F}_i(z)|}{Q(z)} \right)'_{z=1}. \quad (107)$$

### 3.4 SU with multiple channels

Our model can be easily adapted to capture the SU activity utilizing multiple channels concurrently (channel bonding). Depending on the policy the SU follows after losing a channel, different scenarios will arise. In the scenario we will illustrate in the following, after losing the first channel (we assume one PU per channel, but this can easily be generalized) the SU does not initiate the scanning procedure. Rather, she continues transmitting with a lower number of channels, until she loses all of them. We assume that the maximum number of channels the SU can bond together is $M$. After getting all of them together, the SU will transmit at the highest rate. Once the SU starts losing the channels one after another, the data rate will decrease accordingly. After the last available channel is lost, only then will the SU initiate the scanning procedure. In the following, we will consider the interweave mode first.

a) *Interweave mode*: We assume that the distribution of the PU idle periods is identical exponentially distributed with rate $\eta_H$. This is in line with our assumptions in the original single channel model. The arrival process is Poisson with rate $\lambda$. The scanning time is exponentially distributed (but now to bond $M$ channels) with rate $\eta_L$. The scanning time can be also Erlang or hyperexponential, with no major changes in the analysis. The Markov chain to mimic this system has more than two states vertically. Each state is described by the couple $\{j, i\}$, where $j$ is the number of files in the SU buffer, and $i$ represents the number of channels the SU is currently transmitting on. The 2D Markov chain is shown in Fig. 11. There are a number of possible transitions corresponding to the following events:
New arrival: From any state of any level, the chain moves to the right (horizontally) with rate $\lambda$.
File finishes transmission: From any state $\{j, i\}, (i > 0)$, the chain moves to the left (horizontally) with a different rate. This rate depends on the number of channels the user is currently having. If the rate of a channel is $\mu_H$, then with $i$ channels having identical exponential distribution, the transition rate is $i\mu_H$. So, going from the first row down, these rates are $M\mu_H$, $(M-1)\mu_H, \ldots, \mu_H$. Needless to say, when all the channels are lost, the SU initiates scanning and there is no transmission. Hence, no transition backwards in the last row (for states $\{j, 0\}$).
Losing a channel due to PU return in that channel: When such an event occurs the chain moves vertically from level $i$ (state $j, i$) to level $i-1$ (state $\{j, i-1\}$) (transitions only to the neighboring states are possible vertically). The rate at which the chain moves vertically depends on the number of channels she is currently using. Namely, if the SU is communicating on all the $M$ channels, then the transition rate to level $M-1$ is $M\eta_H$. The reason is that we are looking at the minimum of $M$ exponentially distributed random variables (the first channel to lose), and the rate of such a random variable is equal to the sum of the rates of the individual random variables, which is $M\eta_H$. When transmitting on $M-1$ channels, the chain moves downwards at rate $(M-1)\eta_H$. Finally, after being left with only one channel, the chain moves to the last level (row) with rate $\eta_H$. Obviously, due to the pol-

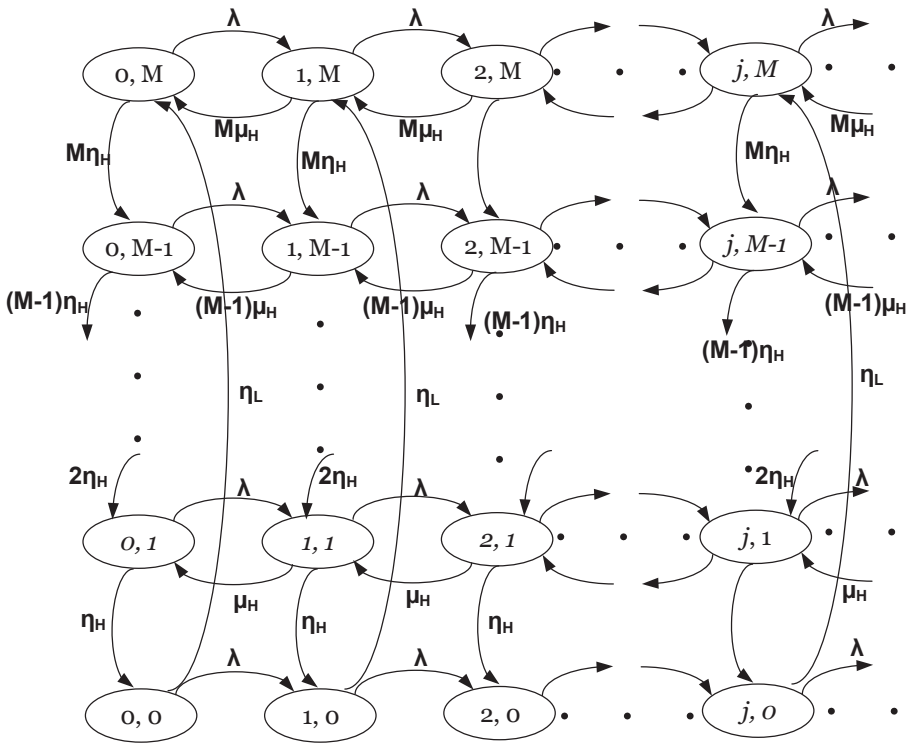


Figure 11: The 2D Markov chain for the multichannel interweave model.

icy used that after losing a channel transmission goes on, it is not possible to have transitions vertically upwards. Only when the chain is in one of the states $\{j, 0\}$ a transition upwards is possible. Namely, then the scanning for $M$ channels starts, and the chain moves to the first level after bonding $M$ channels together. We assume that this happens according to an exponential distribution with rate $\eta_L$. This model can also capture, with a slight modification, the Erlang and hyperexponentially distributed scanning times too.

Having introduced all the possible transitions with the corresponding transition rates, we can proceed to solve the Markov chain. Its solution will lead to the average file delay in the multichannel interweave mode.

If we look carefully at the chain in Fig. 11, we can infer that this chain is actually a special case of the more general chain we have already solved. That is the chain corresponding to the $M$ different power levels in the underlay mode (Fig. 10).

Our chain is a special case of the chain in Fig. 10 with the following rates: $\mu_M = M\mu_H, \mu_{M-1} = (M-1)\mu_H, \ldots, \mu_1 = 0, \eta_{1,M} = \eta_L, \eta_{i,i-1} = i\eta_H$, and all the other $\eta_{i,j} = 0$.

b) *Underlay mode*: The Markov chain for the multichannel underlay model is slightly different compared to the interweave model (Fig. 11). There is no scanning time involved here. The rate at which the PU leaves the channel, in line with the single channel case, is exponentially distributed with rate $\eta_L$. The SU uses always the same set of $M$ channels. Once a channel is lost due to PU return, the rate contribution from that channel is $\mu_L$ and $(M-1)\mu_H$ from the idle channels. The 2D Markov chain is shown in Fig. 12. There are a number of possible transitions corresponding to the following events:

New arrival: From any state of any level, the chain moves to the right (horizontally) with rate $\lambda$.

File finishes transmission: From any state $\{j, i\}, (i > 0)$, the chain moves to the left (horizontally) with a different rate. This rate depends on the number of channels the user is currently having idle and busy. If the rate of a channel is $\mu_H$, then with $i$ idle channels the transition rate is $i\mu_H + (M-i)\mu_L$. The reason is that if there are $i$ idle channels, the SU will transmit at full power on those, and on the remaining $M-i$ channels that are busy, she can transmit at the lower rate of $\mu_L$. Hence, the total rate in this case is $i\mu_H + (M-1)\mu_L$.

Losing (getting) a channel due to PU return (departure) in (from) that channel: Since we are dealing with the underlay mode, there is no scanning of other channels involved. The SU is stuck with a given set of $M$ channels. When an event of losing a channel occurs the chain moves vertically from level $i$ (state $\{j, i\}$) to level $i-1$ (state $\{j, i-1\}$) (transitions only to the neighboring states are possible vertically). The rate at which the chain moves vertically depends on the number of channels she is currently using. Namely, if the SU is communicating on all the $M$ channels, the transition to level $M-1$ is $M\eta_H$. The reason is that we are looking at the minimum of $M$ exponentially distributed random variables (the first channel to lose), and the rate of such a random variable is equal to the sum of rates of the individ-

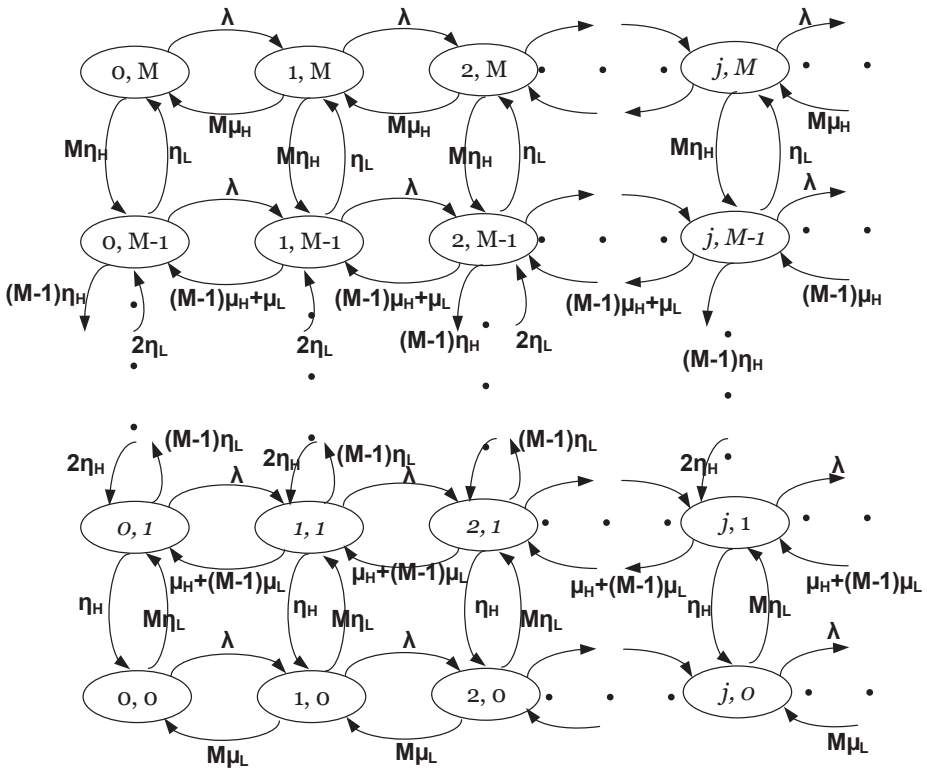


Figure 12: The 2D Markov chain for the multichannel underlay model.

ual random variables, which is $M\eta_H$. When transmitting on $M-1$ channels, the chain moves downwards with rate $(M-1)\eta_H$. Finally, after being left with only one channel, the chain moves to the last level (row) with rate $\eta_H$.

In case a channel becomes idle while being at state $\{j,i\}$, the chain moves vertically upwards to state $\{j,i+1\}$. The rate at which the chain moves vertically depends on the number of channels that are occupied by the corresponding PUs. For instance, if 5 channels are busy, then the rate at which the first one leaves is exponentially distributed with rate $5\eta_L$ (the minimum of 5 identical exponentially distributed variables). So, if the chain is in state $\{j,i\}$, the transition rate to state $\{j,i+1\}$ is $i\eta_L$. Having introduced all possible transitions with the corresponding transition rates, we can proceed solving this Markov chain. Its solution will lead to the average file delay in the multichannel underlay mode.

This chain is also a special case of the chain of Fig. 10, with the following rates: $\mu_M = M\mu_H, \mu_{M-1} = (M-1)\mu_H + \mu_L, \ldots, M\mu_L, \eta_{i,i-1} = i\eta_H, \eta_{i,i+1} = (M-i)\eta_L$. The solution to this chain is then similar to that of Fig. 10.

We would like to point out again that depending on the scanning policy, different chains will arise. Nevertheless, most of them will be special cases of the $M$-level model.

### 3.5 Miss-detections and false alarms

So far, we have been assuming that spectrum sensing is perfect. However, in practice that is not the case, since often the SU cannot detect the presence of a PU signal. When that happens, we say that there is a *miss-detection*. We assume that the probability of miss-detection is equal to $p_{md}$ [15]. On the other hand, if the SU is very sensitive and can detect even a very low power signal (not coming from a PU), the SU will perceive it wrongly as PU signal. This is known as *false alarm*, and the corresponding probability is denoted by $p_{fa}$.

Let's consider the underlay mode (Fig. 8). We assume that an SU periodically senses the channel. Let's denote with $E[T_{is}]$ the average inter-sensing time (the average time between two consecutive sensing instants). We assume w.l.o.g. that it is constant, and each trial is Bernoulli with a probability of "success" (the probability of a PU arriving to the channel) $p = \frac{E[T_{OFF}]}{E[T_{ON}]+E[T_{OFF}]}$. In the ideal case (no false alarms nor miss-detections), the average ON period would be

$$E[T_{ON}] = \frac{1}{p} E[T_{is}]. \tag{108}$$

In case we consider both miss-detections and false alarms, then the probability of a PU arrival as perceived by the SU is $p(1-p_{md})$, i.e., there is an arrival and it is correctly detected. In order the SU to remain in the high period, there must be *no arrivals nor false alarms*. The probability for this to happen is $(1-p\,(1-p_{md}))\,(1-p_{fa})$. From this we get the probability of an SU leaving the high state at the moment of sensing as $1-(1-p\,(1-p_{md}))\,(1-p_{fa})$, which

after rearranging leads to

$$p^{'} = p_{fa} + (1 - p_{md}) \cdot p \cdot (1 - p_{fa}) \,. \tag{109}$$

The actual average duration of an ON period when we take into account these imperfections is

$$E[T_{ON,im}] = \frac{1}{p^{'}} E[T_{sense}] = \frac{p}{p^{'}} E[T_{ON}], \tag{110}$$

$$E[T_{ON,im}] = \frac{p}{p_{fa} + (1 - p_{md}) \cdot p \cdot (1 - p_{fa})} \cdot E[T_{ON}], \tag{111}$$

and for the transition rate out of an ON period

$$\eta_H^{'} = \frac{1}{E[T_{ON,im}]} = \frac{p_{fa} + (1 - p_{md}) \cdot p \cdot (1 - p_{fa})}{p} \cdot \eta_H. \tag{112}$$

We consider next the average duration of an OFF period when imperfections are considered. In the ideal scenario, the average duration of the OFF period would be

$$E[T_{OFF}] = \frac{1}{q} E[T_{is}], \tag{113}$$

where $q = \frac{E[T_{ON}]}{E[T_{ON}] + E[T_{OFF}]}$ is the probability that at a given sensing time the PU has left (equivalent to no PU).

As opposed to the ideal scenario, the probability of a PU leaving the observed channel as perceived by the SU is $q \cdot (1 - p_{fa})$, i.e., there is a PU "departure" and the SU can "see" it. The event of moving out of a low (OFF) period can be due to a miss-detection or to the fact that PU has really left. The probability for an SU to stay in the OFF period is simply $(1 - p_{md})\,(1 - q \cdot (1 - p_{fa}))$, while the probability to leave is $1 - (1 - p_{md})\,(1 - q \cdot (1 - p_{fa}))$, leading to

$$q^{'} = p_{md} + (1 - p_{fa}) \cdot q \cdot (1 - p_{md}) \,. \tag{114}$$

For the average duration of the low periods, we obtain

$$E[T_{OFF}^{'}] = \frac{1}{q^{'}} E[T_{is}] = \frac{q}{q^{'}} E[T_{OFF}], \tag{115}$$

and

$$E[T_{OFF}^{'}] = \frac{q}{p_{md} + (1 - p_{fa}) \cdot q \cdot (1 - p_{md})} \cdot E[T_{OFF}]. \tag{116}$$

From Eq.(116) we get the transition rate out of a low period as

$$\eta_L^{'} = \frac{1}{E[T_{OFF}^{'}]}, \tag{117}$$

resulting in

$$\eta_L^{'} = \frac{p_{md} + (1 - p_{fa}) \cdot q \cdot (1 - p_{md})}{q} \cdot \eta_L. \tag{118}$$

The other transition rate parameters of Fig. 8 remain unchanged. Although the result expressed through Eq.(79) will not be the same, the procedure is identical. Hence, the obtained results will still be in closed-form. The same conclusion can be drawn for the interweave access mode as well.

### 3.6 Analytical comparison of delays in underlay and interweave mode

Having derived the formulas for the mean delay in underlay and interweave CRNs in Sections 3.1 and 3.2, we are able to compare the delays incurred in each of them. As could have been noticed, the delay depends on the statistics of the PU activity, data rate, traffic intensity, and scanning time. In a first scenario, we assume that the SU has to decide at the beginning which of the access modes to use: underlay (i.e. always stay on same channel and transmit with the permitted power), or interweave (i.e. become silent whenever a PU arrives on the channel and scan for a new one). We will refer to this simply as "the static policy". While not a real policy per se (in practice, a node will always be able to scan and switch channels eventually), it allows us to gain some insights as to the relative parameters affecting the performance in each case. In Section 3.7, we will consider a more realistic, *dynamic policy*.

In general, for interweave access to outperform underlay access, the expected scanning time $E[T_s]$ should be short enough to ensure that the opportunity cost of not transmitting/receiving any data for some time (which is allowed in underlay) is amortized by the quick discovery of a new white space. In Table 2, we provide analytical expressions for the maximum $E[T_s]$ values for which interweave access has lower delays. As can be seen from Table 2, there is a complex dependency on the various system parameters. What is more, this "boundary" point further depends on the variability of the scanning time.

From Table 2 we can observe the following relations: $B_2 = 2kA_2$, $B_3 = 2kA_3$, and $B_1 < 2kA_1$. So, for the crossing point of the Erlang distributed scanning time, we have

$$\frac{-B_2+\sqrt{B_2^2-4B_1B_3}}{2B_1} = \frac{-2kA_2+\sqrt{4k^2A_2^2-4B_1\cdot 2kA_3}}{2B_1} >$$

$$> \frac{-2kA_2+\sqrt{4k^2A_2^2-8kA_1\cdot 2kA_3}}{2B_1} > \frac{-2kA_2+\sqrt{4k^2A_2^2-16k^2A_1A_3}}{4kA_1} =$$

$$= \frac{-A_2+\sqrt{A_2^2-4A_1A_3}}{2A_1} \tag{119}$$

From Eq.(119) we can observe that for the interweave mode to outperform underlay access, in the case of exponentially distributed scanning time a smaller scanning time is needed compared to the case of Erlang distributed scanning time. Similar conclusions can be drawn by comparing the parameters of hyperexponential distribution with the two previous ones, i.e., for hyperexponential distributed scanning time the scanning time is required to have the lowest value, so that the

Table 2: The analytical comparison of underlay and interweave modes.

| $T_s$ | Condition | Notation |
|---|---|---|
| Erlang | $E[T_s] < \frac{-B_2+\sqrt{B_2^2-4B_1B_3}}{2B_1}$ | $B_1 = 2\eta_H^2 k + \eta_H(k+1)\mu_H + 2kE[T_u]\lambda\eta_H^2$<br>$B_2 = \eta_H(4k - 2k\mu_H E[T_u] + 4kE[T_u]\lambda)$<br>$B_3 = 2k(1-(\mu_H-\lambda)E[T_u])$ |
| Exponential | $E[T_s] < \frac{-A_2+\sqrt{A_2^2-4A_1A_3}}{2A_1}$ | $A_1 = \eta_H(\mu_H+\eta_H) + \lambda\eta_H^2 E[T_u] > 0$<br>$A_2 = 2\eta_H - \eta_H(\mu_H - 2\lambda)E[T_u]$<br>$A_3 = 1-(\mu_H-\lambda)E[T_u]$ |
| Hyperexponential | $E[T_s] < \frac{-C_2+\sqrt{C_2^2-4C_1C_3}}{2C_1}$ | $C_1 = \eta_L\eta_V\eta_H^2(1+\lambda E[T_u])$<br>$C_2 = \eta_H[\mu_H(\eta_V+\eta_L) + \eta_L\eta_V(2+2\lambda E[T_u] - \mu_H E[T_u])]$<br>$C_3 = \eta_L\eta_V - \eta_H\mu_H - \eta_L\eta_V(\mu_H-\lambda)E[T_u]$ |

interweave mode still outperforms the underlay access. We can observe that the variability of the scanning time has an important impact on the boundary scanning time. The higher the variability of the scanning time, the lower the scanning time needed for the interweave mode to outperform the underlay access.

### 3.7 The delay minimization policy

In the previous section, we have compared underlay and interweave access, in a "static" context, where the decision between the two is made once, at the beginning. In practice, a node with a cognitive radio will normally be able to choose to stay at the current channel and transmit at low(er) power, or scan for a new white space *at any time*. Such a hybrid policy might lead to a further improvement in performance, if designed properly. We next define such a hybrid policy, identify the conditions under which it offers gains, and derive an optimal switching rule (from the one mode to the other).

**Definition 1. Delay minimization policy.**

- *The SU will reside on the current channel if it is idle (no PU activity) and continue its activity there.*
- *If a PU is detected, the SU will continue transmitting with lower power, until a time $t$, called the "turning point".*
- *If the PU does not release the channel by time $t$, then the SU ceases transmission and starts scanning for a new idle channel.*
- *If the PU leaves the channel before $t$, then the SU resumes transmitting at higher power, and resets the turning point to $t$ time units ahead.*

The above policy is generic. Our goal is to find an optimal value for $t$. Let us consider some cases, to better understand the tradeoffs involved. First, if the static interweave policy, as described in the previous section is better than the static underlay policy, then it is easy to see that the optimal value of $t$ is 0: it is always

better to start scanning immediately when a PU arrives. Hence, we are interested only in cases where the underlay is better *on average* (i.e. the respective condition in Table 2 *not* satisfied), but there are instances when the current channel remains busy for too long and it then becomes better to start scanning instead.

In the above context, assume that the PU activity (OFF) periods are exponentially distributed. Assume further that a PU arrived at the current channel and $t$ units have already elapsed and the channel is still busy. Due to the memoryless property of the exponential distribution, the remaining time until the PU leaves is still the same, as in the beginning (when the PU just arrived), i.e. equal to $E[T_{OFF}]$. Hence, if at time 0 it was better to stay on the channel and transmit at lower rate rather than initiate scanning (which is what we assumed above), for any elapsed time $t$ *it is still better to stay on the channel and not start scanning*. A similar conclusion can be drawn for PU activity periods with *increasing failure rate (IFR)*[4], i.e. lower variability than exponential. There, if at $t = 0$ one cannot gain by scanning (i.e. static underlay is better on average), then as $t$ increases, the expected gain from staying in the underlay mode in fact increases.

Hence, we can conclude that a dynamic policy (i.e. an optimal value of $t$ strictly larger than 0) may offer gains *only for PU activity periods with decreasing failure rate (DFR)*. There, although at the beginning, when the PU arrives, it might be on average better to do underlay, as time elapses, the expected remaining PU busy time keeps increasing (above the average), until at some point it becomes profitable to stop and scan for a new empty channel. This allows the dynamic policy to outperform any of the static policies, as we show later, by essentially "pruning" the long OFF periods from the underlay mode.

In deriving the optimal turning point for the dynamic delay policy we make the assumption that files are not excessively large (average size $\Delta$), and that they can be transmitted in 1-2 ON and OFF periods and that the arrival rate is not high. We use these approximations to keep the analysis tractable. Under these assumptions, the total transmission delay is almost equal to the average service time. If by $p_{ON} = \frac{E[T_{ON}]}{E[T_{ON}]+E[T_{OFF}]}$ we denote the probability that an arriving file will find the system in an ON state, and by $p_{OFF} = \frac{E[T_{OFF}]}{E[T_{ON}]+E[T_{OFF}]}$ the probability of finding the system in an OFF state, then the average service time would be

$$E[S] = p_{ON}E[T_{X,ON}] + p_{OFF}E[T_{X,OFF}]. \tag{120}$$

In Eq.(120), $E[T_{X,ON}]$ ($E[T_{X,OFF}]$) denotes the average service time of a file that arrives during an ON (OFF) period. Each arriving file (in an ON period) will be partially transmitted during the time $T_{ON}^{e}$, which is the remaining (excess) duration of the ON period. Then, at the beginning of the OFF period, the remaining file size is $\Delta_0$. There are two options for the OFF periods (larger or smaller than $t$). With probability $P[T_{OFF} < t] = F_{OFF}(t)$ ($F_{OFF}(t)$ is the cumulative distribution function (CDF) of the OFF period at $t$), the next OFF period will be short enough,

[4]The distributions with increasing (decreasing) failure rate are those for which $\frac{f(x)}{1-F(x)}$ is an increasing (decreasing) function of $x$.

so the SU will reside on the current channel and transmit with low power. In that case, the file will be completely transmitted in the next ON period. On the other hand, with probability $P\left[T_{OFF} > t\right] = \bar{F}_{OFF}(t)$, the next OFF period will be larger than the turning point. Note that $P\left[T_{OFF} > t\right] = 1 - F_{OFF}(t) = \bar{F}_{OFF}(t)$ is the complementary cumulative distribution function (CCDF) of the OFF period at $t$. After that, the SU will initiate the scanning procedure and will start looking for another available channel (for $T_s$ time units). During that period of time, there will be no transmissions. After finding an available channel, the SU will transmit with rate $c_H$. So, the average service time of a file arriving during an ON period is

$$\begin{aligned} E[T_{X,ON}] = E[T_{ON}^{(e)}] &+ F_{OFF}(t)\left\{E[T_{OFF}|T_{OFF} < t] + \frac{\Delta_0 - c_L E[T_{OFF}|T_{OFF} < t]}{c_H}\right\} \\ &+ \bar{F}_{OFF}(t)\left\{t + E[T_s] + \frac{\Delta_0 - c_L t}{c_H}\right\} + \Omega_{ON}. \end{aligned} \quad (121)$$

Note that in Eq.(121), the term $\Omega_{ON}$ represents the contribution to the average delay of other scenarios not included in the other term (the file transmitted during the first ON period, during the first OFF period, or eventually if it needs more ON and OFF periods to complete the transmission). However, as we have assumed that file sizes are exponentially distributed, and hence with a low variance, and that their sizes are such that in most cases it will suffice 1-2 ON-OFF periods to complete the transmission, the term $\Omega_{ON}$ is negligible ($\Omega_{ON} -> 0$).

For files arriving during an OFF period, there are also two possibilities. They could either arrive to an OFF period whose remaining time, $T_{OFF}^{(e)}$, is shorter than the turning point, or to an OFF period with excess time larger than $t$. The probability for the first scenario is $P\left[T_{OFF}^{(e)} < t\right] = F_{OFF}^{(e)}(t)$ , while for the second one it is $P\left[T_{OFF}^{(e)} > t\right] = \bar{F}_{OFF}^{(e)}(t)$. In the first case, the SU will remain on the current channel and continue its transmission with rate $c_L$, and the file will be transmitted during the next ON period (when the rate is $c_H$). In the second case, after time $t$, the SU will initiate the scanning process (no transmission) that will last $T_s$ until an idle channel is found, and then will transmit with rate $c_H$. The file will be transmitted during that ON period. So, the average service time of a file arriving during an OFF period can be expressed as

$$\begin{aligned} E[T_{X,OFF}] &= F_{OFF}^{(e)}(t)\left\{E\left[T_{OFF}^{(e)}|T_{OFF}^{(e)} < t\right] + \frac{\Delta - c_L E\left[T_{OFF}^{(e)}|T_{OFF}^{(e)} < t\right]}{c_H}\right\} \\ &+ \bar{F}_{OFF}^{(e)}(t)\left\{t + E[T_s] + \frac{\Delta - c_L t}{c_H}\right\} + \Omega_{OFF}. \end{aligned} \quad (122)$$

For the same reasons as for $\Omega_{ON}$ in Eq.(121), the term $\Omega_{OFF}$ in Eq.(122) can be neglected.

In Eq.(121), the following two terms are equivalent to

$$E\left[T_{ON}^{(e)}\right] = \frac{E[T_{ON}^2]}{2E[T_{ON}]}, \quad (123)$$

$$E[T_{OFF}|T_{OFF} < t] = \int_0^t \frac{x f_{OFF}(x)}{F_{OFF}(t)} dx. \tag{124}$$

Similarly, we have the following relations for the terms in Eq.(122)

$$F_{OFF}^{(e)}(t) = \int_0^t f_{OFF}^{(e)}(x) dx, \tag{125}$$

$$E\left[T_{OFF}^{(e)}|T_{OFF}^{(e)} < t\right] = \int_0^t \frac{x f_{OFF}^{(e)}(x)}{F_{OFF}^{(e)}(t)} dx. \tag{126}$$

In the two previous equations we have

$$f_{OFF}^{(e)}(x) = \frac{1 - F_{OFF}(x)}{E[T_{OFF}]}. \tag{127}$$

Replacing Eq.(121) and Eq.(122) into Eq.(120), we find the average service time. We are assuming that the average transmission delay is almost identical to the average service time, i.e. $E[T] \approx E[S]$. Further, to find the value of turning point that minimizes the delay, we need to solve the following equation

$$\frac{\partial E[T]}{\partial t} = 0. \tag{128}$$

So, the solution to Eq.(128) is the optimal turning point. For illustration purposes, after solving Eq.(128), the following result provides an analytical expression for the case of Pareto distributed OFF periods (with parameters $L, \alpha$), a popular distribution with decreasing failure rate, and for exponential ON periods.

**Result 6.** *The optimal turning point, $t_{opt}$, in a Cognitive Radio Network can be found as the solution to*

$$\begin{aligned} &\frac{\eta_H}{\eta_H + \eta_L} \cdot \frac{c_H - c_L}{c_H} \left( \left(1 - \frac{1}{\alpha}\right) t^{1+\alpha} + \frac{1}{\alpha} L^{\alpha-1} t^2 \right) \\ &+ \frac{\eta_H}{\eta_H + \frac{c_H}{\Delta}} \cdot \frac{\eta_L}{\eta_H + \eta_L} \left( \frac{c_H - c_L}{c_H} L^{\alpha} t - E[T_s] \alpha L^{\alpha} \right) = 0. \end{aligned} \tag{129}$$

The solution to Eq.(129) can be obtained numerically. Table 3 summarizes all the possible scenarios.

Table 3: Summary of the delay policies.

| Scenario | Optimal dynamic policy |
|---|---|
| Static interweave better | Static interweave ($t_{opt} = 0$) |
| Static underlay better + IFR OFF | Static underlay ($t_{opt} = \infty$) |
| Static underlay better + Exp. OFF | Static underlay ($t_{opt} = \infty$) |
| Static underlay better + DFR OFF | Dynamic policy with $t_{opt} \in (0, \infty)$ |

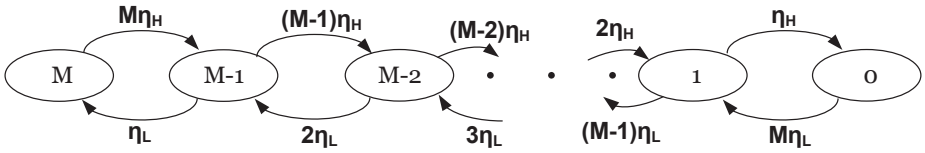


Figure 13: The Markov chain illustrating the dynamics of the number of idle channels.

It is interesting to note that, unlike the variability of OFF periods for underlay access, the variability of the scanning time distribution (the "OFF" periods in the interweave mode) does not affect the dynamic policy decisions. It only enters the picture for the comparison between the static underlay and interweave modes (Table 2).

It is also positive that in all but few cases the optimal policy is just the static one. This reduces the complexity of the algorithm significantly, as one needs to make this decision only once. In practice, some recalculation of this threshold (and thus the optimal mode) might still be necessary periodically, in order to account for qualitative (e.g. non-stationarity) changes in the behavior of the system and estimated statistics.

# 4 Throughput analysis of underlay and interweave access

## 4.1 Analytical comparison of throughput

The content of this subsection can be found in our journal submission.

## 4.2 Throughput model with multiple channels

Our approach based on the renewal-reward theory still holds. The policy is the same as for the delay models presented above. We will consider first the interweave model. For the sake of simplicity we will show the analysis assuming exponentially distributed OFF periods. If at first the user transmits on $M$ channels, assuming the same capacity on all of them, the average amount of data sent during the time the user has all the $M$ channels available is $\frac{Mc_H}{M\eta_H} = \frac{c_H}{\eta_H}$. From the moment losing the first channel to the moment losing the second one, the average amount of transmitted data is the same, because $\frac{(M-1)c_H}{(M-1)\eta_H} = \frac{c_H}{\eta_H}$. In the second case, we have the minimum of $(M-1)$ exponentially distributed random variables, hence

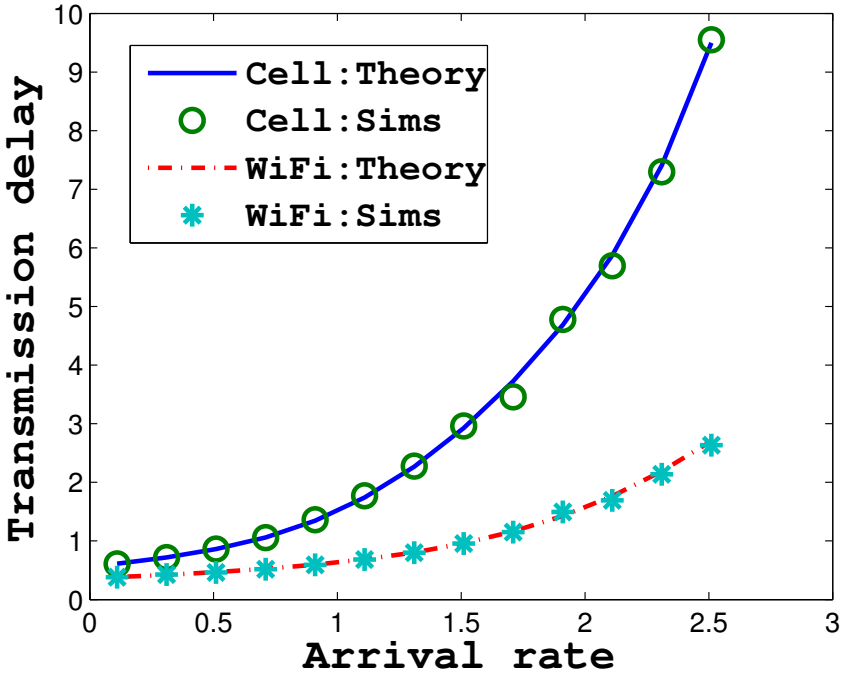


Figure 14: The delay for underlay spectrum access.

the factor $(M-1)$ in the denominator. The same result propagates to any number of available channels. During the scanning phase, there is no transmission. The average duration until the first channel is lost is $\frac{1}{M\eta_H}$, then for the second channel to become occupied after the first one is already busy is $\frac{1}{(M-1)\eta_H}$, and so on. The average duration between losing the penultimate and the last channel is $\frac{1}{\eta_H}$. The average duration of the scanning time is $\frac{1}{\eta_L}$. So, the average cycle duration is

$$E[T_{i,cycle}] = \frac{1}{\eta_H}\left(\frac{1}{M} + \frac{1}{M-1} + \ldots + 1\right) + E[T_s].$$

For the average reward rate (average throughput) during a cycle for the multi-channel interweave mode we have

$$E[X_i] = \frac{E[R_i]}{E[T_{i,cycle}]} = \frac{M\frac{c_H}{\eta_H}}{\frac{1}{\eta_H}\sum_{i=1}^{M}\frac{1}{i} + E[T_s]}. \tag{130}$$

As far as the throughput model for the multichannel underlay mode is concerned, the situation is slightly different, although the approach relies on renewal-reward theory. As a first step, we need to define the cycle duration and the amount of data transmitted during a cycle. To that direction, we define the cycle as the duration from the moment the SU collects all the $M$ channels until the next moment

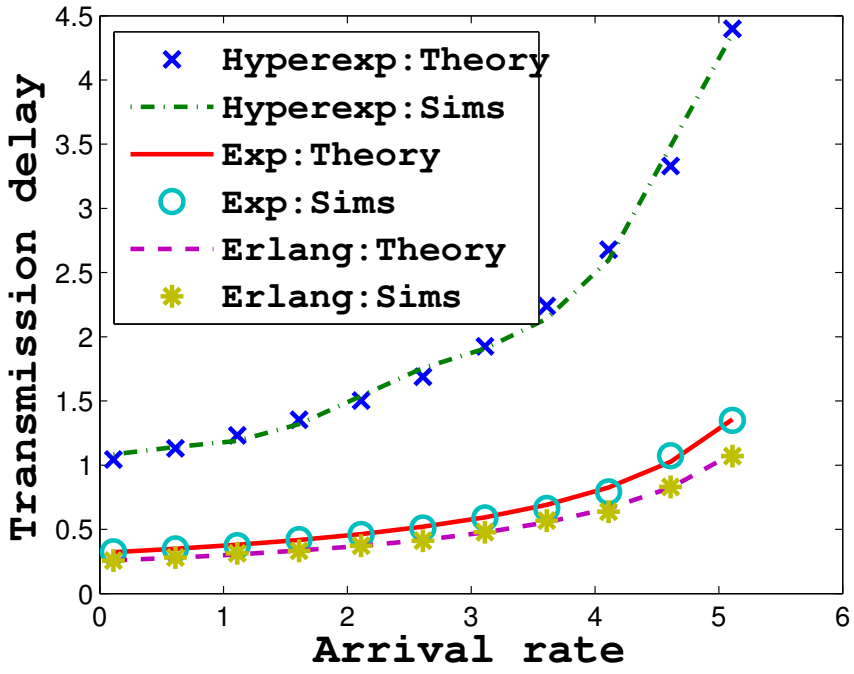


Figure 15: The delay for interweave spectrum access in a cellular scenario.

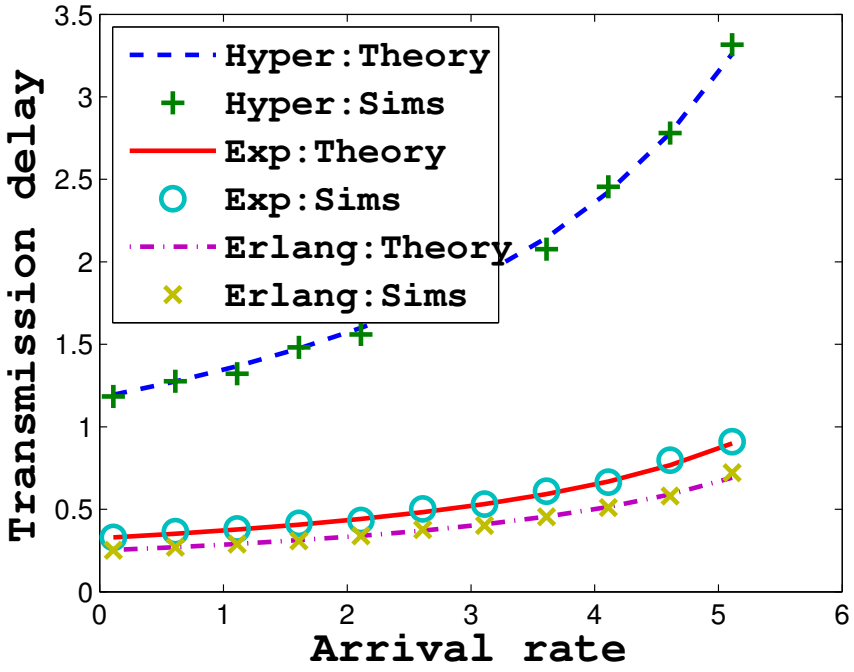


Figure 16: The delay for interweave spectrum access in a WiFi scenario.

when she will have collected $M$ channels. Since we are dealing with exponentially distributed random variables and Markov chains, due to the strong Markov property, the moment the SU collects all the $M$ channels can be considered to be the point when the system probabilistically restarts itself. Hence, that moment can be considered to be *a renewal*. We have already established the transition rates of a given state depending on the fact whether there is a channel gained or lost (see Fig. 12). The corresponding Markov chain is illustrated in Fig. 13. Basically, this chain is the same as any of the vertical chains of Fig. 12. The state $i, i = 0, \ldots, M$ denotes the number of idle channels the SU is transmitting on.

Before obtaining the average cycle duration, we need to determine the stationary probabilities of the states of the chain, i.e., $\pi_i$, for $i = 0, \ldots, M$.

The local balance equations for this chain are:
1) for states 0 and 1: $\eta_H \pi_1 = M\eta_L \pi_0 => \pi_1 = \frac{M\eta_L}{\eta_H}\pi_0$;
2) for states 1 and 2:
$2\eta_H \pi_2 = (M-1)\eta_L \pi_1 => \pi_2 = \frac{(M-1)\eta_L}{2\eta_H}\pi_1 = \frac{M(M-1)}{2}(\frac{\eta_L}{\eta_H})^2\pi_0$;
3) for states 2 and 3:
$3\eta_H \pi_3 = (M-2)\eta_L \pi_2 => \pi_3 = \frac{(M-2)\eta_L}{3\eta_H}\pi_2 = \frac{M(M-1)(M-2)}{3\cdot 2}(\frac{\eta_L}{\eta_H})^3\pi_0 = \binom{M}{3}(\frac{\eta_L}{\eta_H})^3\pi_0$;
...
4) for states M-1 and M: $M\eta_H \pi_M = (M-1)\eta_L \pi_{M-1} => \pi_M = \frac{\eta_L}{M\eta_H}\pi_{M-1} = (\frac{\eta_L}{\eta_H})^M\pi_0 = \binom{M}{M}(\frac{\eta_L}{\eta_H})^M\pi_0$.

Carefully observing the series of equations in 1)-4), it can be deduced that the stationary probability of any state $i$ can be written as $\pi_i = \binom{M}{i}(\frac{\eta_L}{\eta_H})^i\pi_0$.

From the normalization condition $\sum_{i=0}^{M}\pi_i = 1$, we obtain $\pi_0$ from $\sum_{i=0}^{M}\binom{M}{i}(\frac{\eta_L}{\eta_H})^i\pi_0 = 1 => \pi_0 = \frac{1}{\sum_{i=0}^{M}\binom{M}{i}(\frac{\eta_L}{\eta_H})^i} = \frac{1}{\left(1+\frac{\eta_L}{\eta_H}\right)^M} = \left(\frac{\eta_H}{\eta_L+\eta_H}\right)^M$.

Hence, for the equilibrium distribution of any state we have $\pi_i = \left(\frac{\eta_H}{\eta_L+\eta_H}\right)^M \binom{M}{i}\left(\frac{\eta_L}{\eta_H}\right)^i = \binom{M}{i}\left(\frac{\eta_L}{\eta_L+\eta_H}\right)^i\left(\frac{\eta_H}{\eta_L+\eta_H}\right)^{M-i}$, for $i = 0, \ldots, M$. So, the probability of being in one of the states is binomially distributed, with parameter (probability) $\frac{\eta_L}{\eta_L+\eta_H}$.

If the average duration of a cycle is $E[T_{u,cycle}]$, then during this cycle the SU will have $i$ idle channels for $\pi_i E[T_{u,cycle}]$ time units. Given that when there are $i$ idle channels, there are $M-i$ busy channels, the transmission rate during this time is $ic_H + (M-i)c_L$, whereas[5] the total amount of transmitted data during a cycle is given by

$$E[R_u] = E[T_{u,cycle}]\sum_{i=0}^{M}(ic_H + (M-i)c_L)\pi_i, \tag{131}$$

[5]Note that, as mentioned in previous sections, $\mu_H = \frac{c_H}{\Delta}$ and $\mu_L = \frac{c_L}{\Delta}$, where $\Delta$ is the average file size.

which leads to an average throughput of

$$E[X_u] = \frac{E[R_u]}{E[T_{u,cycle}]} = \sum_{i=0}^{M}(ic_H + (M-i)c_L)\pi_i = Mc_L + (c_H - c_L)\sum_{i=0}^{M} i\pi_i.$$

The term $\sum_{i=0}^{M} i\pi_i$ is apparently the expectation of the Binomial distribution with parameter $\frac{\eta_L}{\eta_L+\eta_H}$, which is known to be $M\frac{\eta_L}{\eta_L+\eta_H}$. Finally, for the average throughput, after some simple algebra, we have

$$E[X_u] = \frac{M}{\eta_H + \eta_L}(\eta_H c_L + \eta_L c_H). \tag{132}$$

Comparing Eq.(130) and Eq.(132), we can find for what values of $E[T_s]$ the interweave mode outperforms the underlay mode in the multichannel case.

Obviously, our models can be extended to the multichannel case for both metrics, the delay and throughput, relying partly on the single channel theory. Hence, the significant advantage our models offer in predicting the performance, showing that the contribution is not limited.

### 4.3 Throughput maximization policy

**Theorem 7.** *The optimal value of the turning point that maximizes the average throughput for Pareto OFF, and exponential ON periods is the solution to*

$$\frac{c_H - c_L}{\eta_H}t^{\alpha} - \alpha E[T_s]\left(\frac{c_H}{\eta_H} - c_L L\frac{\alpha}{1-\alpha}\right)t^{(\alpha-1)} - \frac{c_L E[T_s]L^{\alpha}}{1-\alpha} = 0. \tag{133}$$

*Proof.* We will show here only the part of the proof that was omitted in the original journal submission. Up to that point, the procedure is the same.

The term $E\left[T_{ON}^{(e)}\right]$ is given by Eq.(123), and for an exponential ON period, due to its memoryless property, it reduces to $E[T_{ON}]$.

For Pareto distributed OFF periods, their probability density function is

$$f_{OFF}(x) = \frac{\alpha L^{\alpha}}{x^{\alpha+1}}, x \geq L.$$

The CDF of Pareto distribution is

$$F_{OFF}(x) = 1 - \left(\frac{L}{x}\right)^{\alpha}, x \geq L.$$

After some algebra, we can also find the conditional expectation $E[T_{OFF}|T_{OFF} < t]$ for a Pareto distribution, as

$$E[T_{OFF}|T_{OFF} < t] = \int_0^t \frac{x f_{OFF}(x)}{F_{OFF}(t)}dx = \frac{\alpha L^{\alpha}}{(1-\alpha)\left(1-\frac{L^{\alpha}}{t^{\alpha}}\right)}\left(t^{1-\alpha} - L^{1-\alpha}\right). \tag{134}$$

The average data rate is given by

$$E[X] = \frac{E[R]}{E[T_{cycle}]}, \tag{135}$$

where

$$E[R] = (c_L E[T_{OFF}|T_{OFF} < t] + c_H E[T_{ON}])\, F_{OFF}(t) + \left(c_L t + c_H E[T_{ON}^{(e)}]\right) \bar{F}_{OFF}(t)$$

and

$$E[T_{cycle}] = (E[T_{OFF}|T_{OFF} < t] + E[T_{ON}])\, F_{OFF}(t) + \left((t + E[T_s]) + E[T_{ON}^{(e)}]\right) \bar{F}_{OFF}(t).$$

The optimal value of the turning point $t$ that maximizes the throughput can be found by solving the equation

$$\frac{\partial E[X]}{\partial t} = 0. \tag{136}$$

Solving Eq.(136), we find the optimal turning point that maximizes the throughput in our dynamic throughput policy. □

# 5 Multi-SU scenario

Considering multiple SUs unavoidably involves the contention or scheduling phase, and because of that, the issue of MAC scheduling needs to be addressed. Consider a specific secondary user. She can be interrupted either by a PU or by another SU. This is oblivious to the given SU, and she perceives it only as if the channel is busy with only PU activity. When this SU gets channel access, she resumes file transmission from the point of interruption and the new files (or flows) that have arrived in the meantime will be served conforming to an FCFS scheduling discipline. Consequently, we can either assume a MAC protocol providing the specific SU exclusive access to the medium for the entire flow, or a TDMA scheme where each SU can utilize only a fraction of the ON (high) period for transmission, leading to larger OFF periods for that SU. Nevertheless, this assumption preserves the validity of our model.

We would like to point out that the common assumption made in the rich body of literature is that at any given time and over any given channel, it is assumed at most one SU active link in a given neighborhood [16]. This means that different SU links do not interfere with each other. Various spectrum access protocols have been proposed to handle SU-SU interference. For more details see [17, 18].

SU-SU interference is eliminated by enforcing an "exclusive channel occupancy" policy among SUs. Specifically, a channel is allocated to only a single SU link in a given geographical area. This is done using a modified version of

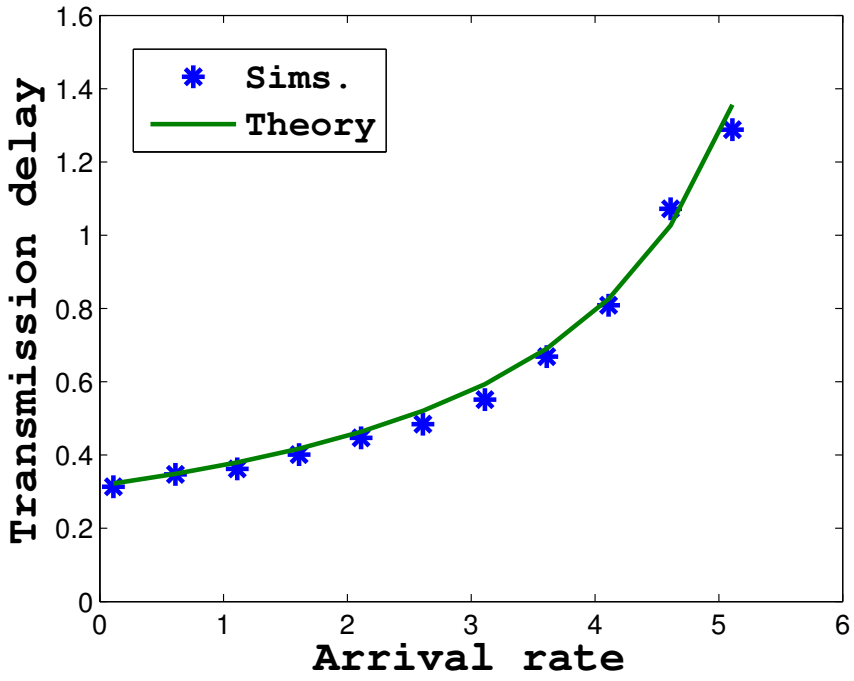


Figure 17: The delay for generic interweave spectrum access.

contention-based channel access approaches (e.g., CSMA/CA). Further, according to [16], SUs can communicate over a non-dedicated common control channel (CCC) and perform a threeway handshake to exchange different control information. During this handshake, the channel assignment and transmission duration are announced. Neighboring SUs defer from using the channel until the ongoing transmission ends. The latter is in line with our TDMA scheme mentioned above.

# 6 Simulation results

The first goal of this section is to validate the various analytical expressions we have derived, against simulated scenarios, including ones where one or more of the assumptions do not hold. We will also show the improvements offered by the dynamic policy.

In the first scenario, we will assume that the average ON and OFF durations correspond to those measured in [19] and are equal to $E[T_{OFF}] = 10$ s ($\eta_L = 0.1\ s^{-1}$), and $E[T_{ON}] = 5$ s ($\eta_H = 0.2\ s^{-1}$). We will refer to this as the cellular network scenario. In the second scenario, we fit the average ON and OFF durations to the values observed in [20], with $E[T_{OFF}] = 9$ s ($\eta_L = 0.11\ s^{-1}$), and $E[T_{ON}] = 4$ s ($\eta_H = 0.25\ s^{-1}$). We will refer to this as the WiFi scenario. For both scenarios, unless otherwise stated, we assume exponential distributed periods.

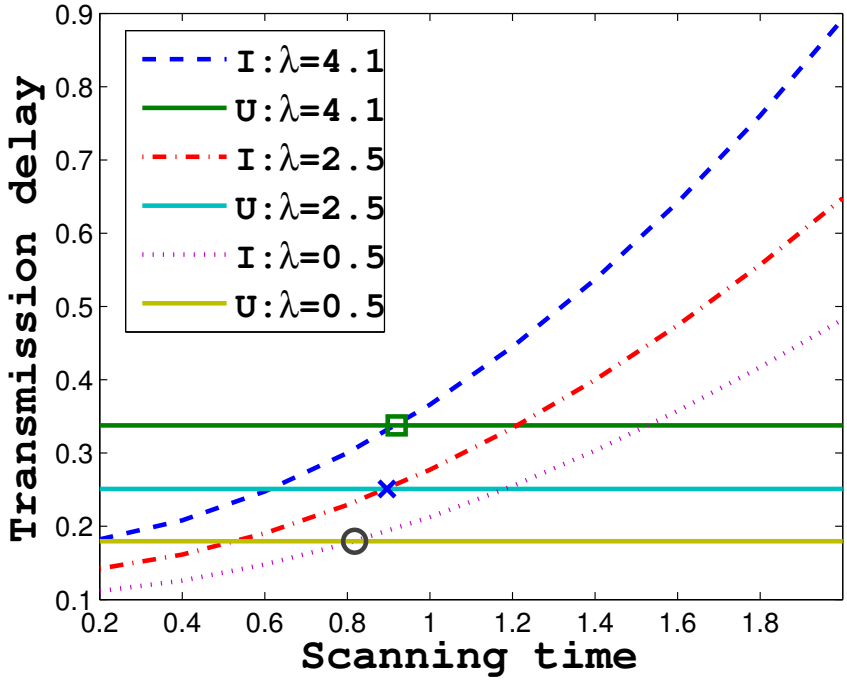


Figure 18: The static delay policy for different $\lambda$ and exp. scanning times.

We consider other distributions later in Sections 6.1 and 6.2. The data rates for the cellular scenario are $c_L = 1.2$ Mbps and $c_H = 8$ Mbps. For the WiFi scenario the data rates are $c_H = 10$ Mbps, and $c_L = 2$ Mbps.[6] Finally, we assume that file arrivals at the SU are Poisson distributed with rate $\lambda$, and file sizes exponentially distributed with mean size 125 KB.[7]

### 6.1 Validation of the delay models

Fig. 14 compares simulation results to our analytical model predictions for the average delay of SU files as the file arrival rate increases. The system parameters correspond to the cellular scenario. As can be seen, our theoretical results match with the results obtained from simulations. As is expected in queueing systems, the delay increases when the arrival rate (and thus the utilization of the system)

[6]These values are taken to be of the same order of magnitude as the actual values encountered in practice [21]. Although these correspond to PUs, and the actual data rates for SUs depend on the distance of the SUs from the BTS or WiFi AP, channel width, channel conditions, modulation/coding, etc., we assume w.l.o.g. that the data rates of the SU in a WiFi network are higher than in a cellular network.

[7]This value is normalized for the arrival rates considered, to correspond approximately to the traffic intensities reported in [19] and [20]. We have also considered other values with similar conclusions drawn.

increases. The delay incurred in the cellular scenario is larger, since WiFi data rates are considered to be higher and the PU is less active there.

We move next to validating our analytical predictions for the interweave scenario. Fig. 15 and 16 show the theoretical vs. simulated results for the cellular and WiFi network scenarios for three types of scanning time distributions (exponential, 4-stage Erlang, hyperexponential), all with the same mean $E[T_s] = 1$ s. For the hyperexponential scanning time, we take $\eta_L = 1.9\ s^{-1}$ and $\eta_V = 0.1\ s^{-1}$. The probability of having a large scanning time (far away channel) is $0.05$. The coefficient of variation for this distribution is around 3.

As the plot shows, the theory is correct and provides an excellent match with simulations for different scanning time distributions. Another outcome is that the average delay has the lowest value for Erlang distributed scanning time, while the worst performance is achieved for hyperexponential distribution. The above conclusion is in line with our analytical outcome of Section 3. As our analysis suggest, higher (lower) variability in scanning time leads to higher (lower) variability in the service time, which leads further to higher (lower) delays. Observing the curve corresponding to the cellular case in Fig. 14 and Fig. 15, it can be noticed that the delay in the interweave access is lower than in the underlay. For the cellular scenario with exponential scanning time, and $\lambda = 1\ s^{-1}$, Table 2 suggests that the maximum average scanning time should be 2.8 s. In our case $E[T_s]$ is much smaller (1 s). Hence, the interweave mode is superior.

In the previous scenarios we have used realistic values for the transmission rates and WiFi availabilities, but we have assumed exponential distributions for ON and OFF periods, according to our model. While the actual distributions are subject to the PU activity pattern, measurement studies [19, 20] suggest these distributions to be "heavy-tailed". It is thus interesting to consider how our model's predictions fare in this (usually difficult) case. To this end, we consider a scenario with "heavy-tailed" ON/OFF distributions (Bounded Pareto-BP), with parameters $L_{ON} = 1.31, H_{ON} = 200, \alpha_{ON} = \alpha_{OFF} = 1.2, L_{OFF} = 2.9, H_{OFF} = 200$. Due to space limitations, we focus on the cellular scenario. The other parameters are the same as for the scenarios of Fig. 14 and 15. Figure 17 compares the average file delay for the interweave access against our theoretical prediction. The scanning time is exponential with mean 1 s. Interestingly, our theory still offers a reasonable prediction accuracy, despite the considerably higher variability of ON/OFF periods in this scenario. Although we cannot claim this to be a generic conclusion for any distribution, the results emphasize the utility of our models in practice.

### 6.1.1 Validation of the multiple power level model

Finally, we would like to consider the case of multiple power levels ($M > 2$) in the underlay spectrum access technique. Although the assumption of two power levels was made for analytical tractability, in most practical scenarios there would be more than two levels. In that case, the corresponding Markov chain would have more states vertically. The solution of such a Markov chain can be obtained

numerically, as we showed in this report (Eq.(107)). Another way of dealing with this problem would be by lumping the $\lfloor \frac{M}{2} \rfloor$ levels (if there are $M$ possible power levels) with lower power into a single level. The data rate in this new level would be the weighted average of the corresponding data rates, while the average time spent in this level would be the sum of the average times of all the levels that get lumped into this new level. The same approach can be followed for the remaining $M - \lfloor \frac{M}{2} \rfloor$ higher power levels. In this way we would obtain the Markov chain with 2 levels (Fig. 8), with other transition rates, whose solution is, as shown, in closed-form.

To illustrate the above mentioned approach, we consider the scenario with $M = 4$ power levels, where the data rates for each one of them are shown in Table 4. In Table 4 we also give the average time a SU has a given data rate. These are exponentially distributed. Hence, the rate of leaving a level is $\eta_i = \frac{1}{E[T_i]}$. We sort the levels according to the data rates (from the highest to the lowest) a SU experiences in each one of them. Then we place these levels into 2 groups: in the first group there would be the 2 "best" levels (1,3), and in the other one the 2 "worst" levels (2,4). For each of these groups we calculate the average data rate, and the average time spent there. So, for group 1 we have:

$$E[R_{g1}] = \frac{1}{3} \cdot 2.5 + \frac{2}{3} \cdot 2 = 2.17 \text{ Mbps},$$

$$E[T_{g1}] = 10 + 5 = 15 \text{ s}.$$

Similarly, for the second group we get

$$E[R_{g2}] = 1.19 \text{ Mbps},$$

$$E[T_{g2}] = 8 \text{ s}.$$

In the next step we use the Markov chain of Fig. 8 to capture approximately the scenario with multiple levels. We keep the same average size of the files as before, i.e., $\Delta = 125$ KB. For the values considered here, we have the following transition rates for the new levels: $\mu_H = \frac{E[R_{g1}]}{\Delta} = 2.17\ s^{-1}$, $\eta_H = \frac{1}{E[T_{g1}]} = 0.07\ s^{-1}$, $\mu_L = \frac{E[R_{g2}]}{\Delta} = 1.19\ s^{-1}$, and $\eta_L = \frac{1}{E[T_{g2}]} = 0.125\ s^{-1}$. Fig. 19 shows the *actual* average system time for this scenario vs. the result obtained from Eq.(79), which is an approximation.

As it can be observed from Fig. 19, there is a good match between simulations and the theoretical approximation with the mismatch not exceeding 7-8%. Hence, even if we have multiple power levels, we can use our result. We just need to lump all the best levels in one new level, and all the worst levels in another level, and then use the result for 2 levels.

For a higher accuracy of the multi-power level, we need to use the M-level Markov chain model of Section 3.3. We consider a scenario with 4 different power levels, leading to 4 different data rates: $c_1 = 2$ Mbps, $c_2 = 1$ Mbps, $c_3 = 1.5$ Mbps, and $c_4 = 10$ Mbps. The corresponding average times an SU sees

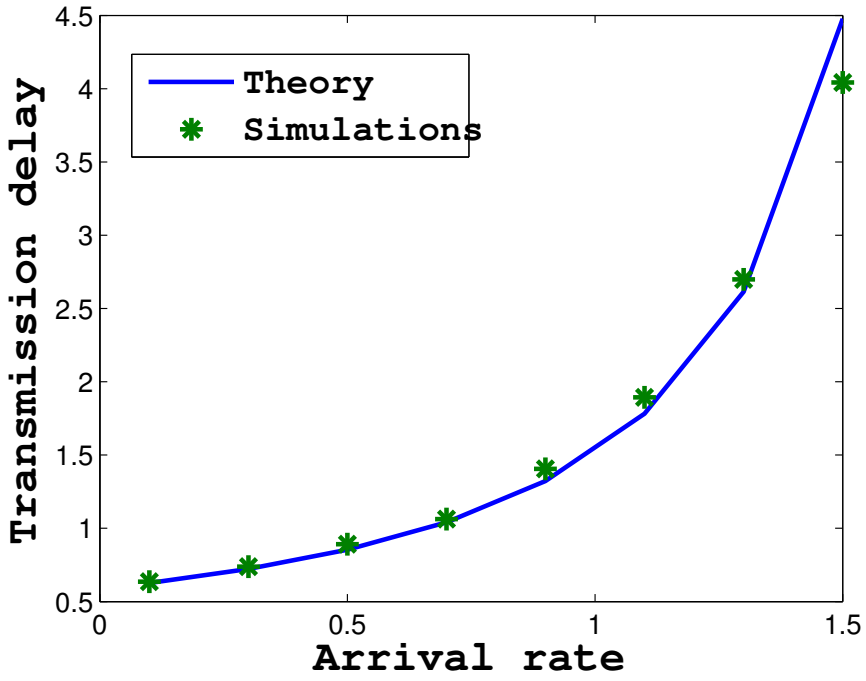


Figure 19: The average delay for multiple power levels.

a given rate are: 10, 3, 5 and 5 s, respectively. Files are exponentially distributed with average size of 125 kB, and the arrival process is Poisson. The probability of moving to any specific level is 1/3. Fig. 20 shows the average file delay for this system. As we can see from Fig. 20, our theoretical result of Eq.(107) matches the simulated result.

Table 4: The data rate and average spending time for each level.

| Level | Data rate (Mbps) | Average time spent in the level (s) |
|---|---|---|
| 1 | $c_1 = 2$ | $E[T_1] = 10$ |
| 2 | $c_2 = 1$ | $E[T_2] = 5$ |
| 3 | $c_3 = 2.5$ | $E[T_3] = 5$ |
| 4 | $c_4 = 1.5$ | $E[T_4] = 3$ |

## 6.2 Validation of the throughput models

See the journal submission for this section.

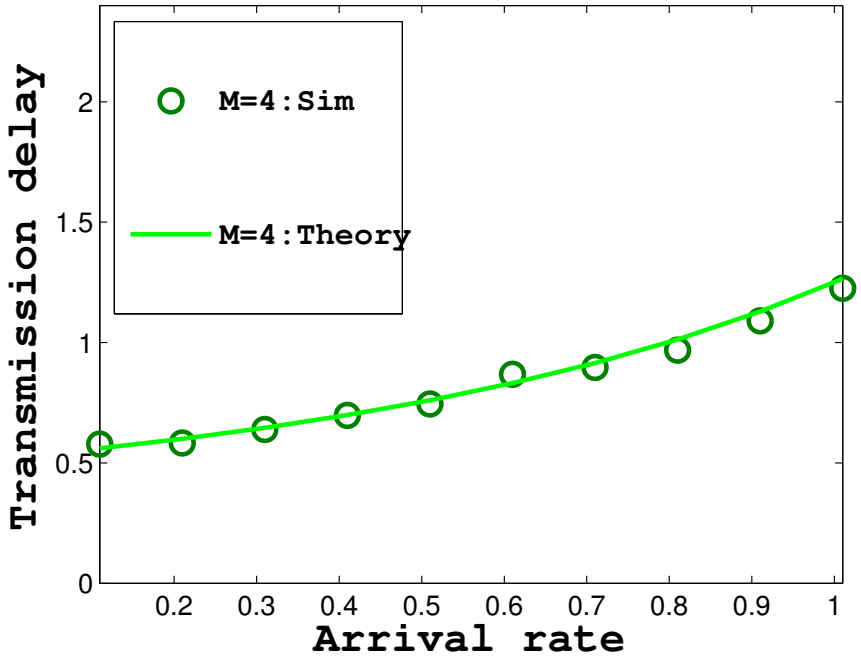


Figure 20: M-level model validation.

### 6.3 Delay minimization policies

In this section, we would like to perform a more detailed comparison of the underlay and interweave access modes. Our first goal is to examine the "static" version of the two policies, and validate our analytical predictions regarding when the one or the other will perform better. Our second goal is to consider the dynamic policy, and see if and when it can outperform both simple policies.

As another interesting scenario, we consider the underlay access with parameters: $\eta_H = 0.1\ s^{-1}$, $\eta_L = 1\ s^{-1}$, $c_H = 10$ Mbps, $c_L = 0.5$ Mbps. Fig. 18 shows the average file delay (denoted by I) against different average scanning times (exp. distributed), for three different traffic intensities (low, medium, high). On the same plot, for each traffic intensity the corresponding underlay delay (denoted by U) is shown as well. Finally, the theoretical maximum values for the expected scanning times (Table 2), for which the interweave access mode outperforms underlay access are depicted with small circles. For the sparse traffic case, the interweave starts to become better for scanning times lower than 0.8 s. The first thing to observe is that the predicted maximum value for the scanning time (i.e. the crossing point) is correct. The second important outcome is that this boundary is higher when the load increases. This is due to the fact that for higher loads the queueing delay is the largest delay component. Hence, it is worth waiting some time, find an idle channel and then get rid of the queued data at higher rate. We have also noticed that

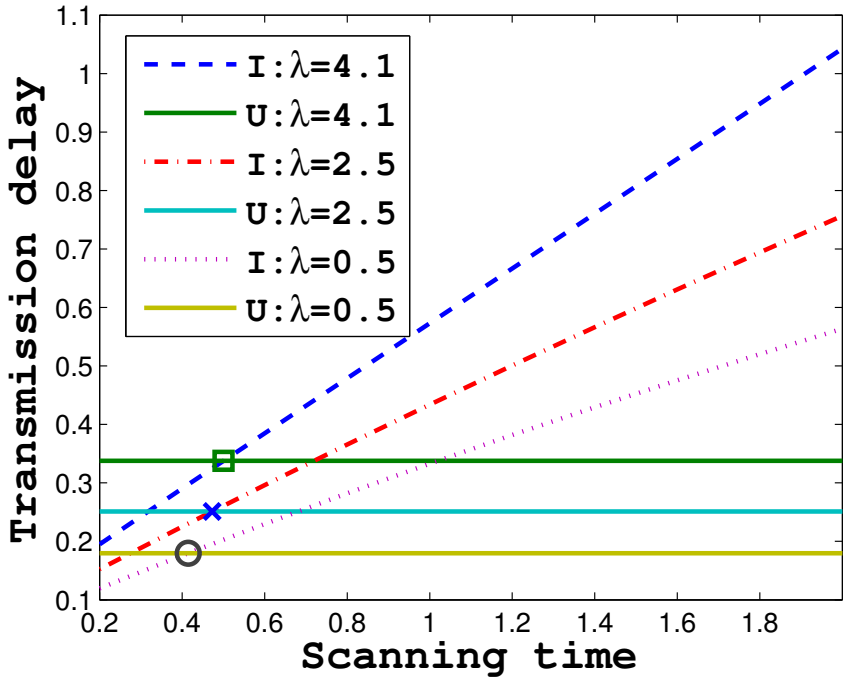


Figure 21: The static delay policy for different $\lambda$ and hyperexp. scan. times.

increasing the load further leads to smaller and smaller increases of this crossing point.

Next, we consider the hyperexponential distribution for the scanning times with parameters $\eta_L = 6\ s^{-1}$, $\eta_V = 0.4\ s^{-1}$, and the probability $p_0$ taking values such that a given average scanning time is achieved. The observed coefficient of variation observed is in the range (2,2.5). The other parameters are identical as for the previous scenario. Fig. 21 shows the average delay. Due to the higher variance of the scanning time, the crossing point between underlay and interweave are lower compared to the scenario of Fig. 18.

**Dynamic delay policy.** As discussed in Section 3.7, the dynamic policy can offer additional performance benefits, when the PU activity periods (i.e. the OFF periods in the underlay mode) are subject to a probability distribution with a decreasing failure rate (i.e. with very high variability). We consider a scenario where the low (OFF) periods have Bounded Pareto distribution, with parameters $L = 0.2$, $H = 100$, $\alpha = 1.2$. The average scanning time is $E[T_s] = 1$s. The other parameters are the same as for the cellular scenario. Fig. 22 shows the average delay vs. the arrival rate. According to the static policy, the underlay mode is better than the interweave. However, the best result is achieved with the dynamic policy, which offers an additional delay reduction of 20-50%.

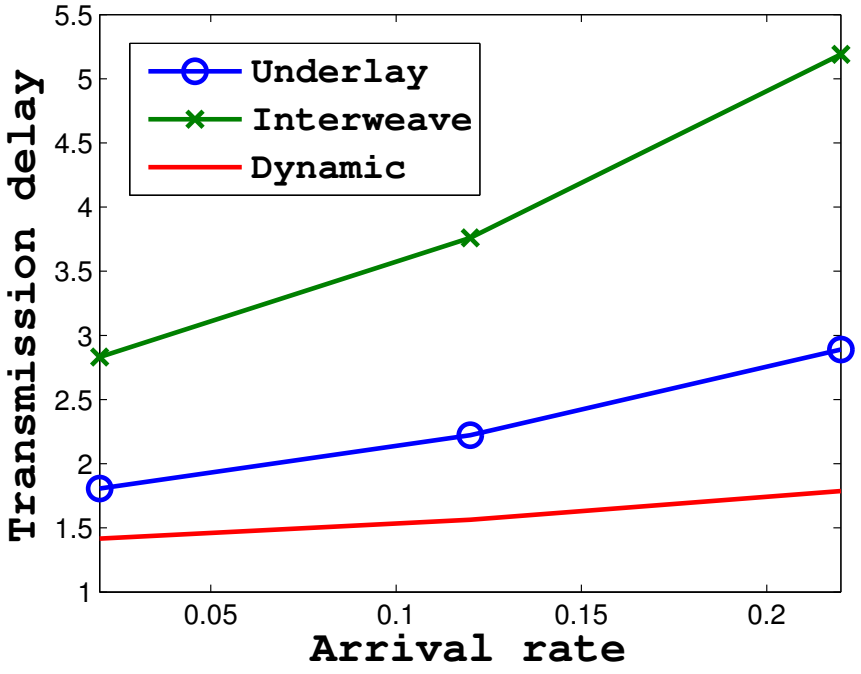


Figure 22: The dynamic delay policy for Pareto OFF periods.

## 7 Related Work

Much fewer studies exist about the per file/flow delay in such systems [22–24]. In [25], authors propose an M/G/1 queueing system with finite buffer and timeout, and derive different metrics, among which the delay as well. However, their results can be obtained only numerically and as such are difficult to be interpreted and used in solving different optimization problems. A similar conclusion can be drawn for [26]. The authors there also model the SU activity with an M/G/1 queue. However, they do not show how to find the second moment of the service time in the P-K formula. On the other hand, we propose an analytical queueing model that leads to a closed-form expression for delay, which not only provides more insight into the effect of system parameters, but also allows to analytically compare and optimize the policies.

As far as interweave CRNs are concerned, there exist more analytical works that aim to derive the average packet delay. Most of the works model the PU activity with stochastic ON-OFF process. Some recent work [22,23] have capitalized on the measurement-based study of [27], in which the Poisson approximation seems to be decent for call arrivals, but call duration is generically distributed. These works model SUs together with PUs, as an M/G/1 system with priorities and preemption. Nevertheless, there are some important caveats in the above models. *First*, they consider the problem in the packet level and model the problem as an M/G/1 sys-

tem with preemptive-resume, while in reality SU packets will collide with a PU when it reclaims channel back, and have to retransmit in the next available period. On the other hand, our model can capture both the resume and retransmitting feature of real wireless systems. *Second*, the M/G/1 systems with priorities can capture only exponential scanning times. Our interweave model holds for generic scanning times. We do not need the Poisson assumption for the PU traffic, as opposed to the priority models, since in our model the time between two PU arrivals is the sum of an exp. (ON period) and a generic (scanning time) random variable, which is generic.

Very few studies exist that directly compare the performance between the access modes. Comparing results from different papers is not straightforward due to different assumptions, non-closed form expressions, etc. In this paper, we propose models that enable us to do analytical comparisons between the modes. A study closer to ours in its aim is [28], where a hybrid CR system is investigated, in which a SU probabilistically changes its mode of operation for throughput optimization. However, the delay metric is not considered there, and the arrival process at the PU is quite restrictive (Bernoulli). On the other hand, we propose policies that are able to optimize the delay, and our models hold for generic PU arrivals.

Summarizing, the main novelties of this paper compared to different related works revolve around the following key points: (i) we make a direct analytical comparison of interweave and underlay access delays; (ii) we provide closed form expressions for all cases; (iv) we use our results to propose an optimal hybrid policy.

# 8 Conclusion

In this paper, we have proposed queueing analytic models for the delay analysis of interweave and underlay spectrum access, and we validated them against realistic scenarios. Besides that, we have proposed models relying on renewal-reward theory to determine the average throughput in both modes. We have also provided the bounds on the scanning times for which the interweave access outperforms the underlay. We have also proposed a dynamic policy to further improve the performance (up to additional 50%), and we have validated our results extensively for realistic scenarios. In future work, we plan to extend our model to capture generic file sizes.

# References

[1] http://share.cisco.com/internet-of-things.html.

[2] B. Wang and K. Liu, "Advances in cognitive radio networks: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, 2011.

[3] I. F. Akiyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Comm. Mag.*, vol. 46, no. 4, apr. 2008.

[4] N. Mahmood, F. Yilmaz, G. Oien, and M. Alouini, "On hybrid cooperation in underlay cognitive radio networks," *IEEE Tran. Wireless Comunications*, vol. 12, no. 9, 2013.

[5] H. Kim and K. Shin, "Fast discovery of spectrum opportunities in cognitive radio networks," in *Proc. of IEEE DySPAN*, 2008.

[6] Z. Yang, X. Xie, and Y. Zheng, "A new two-user cognitive radio channel model and its capacity analysis," in *Proc. of ISCIT*, 2009.

[7] S. M. Ross, *Stochastic Processes*. John Wiley & Sons, 1996.

[8] M. F. Neuts, *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach*. John Hopkins University Press, 1981.

[9] F. Mehmeti and T. Spyropoulos, "To scan or not to scan: The effect of channel heterogeneity on optimal scanning policies," in *Proc. of IEEE SECON*, 2013.

[10] D. Gozupek, S. Buhari, and F. Alagoze, "A spectrum switching delay-aware scheduling algorithm for centralized cognitive radio networks," *IEEE Tran. Mobile Computing*, vol. 12, no. 7, 2013.

[11] L. Kleinrock, *Queueing theory, Volume I: Theory*. John Wiley & Sons, 1975.

[12] U. Yechiali and P. Naor, "Queueing problems with heterogeneous arrivals and service," *Operations Research*, 1971.

[13] F. Mehmeti and T. Spyropoulos, "Performance analysis of mobile data offloading in heterogeneous networks," *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, 2017.

[14] I. Mitrany and B.Avi-Itzhak, "A many-server queue with service interruptions," *Operations Research*, vol. 16, no. 3, 1968.

[15] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio networks," in *Proc. of IEEE DySPAN*, 2005, pp. 137–143.

[16] W. Afifi and M. Krunz, "TSRA: An adaptive mechanism for switching between communication modes in full-duplex opportunistic spectrum access systems," *IEEE Trans. Mob. Comput.*, vol. 16, no. 6, 2017.

[17] H. B. Salameh and M. Krunz, "Channel access protocols for multihop opportunistic networks: Challenges and recent developments," *IEEE Network*, vol. 23, no. 4, 2009.

[18] M. Abdel-Rahman, H. Rahbari, and M. Krunz, "Multicast rendez-vous in fast-varying DSA networks," *IEEE Trans. Mob. Comput.*, vol. 14, no. 7, 2015.

[19] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access," *IEEE Comm. Mag.*, vol. 47, no. 3, mar 2009.

[20] S. Geirhofer and L. Tong, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Comm. Mag.*, vol. 45, 2007.

[21] R. Research, "Beyond LTE: Enabling the mobile broadband explosion," 2014.

[22] I. Suliman and J. Lehtomaki, "Queueing analysis of opportunistic access in cognitive radios," in *Proc. of CogART*, 2009.

[23] T. Q. D. T. Hung and H.-J. Zepernick, "Average waiting time of packets with different priorities in cognitive radio networks," in *Proc. of IEEE ISWPC*, 2009.

[24] L. Wang, C. Wang, and C. Chang, "Modeling and analysis for spectrum handoffs in cognitive radio networks," *IEEE Tran. Mob. Computing*, vol. 11, no. 9, sep. 2012.

[25] T. Chu, H. Phan, and H. Zepernick, "On the performance of underlay cognitive radio networks using M/G/1/K queueing model," *IEEE Comm. Letter*, vol. 17, no. 5, 2013.

[26] L. Sibomana, H. Zepernick, H. Tran, and C. Kabiri, "Packet transmission time for cognitive radio networks considering interference from primary user," in *Proc. of IWCMC*, 2013.

[27] L. T. S. Geirhofer and B. Sadler, "Cognitive medium access: Constraining interference based on experimental models," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, Jan. 2008.

[28] H. Song, J. P. Hong, and W. Choi, "On the optimal switching probability for a hybrid cognitive radio system," *IEEE Tran. Wireless Comm.*, vol. 12, no. 4, 2013.