Performance Analysis, Comparison, and Optimization of Interweave and Underlay Spectrum Access in Cognitive Radio Networks

Fidan Mehmeti and Thrasyvoulos Spyropoulos, *Member, IEEE*

***Abstract*—Cognitive radio networks (CRNs) have been proposed to exploit licensed bands opportunistically, with secondary users' (SU) activity being subordinated to primary users (PU). The two most popular types of spectrum access in the CRN are *interweave* and *underlay*. There is no clear consensus which one provides better results, for different metrics, despite the fact that there has been a lot of research dedicated to each mode. In this paper, we analyze this problem theoretically, providing formulas that are in closed form. These expressions allow the performance comparison of interweave and underlay modes under a unified network setup. Our focus are two metrics, *throughput* and *delay*, which we analyze relying on the renewal-reward theory and queueing theory, respectively. These results enable an SU to decide what mode to use depending on what the optimization objective is, given the key network parameters. Furthermore, relying on the results of our analysis, we propose hybrid policies, which are dynamic, and allow the SU to switch between the two modes at any point. These policies offer an additional performance improvement of up to 50%. We validate our results with extensive realistic simulations.**

***Index Terms*—Cognitive networks, underlay, interweave.**

I. Introduction

OVER the past years, we have been witnessing a dramatic expansion of different types of ultra modern mobile devices, such as very thin tablets, powerful smartphones, etc., in the everyday use, together with a vast number of services and applications that they offer [2]. This has resulted in an increased spectrum demand, forcing operators to work very close to their capacity limits [3].

Due to this explosion in network traffic, and because of the non-dynamic policies of spectrum assignment followed by telecommunication regulation bodies worldwide, spectrum scarcity has become a serious concern in wireless communications nowadays. Contradictorily, measurements of wireless spectrum utilization reveal a severe underutilization of wireless spectrum bands, with very high variability across time, space, and frequency [4].

Manuscript received January 6, 2018; revised March 3, 2018; accepted April 12, 2018. Date of publication April 18, 2018; date of current version August 13, 2018. This paper was presented in part at the ACM MSWiM 2016 as a poster and has been published in Proceedings of ACM MobiWac 2016 [1]. The review of this paper was coordinated by Dr. Z. Fadlullah. *(Corresponding author: Fidan Mehmeti.)*

F. Mehmeti is with the Penn State University, State College, PA 16801 USA (e-mail: fidan.mehmeti@psu.edu).

T. Spyropoulos is with the Institute EURECOM, Biot 06410, France (e-mail: spyropou@eurecom.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2828090

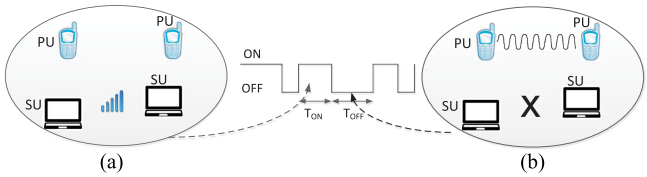

Fig. 1. Interweave mode illustration with: (a) silent PU (high periods), (b) active PU (low periods).

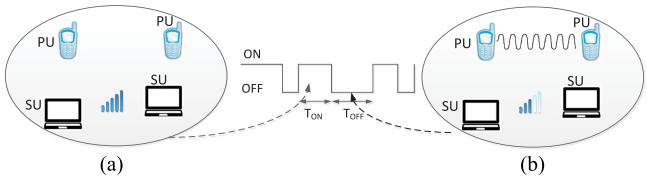

Fig. 2. Underlay mode illustration with: (a) silent PU (high periods), (b) active PU (low periods).

To properly address this issue, dynamic spectrum access techniques, with cognitive radio (CR) as their key enabling technology, have been proposed recently [3]. There exist two types of users in a cognitive radio network: *primary* users (PU), which are licensed (by regulation authority) to access the spectrum, and *cognitive* or *secondary* users (SU) that are unlicensed and utilize the spectrum in an opportunistic fashion. The activity of the latter is subordinated to PUs. So, they must tune their transmission parameters accordingly, in order not to deteriorate PU communications.

Spectrum access is one of the most important functions of cognitive radios [3]. It helps in preventing potential collisions among the PUs and SUs. Spectrum access techniques can be broadly classified into three categories: *interweave*, *underlay*, and *overlay*. Here, we only focus on the first two types. In interweave access (Fig. 1), an SU is able to transmit only if there isn't any active PU on the same channel. When that is case, it uses the maximum power according to the spectral mask. When a PU becomes active, the SU must stop its communication right away, and start searching for a *white space*, which is a PU-free portion of the spectrum. In underlay access, illustrated in Fig. 2, an SU decreases its power level when the channel is being utilized by a PU. This reduction is concordant to the maximum value of interference a PU can tolerate. In overlay access, the PU uses the SU as a relay. In turn, the SU is allowed to have access to a slice of PU spectrum. However, the need for

complete channel state information from both SU and PU results in significant increase in complexity, making this mode far less attractive.

There is hardly any consensus about the more suitable mode for SUs, although there exists a rich body of literature for interweave and underlay modes separately [5], [6], providing some arguments in favor of one or the other (related mostly to the potential harm caused to PU activity [7]).

While the possibility of interruption-less transmissions (usually causing issues to higher layer protocols) is certainly an advantage the underlay mode provides, there are some disadvantages related to it as well. In this mode, an SU is allowed to transmit with maximum power only if there is no PU being active on that channel. These idle periods can be much shorter than the PU busy periods. This is particularly emphasized when there are highly active PUs. Even more, if the PU is located close to the SU, the SU's data rate will be reduced considerably, diminishing the data rate and the resulting throughput significantly.

As opposed to this, the interweave access allows the SU to search for another available channel (possibly less utilized) upon the PU arrival, and after finding one, to resume communication at high power again, possibly leading to an improvement in the average throughput. Yet, the intermittent nature of interweave access for SUs can delay some application files/flows, such as fetching a web page, a request to transmit a short file, etc., significantly, if they happen to arrive when the SU is looking for a new idle channel. Such delays can be aggravated not only if the SU operates in a part of the spectrum with highly active PUs or a large number of PUs (e.g., metropolitan areas in "peak" hours), but also if the scanning time is highly variable (sometimes an available channel is found promptly, but there are instances when the SU might spend a lot of time until an idle channel is found).

Following the discussion above, we can say that it is very difficult, if not impossible, to determine *a priori* which spectrum access mode to use for a given scenario. There is a close dependence of the performance on network parameters (allowed transmission power, PU duty cycle), metric of interest (delay, throughput), SU traffic (sporadic, intense), flow sizes, and higher-order statistics of some of the parameters, such as the variance of the time to find a new available channel.

To this end, in this paper we approach this problem analytically, characterizing the performance of underlay and interweave modes quantitatively, and comparing them in a unified setup. Our focus are two metrics: average delay and long-term throughput. We analyze the former using a queueing theoretic framework, and the latter using renewal-reward theory. We can summarize our contributions as follows:

- We derive closed-form expressions for the average file delay of interweave and underlay modes as a function of network parameters (mean PU idle time, data rates, statistics of the scanning time), and of traffic statistics (arrival rate, flow sizes). This allows a direct performance comparison. We derive the conditions under which the interweave mode is better. Finally, we propose a "dynamic" policy with which the SU can switch between the two modes dynamically, improving the performance considerably;
- We perform the throughput analysis in both modes, provide closed-form expressions, compare them analytically, and optimize;
- We validate our analyses with extensive realistic simulations, explore conditions under which one of the modes is better, and show under what conditions the dynamic policies can further improve the performance.

The paper is organized as follows. Some related work is discussed in the next section. Then, in Section III, we introduce the problem setup for underlay and interweave access. We perform the delay analysis in Section IV. Then, we continue with the throughput analysis in Section V. In Section VI, we validate our theoretical results vs. realistic simulations, and finally, in Section VII, we conclude this work.

II. Related Work

Some works focusing mostly on algorithm design of the underlay mode are [5], [8], [9], whereas focusing on the interweave mode are [6], [7], [10]. Underlay access analysis was conducted in [11]–[13]. Works containing analysis related to interweave mode are [14]–[17]. A survey on capacity limits of the three modes (including overlay mode) is presented in [18].

Still, most of performance analysis studies in underlay mode are information-theoretic, focusing on capacity only, and relying on interference temperature [11], [12], [19]. These papers are orthogonal works to ours. Their concern is the maximum power (equivalent to data rate in our work), which they determine by estimating the quality of the channels between SUs and PUs. In addition, those results can be used to determine the data rate in our analysis.

The achievable capacity for overlay and underlay mode is studied in [20]. However, the interweave mode is not considered, while in our work we compare the performance between the two modes for both the delay and throughput. Throughput maximization efforts were conducted in [21], by trying to design the optimal sensing time and power allocation strategy. However, there is no delay analysis.

In contrast, the file delay has been studied less for the underlay mode. An M/G/1 queue with finite buffer and timeouts has been proposed in [22], where the authors derive different metrics (including the delay). Those are numerically obtained results, and as such are difficult to interpret. A similar conclusion holds for [23], where the SU activity is also modeled as an M/G/1 queue. But, the procedure to derive the service time variance (needed in the P-K formula) is not shown there. Then again, we propose analytical models that lead to closed-form expressions for both the throughput and delay, which not only provide detailed insights on the effect of different network parameters, but also allow us to perform analytical comparisons and policy optimizations.

There have been considerably more works focusing on the expected packet delay in interweave access. An example is [24], which includes the analysis for both metrics. However, the underlay mode isn't analyzed, and as a consequence there can be no comparison between them. In most works, the PU activity is modeled as an alternating renewal process. According to [25], the Poisson process can capture reliably the call arrival process, but the durations of PU calls are generally distributed. Some works [14], [15] have used the outcomes of [25] to propose queueing models. However, those models suffer from some important caveats. *Firstly*, the problem is considered in the packet level and is modeled as a preemptive-resume M/G/1 queue, whereas in reality there will be a collision between SU and PU packets, when the latter becomes active again. As a result, those

packets have to be retransmitted when there is no PU. Contrary to this, our model enables capturing both the retransmitting and resume characteristic of real wireless systems. *Secondly*, the M/G/1 systems with priorities are only able to capture exponentially distributed scanning times, whereas our (different) interweave models hold for non-exponential scanning times, too. Even more, contrary to priority models, there is no need for the Poisson assumption for the PU traffic generation because the interval between consecutive PU arrivals is the sum of an exponential (ON period) and a generic (scanning time) random variable. That sum is a generally distributed random variable. This is what makes our model far more generic.

An interesting work is [26], in which some channel selection schemes are proposed to increase the throughput potential of the three modes (plus overlay). However, it suffers from a lack of performance comparison between the modes. A policy for throughput improvement is proposed in [27]. While the work itself is interesting, there is no delay analysis. As opposed to our work, where for the throughput model we consider a generic model, they assume the PU activity to be memoryless. A multiuser multi-channel setup is considered in [28]. Using dynamic programming, the authors find the most convenient time instant to switch to another channel, trying to maximize the throughput. However, there is no comparison between the modes, and the delay is not considered at all. Further, a hybrid mode is proposed in [29] together with an analysis on QoS provisioning. However, there is no performance comparison between the modes, which is one of the main foci of our work.

Our approach (considering different scanning time distributions) is the most comprehensive one, and the theory we develop here can capture even the full duplex technology, which for CRN is considered in [30], [31]. Assuming scanning time distributions other than exponential is justified in [32], [33]. In [32], the authors consider two scanning policies. In the first one (greedy policy), the scanning time is geometric (the discrete counterpart of exponential distribution). The second policy there is the optimal one, and it consists in choosing the channel to transmit on only after having scanned the whole channel pool. Obviously, the scanning time distribution for this policy is deterministic. As we know, we can model that scanning time with an Erlang distribution with a large number of stages. Further, in the very interesting work with thorough analysis [33], the authors suggest that the black hole distribution (equivalent to the scanning time in our work) in some scenarios can be exponential, and in some others it can be hyperexponentially distributed.

Directly comparing the performance of the two modes has been the focus of very few studies. One such study is [34], but the comparison is only regarding the outage probability, and not in terms of other metrics. Comparing results from different papers is not straightforward because of the different setups and assumptions made, non-closed form and implicit expressions, etc. In this work, we propose models that enable direct (analytical) performance comparison between the two modes. [35] is a study that is closer in spirit to ours. There, a hybrid cognitive system is studied, where an SU changes its operation mode in a probabilistic fashion. The goal is to optimize the throughput. There is no delay analysis, and the PU arrival process is rather restrictive (Bernoulli process). Our policies can optimize both the throughput and average delay. The analysis of the three modes (underlay, interweave, and overlay) is the focus of [36], where the authors provide theoretical limits for the 3 schemes, considering AWGN and power constraints. It is worth mentioning that the comparison (of the capacity only) is not performed analytically, but only simulating few scenarios. In this work, on the other hand, we provide delay and throughput analysis for both modes, under a wide variety of scanning time distributions. Our analysis enables performance comparison and optimal policy design.

On a similar note, we have compared the two modes only in terms of the mean file delay in [1]. Additionally, we have introduced throughput analysis, comparison, and optimization in this extended work.

Summarizing, the main novelties of our work are: (i) we perform a direct analytical comparison of underlay and interweave mode; (ii) we consider two metrics of interest, average delay and long-term throughput; (iii) we derive closed-form expressions in all scenarios; (iv) using our results, we propose optimal policies for the two metrics.

III. Performance Modeling of Spectrum Access

Two metrics are our focus: the *delay* and *throughput*. When analyzing the former, we assume that traffic flows are generated with rate λ, and underlie a Poisson process. We also assume that queues are of infinite length. We assume exponentially distributed file sizes. A file will be queued if when arriving in the system finds other files, and it will be served according to the First Come First Served (FCFS) discipline. The total time a file spends in the system is the sum of the service time and queueing delay. We refer to it as the *system time* or *transmission delay*.

When it comes to the second metric of interest of this paper, throughput is defined as the long-term average data rate. When analyzing the throughout, we make the usual assumption encountered in almost all works dealing with capacity analysis, namely that SUs have always some backlogged traffic.

A. Problem Setup for Underlay Access

If there is no PU using the channel, an SU operating in underlay mode can perform its transmission at full power until the moment when PU claims the channel back. Then, the SU has to decrease its power, in order not to deteriorate the PU QoS. For simplicity, we assume that the SU power varies between two levels: a "low" level, with some sort of activity by PU, and a "high" level with no PU activity. As a matter of fact, the permitted power level is a function of the quality of the primary and interfering channel (LoS, distance). Therefore, there would be more power levels in reality. In Section IV-C, we explain how the multi-power level scenario reduces to the two-level case.

Consider a channel that is intermittently used by one or several licensed users. Its occupancy can be modeled with an ON-OFF stochastic process (alternating renewals) [37] $(T^i_{\mathrm{ON}}, T^i_{\mathrm{OFF}})$, $i \geq 1$. This is illustrated in Fig. 2,[1] where an ON period represents *no PU activity*, whereas an OFF period denotes the existence of PU activity. i is the number of elapsed ON and OFF periods until the moment t. Unless stated otherwise, ON periods (T_{ON}) are assumed to be exponential with rate η_H. They are independent of each other and also independent of the duration of OFF periods, whose durations are also exponentially distributed, but with rate η_L.

[1] This is a more general approach than using a Poisson process for the PU activity modeling. For more details, see Section II.

While this is a necessary assumption for analytical tractability when considering the delay, as it will be shown in Section V, we consider generally distributed ON and OFF periods when it comes to throughput analysis. ON/OFF periods subject to generic distributions could also be introduced in our delay analysis. This is done using phase-type distributions for ON/OFF periods, and then using matrix-analytic methods [38]. But, those methods provide numerical solutions only, which in turn make it impossible for any direct comparison analytically. Furthermore, a number of simulation results (Section VI) suggest that, even for generally distributed ON and OFF periods, our theoretical predictions provide a satisfying accuracy in both setups.

The data rates during ON and OFF periods are denoted by c_H and c_L, respectively. It should be mentioned that the actual values of these rates depend on channel bandwidth, radio access technology, modulation and coding scheme, etc. Nevertheless, having a smaller value of the transmission power in OFF periods results in $c_L < c_H$.[2]

B. Problem Setup for Interweave Access

As opposed to underlay access, an SU with interweave access is able to realize transmission only if there is no PU active on that channel, i.e., during ON periods. Again, we assume *no PU periods* to be exponential with rate η_H, and the data rate is again denoted by c_H.

At the end of an ON period (PU arrival), the SU stops its transmission activity, and starts searching for an idle channel. At the moment it finds one, the transmission is resumed with full power (rate c_H). Consequently, this system can be again modeled with alternating renewals. Now, OFF intervals correspond to scanning periods where data transmission isn't allowed, i.e., $c_L = 0$. So, we say that SU switches its operation to the *scanning* mode. In this mode (OFF period), the cognitive user advances to a new channel and senses it shortly. If it is available, it stays there and transmission mode is restored. Otherwise, the SU switches to another frequency to sense a new channel. This is done until an idle channel is found.[3] Then, transmission resumes. As can be seen, per-channel scanning time is the sum of the switching delay (T_{switch}) and sensing time (T_I).[4] The total scanning time is

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

with L being a random variable that denotes the total number of channels the cognitive user needs to sense before finding the first idle channel. Per-channel sensing time is considered to be much shorter than the duration of ON/OFF periods.

When moving from frequency f_s to another channel with frequency f_d, there is a switching delay introduced, which is [39]

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

where β is a hardware-dependent parameter [39] denoting the time it takes to move to a neighboring channel, and δ represents the frequency distance between two contiguous channels.

Carefully observing Eq. (1), we can infer as follows. If the channel availabilities are independent with almost the same duty cycle (% of time the PU is being active on a frequency band), the random variable L can be considered to underlie a geometric distribution. Further, if the switching delay is always the same when moving between any two contiguous channels, knowing that the geometric distribution is obtained by rounding the exponential, we can deduce that an exponential distribution can approximate the scanning time in this scenario. Therefore, the exponentially distributed scanning time is considered first.

Other scenarios might arise as well. The scanning time nature depends heavily on the backup channels availability and on the frequency separation between them. In most cases of interest (very low duty cycle in all channels, high discrepancy between different channels' duty cycles, idle channels found only in the most remote spectrum parts, etc.), the exponential distribution will not be able to capture that scanning time behavior. Hence, we also perform analysis for high- and low-variability scanning time distributions.

Let's consider the scenario with a pool of channels with approximately the same very low duty cycle, which are close to each other. In this scenario, we expect a roughly constant time to find an idle channel (most of the time the available channel is found immediately), close to the average value $E[T_s]$. In such a scenario, the distribution of scanning time will be of *low variability* (lower than exponential). A very convenient way of modeling low variability distributions is to use k-stage Erlang distributions [40].

Contrary to the previous scenario, we might have the situation with a number of high duty-cycle channels (close to each other), and few other low duty-cycle channels far away in the spectrum. It may happen, with a small probability though, that the "neighboring" channels are all busy, forcing the SU to scan far away channels. Being proportional to the frequency distance, the scanning time in this case will have a significantly higher value. As a result, some of the samples might deviate from the mean value considerably. When that is the case, the distribution of scanning time is considered to be heavy-tailed. That behavior can't be captured by the exponential distribution. Hence, in that case, we use a hyperexponential distribution (with two branches) to model the scanning time. Regarding the latter, it can either be exponentially distributed with rate η_L (with probability p_0), or exponentially distributed with rate η_V (with a very small probability $1 - p_0$), where $\eta_L >> \eta_V$.

We consider two scenarios for supporting our aforementioned claims. In the first one, there are 20 channels next to each other in the spectrum. They all have low duty cycles (0.2), i.i.d. PU activities, $T_I = 1$ ms, and $T_{\text{switch}} = 10$ ms. The complementary cumulative distribution function (CCDF) of the duration of scanning time is shown in Fig. 3. On the same plot, we also show CCDFs of Erlang (for $k = 3$ and $k = 6$), and exponential distributions. Observing Fig. 3, we can conclude that the system behavior cannot be captured by an exponential distribution. Instead, we need to use an Erlang distribution. The same conclusion stems from Fig. 4 about the exponential distribution not

[2]We consider the interference and power constraints implicitly through the maximum allowed transmission rate in the model (both for underlay and interweave).

[3]We assume that channels are sequentially scanned from a list [39].

[4]Scanning consists of several processes. In our work, these processes are sensing and switching. Since our approach can capture different scanning time distributions (depending on the variability), it can be considered generic. Being generic, it can include other processes as well, such as rendezvous. Nevertheless, it is well beyond the scope of this paper to consider a specific rendezvous protocol. Any existing rendezvous protocol could be used, without altering the outcomes.

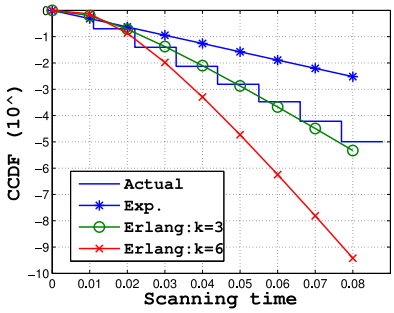

Fig. 3. The scanning time distribution for low PU duty cycles.

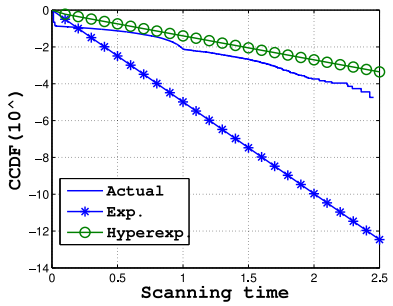

Fig. 4. The scanning time distribution for two channel groups.

being able to capture the scanning time. Contrary to the previous scenario, the channels are now with very high duty cycles (0.8), i.e., with highly active PUs. The switching time between this group and the group with 2 channels (with a low duty cycle of 0.2) is 500 ms. The hyperexp. distribution, also shown on the plot, is with parameters: $p_0 = 0.2$, $\lambda_1 = 90$, $\lambda_2 = 3$. Obviously, the hyperexp. distribution can capture this scenario more reliably.[5]

The results expressed through Figs. 3 and 4 emphasize the need to also consider other scanning time distributions in the analysis. Yet, closed-form results can still be found even for more general scanning times.

Practical considerations: We assume that a single radio and antenna are used in both modes. Hence, an SU cannot simultaneously scan and transmit. In order to detect that the PU is active again, one could simply switch the radio to receive mode from time to time, take a short-time sample (in the order of μs), and perform energy detection to check if a PU signal is present [10]. Assuming (as we do) that sensing time is much shorter compared to the actual duration of ON and OFF periods, we can safely ignore these sensing instants. On the other hand, the switching time is usually much higher than the sensing period (order of *ms* or even seconds), and cannot be neglected. It is worth mentioning that our approach is fully compliant with the full duplex technology [41], where three of its four modes: TS, SO, and CS [30] are related to our work, and lead to a higher sensing accuracy, which in turn, is captured by our perfect sensing scenarios (Section IV-A-IV-B). The fourth full-duplex mode (TR) is beyond the scope of this paper.

To better elucidate our model and subsequent analysis, we have presented our results implying a single SU pair using a single channel. This is an assumption that can be encountered in other works as well [31]. Nevertheless, our model can be easily adapted to capture both the multiuser and multichannel setups. For more details, see [42].

Table I summarizes some of the notation that will be used in the paper.

[5]It is worth mentioning that the aforementioned three types of distributions can also mimic, with a satisfying accuracy, the scanning time over channels whose occupancy exhibit a certain degree of correlation.

TABLE I
NOTATION

Variable	Definition
λ	Average SU file generation rate
T_{ON}	PU idle period duration
T_{OFF}	PU busy period duration or scanning time
$\pi_{i,L}$	Prob. of finding i files while in an OFF (low) period
$\pi_{i,H}$	Prob. of finding i files while in an ON (high) period
$\pi_{i,V}$	Prob. of finding i files while in an OFF (V-state) period
$\eta_L(\eta_H)$	Rate of leaving an OFF (ON) state
η_V	Rate of leaving a V-state
Δ	Mean file size
$\mu_L = \frac{c_L}{\Delta}$	Transition (service) rate while being in a low state
$\mu_H = \frac{c_H}{\Delta}$	Transition (service) rate while being in a high state
$E[T]$	Average system (transmission) time
p_{md}	Probability of miss-detection
p_{fa}	Probability of false alarm

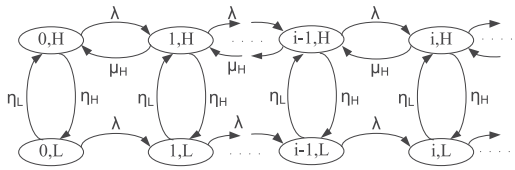

Fig. 5. The Markov chain for interweave access with exponential scanning time.

IV. Delay Analysis of Interweave and Underlay Modes

Next, we derive the expressions for the average file delay of underlay and interweave modes. For the latter, the analysis is performed over different distributions of scanning time: k-stage Erlang, exponential, and hyperexponential. We use various 2D Markov chains to model different scenarios, and the approach we follow here to obtain the results is based on Probability Generating Functions (PGF).

A. Delay Analysis for Interweave Access

Exponentially distributed scanning time: In Section III-B, we have assumed that ON periods are exponential, file sizes are exponential and generated according to a Poisson process. If the scanning time is also exponential, the system can be modeled as a 2D Markov chain, which is infinite in one of the dimensions (infinite buffer size), as depicted in Fig. 5. In this chain, each state is marked with a combination of the number of SU files, and the activity (L- states) or inactivity (H-states) of the PU. Note that OFF periods denote the scanning time in this case. $\pi_{i,H}$ denotes the stationary probability of SU having i files when there isn't any active PU on that channel. $\pi_{i,L}$ is the stationary probability of SU having i files while scanning. As we mentioned above, ON/OFF periods are exponential. Their parameters are η_H and η_L, respectively.

While in the upper portion of the chain (H-states) there are transitions going backwards from states $\{i, H\}$ to states $\{i-1, H\}$, with rates $\mu_H = \frac{c_H}{\Delta}$, in the lower portion, L-states (scanning), there are no transitions going backwards from states

$\{i, L\}$ to states $\{i-1, L\}$. The reason for this is the lack of ability of the SU to communicate while scanning. The average scanning time is $E[T_s] = \frac{1}{\eta_L}$.

Theorem 1: The average file delay in the interweave mode with exponential 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: It is obvious that except for $i = 0$, all the $\{i, H\}$ states have the same transition rates into and out of that state. The same holds for $\{i, L\}$ states, for which $i \geq 1$. Needless to say, but for $i = 0$ the states can't transition backwards (not possible to have negative number of files). Hence, the transitions into/out of these states are different compared to states with $i \geq 1$. As a result, we need to consider a system of 4 balance equations: 2 for high and low states corresponding to $i = 0$, and 2 other for high and low states corresponding to $i \geq 1$. We have [42]

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

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

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

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

The PGFs for this chain are defined as

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

We multiply Eq. (5) by z^i and add it to Eq. (4) to obtain

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

This leads to

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

Next, we multiply Eq. (7) by z^i and add it to Eq. (6):

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

Eq. (10) reduces to

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

The solution of the system of Eq. (9) and (11) leads to

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

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

$\pi_{0,H}$ (the prob. of SU having zero files with no PU activity) is the only unknown in Eq. (12). To find it, we need the balance equation across the vertical cut between states $\{i, H\}$ and $\{i, L\}$ on one side, and states $\{i, H + 1\}$ and $\{i, L + 1\}$ on the other, leading to

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

Summing over all i (from 0 to ∞) gives

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

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

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

Replacing $z = 1$ into Eq. (9) yields $G_L(1) = \frac{\eta_H}{\eta_L} G_H(1)$. Obviously, it holds $G_L(1) + G_H(1) = 1$, leading to

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

Substituting Eq. (17) into Eq. (16) enables finding $\pi_{0,H}$:

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

After finding $\pi_{0,H}$ and substituting it into Eq. (12), and the latter into Eq. (13), we obtain $G_L(z)$ and $G_H(z)$.

Finding the expected number of SU files is the next step. It is the sum of derivatives of partial PGFs at $z = 1$:

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

Differentiating Eq. (12) with respect to z leads to

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

In the previous equation, $F(z) = A(z)B(z) - \eta_L z$, 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 be easily shown that Eq. (20) at $z = 1$ is of the form $\frac{0}{0}$. Applying L'Hopital's rule twice, we obtain

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

Based on Eq. (21), we get

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

Performing some simple algebra, we get

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

Finding $E[N_H]$ is the next step. In that direction, Eq. (13) is differentiated, yielding

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

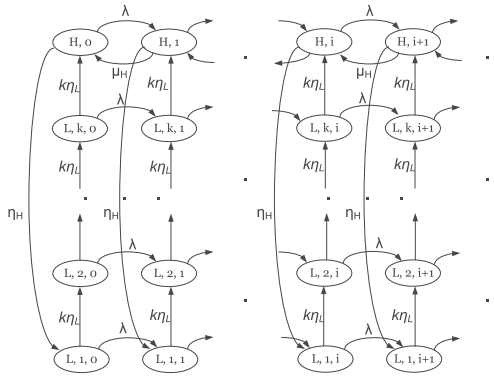

Fig. 6. The Markov chain for interweave access with Erlang scanning time.

Knowing that $E[N_H] = \lim_{z\to 1} G_H^{'}(z)$, and substituting $z = 1$ into Eq. (24) leads to

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

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

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

Performing some algebra, the average number of files in the system is found to be $E[N] = E[N_L] + E[N_H]$.

Finally, Little's law, $E[N] = \lambda E[T]$ [37], leads to the expected file delay (Eq. (3)) for the interweave mode. ■

Low-variability scanning time: We have derived the average delay for exponential scanning times in the previous section. Further, as described in Section III-B, there may exist some scenarios where the scanning time has an even lower variability. An Erlang k-stage distribution is used to capture this low variability. Despite the difference, we can still model the system with a 2D Markov chain, as shown in Fig. 6. As opposed to transitioning directly to the high state in the exponential scenario (Fig. 5), now a transition from a low state (scanning) to a high state (finding and using a new available channel) would need to go through additional $k-1$ intermediate states (vertically). The transition rates between these intermediate states are $k\eta_L$. Having k stages, the expected scanning time is $E[T_s] = k\frac{1}{k\eta_L} = \frac{1}{\eta_L}$. Making transitions backwards while scanning is impossible (no transmission). Solving analytically these sorts of Markov chains is very difficult, and we need to use numerical, matrix-analytic methods [38]. However, numerical methods provide no insights on the dependency of the solution on input parameters.

Interestingly enough, we are still able to derive an analytical expression in closed form. This is because of the particular structure of the 2D Markov chain at hand.

Although there is a higher number of states vertically, we can still use our approach to solve a system of $k+1$ equations in partial PGFs. The higher number of partial PGFs represents the crucial difference from the exponential scanning time scenario. The following theorem provides the average delay for this case.

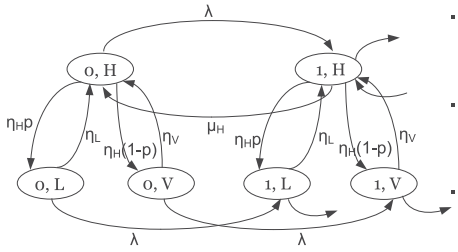

Fig. 7. The Markov chain for interweave access with hyperexp. scanning time.

Theorem 2: The average file delay in the interweave mode with Erlang scanning time is

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

We can see from Fig. 6 that we need to write two different balance equations for high states: one for $\{H,0\}$, and the other one for $\{H,i\}$, $i \geq 1$. The same holds for the k low levels. Hence, in this case we need to solve a system of $2(k+1)$ balance equations, leading to $k+1$ equations in partial PGFs. We omit the proof here because of space limitations. It can be found in [42].

Comparing Eq. (27) with Eq. (3), we can observe that the expected file delay for exponentially distributed scanning time is always higher than for Erlang-distributed scanning time, because $\frac{k+1}{2k} < 1, \forall k > 1$.

High-variability scanning time: The highly variable scanning time scenario is the next one considered here. It is modeled by a hyperexponentially distributed random variable with two branches. These will be mapped into two separate states (specified as *L* and *V*). The corresponding Markov chain is illustrated in Fig. 7. When scanning, the SU can be in one of the $\{i,L\}$ states (quickly finding an available channel) with probability p_0, or in one of the $\{i,V\}$ states (long time to find an available channel) with probability $1-p_0$. The average scanning time 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 high variability, we pick a very small value for $1-p_0$ (e.g., around 0.01), and $\eta_L >> \eta_V$.

Again, the structure of this chain allows avoiding matrix-analytic methods, which are numerical. Instead, we apply the methodology based on PGFs to derive a formula that is in closed form, which is presented in the following theorem, whose detailed proof can be found in [42].

Theorem 3: The average file delay in interweave mode with hyperexponential scanning time is given by

$$E[T_{hyp}] = \frac{(\eta_H E[T_s])^2 + \eta_H\mu_H\left(\frac{p_0}{\eta_L^2} + \frac{1-p_0}{\eta_V^2}\right) + 2\eta_H E[T_s] + 1}{(1+\eta_H E[T_s])(\mu_H - \lambda - \lambda\eta_H E[T_s])}. \tag{28}$$

The number of balance equations in the system that describes this Markov chain is 6: we need 3 equations for states $\{0,H\}$, $\{0,L\}$, $\{0,V\}$, since they have fewer transitions out/into, and 3 for the other states $\{i,H\}, \{i,L\}, \{i,V\}, i \geq 1$. Considering that the total flow rate out of a given state should be equal to the flow rate into that state, and knowing the corresponding

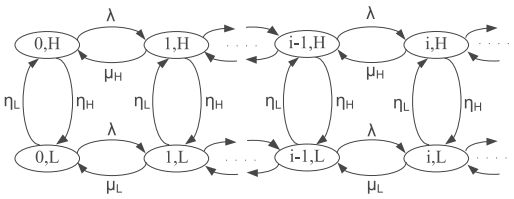

Fig. 8. The 2D Markov chain for the underlay mode.

transition rates, we can write these equations. See [42] for complete derivation.

It can be easily proven [42] that $\frac{p_0}{\eta_L^2}+\frac{1-p_0}{\eta_V^2}=\frac{1}{\eta_V^2}+(\frac{1}{\eta_L}+\frac{1}{\eta_V})(E[T_s]-\frac{1}{\eta_V})>\frac{1}{\eta_V^2}+E[T_s](E[T_s]-\frac{1}{\eta_V})=(E[T_s])^2+\frac{1}{\eta_V}(\frac{1}{\eta_V}-E[T_s])>(E[T_s])^2$. This leads to $E[T_{hyp}]>E[T_{\exp}]$.

B. Delay Analysis for Underlay Access

As mentioned in Section III-A, the SU can communicate at all times (whenever it has something to send/receive) when operating in the underlay mode, irrespective if there is or isn't an active PU on that channel. However, the cognitive user will experience different rates, depending on the fact whether there is/isn't an active PU.

Once more, we model the system behavior as a 2D Markov chain (Fig. 8). Note the main difference from Fig. 5, where transitions backwards in the low states do not exist. These transitions do exist in the underlay mode. Their rate is $\mu_L=\frac{c_L}{\Delta}$.

It should be mentioned that $\pi_{0,L}$ ($\pi_{0,H}$) denote, as before, the stationary probability of no files at the SU, when there is (not) an active PU.

Theorem 4: The average file delay in underlay access is given by Eq. (29) shown at the bottom of this page.

Proof: Equating the flow rates in and out of states (similarly to interweave mode) $\{0,L\},\{0,H\},\{i,L\},\{i,H\}$, we get the following balance equations

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

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

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

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

We define the PGFs for both high and low states as before:

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

We multiply Eq. (32) and Eq. (33) by z^i, and then add them to Eq. (30) and Eq. (31), respectively. After rearranging, we obtain

$$(\lambda+\eta_L+\mu_L)G_L(z)=\lambda zG_L(z)+\eta_HG_H(z)+\frac{\mu_L}{z}(G_L(z)-\pi_{0,L})+\pi_{0,L}\mu_L, \tag{34}$$

$$(\lambda+\eta_H+\mu_H)G_H(z)=\lambda zG_H(z)+\eta_LG_L(z)+\frac{\mu_H}{z}(G_H(z)-\pi_{0,H})+\pi_{0,H}\mu_H. \tag{35}$$

The solution to the system of Eq. (34)–(35) leads to

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

where

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

There is only one root in $(0,1)$ of the polynomial given by Eq. (37) [43]. We denote this root as z_0. Setting $z=z_0$ in Eq. (36), and after some elementary algebra, we obtain $\pi_{0,H}$ and $\pi_{0,L}$, as

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

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

The procedure onwards is similar as in interweave access. Finally, using Little's law $E[N]=\lambda E[T]$ [37], we get Eq. (29). ■

Queue Stability: In order our results to hold, the queues must be stable. Stability also implies the ergodicity of our Markov chains of interest. The condition for stability of the interweave access queue is $\lambda<\frac{E[T_{\rm ON}]}{E[T_{\rm ON}]+E[T_s]}\mu_H$. On the other hand, for the underlay access queue the condition for stability is $\lambda<\frac{E[T_{\rm ON}]}{E[T_{\rm ON}]+E[T_{\rm OFF}]}\mu_H+\frac{E[T_{\rm OFF}]}{E[T_{\rm ON}]+E[T_{\rm OFF}]}\mu_L$.

C. Underlay Access With Higher Number of Levels

While the two-level power assumption was made for tractability purposes, there will be a higher number of levels in most practical scenarios of interest. When that is the case, there will be more than 2 states in vertical direction of the corresponding 2D Markov chain. The solution to that chain can be obtained numerically [42]. Another way would be to lump the $\lfloor\frac{M}{2}\rfloor$ levels (assuming there are M power levels in total) with lowest power into a single level, whose data rate is the weighted average of the corresponding levels' data rates. The mean time in this level is the sum of the expected times of staying in all the levels that are lumped into this new level [42]. The same approach follows for the $M-\lfloor\frac{M}{2}\rfloor$ highest power levels. The Markov chain of

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

Fig. 8 is obtained this way, but with other transition rates [42]. Its solution is, as shown, in closed form.

D. Miss-Detections and False Alarms

So far, we have been assuming a perfect sensing capability of the SU. In case of imperfect sensing, two phenomena might arise: *miss-detection* and *false alarm* [44].

When taking into account sensing imperfections, the analysis needs to be slightly modified. Due to space limitations, we refer the interested readert to our tech report [42] for more details.

The miss-detections can affect the performance of licensed users. As a result, the PU will experience a higher interference. The percentage of the PU busy period interfered by an SU, due to miss-detections, can easily be computed. However, the impact on the SU is what we are concerned with in this work. Analyzing the imperfections effect on the PU is beyond the scope of this work.

E. Analytical Comparison of Delays in Interweave and Underlay Access

Having derived the formulas for average file delay in interweave and underlay modes in Sections IV-A and IV-B, we can compare the delays incurred in both of them. The average delay depends on traffic intensity, scanning time, the PU activity statistics, and data rate. Initially, the SU has to decide at the very beginning which mode to use: interweave (stop transmitting when a PU arrives, and start scanning for a new idle channel), or underlay (keep staying on the same channel and transmit with the allowed power level). We call this "the static policy", although it is not a real policy *per se* (in practice, an SU will be able to switch channels and scan at any time). Nevertheless, it allows to attain a deeper understanding on the parameters that affect the performance in every case. We consider a *dynamic policy*, which is more realistic, in Section IV-F.

In general, for the interweave to outperform underlay mode, the average scanning time, $E[T_s]$, should be short enough to make sure that the opportunity cost of not communicating any data for some time is amortized by the quick discovery of an available (idle) channel. In Table II, we provide the maximum values of average scanning time for which the interweave mode still outperforms the underlay. All expressions in Table II indicate a complex relationship between $E[T_s]$ and different system parameters. Even more, this "boundary" point further depends on the scanning time variability. For instance, we can observe that $B_1 > A_1, B_2 > A_2$ [42]. Also, for exponentially distributed scanning time, a smaller value of the scanning time is needed in order for the interweave mode to perform better than for the case of Erlang scanning time. Comparing the parameters of the hyperexp. distribution with the other two (exp. and Erlang), we can get a similar conclusion. This shows the impact that the variance of the scanning time has on the boundary value (crossing point). If the scanning time has lower variability, the crossing point is higher.

F. Delay Minimization Policy

Previously, we compared interweave and underlay modes in a "static" context, where the decision which mode to choose is made at the beginning, and that mode is then used at all times. In practice, an SU will normally be able to choose to either stay at the current channel and transmit with lower power, or search for a new idle channel (scan) *at any time*. If that policy is designed properly, it can lead to further performance improvements. In the following, we define a hybrid policy, and identify the conditions under which this policy offers some gains. We also derive an optimal switching rule (between modes).

Definition 1: Delay minimization policy:

- The SU will keep residing 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, the SU interrupts transmission and initiates scanning for a new white space.
- If the PU leaves the channel before time t, the SU resumes transmission at higher power, and resets the turning point to t time units ahead.

This policy is generic. Finding an optimal t is our goal. Let us consider some possible scenarios, so that we can understand better the trade offs involved. First, if the static underlay policy is worse than static interweave policy, we can easily infer that the optimal turning point value is 0: after the PU arrival, it is always better for the SU to start scanning immediately. Therefore, we are only interested in cases in which the underlay is better *on average* (i.e., the respective condition in Table II is *not* fulfilled), but with some OFF periods that are too large, when it is much better to initiate scanning instead (after some time).

In the above context, let's assume first that we have exponential OFF periods. Further, assume also that a PU became active in the current channel and t units have already elapsed with the channel still not becoming idle. As the exponential distribution is memoryless, the average remaining time of the PU busy period is still the same as at the moment when the PU just arrived, i.e., equal to $E[T_{\mathrm{OFF}}]$. Therefore, if at time 0 it was better to reside on the channel and communicate at lower rate rather than start scanning, for any elapsed time $t > 0$ *it is still better to reside on that channel and not initiate scanning*. The same conclusion holds for PU busy periods that are subject to *increasing failure rate (IFR)* distributions,[6] i.e., distributions with lower variance than the exponential distribution. There, if SU can't gain by scanning at $t = 0$ (static interweave is worse on average), as t increases, the expected gain from staying in the underlay mode in fact increases.

As a result, we can infer that a dynamic policy (i.e., a strictly positive optimal value of t) can offer gains *only for PU activity periods that are subject to decreasing failure rate (DFR) distributions*. In those cases, when the PU becomes active, at the beginning it will be, on average, better to do underlay, but as time elapses, the average remaining PU busy period keeps increasing, until at some point it becomes beneficial to stop transmitting and start scanning for a new idle channel. This allows dynamic policy to outperform both static policies, as we show in Section VI, by essentially "pruning" the long OFF periods from the underlay mode.

Next, we provide an expression for the optimal turning point in the scenario with exponentially distributed ON periods, and OFF periods that are Pareto distributed (minimum value L and shape parameter α) [42]. The latter is a popular DFR distribution.

Result 5: The optimal turning point that minimizes the average file delay, for exponentially distributed ON periods and

[6] Decreasing (increasing) failure rate distributions are those for which $\frac{f(x)}{1-F(x)}$ is a decreasing (increasing) function in x.

TABLE II
THE MAXIMUM VALUES OF THE SCANNING TIME FOR WHICH THE INTERWEAVE MODE IS SUPERIOR (CROSSING POINT)

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	$E[T_s] < \frac{-C_2+\sqrt{C_2^2-4C_1C_3}}{2C_1}$	$C_1 = \eta_L\eta_V\eta_H^2(1+\lambda E[T_u])$ $C_2 = \eta_H[\mu_H(\eta_V+\eta_L) + 2\eta_L\eta_V + 2\lambda\eta_L\eta_V E[T_u] - \eta_L\eta_V\mu_H E[T_u]]$ $C_3 = \eta_L\eta_V - \eta_H\mu_H - \eta_L\eta_V(\mu_H-\lambda)E[T_u]$

TABLE III
SUMMARY OF THE DELAY POLICIES

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

Pareto OFF periods, is the solution to

$$\frac{\eta_H}{\eta_H+\eta_L}\cdot\frac{c_H-c_L}{c_H}\left(\left(1-\frac{1}{\alpha}\right)t^{1+\alpha}+\frac{1}{\alpha}L^{\alpha-1}t^2\right) + \frac{\eta_H}{\eta_H+\frac{c_H}{\Delta}}\cdot\frac{\eta_L}{\eta_H+\eta_L}\left(\frac{c_H-c_L}{c_H}L^\alpha t - E[T_s]\alpha L^\alpha\right)=0. \tag{40}$$

We can solve Eq. (40) numerically. All the possible scenarios are summarized in Table III.

Interestingly, as opposed to the variability of OFF periods in underlay mode, the scanning time variability (the variance of "OFF" periods in the interweave access) has no impact on the dynamic policy decisions. It only "gets involved" when comparing the static interweave and underlay access directly (Table II).

It is also very positive that the static policy in almost all cases, except one, is the optimal one. This leads to a significant reduction in algorithm complexity, since the SU needs to decide only once, at the beginning.

V. THROUGHPUT ANALYSIS OF INTERWEAVE AND UNDERLAY MODES

While file delay is a relevant metric of interest for applications that are interactive, a number of other applications, such as one-way streaming, peer-to-peer file exchange, etc., are throughput-sensitive. For these types of applications, the SU will strive to maximize the throughput. In the scenario having delay as the metric of interest, transmitting the message right away, although at a low rate, may be the better option, so that the message is not delayed during the scanning procedure. In that case, underlay access is more preferable. On the other hand, if SU is more interested in throughput optimization, it simply might experience higher gains by looking for a new high throughput channel. Obviously, there is a tradeoff involved between scanning for a new (better) available channel, and transmitting at a low rate.

At first, as before, we compare the two static modes, deciding on the more suitable among them at the very beginning, given the input statistics. Afterwards, we propose a hybrid policy, which is dynamic, and also stipulate the conditions under which it improves the performance throughput-wise.

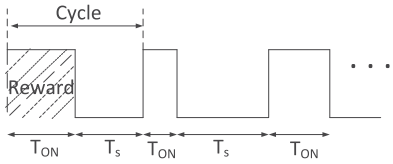

Fig. 9. Renewals in interweave mode.

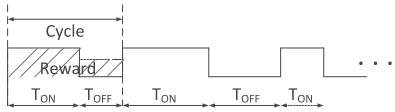

Fig. 10. Renewals in underlay mode.

A. Analytical Comparison of Throughput

Modeling the channel occupancy (in both modes) with alternating renewal processes enables us to use the approach based on renewal-reward theory to calculate the throughput, as a first step. Then, we can compare the modes directly. In both of them, an ON and an OFF period comprise a *cycle*, whereas the amount of data transmitted/received during a cycle is the *reward*. Those rewards are illustrated in Fig. 9 (interweave) and Fig. 10 (underlay), respectively. There are two distinctions: (i) OFF periods in the interweave mode depend on the scanning time statistics, and not on the PU busy periods directly; (ii) there is a positive reward during OFF periods in underlay access.

The condition under which the interweave mode provides higher throughput than underlay access is given as follows.

Theorem 6: The interweave mode provides a higher throughput than the underlay mode under the following condition

$$\frac{E[T_s]}{E[T_{\rm OFF}]} < \frac{1-\frac{c_L}{c_H}}{1+\frac{c_L}{c_H}\frac{E[T_{\rm OFF}]}{E[T_{\rm ON}]}}. \tag{41}$$

From Eq. (41), if e.g., $c_L=\frac{1}{2}c_H$ and $E[T_{\rm OFF}]=E[T_{\rm ON}]$, the interweave mode will provide higher throughput for $\frac{E[T_s]}{E_{T_{\rm OFF}}} < \frac{1}{3}$. For a higher difference between data rates, e.g., $c_L=\frac{1}{5}c_H$, Eq. (41) leads to the condition $\frac{E[T_s]}{E[T_{\rm OFF}]}<\frac{2}{3}$. As can be seen from these two examples, for a higher ratio of $\frac{c_H}{c_L}$, the interweave mode will be superior even for higher $E[T_s]$. Similarly,

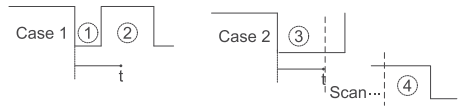

Fig. 11. Possible cases in optimal throughput policy.

a higher $\frac{E[T_{\rm ON}]}{E[T_{\rm OFF}]}$ leads to a higher relative threshold, in line with Eq. (41).

B. Throughput Maximization Policy

Similarly to the delay minimization policy, if the interweave access provides a higher throughput in a given scenario, there is no need for the SU to use the underlay access. But, if the underlay is better *on average*, the SU may experience an improvement by deciding to scan and switch to another available channel in some instances. Here we present such a dynamic (hybrid) policy that predicts the potential gains "on-the-fly", and may choose at any time one of the two options: to transmit at lower power, or to scan. We determine the optimal *turning point*, t_{opt}, which is the moment the SU should switch to the interweave mode, so that the throughput is maximized. The definition of the policy is:

Definition 2: Throughput maximization policy:

- Identical to the dynamic delay policy, but different optimal turning point.

As stated in policy definition, the different value of the optimal turning point, t_{opt}, is the only difference between the dynamic policies for delay minimization and throughput maximization. To find the turning point, we use again renewal-reward theory, noting that low periods can either be a combination of a low rate period and scanning time, or a pure low period (like in the underlay mode).

As before, the decision about the switching moment lies in the uncertainty of the remaining duration of the current OFF period (PU activity). Knowing the average value of the PU activity duration (one of the factors for choosing the static policy) is not sufficient, because the variance also plays a very important role. The same reasoning, as for delay optimization, can be made to show that any performance improvement beyond the preferable mode is only possible if and only if the underlay mode is superior *on average*, and if OFF periods are subject to a distribution with decreasing failure rate. Consequently, Table III conclusions are valid for the maximum throughput policy, too. The only difference, as mentioned, is that the value of t_{opt} is different now. The optimal value of t_{opt} for Pareto distributed OFF periods with minimum value L, and shape parameter α, is given as follows.

Theorem 7: The optimal turning point that maximizes the throughput for exponentially distributed ON periods and Pareto OFF periods is the solution to

$$\frac{c_H - c_L}{\eta_H} t^{\alpha} - \alpha E[T_s] \left(\frac{c_H}{\eta_H} - \frac{\alpha}{1-\alpha} c_L L \right) t^{(\alpha-1)} - \frac{c_L E[T_s] L^{\alpha}}{1-\alpha} = 0. \quad (42)$$

Proof: The cycle consists of two possible combinations of ON and OFF periods (see Fig. 11): (1) An OFF period that is smaller compared to the turning point, and a "regular" ON period (Case 1). This corresponds to a "pure" underlay mode; (2) An OFF period larger than the turning point $T_{\rm OFF} > t$, and an ON period that corresponds to the remaining available time of the newly found channel, $(T_{\rm ON}^{(e)})$ (Case 2).

In Case 2, The OFF period is $t + T_s$. The probability for the occurrence of Case 1 is $P[T_{\rm OFF} < t] = F_{\rm OFF}(t)$. The average cycle duration and the expected amount of transmitted data per cycle are given by

$$E[T_{\rm cycle}] = \left(E[T_{\rm OFF}|T_{\rm OFF} < t] + E[T_{\rm ON}]\right) F_{\rm OFF}(t) + \left((t + E[T_s]) + E[T_{\rm ON}^{(e)}]\right) \bar{F}_{\rm OFF}(t), \quad (43)$$

$$E[R] = \left(c_L E[T_{\rm OFF}|T_{\rm OFF} < t] + c_H E[T_{\rm ON}]\right) F_{\rm ON}(t) + \left(c_L t + c_H E[T_{\rm ON}^{(e)}]\right) \bar{F}_{\rm ON}(t). \quad (44)$$

The average OFF period duration conditioned on having a value smaller than the turning point t, $E[T_{\rm OFF}|T_{\rm OFF} < t]$, can be found because we know the channel statistics. We find it as $E[T_{\rm OFF}|T_{\rm OFF} < t] = \int_0^t \frac{x f_{\rm OFF}(x)}{F_{\rm OFF}(t)} dx$.

The turning point value, t_{opt}, that maximizes the throughput can be found as the solution to $\frac{d\left(\frac{E[R]}{E[T_{\rm cycle}]}\right)}{dt} = 0$. Its solution results in Eq. (42). See [42] for more details. ■

We will see in Section VI that the optimal throughput policy indeed increases considerably the long term throughput. This increase comes with a little extra cost to keep track of the estimates for PU variability on every channel. Compared to the "static" policies, this is an addition to the first order statistics (duty cycle) needed. However, as we have illustrated in Table III, in most cases, the static policy is the optimal one.

VI. Simulation Results

In this section, we validate first various analytical expressions against realistic simulations, including several cases where we relax one or more of the theoretical assumptions. Then, we show the improvements that are achieved using different policies.

We will consider two types of networks. In the first one, we assume that the mean duration of ON and OFF periods correspond to those measured in [45], with mean values $E[T_{\rm ON}] = 5$ s ($\eta_H = 0.2\ {\rm s}^{-1}$), and $E[T_{\rm OFF}] = 10$ s ($\eta_L = 0.1\ {\rm s}^{-1}$). This is the cellular scenario. In the second one, we fit the average ON/OFF duration to the average values observed in [46], where $E[T_{\rm ON}] = 4$ s ($\eta_H = 0.25\ {\rm s}^{-1}$), and $E[T_{\rm OFF}] = 9$ s ($\eta_L = 0.11\ {\rm s}^{-1}$). This is the WiFi scenario. Unless stated otherwise, we assume exponential ON/OFF periods in both scenarios. In Section VI-A and VI-B we consider other distributions, too. The data rates in cellular scenario are $c_H = 8$ Mbps and $c_L = 1.2$ Mbps. The WiFi data rates are $c_L = 2$ Mbps and $c_H = 10$ Mbps.[7] Finally, the SU files are generated according to a Poisson process, with average size 125 KB,[8] and are exponentially distributed.

A. Validation of the Delay Models

We depict in Fig. 12 theoretical model predictions vs. simulation results of the average file delay for a cellular and WiFi

[7]These rates are similar to the values encountered in practical settings [47]. Although they correspond to PU rates, while the actual SU rates depend on a number of factors, such as: SU distance from AP/BTS, modulation and coding scheme, channel bandwidth, etc., still w.l.o.g. we assume higher WiFi SU rates (than cellular SU rates).

[8]This is a normalized value so that arrival rates roughly correspond to traffic intensities in [45] and [46], although other values lead to similar conclusions.

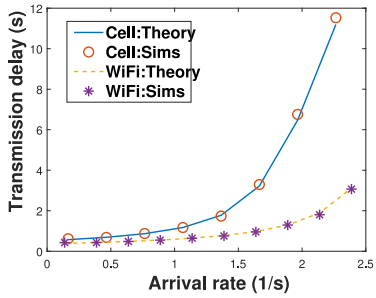

Fig. 12. The delay for underlay mode (imperfect sensing).

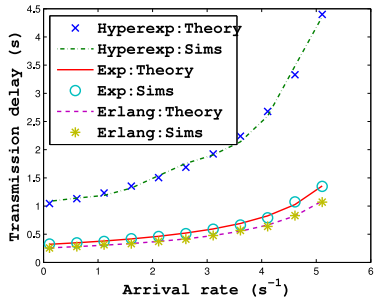

Fig. 13. The delay for interweave mode in a cellular scenario (perfect sensing).

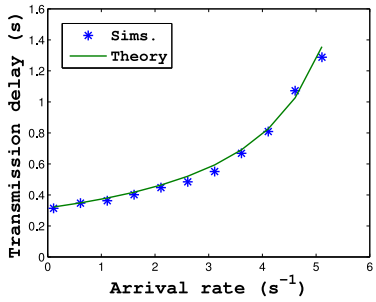

Fig. 14. The delay for generic interweave mode.

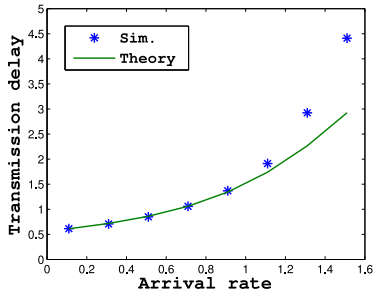

Fig. 15. The delay for generic underlay mode.

scenario. In the cellular scenario, $p_{\rm md} = p_{\rm fa} = 0.05$, whereas in the cellular scenario $p_{\rm md} = p_{\rm fa} = 0.1$. As can be noticed, theoretical results match those from simulation. The delay increases with the traffic intensity (arrival rate), which is common for queueing systems. There is a higher delay in the cellular setup because the considered cellular rates are lower, and also the PU is more active.

Next, we validate analytical predictions for the interweave scenario. We do this assuming perfect sensing. In the following, unless otherwise stated, we assume that sensing is perfect. Fig. 13 illustrates the theoretical vs. simulation results for the cellular scenario with three different scanning type distributions (exp., 4-stage Erlang, hyperexponential), all of them having $E[T_s] = 1$ s. We take $\eta_L = 1.9$ s^{-1} and $\eta_V = 0.1$ s^{-1} for the hyperexp. distribution. The probability of having a large scanning time (scanning a channel that is far away in the spectrum) is 0.05, resulting in a coefficient of variation with a value around 3.

As can be observed from Fig. 13, there is an excellent match between theoretical and simulated results for different distributions of scanning time. Another outcome out of this is that the lowest average delay is achieved for Erlang-distributed scanning time, whereas hyperexponential scanning time leads to the worst result. This is in accordance with the analytical results of Section IV. According to our results, lower (higher) variability in scanning time results in lower (higher) variability in the service time, which further leads to lower (higher) delays. Looking at the curves of the cellular scenario in Fig. 12 and Fig. 13, we can infer that the interweave mode incurs a lower delay. For the cellular setup with an arrival rate of $\lambda = 1$ s^{-1}, and exponentially distributed scanning time, Table II predicts a maximum expected scanning time of 2.8 s. In the scenario we are considering, $E[T_s]$ is considerably smaller (1 s). Also, there are no sensing errors in the scenario of Fig. 13. Hence, the lower delay in interweave access.

Previously, we have used real-life values for channel availabilities and data rates, but we have assumed memoryless ON/OFF periods, in line with our models. While the actual distributions are subject to the PU activity pattern, measurements [45], [46] indicate that these distributions are "heavy-tailed". Hence, it is interesting to see how our models' predictions cope in this case. In that direction, we consider a scenario with Bounded Pareto (BP) ON and OFF periods, having the following parameters $\alpha_{\rm ON} = \alpha_{\rm OFF} = 1.2, L_{\rm ON} = 1.31, H_{\rm ON} = 200, L_{\rm OFF} = 2.9, H_{\rm OFF} = 200$. We focus our attention on the cellular scenario because of space limitations. The rest of the parameters do not change from the scenarios of Figs. 12 and 13. Fig. 14 illustrates the mean file delay for the interweave mode vs. our analytical prediction, whereas Fig. 15 does the same thing for underlay access. The scanning time in the former is exponentially distributed with an average value of 1 s. Despite the much higher variability of ON and OFF periods, our theory still offers a satisfying prediction accuracy. These results highlight the practical utility of the models.

B. *Validation of the Throughput Models*

Validating the throughput theoretical results (Section V) is what we proceed next with. The advantage of our approach is that its valid for general ON/OFF distributions. We consider first the interweave mode throughput, and two scenarios: WiFi and cellular. For both networks, the data rates and ON period statistics are identical to those in Section VI-A. We assume a Bounded Pareto scanning time distribution with mean 0.5 s. Fig. 16 illustrates the throughput vs. *channel availability*. The latter denotes the % of time the SU is able to communicate, and is expressed as $\frac{E[T_{\rm ON}]}{E[T_{\rm ON}]+E[T_s]}$. This is a more appropriate metric for this mode than the duty cycle, which is relevant only for the underlay, where the SU keeps communicating on the same channel. Different values of duty cycle characterize different channels.

Next, we consider the underlay mode throughput. The parameters are the same as in Section VI-A. Fig. 17 depicts the throughput for different duty cycle values. For this mode too, the theoretical predictions match the simulation results. As ex-

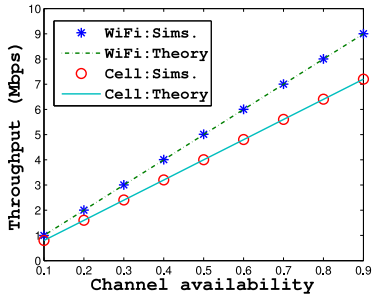

Fig. 16. The interweave mode throughput.

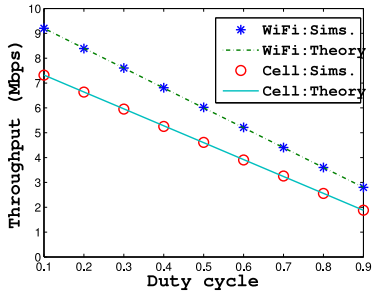

Fig. 17. The underlay mode throughput.

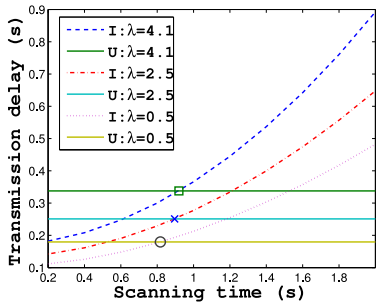

Fig. 18. The static delay policy for different λ and exp. scanning times.

pected, the throughput decreases for more active PUs. The decrease exhibits a linear dependency on duty cycle. The same conclusions hold for the WiFi results, which are also shown on the plot.

The results in Figs. 16 and 17 can be related by comparing the value for channel availability $1 - X$ in Fig. 16 to the corresponding duty cycle of X in Fig. 17. The values of throughput in Fig. 16 are a little lower than in Fig. 17.

C. Delay Minimization Policies

We perform next a detailed comparison of the two access modes. The first goal is to assess the "static" versions of the two policies, together with validating the theoretical predictions when one of them outperforms the other. The other goal is to see if the dynamic policy outperforms the two modes.

As an interesting case, we consider the scenario with the following parameters: $c_H = 10$ Mbps, $c_L = 0.5$ Mbps, $\eta_H = 0.1$ s^{-1}, $\eta_L = 1$ s^{-1}. Fig. 18 illustrates the mean file delay (denoted by I) vs. different mean scanning times, which are exponentially distributed, for three different traffic intensities (sparse, moderate, intensive). The corresponding average delay of underlay mode (denoted by U) is depicted on the same plot, too. Further, we depict with small circles the theoretical boundary values (crossing points) of the average scanning times (Table II). In the sparse traffic scenario, the interweave mode is worse for

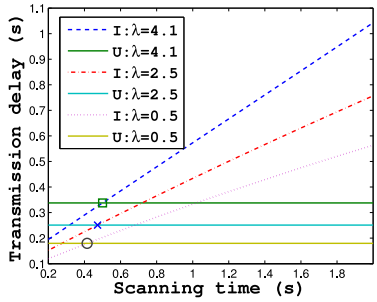

Fig. 19. The static delay policy for different λ and hyperexp. scan. times.

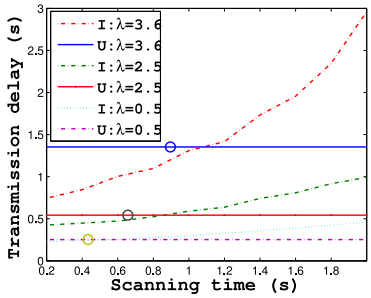

Fig. 20. The crossing point for static generic delay policy (different λ).

scanning times larger than 800 ms. The first important outcome is that the predicted boundary value of the scanning time is correct. The other thing to observe is that this boundary inclines for higher intensity traffic. The reason for this effect lies in the fact that for higher loads the largest delay component is the queueing delay. Consequently, it is worth waiting for a while, find an available channel, and only then get rid of the queued data (at a higher rate). A further increase in the load results in smaller increases of the crossing point.

We proceed with the hyperexponentially distributed scanning time, with rate parameters $\eta_L = 6$ s^{-1}, $\eta_V = 0.4$ s^{-1}, and the probability p_0 taking values to maintain a given mean scanning time. The coefficient of variation, observed in this scenario, falls in the range (2, 2.5). There are no changes on the other parameters. Fig. 19 illustrates the average delay. The crossing points between interweave and underlay mode are lower than in the scenario of Fig. 18 because of the higher variance of the scanning time.

The correctness of the crossing-point theoretical expression between interweave and underlay delay was shown previously. We proceed showing the (broader) practical importance of the crossing-point formula. In that direction, we consider a scenario with generally distributed ON and OFF periods. Three different traffic patterns (sporadic, medium, intense) are again considered. The input data are: $\eta_L = 1$ s^{-1}, $\eta_H = 0.1$ s^{-1}, $c_L = 2$ Mbps, and $c_H = 5$ Mbps. We consider Bounded Pareto ON periods with shape parameter $\alpha = 1.2$, minimum value of $L = 2.375$, and maximum value of $H = 1000$. Fig. 20 illustrates the mean (file) delay for different exponentially distributed scanning times with different traffic intensities vs. the underlay scenario with the same traffic intensity. The small circles depict the maximum (theoretical) value of the scanning time for which the interweave access is superior, i.e., the crossing point. It can be observed from Fig. 20 that for low arrival rates our result is accurate in predicting the bound. For higher arrival rates, the mismatch is in the order of 10%. Apparently, our model can be used to predict the performance with a very high accuracy, even

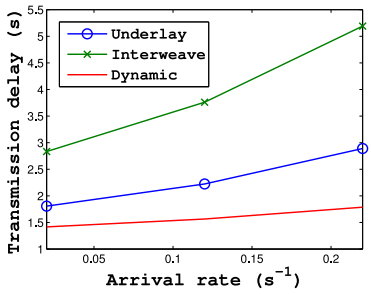

Fig. 21. The dynamic delay policy for heavy-tailed OFF periods.

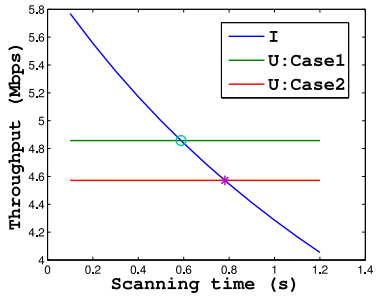

Fig. 22. The throughput static policy for two different low period data rates.

under generic distributions (different from the assumptions in the theory).

Dynamic delay policy: We have seen in Section IV-F that the dynamic policy can offer a significant gain when the PU busy periods (OFF periods in underlay access) underlie a decreasing failure rate distribution (heavy-tailed). In the scenario we consider, OFF periods are Bounded Pareto, with extreme values $H = 100$, $L = 0.2$, shape parameter $\alpha = 1.2$, and a mean scanning time of $E[T_s] = 1$ s. The rest of the parameters are like in the cellular scenario. The average delay vs. the arrival rate, for this scenario, is shown in Fig. 21. Based on the results of the static policy, the underlay outperforms interweave access. However, the dynamic policy offers the best result, with an extra reduction in average delay of 20–50%.

D. Throughput Maximization Policies

We consider throughput optimization policies in this final subsection. A thorough comparison of the two modes will be performed. As a first step, we investigate the "static version" of both policies, validating the theoretical predictions when one of the modes is superior. Then, we demonstrate the absolute superiority of the dynamic policy against the two simple policies. ON/OFF periods are exponential with parameters $\eta_L = 1$ s^{-1} and $\eta_H = 0.4$ s^{-1}. The data rates are $c_L = 2$ Mbps, and $c_H = 6$ Mbps. This is Case 1. In Case 2, all the parameters, except $c_L = 1$ Mbps, are the same. We put these two cases against the interweave scenario with $\eta_H = 0.4$ s^{-1}, $c_H = 6$ Mbps, and scanning times that are exponentially distributed. Fig. 22 shows the throughput for expected scanning times in the range 0.1–1.2 s. Since the underlay access has no dependency on scanning, its performance does not change (constant value of throughput in Case 1 and Case 2).

In Fig. 22, small circles depict the theoretical boundary values for scanning time (crossing point). Obviously, there is a match. For scanning times lower than 600 ms in Case 1, and 800 ms in Case 2, the interweave mode gives higher throughput. For higher scanning time values, underlay mode should be used by the SU. For deteriorated SU channel conditions (related to its distance to the BTS/AP, higher PU transmission power, etc.), the value of the crossing point inclines, too. Finally, there is a lower throughput in Case 2, compared to Case 1, because of the lower data rate when a PU is present.

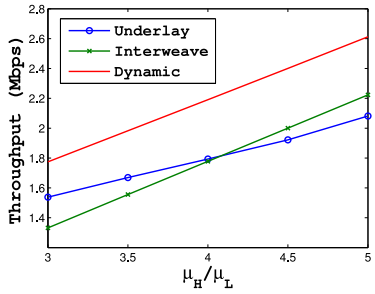

Fig. 23. The throughput dynamic policy for heavy-tailed OFF periods.

Throughput dynamic policy: Our theoretical results suggest that the dynamic policies, for both metrics (delay and throughput), provide better results for OFF periods that underlie a distribution with decreasing failure rate (higher variability). Let's consider a scenario with OFF periods that are Bonded Pareto distributed, with minimum value $L = 0.2$, maximum value $H = 100$, and shape parameter $\alpha = 1.2$. The average ON/OFF duration, and mean scanning time remain the same as previously. Fig. 23 shows the throughput for the considered policies. As illustrated in Fig. 23, the dynamic policy further improves the throughput by an additional 20%, consistently with our assertions of Section V-B that the throughput can be further improved when the underlay is superior to interweave access. Since we are dealing with highly variable OFF periods, some of them will be excessively large, and these will have a crucial impact on throughput. When that is the case, finding a new available channel will be the right thing to do.

VII. Conclusion and Future Work

We have analyzed underlay and interweave access modes, deriving expressions for the throughput and expected file delay. Then, we have validated them against realistic simulations. Another contribution is the analytical comparison between the two modes (for both metrics). Finally, we have proposed dynamic policies which enable performance improvements of up to 50%. We plan to analyze the use of Non-Orthogonal Multiple Access [48] in CRNs in the future.

References

[1] F. Mehmeti and T. Spyropoulos, "Stay or switch? Analysis and comparison of delays in cognitive radio networks with interweave and underlay spectrum access," in *Proc. ACM Int. Symp. Mobility Manage. Wireless Access*, 2016, pp. 9–18.

[2] [Online]. Available: http://share.cisco.com/internet-of-things.html

[3] I. F. Akiyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[4] B. Wang and K. Liu, "Advances in cognitive radio networks: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 5–23, Feb. 2011.

[5] N. Mahmood, F. Yilmaz, G. Oien, and M. Alouini, "On hybrid cooperation in underlay cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4422–4433, Sep. 2013.

[6] H. Kim and K. Shin, "Fast discovery of spectrum opportunities in cognitive radio networks," in *Proc. IEEE Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2008, pp. 1–2.

[7] Z. Yang, X. Xie, and Y. Zheng, "A new two-user cognitive radio channel model and its capacity analysis," in *Proc. Int. Symp. Commun. Inf. Technol.*, 2009, pp. 410–414.
[8] H. Hakim, H. Boujemaa, and W. Ajib, "Performance comparison between adaptive and fixed transmit power in underlay cognitive radio networks," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4836–4846, Dec. 2013.
[9] M. Seyfi, S. Muhaidat, and J. Liang, "Relay selection in underlay cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2012, pp. 283–288.
[10] F. Mehmeti and T. Spyropoulos, "To scan or not to scan: The effect of channel heterogeneity on optimal scanning policies," in *Proc. IEEE Int. Conf. Sens. Commun., Netw.*, 2013, pp. 264–272.
[11] M. Ki, H. Lee, and J. Song, "Performance analysis of distributed cooperative spectrum sensing for underlay cognitive radio," in *Proc. Int. Conf. Adv. Commun. Technol.*, 2009, pp. 338–343.
[12] B. Wang and D. Zhao, "Performance analysis in CDMA-based cognitive wireless networks with spectrum underlay," in *Proc. IEEE Global Telecommun. Conf.*, 2008, pp. 1–6.
[13] H. Chamkhia, M. Hasna, R. Hamila, and S. Hussain, "Performance analysis of relay selection schemes in underlay cognitive networks with decode and forward relaying," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.*, 2012, pp. 1552–1558.
[14] I. Suliman and J. Lehtomaki, "Queueing analysis of opportunistic access in cognitive radios," in *Proc. Cognitive Radio Adv. Spectr. Manage.*, 2009, pp. 153–157.
[15] H. Tran, T. Q. Duong, and H.-J. Zepernick, "Average waiting time of packets with different priorities in cognitive radio networks," in *Proc. IEEE Int. Symp. Wireless Pervasive Comput.*, 2009, pp. 122–127.
[16] 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.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.
[17] P. Wang, D. Niyato, and H. Jiang, "Voice-service capacity analysis for cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1779–1790, May 2010.
[18] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.
[19] K. Tourki, K. Qaraqe, and M. Alouini, "Outage analysis for underlay relay-assisted cognitive networks," in *Proc. IEEE Global Commun. Conf.*, 2012, pp. 1248–1253.
[20] M. Khoshkholgh, K. Navaie, and H. Yanikomeroglu, "Access strategies for spectrum sharing in fading environment: Overlay, underlay, and mixed," *IEEE Trans. Mobile Comput.*, vol. 9, no. 12, pp. 1780–1793, Dec. 2010.
[21] S. Stotas and A. Nallanathan, "Optimal sensing time and power allocation in multiband cognitive radio networks," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 226–235, Jan. 2011.
[22] T. Chu, H. Phan, and H. Zepernick, "On the performance of underlay cognitive radio networks using M/G/1/K queueing model," *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 876–879, May 2013.
[23] L. Sibomana, H. Zepernick, H. Tran, and C. Kabiri, "Packet transmission time for cognitive radio networks considering interference from primary user," in *Proc. Int. Wireless Commun. Mobile Comput. Conf.*, 2013, pp. 791–796.
[24] M. Rashid, J. Hossain, E. Hossain, and V. Bhargava, "Opportunistic spectrum scheduling for multiuser cognitive radio: A queueing analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5259–5269, Oct. 2009.
[25] S. Geirhofer, L. Tong, and B. Sadler, "Cognitive medium access: Constraining interference based on experimental models," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 95–105, Jan. 2008.
[26] A. Jafar and S. Srinivasa, "The throughput potential of cognitive radio," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 73–79, May 2007.
[27] S. Senthuran, A. Anpalagan, and O. Das, "Throughput analysis of opportunistic access strategies in hybrid underlay-overlay cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2024–2035, Jun. 2012.
[28] Y. Liu and M. Liu, "To stay or to switch: Multiuser multi-channel dynamic access," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 858–871, Apr. 2015.
[29] Y. Wang, P. Ren, F. Gao, and Z. Su, "A hybrid underlay/overlay transmission mode for cognitive radio networks with statistical Quality-of-Service provisioning," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1482–1498, Mar. 2014.
[30] W. Afifi and M. Krunz, "TSRA: An adaptive mechanism for switching between communication modes in full-duplex opportunistic spectrum access systems," *IEEE Trans. Mobile Comput.*, vol. 16, no. 6, pp. 1758–1772, Jun. 2017.
[31] W. Afifi and M. Krunz, "Incorporating self-interference suppression for full-duplex operation in opportunistic spectrum access systems," in *Proc. IEEE Trans. Wireless Comm.*, vol. 14, no. 4, pp. 2180–2191, Apr. 2015.
[32] V. Namboodiri, "Are cognitive radios energy efficient? A study of the wireless LAN scenario," in *Proc. IEEE Int. Perform. Comput. Commun. Conf.*, 2009, pp. 437–442.
[33] J. Misic and V. Misic, "Probability distribution of spectral hole duration in cognitive networks," in *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 2103–2111.
[34] R. Menon, M. Buehrer, and J. Reed, "Outage probability based comparison of underlay and overlay spectrum sharing techniques," in *Proc. IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2005, pp. 101–109.
[35] H. Song, J. P. Hong, and W. Choi, "On the optimal switching probability for a hybrid cognitive radio system," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1594–1605, Apr. 2013.
[36] A. Giorgetti, M. Varrella, and M. Chiani, "Analysis and performance comparison of different cognitive radio algorithms," in *Proc. Cognitive Radio Adv. Spectr. Manage.*, 2009, pp. 127–131.
[37] S. M. Ross, *Stochastic Processes*. New York, NY, USA: Wiley, 1996.
[38] M. F. Neuts, *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Baltimore, MD, USA: John Hopkins Univ. Press, 1981.
[39] D. Gozupek, S. Buhari, and F. Alagoze, "A spectrum switching delay-aware scheduling algorithm for centralized cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 7, pp. 1270–1280, Jul. 2013.
[40] L. Kleinrock, *Queueing Theory, Volume I: Theory*. New York, NY, USA: Wiley, 1975.
[41] M. Amjad, F. Akhtar, M. H. Rehmani, M. Reisslein, and T. Umer, "Full-duplex communication in cognitive radio networks: A survey," *IEEE Commun. Surveys Tut.*, vol. 19, no. 4, pp. 2158–2191, Fourthquarter 2017.
[42] F. Mehmeti and T. Spyropoulos, "Underlay vs. interweave: Which one is better?" EURECOM, Tech. Rep. RR-14-296, 2014. [Online]. Available: http://www.eurecom.fr/en/publication/4443
[43] U. Yechiali and P. Naor, "Queueing problems with heterogeneous arrivals and service," *Oper. Res.*, vol. 19, pp. 722–734, 1971.
[44] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio networks," in *Proc. IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2005, pp. 137–143.
[45] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. 88–95, Mar. 2009.
[46] S. Geirhofer and L. Tong, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66–72, May 2007.
[47] Rysavy Research, "Beyond LTE: Enabling the mobile broadband explosion," Rysavy Res. Rep., 2014. [Online]. Available: http://www.5gamericas.org/files/7514/1021/4070/Beyond_LTE_Enabling_Mobile_Broadband_Explosion_August_2014x.pdf
[48] F. Zhou, Y. Wu, R. Q. Hu, Y. Wang, and K. K. Wong, "Energy-efficient NOMA enabled heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 32, no. 2, pp. 152–160, Mar./Apr. 2018.

Fidan Mehmeti received the graduate degree in electrical and computer engineering from the University of Prishtina, Prishtina, Kosovo, in 2009, and the Ph.D. degree from Eurecom/Telecom ParisTech, France, in 2015. Then, he was a Postdoctoral Fellow with the University of Waterloo, Waterloo, Canada. He is currently working as a Postdoc at Penn State University, State College, PA, USA. His research interests include the area of wireless networks in general.

Thrasyvoulos Spyropoulos received the Diploma degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, CA, USA. He was a postdoc at INRIA, and then, a Senior Researcher with the Swiss Federal Institute of Technology (ETH) Zurich. He is currently an Assistant Professor at Institute EURECOM, Sophia Antipolis, France.