

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/308991243>

Stay or Switch? Analysis and Comparison of Delays in Cognitive Radio Networks with Interweave and Underlay Spectrum Access

Conference Paper · November 2016

DOI: 10.1145/2989250.2989265

CITATIONS

0

READS

19

2 authors:



[Fidan Mehmeti](#)

North Carolina State University

12 PUBLICATIONS 46 CITATIONS

SEE PROFILE



[Thrasyvoulos Spyropoulos](#)

Institut Mines-Télécom

91 PUBLICATIONS 4,631 CITATIONS

SEE PROFILE

Stay or Switch? Analysis and Comparison of Delays in Cognitive Radio Networks with Interweave and Underlay Spectrum Access

Fidan Mehmeti
Department of Electrical and Computer
Engineering
University of Waterloo, Canada
fidan.mehmeti@uwaterloo.ca

Thrasylvoulos Spyropoulos
Department of Mobile Communications
Institute Eurecom, Sophia Antipolis, France
spyropou@eurecom.fr

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 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. We focus on the average delay as the key metric, which we analyze using queueing theory. This allows an SU to decide when one type of access technique provides better performance, as a function of the metric of interest and key network parameters. What is more, based on this analysis, we propose a dynamic (hybrid) policy, that can decide at any point to switch from one type of access to the other, offering up to 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.

Keywords

Cognitive radio networks, Markov chains, Interweave, Underlay.

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

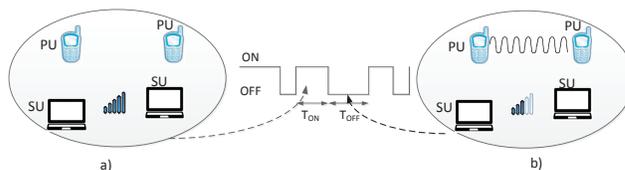


Figure 1: The illustration of the underlay mode with: a) idle PU (high periods), b) active PU (low periods).

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 [24].

To address this issue, dynamic spectrum access techniques have recently been proposed, with cognitive radio (CR) as its key technology [2]. In a cognitive network, there exist licensed users, known as primary users (PU), which are assigned 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.

One of the main functions of CRs is spectrum access [2]. 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 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 [9, 12]. While some arguments for the one or the other exist (often related to the potential harm to PUs [28]), 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 advan-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MobiWac'16, November 13-17, 2016, Malta, Malta

© 2016 ACM. ISBN 978-1-4503-4503-3/16/11...\$15.00

DOI: <http://dx.doi.org/10.1145/2989250.2989265>

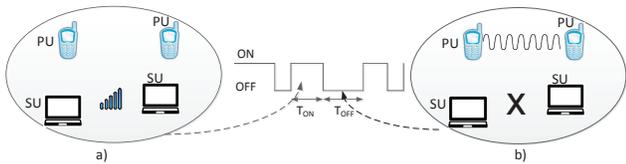


Figure 2: The illustration of the interweave mode with: a) idle PU (high periods), b) active PU (low periods).

tage 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 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 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 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. We focus on the delay, which we analyze using a queueing theoretic framework. 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 (Section 3);
- (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 (Section 4).

2. PERFORMANCE MODELING OF SPECTRUM ACCESS

Our figure of merit is the *delay*. We assume that traffic flows arrive randomly as a Poisson process with rate λ . 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 also use 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 [17] (T_{ON}^i, T_{OFF}^i) , $i \geq 1$, as shown in Fig. 1¹. 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 η_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 η_L . This assumption is necessary for the tractability of the delay analysis. Generic ON/OFF distributions could be introduced also in our delay analysis by considering phase-type distributions and matrix analytic methods [15]. However, such methods lead only to numerical solutions, that do not allow for direct analytical performance comparisons. What is more, simulation results (Section 4) suggest that, even for generic ON/OFF period distributions, the accuracy of our predictions is sufficient for both modes.

The data transmission rate during the ON periods is denoted with c_H , while during OFF periods the data 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 ON 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 η_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., $c_L = 0$.² Hence, it changes its operation to the *scanning* mode. During the scanning mode (i.e., during an

¹This approach is more generic than e.g., modeling the PU activity as a Poisson process. For a detailed discussion on this, see Section 5.

²We 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 μs) and do energy detection, to see if there is a PU signal [13]. Since we assume that the sensing time is much shorter

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 found³, 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}) introduced⁴, and the total scanning time can be written as

$$T_s = L(T_I + T_{switch}), \quad (1)$$

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/channel is much shorter than the ON and OFF periods.⁵

The switching delay while moving from the channel with frequency f_s to the channel with frequency f_d can be expressed as [6]

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

where β is the delay to move to the first contiguous channel, and is hardware dependent [6], while δ 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 with an exponential distribution. Hence, 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 scenarios when the scanning 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

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.

³We assume that the SU chooses channels sequentially from a list [6].

⁴The "scanning time" in our model is a random variable that can capture a number of processes (including the time to rendez-vous), and since we assume it is generic it can include any distribution (sum of switching, scanning and rendez-vous). However, the actual protocol to ensure this rendez-vous is beyond the scope of our paper, and any of the existing schemes can be assumed, without affecting the results.

⁵In scenarios with static PU activity (TV stations), geolocation databases can replace spectrum sensing. However, we consider here PUs with dynamic activity.

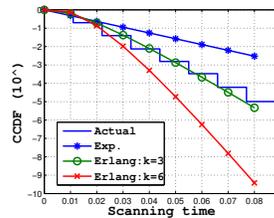


Figure 3: The distribution of the scan-

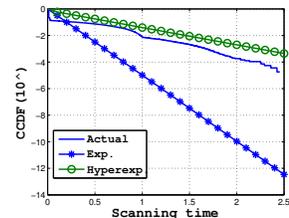


Figure 4: The distribution of the scanning time for two groups of channels.

such low variance distributions is by using a k -stage Erlang distribution [10].

As opposed to the previous scenario, there might be a number of channels with high duty cycles, and there might be 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 the scanning time will be exponentially distributed with parameter η_L , and with a small probability $1 - p$ it will be exponentially distributed with parameter η_V . Note that $\eta_V \ll \eta_L$.

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 scanning time durations. On the same plot, the CCDFs of 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 0.5 s. The hyperexponential distribution (shown also in the plot) has the following parameters: $p = 0.2$, $\lambda_1 = 90$, $\lambda_2 = 3$. 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 into our analysis. Surprisingly, as we shall see, closed-form results for the system delay can still be found for such generic scanning times.

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 hyperexponential. For that purpose

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
λ	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
η_V	The rate of leaving the V-state
Δ	The average file size
$\mu_H = \frac{c_H}{\Delta}$	The service (transition) rate while in a high state
$\mu_L = \frac{c_L}{\Delta}$	The service (transition) rate while in a low state
$E[T]$	The average system (transmission) time

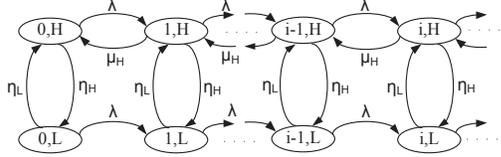


Figure 5: The 2D Markov chain for exponentially distributed scanning time.

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 that ON/OFF periods are exponentially distributed, we can model our system with a 2D Markov chain, as 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 η_H and η_L are the parameters of the exponentially distributed ON and OFF periods. While in the upper parts of the 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 SU inability to transmit while scanning. The transition rate from low to high periods is η_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*

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

PROOF. The balance equations for this chain are

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

$$\pi_{i,L}(\lambda + \eta_L) = \lambda \pi_{i-1,L} + \eta_H \pi_{i,H}, \quad i \geq 1 \quad (5)$$

$$\pi_{0,H}(\eta_H + \lambda) = \eta_L \pi_{0,L} + \mu_H \pi_{1,H} \quad (6)$$

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

We define the PGFs for this chain as

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

Multiplying Eq.(5) with z^i and adding it to Eq.(4), we obtain

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

that leads to

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

Similarly, multiplying Eq.(7) with z^i and summing with Eq.(6), we get

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

Eq.(10) results in

$$\begin{aligned} & [\lambda z(1-z) + \mu_H(z-1) + \eta_H z] G_H(z) - \\ & \eta_L z G_L(z) = \mu_H \pi_{0,H}(z-1). \end{aligned} \quad (11)$$

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

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

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

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

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

After summing over all the values of i , we have

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

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

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

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

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

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

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

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.,

$$E[N] = E[N_L] + E[N_H] = G'_L(1) + G'_H(1). \quad (19)$$

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

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

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

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

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

Based on Eq.(21), we have

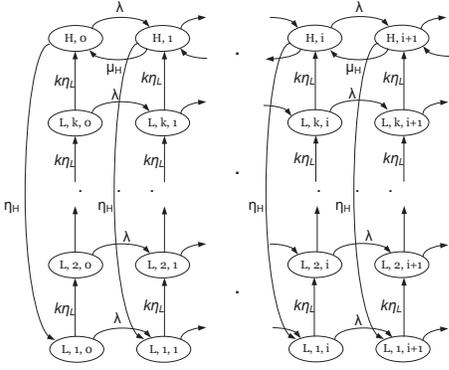


Figure 6: The 2D Markov chain for Erlang-distributed scanning time.

$$E[N_L] = \lim_{z \rightarrow 1} G'_L(z) = \frac{-\mu_H \pi_{0,H} F''(1)}{2F'(1)^2}. \quad (22)$$

After some algebra, we obtain

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

The next step is to find $E[N_H]$ (the average number of files while being in a high state). For that purpose, Eq.(13) is differentiated, giving

$$G'_H(z) = \frac{1}{\eta_H} \left\{ -\lambda G_L(z) + [\lambda(1-z) + \eta_L] G'_L(z) \right\}. \quad (24)$$

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

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

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

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

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 $E[N] = \lambda E[T]$ [17], we obtain the average file delay in the interweave access as in Eq.(3). \square

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 [15]. However, numerical methods do not provide any insight on the nature of the solution and its dependency on certain parameters.

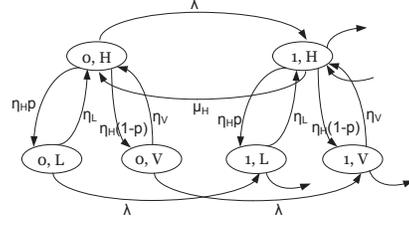


Figure 7: The 2D Markov chain for hyperexponentially distributed scanning time.

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

Due to space limitations, we omit the derivation details here. The interested reader can find them in [14]. 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*

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

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 of 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 a 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 $E[T_s] = \frac{p}{\eta_L} + \frac{1-p}{\eta_V}$. In order to maintain the same $E[T_s]$ as before, but with much higher variability, we choose a very low value for $1 - p$ (e.g., lower than 0.05) and $\eta_V \ll \eta_L$.

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. The interested reader can find the detailed proof in [14].

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

$$E[T_{hyp}] = \frac{(\eta_H E[T_s])^2 + \eta_H \mu_H \left(\frac{p}{\eta_L} + \frac{1-p}{\eta_V} \right) + 2\eta_H E[T_s] + 1}{(1 + \eta_H E[T_s])(\mu_H - \lambda - \lambda \eta_H E[T_s])}. \quad (28)$$

It can be easily proven [14] that $\frac{p}{\eta_L} + \frac{1-p}{\eta_V} = \frac{1}{\eta_V} + \left(\frac{1}{\eta_L} + \frac{1}{\eta_V} \right) \left(E[T_s] - \frac{1}{\eta_V} \right) > \frac{1}{\eta_V} + E[T_s] \left(E[T_s] - \frac{1}{\eta_V} \right) = (E[T_s])^2 + \frac{1}{\eta_V} \left(\frac{1}{\eta_V} - E[T_s] \right) > (E[T_s])^2$. This leads to $E[T_{hyp}] > E[T_{exp}]$, which is expected from queueing systems.

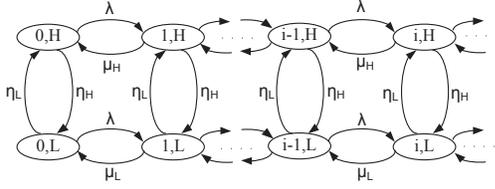


Figure 8: The 2D Markov chain for the underlay model.

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*

$$E[T_u] = \frac{\eta_H + \eta_L + \mu_H(1 - \pi_{0,H}) + \mu_L(1 - \pi_{0,L}) - \lambda + \frac{\mu_L \mu_H}{\lambda}(\pi_{0,L} + \pi_{0,H} - 1)}{\mu_H \eta_L + \mu_L \eta_H - \lambda(\eta_H + \eta_L)} \quad (29)$$

PROOF. The balance equations for this chain are

$$\pi_{0,L}(\lambda + \eta_L) = \pi_{1,L}\mu_L + \pi_{0,H}\eta_H \quad (30)$$

$$\pi_{0,H}(\lambda + \eta_H) = \pi_{1,H}\mu_H + \pi_{0,L}\eta_L \quad (31)$$

$$\pi_{i,L}(\lambda + \eta_L + \mu_L) = \pi_{i-1,L}\lambda + \pi_{i+1,L}\mu_L + \pi_{i,H}\eta_H, (i > 0) \quad (32)$$

$$\pi_{i,H}(\lambda + \eta_H + \mu_H) = \pi_{i-1,H}\lambda + \pi_{i+1,H}\mu_H + \pi_{i,L}\eta_L, (i > 0) \quad (33)$$

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.(33) by z^i and rearranging (in the same direction as we did for interweave CRN), we obtain

$$(\lambda + \eta_L + \mu_L)G_L(z) = \lambda z G_L(z) + \eta_H G_H(z) + \frac{\mu_L}{z}(G_L(z) - \pi_{0,L}) + \pi_{0,L}\mu_L, \quad (34)$$

$$\text{and } (\lambda + \eta_H + \mu_H)G_H(z) = \lambda z G_H(z) + \eta_L G_L(z) + \frac{\mu_H}{z}(G_H(z) - \pi_{0,H}) + \pi_{0,H}\mu_H. \quad (35)$$

Solving the system of equations Eq.(34)-(35) gives

$$f(z)G_L(z) = \pi_{0,H}\eta_H\mu_H z + \pi_{0,L}\mu_L[\eta_H z + (\lambda - z\mu_H)(1 - z)], \quad (36)$$

where

$$f(z) = \lambda^2 z^3 - \lambda(\eta_L + \eta_H + \lambda + \mu_H + \mu_L)z^2 + (\eta_L\mu_H + \eta_H\mu_L + \mu_L\mu_H + \lambda\mu_H + \lambda\mu_L)z - \mu_L\mu_H. \quad (37)$$

It can be proven that the polynomial in Eq.(37) has only one root in the open interval $(0, 1)$ [29]. This root is denoted as z_0 . We omit this proof here due to space limitations. Setting $z = z_0$ into Eq.(36), and after some algebra we get $\pi_{0,L}$ and $\pi_{0,H}$, as

$$\pi_{0,L} = \frac{\eta_H \left(\frac{\eta_H \mu_L + \eta_L \mu_H}{\eta_H + \eta_L} - \lambda \right) z_0}{\mu_L(1 - z_0)(\mu_H - \lambda z_0)}, \quad (38)$$

$$\pi_{0,H} = \frac{\eta_L \left(\frac{\eta_H \mu_L + \eta_L \mu_H}{\eta_H + \eta_L} - \lambda \right) z_0}{\mu_H(1 - z_0)(\mu_L - \lambda z_0)}. \quad (39)$$

Finally, using Little's law $E[N] = \lambda E[T]$ [17], we obtain Eq.(29). \square

While 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 horizontally. The solution of such a Markov chain can be obtained numerically [14]. Nevertheless, we can lump the $\lfloor \frac{M}{2} \rfloor$ levels (if there are M possible power levels) with lower power into a single level, whose data rate would be the weighted average of the corresponding data rates, and the transition rate would be the average transition rate of these levels. The same approach can be followed for the $\lfloor \frac{M}{2} \rfloor$ higher power levels. This way we would obtain the Markov chain of Fig. 8, with other transition rates [14], whose solution is, as shown, in closed-form.

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} [4]. 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 a 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). The 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 constant⁶. Each sensing instant is a Bernoulli trial with 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

$$E[T_{ON}] = \frac{1}{p} E[T_{is}]. \quad (40)$$

In case we consider both miss-detections and false alarms, 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

$$p' = p_{fa} + (1 - p_{md}) \cdot p \cdot (1 - p_{fa}). \quad (41)$$

The actual average duration of an ON period when we consider these imperfections is $E[T_{ON,im}] = \frac{1}{p'} E[T_{is}] = \frac{p}{p'} E[T_{ON}]$, leading to

$$E[T_{ON,im}] = \frac{p}{p_{fa} + (1 - p_{md}) \cdot p \cdot (1 - p_{fa})} \cdot E[T_{ON}], \quad (42)$$

and for the transition rate out of an ON period

$$\eta'_H = \frac{1}{E[T_{ON,im}]} = \frac{p_{fa} + (1 - p_{md}) \cdot p \cdot (1 - p_{fa})}{p} \cdot \eta_H. \quad (43)$$

Following a similar reasoning for low periods, we get for the transition rate out of a low state [14]

$$\eta'_L = \frac{p_{md} + (1 - p_{fa}) \cdot q \cdot (1 - p_{md})}{q} \cdot \eta_L, \quad (44)$$

where $q = \frac{E[T_{ON}]}{E[T_{ON}] + E[T_{OFF}]}$ is the probability of no PU in the channel.

The other transition rate parameters of Fig. 8 remain unchanged. Although the result expressed through Eq.(29) will not be the same,

⁶In case the last performed sensing during a given ON period occurred before (and *not* at the arrival instant of PU) the PU reclaimed back its channel, then we can capture that PU arrival instant with the probability of miss-detection.

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

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

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.

The miss-detections might affect the PU performance. In that case the PU will have higher interference. We can easily calculate the expected percentage of a PU busy period that would be interfered by an SU due to miss-detection. Nevertheless, our main concern in this paper is the impact on the SU. Analyzing the impact of imperfections on the PU is beyond the scope of this paper. Finally, we assume that the SUs are coordinated among themselves, and focus on the interaction between a PU and an SU. The problem of SU impact, if not coordinated (leading to collisions), is interesting, but beyond the scope of this paper.

3.3 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 the 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 parameters affecting the performance in each case. In Section 3.4, 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. For example, we can observe that $B_1 > A_1$, $B_2 > A_2$ [14], and that for interweave to outperform underlay access for exponential scanning time, a smaller scanning time is needed. Similar conclusions can be drawn by comparing the parameters of hyperexponential distribution with the two previous ones. We can observe that the scanning time variability has an important impact on the boundary scanning time. The higher the variability of the scanning time is, the lower scanning time is needed for the interweave mode to outperform the underlay access.

3.4 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 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 then it 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 at 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)*⁷,

⁷The distributions with increasing (decreasing) failure rate are those for which $\frac{f(x)}{1-F(x)}$ is an increasing (decreasing) function of x .

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 dynamic policy to outperform any of the static policies, as we show later, by essentially “pruning” the long OFF periods from the underlay mode.

The following result provides an analytical expression for the case of Pareto distributed OFF periods (with parameters L, α), a popular distribution with decreasing failure rate [14], and for exponential ON periods.

RESULT 5. *The optimal turning point, t_{opt} , in a Cognitive Radio Network can be found as a solution of*

$$\begin{aligned} & \frac{\eta_H}{\eta_H + \eta_L} \cdot \frac{\mu_H - \mu_L}{\mu_H} \left(\left(1 - \frac{1}{\alpha}\right) t^{1+\alpha} + \frac{1}{\alpha} L^{\alpha-1} t^2 \right) \\ & + \frac{\eta_H}{\eta_H + \frac{\mu_H}{\Delta}} \cdot \frac{\eta_L}{\eta_H + \eta_L} \left(\frac{\mu_H - \mu_L}{\mu_H} L^\alpha t - E[T_s] \alpha L^\alpha \right) = 0. \end{aligned} \quad (45)$$

The solution to Eq.(45) can be found numerically. Table 3 summarizes all the possible scenarios.

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 changes in the behavior of the system and estimated statistics.

4. SIMULATION RESULTS

The first goal of this section is to validate the various analytical expressions we have derived, against simulated scenarios, including several cases when one or more of the theoretical 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 [27] 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 [5], 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 exponentially distributed periods. We consider other distributions later in Section 4.1. 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.⁸ Finally, we assume

⁸These values are taken to be of the same order of magnitude as the actual values encountered in practice [16]. 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, chan-

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 $t_{opt} \in (0, \infty)$

nel 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.

4.1 Validation of the delay models

Fig. 9 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 both scenarios (cellular and WiFi). As can be seen, our theoretical results match with the results obtained from simulations. As is expected in queueing systems, the delay increases when arrival rate (and thus the utilization of the system) 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. 10 shows the theoretical vs. simulated results for the cellular network scenario for three types of scanning time distributions (exponential, 4-stage Erlang, hyperexp.), 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 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 suggests, higher (lower) variability in scanning time leads to higher (lower) variability in the service time, which further leads to higher (lower) delays. Observing the curve corresponding to the cellular case in Fig. 9 and Fig. 10, 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 an arrival rate of $\lambda = 1$ 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 [5,27] 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. 9 and 10. Fig-

nel 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.

⁹This value is normalized for the arrival rates considered, to correspond approximately to the traffic intensities reported in [27] and [5]. We have also considered other values with similar conclusions drawn.

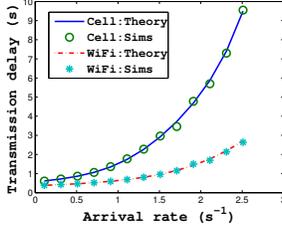


Figure 9: The delay for underlay spectrum access.

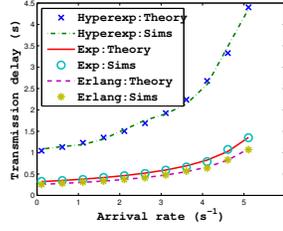


Figure 10: The delay for interweave spectrum access in a cellular scenario.

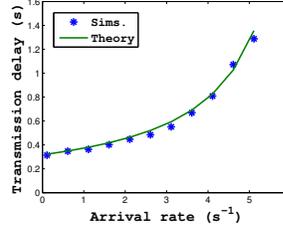


Figure 11: The delay for generic interweave spectrum access.

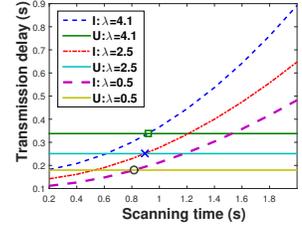


Figure 12: The static delay policy for different λ and exp. scanning times.

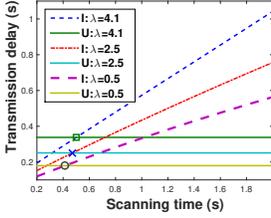


Figure 13: The static delay policy for different λ and hyperexp. scan. times.

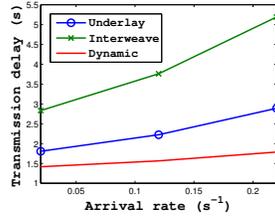


Figure 14: The dynamic delay policy for Pareto OFF periods.

Figure 11 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 high 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.

4.2 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 when one or the other performs 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: $\eta_H = 0.1 \text{ s}^{-1}$, $\eta_L = 1 \text{ s}^{-1}$, $c_H = 10 \text{ Mbps}$, $c_L = 0.5 \text{ Mbps}$. Fig. 12 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, too. Finally, the theoretical maximum values for the expected scanning times (Table 2), for which the interweave 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 for some time, find an idle channel and then get rid of the queued data at a higher rate. We have also noticed that increasing the load further leads to a smaller increase of this crossing point.

Next, we consider the hyperexponential distribution for the scanning time with parameters $\eta_L = 6 \text{ s}^{-1}$, $\eta_V = 0.4 \text{ s}^{-1}$, and the probability p taking values such that a given average scanning time is achieved. The observed coefficient of variation is in the range (2,2.5). The other parameters are identical as in the previous sce-

nario. Fig. 13 shows the average delay. Due to the higher variance of the scanning time, the crossing points between underlay and interweave are lower compared to the scenario of Fig. 12.

Dynamic delay policy. As discussed in Section 3.4, 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 a Bounded Pareto distribution, with parameters $L = 0.2$, $H = 100$, $\alpha = 1.2$. The average scanning time is $E[T_s] = 1 \text{ s}$. The other parameters are the same as for the cellular scenario. Fig. 14 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%.

5. RELATED WORK

There has been a significant amount of research in the areas of interweave and underlay access for CRN. The underlay CRN have been the focus of [7, 8, 12, 19, 25], while the interweave techniques were considered in [9, 11, 13, 30]. None of these works consider the delay as the metric of interest.

Much fewer studies exist about the per file/flow delay in such systems [22, 23, 26]. In [3], 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 [20]. 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 [18], 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 system 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 [21], 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.

6. CONCLUSION

In this paper, we have proposed queuing models for the delay analysis of interweave and underlay access, and we have validated them against realistic scenarios. We have provided the bounds on the scanning time 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.

7. REFERENCES

- [1] <http://share.cisco.com/internet-of-things.html>.
- [2] I. F. Akiyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty. A survey on spectrum management in cognitive radio networks. *IEEE Comm. Mag.*, 46(4), apr. 2008.
- [3] 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*, 17(5), 2013.
- [4] G. Ganesan and Y. Li. Cooperative spectrum sensing in cognitive radio networks. In *Proc. of IEEE DySPAN*, pages 137–143, 2005.
- [5] S. Geirhofer and L. Tong. Dynamic spectrum access in the time domain: Modeling and exploiting white space. *IEEE Comm. Mag.*, 45, 2007.
- [6] D. Gozuepek, S. Buhari, and F. Alagoze. A spectrum switching delay-aware scheduling algorithm for centralized cognitive radio networks. *IEEE Tran. Mobile Comput.*, 12(7), 2013.
- [7] H. Hakim, H. Boujemaa, and W. Ajib. Performance comparison between adaptive and fixed transmit power in underlay cognitive radio networks. *IEEE Tran. Comm.*, 61(12), 2013.
- [8] M. Ki, H. Lee, and J. Song. Performance analysis of distributed cooperative spectrum sensing for underlay cognitive radio. In *Proc. of ICACT*, 2009.
- [9] H. Kim and K. Shin. Fast discovery of spectrum opportunities in cognitive radio networks. In *Proc. of IEEE DySPAN*, 2008.
- [10] L. Kleinrock. *Queueing theory, Volume I: Theory*. John Wiley & Sons, 1975.
- [11] Y. Liu and M. Liu. To stay or to switch: Multiuser multi-channel dynamic access. *IEEE Trans. Mobile Comput.*, 14(4), 2015.
- [12] N. Mahmood, F. Yilmaz, G. Oien, and M. Alouini. On hybrid cooperation in underlay cognitive radio networks. *IEEE Tran. Wireless Communications*, 12(9), 2013.
- [13] 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.
- [14] F. Mehmeti and T. Spyropoulos. Underlay vs. interweave: Which one is better? Technical Report RR-14-296, EURECOM, 2014. <http://www.eurecom.fr/en/publication/4443>.
- [15] M. F. Neuts. *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach*. John Hopkins University Press, 1981.
- [16] R. Research. Beyond LTE: Enabling the mobile broadband explosion, 2014.
- [17] S. M. Ross. *Stochastic Processes*. John Wiley & Sons, 1996.
- [18] L. T. S. Geirhofer and B. Sadler. Cognitive medium access: Constraining interference based on experimental models. *IEEE J. Sel. Areas Commun.*, 26(1), Jan. 2008.
- [19] M. Seyfi, S. Muhaidat, and J. Liang. Relay selection in underlay cognitive radio networks. In *Proc. of IEEE WCNC*, 2012.
- [20] 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.
- [21] H. Song, J. P. Hong, and W. Choi. On the optimal switching probability for a hybrid cognitive radio system. *IEEE Tran. Wireless Comm.*, 12(4), 2013.
- [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] B. Wang and K. Liu. Advances in cognitive radio networks: A survey. *IEEE J. Sel. Topics Signal Process.*, 5(1), 2011.
- [25] B. Wang and D. Zhao. Performance analysis in CDMA-based cognitive wireless networks with spectrum underlay. In *Proc. of IEEE GLOBECOM*, 2008.
- [26] L. Wang, C. Wang, and C. Chang. Modeling and analysis for spectrum handoffs in cognitive radio networks. *IEEE Tran. Mobile Comp.*, 11(9), sep. 2012.
- [27] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary user behavior in cellular networks and implications for dynamic spectrum access. *IEEE Comm. Mag.*, 47(3), mar 2009.
- [28] Z. Yang, X. Xie, and Y. Zheng. A new two-user cognitive radio channel model and its capacity analysis. In *Proc. of ISCIT*, 2009.
- [29] U. Yechiali and P. Naor. Queueing problems with heterogeneous arrivals and service. *Operations Research*, 1971.
- [30] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework. *IEEE J. Sel. Areas Commun.*, 25(3), 2007.